

Unterstellte Zahlungsbereitschaft im Internet: Theorie, Technik, Auswirkungen

Masterarbeit

Eingereicht von: **Florian Pichler, BA**

Matrikelnummer: 01103906

im Fachhochschul-Masterstudiengang Wirtschaftsinformatik
der Ferdinand Porsche FernFH GmbH

zur Erlangung des akademischen Grades

Master of Arts in Business

Betreuung und Beurteilung: Prof.(FH) Priv.-Doz. Mag. Dr. Sascha Sardadvar

Zweitgutachten: Thomas Grabetz, BSc. MA

Wien, August 2024

Ehrenwörtliche Erklärung

Ich versichere hiermit,

1. dass ich die vorliegende Masterarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Inhalte, die direkt oder indirekt aus fremden Quellen entnommen sind, sind durch entsprechende Quellenangaben gekennzeichnet.
2. dass ich diese Masterarbeit bisher weder im Inland noch im Ausland in irgendeiner Form als Prüfungsarbeit zur Beurteilung vorgelegt oder veröffentlicht habe.
3. dass die vorliegende Fassung der Arbeit mit der eingereichten elektronischen Version in allen Teilen übereinstimmt.

Wien, 13.08.2024



Unterschrift

Kurzzusammenfassung: Unterstellte Zahlungsbereitschaft im Internet: Theorie, Technik, Auswirkungen

Die vorliegende Arbeit beschäftigt sich mit dem Thema Preisdiskriminierung auf Basis von Big Data. Dazu wurde folgende Forschungsfrage formuliert: „Wie kommen unterschiedliche Preise für ein und dasselbe Produkt für Verbraucher*innen im Internet zustande und was sind die ökonomischen Auswirkungen davon?“ Um diese Frage zu beantworten, werden die theoretischen Grundlagen einschließlich Definition, technologische Möglichkeiten, ökonomische Auswirkungen und rechtliche Rahmenbedingungen von Preisdiskriminierung erläutert. In einer empirischen Untersuchung wurden die Preise von acht Produkten auf acht E-Commerce-Seiten über einen Zeitraum von fünf Wochen erhoben. Drei Userprofile wurden mittels Web Scraping im normalen und Inkognito-Browsermodus simuliert. Die statistische Auswertung der Daten ergab keine signifikanten Preisunterschiede, die auf den Einsatz von Algorithmen zur Preisdiskriminierung hindeuten. Die Arbeit schließt mit der Überlegung, dass technologische und ethische Bedenken sowie regulatorische Anforderungen Gründe für den Verzicht auf Preisdiskriminierung sein können, gibt Empfehlungen für den Umgang damit und einen Ausblick auf zukünftige Forschung.

Schlagwörter:

Preisdiskriminierung, Big Data, Web Scraping, E-Commerce, Dynamische Preisgestaltung

Abstract: Assumed willingness to pay on the Internet: Theory, technology, effects

This thesis deals with the topic of price discrimination based on big data. The research question was formulated as follows "How do consumers get different prices for the same product on the internet and what are the economic effects?" To answer this question, the theoretical foundations, including the definition, technological possibilities, economic effects and legal framework of price discrimination are explained. In an empirical study, the prices of eight products on eight e-commerce sites were collected over a period of five weeks. Three user profiles were simulated using web scraping in normal and incognito browser mode. Statistical analysis of the data revealed no significant price differences that would indicate the use of price discrimination algorithms. The paper concludes by considering that technological and ethical concerns as well as regulatory requirements may be reasons for not using price discrimination, gives recommendations on how to deal with it, and provides an outlook for future research.

Keywords:

Price discrimination, big data, web scraping, e-commerce, dynamic pricing

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die zum Gelingen dieser Masterarbeit beigetragen haben.

Mein besonderer Dank gilt meinem Betreuer Herrn Prof.(FH) Priv.-Doz. Mag. Dr. Sascha Sardadvar, für seine Unterstützung, seine wertvollen Ratschläge und seine stets konstruktiven Anmerkungen. Sein Fachwissen und sein Engagement haben wesentlich zum Gelingen dieser Arbeit beigetragen.

Ebenso möchte ich mich bei Philip Faber bedanken, der mir in allen Fragen der Softwareentwicklung mit Rat und Tat zur Seite stand. Seine kompetente Unterstützung hat mir geholfen, die technischen Herausforderungen zu meistern.

Ein herzliches Dankeschön geht auch an Sabrina Reithofer, die sich die Zeit genommen hat, meine Arbeit Korrektur zu lesen. Ihre sorgfältigen Anmerkungen und Verbesserungsvorschläge haben die Qualität dieser Arbeit wesentlich verbessert.

Schließlich danke ich meiner Familie, meinen Freunden und allen anderen, die mich auf diesem Weg in vielfältiger Weise unterstützt und begleitet haben.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Arbeitsziel.....	5
1.2	Methodische Vorgangsweise	6
1.3	Aufbau der Arbeit.....	8
2	Grundlagen	9
2.1	Preisdiskriminierung im Allgemeinen.....	9
2.2	Technologische Grundlagen für Big-Data-basierte Preisdiskriminierung	20
2.3	Rechtliche Rahmenbedingungen und ökonomische Auswirkungen	37
3	Methodik	53
3.1	Quantitative standardisierte Beobachtung	53
3.2	Das Web-Scraping-Tool: Aufbau und Funktionsweise	59
3.3	Datenerhebung	69
4	Ergebnisse.....	72
4.1	Auswertung	72
4.2	Präsentation, Diskussion und Limitation.....	87
4.3	Interpretation	89
5	Zusammenfassung und Schlussfolgerung.....	91
6	Literaturverzeichnis	97
7	Abbildungsverzeichnis.....	104
8	Tabellenverzeichnis	106
9	Quellcodeverzeichnis	108

10	Abkürzungsverzeichnis	109
	Anhang A	110

1 Einleitung

In einer zunehmend datengetriebenen Wirtschaft können Unternehmen ihre Preisgestaltung dynamisch an die individuellen Präferenzen und die Kaufkraft potenzieller Kund*innen anpassen. Online-Händler*innen haben Zugang zu einer Vielzahl von Informationen über ihre Kund*innen, die sie zu ihrem eigenen kommerziellen Vorteil nutzen können. Eine dieser Praktiken ist die verhaltensbasierte Preisgestaltung, bei der Online-Unternehmen ihren Kund*innen auf der Grundlage ihres Verhaltensprofils personalisierte Angebote unterbreiten. Ziel ist es, den höchstmöglichen Preis zu erzielen, den ein*e Kund*in zu zahlen bereit ist. Diese Praktiken der Preisdiskriminierung auf der Grundlage von Big Data haben Auswirkungen auf die Fairness und das Wohlergehen der Verbraucher*innen. Aufgrund ihres Erfolgs und ihrer Akzeptanz unter anderem in der Flug- und Hotelbranche sowie der zunehmenden Verfügbarkeit von Daten über das Verhalten, das Engagement und die Einstellungen der Verbraucher*innen werden diese personalisierten und dynamische Preisstrategien immer beliebter (Neubert, 2022; Spann & Skiera, 2020, S. 323; Steinberg, 2020, S. 97 ff.; van Boom, van der Rest, van den Bos & Dechesne, 2020, S. 331 ff.). Scherer und Ross (1990) definieren Preisdiskriminierung als den Verkauf (oder Kauf) unterschiedlicher Einheiten eines Produkts oder einer Dienstleistung zu Preisen, die nicht direkt den Unterschieden in den Angebotskosten entsprechen (Scherer & Ross, 1990, S. 489).

Bereits Mitte der 1990er Jahre wurde damit begonnen, Online-Preissetzungsmechanismen einzusetzen, die den Kund*innen unterschiedliche Preise offerierten. Ein Beispiel hierfür war das Konzept der damals aufkommenden Billigfluglinien, die sich sehr stark über den Preis differenzierten (Spann, Klein, Makhlof & Bernhardt, 2005, S. 53 f.). In den letzten Jahren hat sich der Einsatz dynamischer Preise insbesondere im Online-Handel aufgrund der Verfügbarkeit größerer Datenmengen und leistungsfähigerer IT-Infrastrukturen zunehmend verbreitet (Chen, Mislove & Wilson, 2016, S. 1339 f.). Dadurch hat dynamische Preisgestaltung auch in der öffentlichen Wahrnehmung an Bedeutung gewonnen und wird häufig kritisch betrachtet, obwohl Erfahrungen im Bereich der Flugbranche und die mitunter stark schwankenden Preise an Tankstellen zeigen, dass dynamische Preise akzeptiert werden (Spann & Skiera, 2020 S. 323).

Unternehmen können verschiedene Arten der Preisdiskriminierung anwenden: Theoretisch können sie für jede*n Konsument*in individuelle Preise festlegen oder die Konsument*innen in Gruppen einteilen. Andererseits haben sie auch die Möglichkeit,

eine Preisdiskriminierung über das Produkt anzubieten, z.B. durch verschiedene Varianten eines Produkts. Eine Praxis, die auch im stationären Handel weit verbreitet ist. So gibt es für (fast) identische Produkte, wie z.B. Joghurt, unzählige Varianten in unterschiedlichen Preisklassen. Die Kund*innen entscheiden in diesem Fall selbst, welcher Gruppe sie sich zuordnen. Ziel der Unternehmen ist es dabei, über die individuelle Zahlungsbereitschaft der Kund*innen so viel Konsumentenrente¹ wie möglich zu erhalten und den Gewinn zu steigern. Dynamische Preissetzung kann zeitbasiert, kund*innenbasiert aber auch kanalbasiert angewendet werden (Spann & Skiera, 2020, S. 324 f.; Kollmann, 2020, S. 657 ff.; Konrad & Polk, 2020, S. 793).

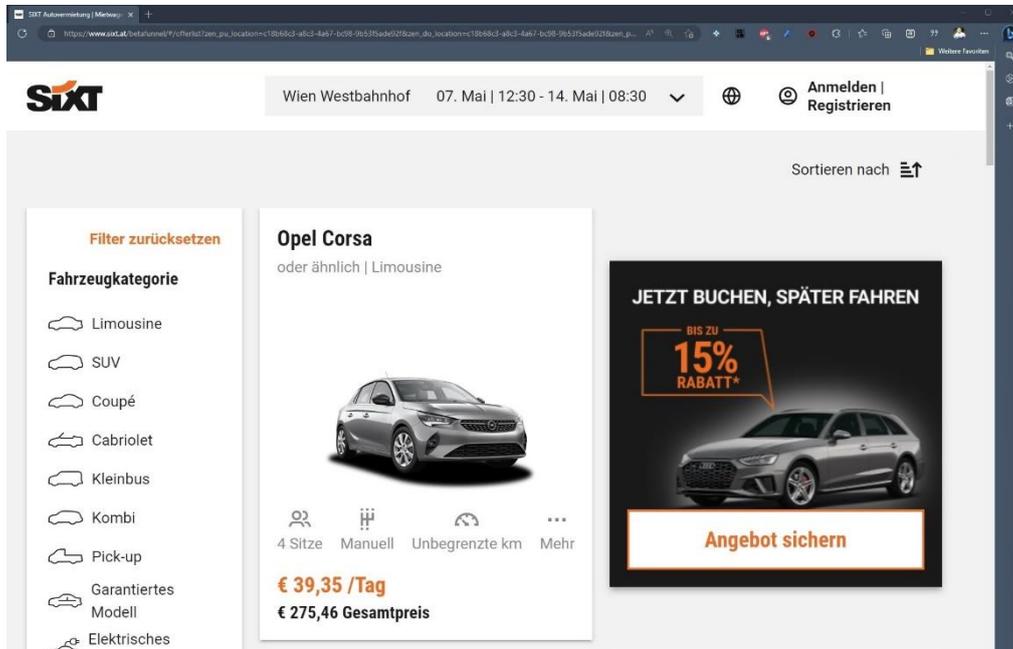
Unternehmen nutzen z.B. Algorithmen des maschinellen Lernens, um eine breite Palette personalisierter Dienste anzubieten, die von personalisierten Nachrichteninhalten über Werbung, Energiemanagement bis hin zu Dating- und Partnervermittlungsdiensten reichen. Amazon gilt als Vorreiter auf diesem Gebiet und nutzt Big Data, um das Einkaufserlebnis zu personalisieren und das Kaufverhalten der Kund*innen zu verstehen. Inzwischen hat sich diese Personalisierung auch auf die Preisgestaltung ausgeweitet, indem Unternehmen auf der Grundlage einer algorithmischen Bewertung der wahrscheinlichen Zahlungsbereitschaft jedes Käufers oder jeder Käuferin unterschiedliche Preise anbieten. Townley, Morrison & Yeung bezeichnen dieses Phänomen als "Algorithmic Consumer Price Discrimination" (ACPD), das in der Vergangenheit zu öffentlichen Protesten geführt hat: Als Beispiel dient hier neuerlich Amazon, das im Jahr 2000 in die Kritik geriet, als es auf seinem Online-Marktplatz DVDs an verschiedene Personen zu unterschiedlichen Preisen verkaufte, oder 2012, als Amazon erneut wegen stark schwankender Preise in die Schlagzeilen der US-Medien geriet. Die digitale Personalisierung des Einzelhandels und insbesondere ACPD kann für die Kund*innen von Vorteil sein, wirft aber auch Bedenken hinsichtlich der Preisfairness und des Datenschutzes auf (Townley, Morrison & Yeung, 2017, S. 683 ff.).

Als praktisches Beispiel für individuelle Preisdiskriminierung können die folgenden Abbildungen (1 & 2) dienen. In Abbildung 1 wurde auf der Webseite eines Autovermieters eine Preisabfrage für ein Auto im Zeitraum 07.05.2023 bis 14.05.2023 (Abholort Wien Westbahnhof) im normalen Modus des Browsers (Microsoft Bing), inklusive Speicherung von z.B. Cookies, Verlaufsdaten oder Suchanfragen durchgeführt. In Abbildung 2 wurde

¹ Die Konsumentenrente ist der Nutzenüberschuss, der sich aus der Differenz zwischen der maximalen Zahlungsbereitschaft und dem tatsächlichen Preis, den ein*e Konsument*in für ein Gut zu zahlen bereit ist, ergibt (Ahlheim & Rose, 1992, S. 34 ff.; Mankiw, 2021, S. 132 ff.).

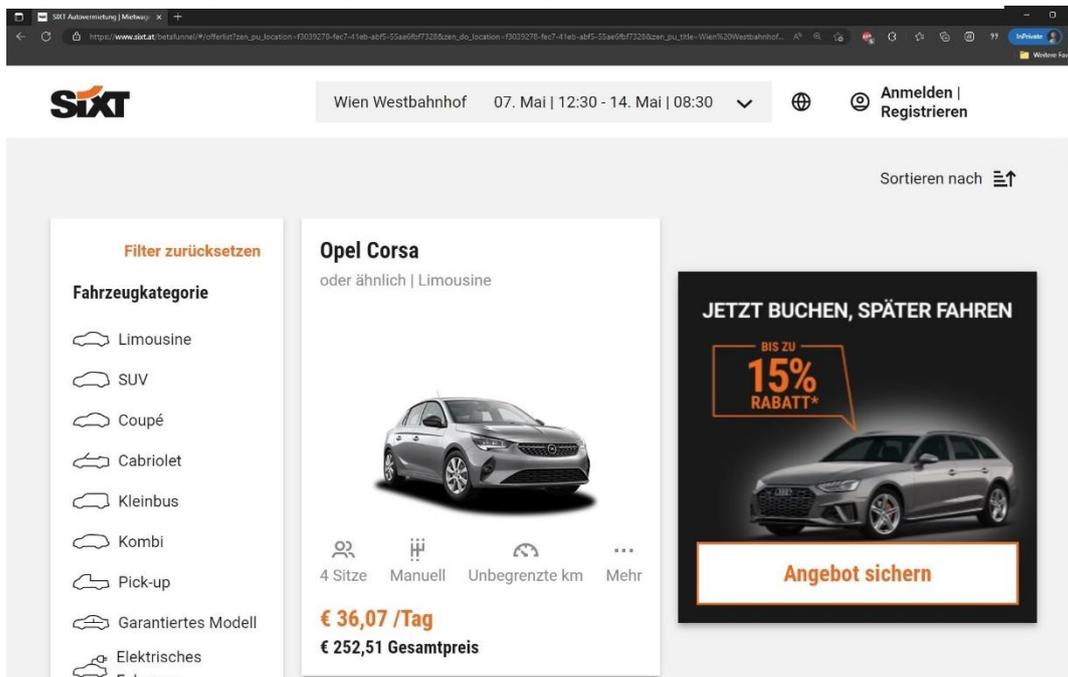
zum selben Zeitpunkt exakt dieselbe Abfrage im Inkognito-Modus (keine Datenspeicherung) desselben Browsers durchgeführt.

Abbildung 1 - Preisabfrage Sixt: normaler Modus (Bing Browser)



Anmerkung: eigener Screenshot (13.04.2023)

Abbildung 2 - Preisabfrage Sixt: Inkognito-Modus (Bing Browser)



Anmerkung: eigener Screenshot (13.04.2023)

Wie aus den beiden Abbildungen ersichtlich ist, kostet das idente Auto im normalen Modus 3,28 EUR (+8,34%) mehr pro Tag als im Inkognito-Modus. Im Gesamtpreis ergibt sich somit ein Unterschied von 21,95 EUR.

Konrad und Polk (2020) argumentieren, dass sich durch personalisierte Preise Überschüsse (Konsumentenrente) von Verbraucher*innen hin zu den Unternehmen verschieben. Indem sie unterschiedliche Preise für verschiedene Kund*innensegmente festlegen, können Unternehmen ihre Gewinne maximieren. Kund*innen mit einer höheren Zahlungsbereitschaft können höhere Preise angeboten werden, während Kund*innen mit einer geringeren Zahlungsbereitschaft niedrigere Preise erhalten können. Darüber hinaus ermöglicht personalisierte Preisdiskriminierung eine bessere Marktsegmentierung, da Unternehmen ihre Angebote und Marketingstrategien auf die Bedürfnisse und Präferenzen einzelner Gruppen von Kund*innen abstimmen können. Dies kann zu einem Wettbewerbsvorteil führen, da Unternehmen flexibler auf Nachfrageänderungen reagieren und ihre Marktanteile erhöhen können. Gleichzeitig kann sich dieses Szenario auch positiv auf die Verbraucher*innen auswirken. Zum einen haben sie die Möglichkeit, Kosten zu sparen, indem sie für Produkte oder Dienstleistungen, die ihrer individuellen Zahlungsbereitschaft entsprechen, niedrigere Preise erhalten. Zum anderen können sie von maßgeschneiderten Angeboten profitieren, die besser auf ihre Bedürfnisse zugeschnitten sind. Auf Basis der Informationen und Daten der Kund*innen, z.B. über das Kaufverhalten, können Unternehmen individuelle Empfehlungen und Rabatte anbieten. Die Nachteile, die sich daraus für die beteiligten Parteien ergeben können, sind jedoch auch der Grund, warum nach den Ergebnissen von Konrad und Polk (2020) individuelle Preisdiskriminierung mit Ausnahme der Reisebranche derzeit eher selten angewendet wird. So stünden Unternehmen vor einem Glaubwürdigkeitsproblem, da Verbraucher*innen Bedenken hinsichtlich des Datenschutzes, des Schutzes ihrer Privatsphäre und der Preisfairness haben können. Darüber hinaus können negative Reaktionen der Kund*innen auftreten, wenn sie das Gefühl haben, ungerecht behandelt zu werden. Transparenz ist daher ein entscheidender Faktor, um die Glaubwürdigkeit und den Ruf der Verlässlichkeit eines Unternehmens in Bezug auf individuelle Preisdiskriminierung zu gewährleisten. Ansonsten kann es dazu führen, dass Kund*innen keine Daten mehr zur Verfügung stellen und ein Imageverlust für das Unternehmen die Folge ist. Dieses Risiko scheint für die Unternehmen derzeit noch schwerer zu wiegen als die möglichen Gewinne durch individuelle Preisdiskriminierung (Konrad & Polk, 2020, S. 793 f. & 797).

Auch Azzolina, Razza, Sartiano und Weitschek (2021) finden in ihrer Studie, in der sie die Preise einer nationalen und einer internationalen Fluggesellschaft für Inlands- und Auslandsflüge über einen Zeitraum von sechs Wochen beobachteten, keine Hinweise auf (individuelle) Preisdiskriminierung. Dennoch argumentieren sie, dass Preisdiskriminierung in diesem Sektor von Bedeutung ist und auch angewendet wird. Darüber hinaus erwähnen die Autoren auch ein Phänomen, bei dem viele Nutzer*innen glauben, dass ihnen Fluggesellschaften aufgrund ihrer häufigen Suche nach Flügen im Internet höhere Preise anbieten als Nutzer*innen, die eine Webseite zum ersten Mal besuchen. In ihrer Studie konnten sie diesen Glauben nicht als Fakt nachweisen und erklären es sich mit der illusorischen Korrelation, bei der Menschen eine direkte Beziehung zwischen zwei Variablen/Ereignissen wahrnehmen, auch wenn dies objektiv nicht der Fall ist. Preisunterschiede, speziell in der Reisebranche, können demnach durch mehrere Faktoren beeinflusst werden, die den Nutzer*innen nicht bekannt sind (Azzolina, Razza, Sartiano & Weitschek, 2021, S. 2299 ff.). Hindermann (2018) bestätigt die Aussage von Azzolina et al., dass Preisdiskriminierung in der Reisebranche am wahrscheinlichsten ist. In seiner Untersuchung mehrerer Studien zum Thema Online-Preisdiskriminierung kommt er jedoch auch zu dem Schluss, dass es wenig bis keine Belege für diese Praxis gibt. Die häufigste Form der Preisdiskriminierung im Internet ist die standortbezogene Preisdiskriminierung, wobei die Standorte durch verschiedene Länder definiert werden. In diesem Fall ist jedoch zu beachten, dass eine (Preis-) Diskriminierung aufgrund der Staatsangehörigkeit oder Nationalität bei grenzüberschreitenden Transaktionen in der Europäischen Union (EU) verboten ist (Das Europäische Parlament und der Rat der Europäischen Union, 2018). Die technische Diskriminierung hängt hauptsächlich vom Betriebssystem ab, die individuelle Diskriminierung von der Tatsache, ob ein*e Nutzer*in ein registriertes Konto auf einer Webseite besitzt oder nicht. Es wird jedoch auch auf eine Studie verwiesen, in der nachgewiesen wurde, dass das Surfverhalten der Nutzer*innen ebenfalls einen Einfluss auf den Preis hat (Hindermann, 2018, S. 7 f.).

1.1 Arbeitsziel

Die Ausführungen in der Einleitung zeigen, dass Preisdiskriminierung auf Basis von Big Data im Internet zwar derzeit noch sehr wenig Anwendung zu finden scheint, aber dennoch immer relevanter wird, da immer mehr Daten von Kund*innen und technische Möglichkeiten zu deren Erhebung zur Verfügung stehen.

Ziel dieser Arbeit ist es, unter anderem einen Überblick über die theoretischen Hintergründe von Preisdiskriminierung im Allgemeinen und aufgrund von Big Data zu geben. Es wird auf die verschiedenen Arten der Preisgestaltung im Internet eingegangen und wie diese zustande kommen. Dabei werden die technischen Möglichkeiten und Modelle unter Berücksichtigung des aktuellen Stands der Technik und der historischen Entwicklung vorgestellt und die ökonomischen Auswirkungen erläutert. Anhand von Fallbeispielen sollen diese Ausführungen untermauert werden. Es soll ein Bewusstsein für individuelle Preisdiskriminierung geschaffen werden und Vorschläge präsentiert werden, wie Verbraucher*innen mit diesem immer relevanter werdenden Thema der Big-Data-basierten Preisdiskriminierung umgehen können. Ebenso wird eine quantitative standardisierte Beobachtung durchgeführt, um herauszufinden, ob Preisdiskriminierung für Produkte oder Dienstleistungen in Österreich durchgeführt wird. Dies dient der Beantwortung folgender Forschungsfrage:

*Wie kommen unterschiedliche Preise für ein und dasselbe Produkt für Verbraucher*innen im Internet zustande und was sind die ökonomischen Auswirkungen davon?*

In diesem Zusammenhang ist auch von Interesse, wie sich individuelle Preisdiskriminierung auf Konsument*innen und Unternehmen auswirkt, welche Vor- und Nachteile damit verbunden sind, welche Bedenken und Gefahren bestehen und wie es generell um die Akzeptanz solcher Praktiken steht.

1.2 Methodische Vorgangsweise

Um das genannte Arbeitsziel erreichen zu können, wird im ersten Teil dieser Arbeit eine umfassende Literaturanalyse zu den theoretischen Grundlagen des Themenfeldes durchgeführt. Dies dient dazu, Definitionen, Hintergründe und Konzepte zu liefern, um ein breites Verständnis für das vorliegende Thema zu schaffen und den aktuellen Stand der Forschung abzubilden. Gleichzeitig soll so ein Bezugsrahmen definiert werden, in dem sich die Arbeit bewegt (Webster & Watson, 2002, S. 20 f.). Dabei wird auf die ökonomischen Grundlagen zur Preisdiskriminierung eingegangen und vorhandene empirische Studien werden vorgestellt. In Bezug auf die Forschungsfrage werden speziell die theoretischen Hintergründe beleuchtet, vor allem die Arten der Preisdiskriminierung und Preisgestaltung im Internet sind in diesem Zusammenhang von Bedeutung, sowie die ökonomischen Auswirkungen für Unternehmen und die Kund*innen.

Im zweiten Teil der Arbeit wird eine empirische Untersuchung mittels quantitativer standardisierter Beobachtung durchgeführt. Bei dieser Methode sind sowohl das Ziel der Beobachtung als auch die Art der Dokumentation genau festgelegt. Die Elemente des Beobachtungsgegenstandes sind bekannt und sollen nun genau erfasst werden. Jedes Element des Gegenstandes wird einer bestimmten Kategorie zugeordnet, deren Vorkommen in der Beobachtungssituation dokumentiert werden soll (Gniewosz, 2011, S. 103). In der vorliegenden Beobachtung werden die Preise von acht, in Österreich erhältlichen Produkten/Dienstleistungen auf ebenso vielen E-Commerce-Seiten im Internet über einen Zeitraum von etwa fünf Wochen regelmäßig (siebenmal pro Woche, mehrmals täglich) von drei verschiedenen Userprofilen mittels Web Scraping² erhoben. Die Userprofile werden durch das verwendete Endgerät definiert. Zur Durchführung des Web Scraping wurde ein Tool entwickelt, das diese Userprofile mit Hilfe von User Agents simuliert und die Beobachtungen aufzeichnet und speichert. Aber nicht nur der Vergleich zwischen den Userprofilen ist dabei von Bedeutung auch der Vergleich der einzelnen Profile mit sich selbst, einmal mit (Normalmodus des Browsers) und einmal ohne (Inkognito-Modus des Browsers) Verwendung von Cookies, ist für die Auswertung von Interesse. Die so gesammelten Daten werden mithilfe statistischer Methoden deskriptiv ausgewertet und anschließend basierend auf den gewonnenen Erkenntnissen interpretiert. Damit soll untersucht werden, ob Unterschiede zwischen und innerhalb der Userprofile festzustellen sind. Ziel dieser standardisierten quantitativen Beobachtung ist es daher, herauszufinden, ob diese Userprofile auf den untersuchten Webseiten zur Preisdiskriminierung im Internet beitragen und somit der Beantwortung der Forschungsfrage dienen. Darüber hinaus werden folgende Hypothesen aufgestellt:

- H0 (Nullhypothese): Es gibt keine Unterschiede in den Preisen zu den verschiedenen Zeitpunkten zwischen den verschiedenen User Agents und den verschiedenen Browsermodi.
- H1 (Alternativhypothese): Es gibt Unterschiede in den Preisen zu den verschiedenen Zeitpunkten zwischen den verschiedenen User Agents und den verschiedenen Browsermodi.

² Web Scraping ist eine Technik zur Extraktion von Daten aus dem Internet. Diese können z.B. für spätere Analysen oder andere Verwendungen in einer Datei oder einer Datenbank gespeichert werden (Zhao, 2017, S. 1).

1.3 Aufbau der Arbeit

Die vorliegende Arbeit gliedert sich in mehrere aufeinander aufbauende Kapitel, die sowohl theoretische als auch empirische Aspekte der Preisdiskriminierung im Online-Handel untersuchen.

In Kapitel 1 wird ein Überblick über das Thema gegeben und die Forschungsfrage vorgestellt. Die Relevanz von Preisdiskriminierung im digitalen Zeitalter wird diskutiert und die Zielsetzung der Arbeit definiert. Des Weiteren werden das methodische Vorgehen und der Aufbau der Arbeit beschrieben.

In Kapitel 2 werden die Definition von Preisdiskriminierung sowie die Abgrenzung zu verwandten Begriffen wie Preisstreuung und Dynamic Pricing erläutert. Des Weiteren werden die technologischen Grundlagen der Big-Data-basierten Preisdiskriminierung sowie die rechtlichen Rahmenbedingungen und ökonomischen Implikationen behandelt.

Kapitel 3 beschreibt das empirische Vorgehen der Arbeit. Es wird die Methode der quantitativen standardisierten Beobachtung erläutert, bei der die Preise von acht Produkten bzw. Dienstleistungen auf verschiedenen E-Commerce-Seiten über einen Zeitraum von fünf Wochen erhoben wurden. Der Einsatz eines eigens entwickelten Web Scrapers zur Datenerhebung und die Simulation verschiedener Nutzerprofile werden detailliert beschrieben.

Die Ergebnisse der empirischen Untersuchung werden in Kapitel 4 dargestellt. Die erhobenen Daten werden statistisch ausgewertet und die Preisunterschiede zwischen verschiedenen Userprofilen und Browsermodi analysiert. Diese Ergebnisse werden im Hinblick auf die Forschungsfrage interpretiert und diskutiert, wobei auch die Grenzen der Studie aufgezeigt und mögliche Erklärungen für die beobachteten Phänomene gegeben werden.

Das abschließende Kapitel 5 fasst die wichtigsten Ergebnisse der Arbeit zusammen, zieht Schlussfolgerungen und gibt einen Ausblick auf mögliche zukünftige Forschungsarbeiten.

2 Grundlagen

Im deutschsprachigen Raum werden die Begriffe *Preisdiskriminierung* und *Preisdifferenzierung* häufig synonym verwendet. Im englischen Sprachgebrauch ist dagegen ausschließlich von Preisdiskriminierung (engl. *price discrimination*) die Rede (o. V., 2005, S. 206; Kalka & Krämer, 2020, S. 78). Aufgrund dieser Argumentation wird in dieser Arbeit der Begriff *Preisdiskriminierung* als übergeordneter Begriff verwendet.

2.1 Preisdiskriminierung im Allgemeinen

Wenn Verkäufer*innen verschiedenen Käufer*innen zum selben Zeitpunkt unterschiedliche Preise für identische Produkte berechnen, wird von Preisdiskriminierung gesprochen. Die Definition von Preisdiskriminierung ist komplex, aber im Wesentlichen bedeutet sie den Verkauf von Einheiten eines Gutes oder einer Dienstleistung zu Preisunterschieden, die nicht direkt den Unterschieden in den Herstellungskosten entsprechen. Dies umfasst nicht nur den Verkauf identischer Produkte zu unterschiedlichen Preisen an unterschiedliche Personen, sondern auch den Verkauf identischer Einheiten an denselben oder dieselbe Käufer*in zu unterschiedlichen Preisen (z.B. wenn Telekommunikationsunternehmen weniger Geld für zusätzliches Datenvolumen berechnen) und das Verlangen desselben Preises für Transaktionen mit unterschiedlichen Kosten (z.B. wenn ein Restaurant sowohl á la carte-Gerichte als auch ein festes Tagesmenü anbietet, kann es für beide Optionen denselben Preis verlangen, obwohl die Kosten für die Zubereitung der á la carte-Gerichte in der Regel höher sind als die Kosten für die Menükomponenten). Auch kann die Definition von Preisdiskriminierung teilweise irreführend sein, da Preisdiskriminierung demnach auch dann vorliegen würde, wenn ein*e Anbieter*in von Kund*innen unterschiedliche Preise verlangt, weil er oder sie tatsächlich unterschiedliche Kosten hat. Ein Beispiel hierfür wären höhere Versandkosten, weil Kundin A, im Gegensatz zu Kunden B, im Ausland ansässig ist. Im Regelfall wird dieses Beispiel aber nicht unter der Definition von Preisdiskriminierung verstanden. Unter Preisdiskriminierung werden in der Regel Situationen verstanden, in denen Anbieter*innen zur Gewinnmaximierung systematisch Preisbildungsmechanismen einsetzen, um von einzelnen Kund*innen den höchstmöglichen Preis zu erhalten, den diese für ein bestimmtes Gut oder eine bestimmte Dienstleistung zu zahlen bereit sind (Reservationspreis) (Scherer & Ross, 1990, S. 489; Wiedemann, 2023, S. 83 f.).

Preisdiskriminierung im Kontext einer digitalisierten Welt kann als einer von zwei Aspekten der dynamischen Preisgestaltung gesehen werden. Dynamische Preisgestaltung beschreibt die flexible und schnelle Anpassung der Preise an Marktbedingungen wie beispielsweise Änderungen von Angebot und Nachfrage oder das Verhalten von Wettbewerber*innen. Der andere Aspekt ist die Preisstreuung, die entweder räumlich oder zeitlich begrenzt ist. Bei der räumlichen Preisstreuung werden bestimmte Produkte von mehreren Verkäufer*innen an verschiedenen Orten oder Märkten zu unterschiedlichen Preisen angeboten. Die zeitliche Preisstreuung bezieht sich auf die Veränderung der Preise für eine bestimmte Ware oder Dienstleistung im Laufe der Zeit. Der Preis für ein Produkt kann sich also je nach Angebot und Nachfrage über den Zeitpunkt des Verkaufs definieren. Eine solche dynamische Preisgestaltung wird beispielsweise von Fahrgemeinschafts- oder Car-Sharing-Plattformen genutzt, die ihre Preise bei hoher Nachfrage erhöhen und bei geringer Nachfrage senken. Obwohl sich beide Begriffe, dynamische Preisgestaltung und Preisdiskriminierung, auf die Preisgestaltung beziehen, unterscheiden sie sich in ihrer Anwendung und ihrem Fokus. Dynamische Preisgestaltung ist marktorientiert, während Preisdiskriminierung verbraucher*innenorientiert ist. Es ist jedoch wichtig anzumerken, dass die beiden Konzepte in der Praxis oft Hand in Hand gehen können, insbesondere in der digitalen Wirtschaft, in der Algorithmen und künstliche Intelligenz für die Preisgestaltung eingesetzt werden (Grochowski, Jablonowska, Lagioia & Sartor, 2022, S. 36 ff.; Narahari, Raju, Ravikumar & Shah, 2005, S. S. 233 f.).

Voraussetzungen für Preisdiskriminierung

Damit ein*e Anbieter*in Preisdiskriminierung gewinnbringend betreiben kann, müssen drei Bedingungen erfüllt sein: Marktmacht, Kenntnis des Reservationspreises und Verhinderung von Arbitrage (Scherer & Ross, 1990, S. 489; Wiedemann, 2023, S. 89).

Marktmacht

Der*die Anbieter*in muss eine gewisse Preiskontrolle ausüben können, d.h. er*sie muss über eine gewisse Marktmacht (z.B. Monopol) verfügen. Bei einem perfekten Wettbewerb kann ein Unternehmen nicht dauerhaft einen Preis verlangen, der vom Marktpreis abweicht. In diesem hypothetischen Modell des vollkommenen Wettbewerbs hat kein Unternehmen Marktmacht und damit keinen Einfluss auf den Preis. Der Marktpreis entspricht in diesem Fall den Grenzkosten³. So kann ein rein wettbewerbsorientiertes

³ Grenzkosten sind die zusätzlichen Kosten, die einem Unternehmen entstehen, wenn es eine zusätzliche Einheit eines Produkts herstellt. Sie geben an, wie sich die Gesamtkosten ändern, wenn

Unternehmen keine profitable Preisdiskriminierung betreiben, da es eben keine Kontrolle über den Preis hat. Verlangt es höhere Preise für ein Produkt, werden die Kund*innen einfach zur Konkurrenz abwandern. Natürlich kann das Unternehmen einige Einheiten unter dem Marktpreis verkaufen (z.B. aus altruistischen Gründen) und damit auf einen Teil des Gewinns verzichten, aber es ist nicht verpflichtet, dies zu tun, um so viele Einheiten wie gewünscht zu verkaufen (Scherer & Ross, 1990, S. 489; Wiedemann, 2023, S. 89).

Kenntnis des Reservationspreises

Der*die Verkäufer*in muss in der Lage sein, seine*ihre Kund*innen in Gruppen mit unterschiedlicher Preiselastizität der Nachfrage oder in diskrete Klassen mit unterschiedlichen Reservationspreisen einzuteilen. Es muss zumindest annähernd erkennbar sein, wie viel bestimmte Gruppen von Kund*innen bereit sind, für ein Produkt oder eine Dienstleistung auszugeben. Daraus folgt, dass Preisdiskriminierung nur erfolgreich sein kann, wenn Kenntnis über die eigenen Kund*innen vorliegt. Sind diese hingegen anonym und weisen ein homogenes Kaufverhalten auf, ist Diskriminierung über den Preis nicht möglich (Scherer & Ross, 1990, S. 489; Wiedemann, 2023, S. 91 f.)

Verhinderung von Arbitrage

Die Möglichkeiten der Arbitrage, d.h. des Weiterverkaufs von Produkten durch Kund*innen an andere Käufer*innen zu höheren Preisen, müssen begrenzt bzw. unmöglich gemacht werden. Arbitrage ist in Dienstleistungssektoren wie Gesundheit und Bildung kaum möglich, während sie bei den meisten physischen Gütern leichter zu erreichen ist (Scherer & Ross, 1990, S. 489). Das Bestreben diskriminierender Unternehmen, die Produzentenrente⁴ durch Abschöpfung der Konsumentenrente zu maximieren, kann langfristig nicht erfolgreich sein, wenn Kund*innen die Möglichkeit haben, ein Gut (oder eine Dienstleistung), das sie zu einem bestimmten Preis erworben haben, an Dritte zu einem höheren Preis weiterzuverkaufen, der jedoch unter dem Preis liegt, den sie an den*die Anbieter*in zahlen müssten. Dies gilt für sämtliche Arten der Preisdiskriminierung. Gewährt ein*e Anbieter*in beispielsweise Mengenrabatte (Preisdiskriminierung zweiten Grades), so kann der für ihn*sie langfristig positive Effekt

die Produktionsmenge um eine Einheit erhöht wird. Konkret sind die Grenzkosten der Betrag, um den die Gesamtkosten steigen, wenn das Unternehmen die Produktion um eine Einheit erhöht (Mankiw, 2021, S. 251 f.).

⁴ Die Produzentenrente ist der Betrag, den ein*e Verkäufer*in für ein Produkt erhält, abzüglich der Kosten für die Produktion dieses Produkts. Sie misst den Vorteil oder Nutzen, den Verkäufer*innen aus der Teilnahme am Markt ziehen (Mankiw, 2021, S. 137 ff.).

gemindert oder ganz aufgehoben werden, wenn einzelne Kund*innen große Mengen des Produkts abnehmen. Dadurch sinkt der Stückpreis für diese Kund*innen und sie können das Produkt stückweise zu einem Preis weiterverkaufen, der zwar über dem von ihnen gezahlten Stückpreis, aber unter dem Einzelpreis des*der Anbieter*in liegt. Um Arbitrage unmöglich zu machen oder zumindest einzuschränken, ist die Art des verkauften Gutes von Bedeutung. Bei beweglichen Gütern ist die Unterbindung von Arbitrage generell schwierig und der Weiterverkauf ohne größere Hindernisse möglich, was für die Anbieter*innen ein durchaus großes Hindernis darstellt, Preisdiskriminierung zu praktizieren. Lediglich das vertragliche Verbot des Weiterverkaufs oder die Einschränkung von Service- oder Garantieleistungen sind hier Möglichkeiten, dem entgegenzuwirken. Bei verderblichen Gütern oder solchen, die mit hohen Transaktionskosten⁵ verbunden sind, ist Arbitrage dagegen von vornherein eingeschränkt oder sogar nahezu unmöglich. Fluggesellschaften oder auch Eventveranstalter*innen bekämpfen Arbitrage, indem sie ihre Tickets personalisieren und so einen Weiterverkauf unmöglich machen, obwohl die Personalisierung für die Leistungserbringung nicht relevant ist (Wiedemann, 2023, S. 90).

Die drei Grade der Preisdiskriminierung

In Anlehnung an Arthur Cecil Pigou (1920) wird heute üblicherweise von drei Graden der Preisdiskriminierung gesprochen (Pigou, 1932, S. 278 f.; Scherer & Ross, 1990, S. 489).

Preisdiskriminierung ersten Grades

Bei der Preisdiskriminierung *ersten Grades*, auch perfekte Diskriminierung genannt, verkauft das Unternehmen jedes Gut oder Dienstleistung zum höchstmöglichen Preis, den der oder die Kund*in bereit ist zu zahlen, wodurch kein Überschuss an Kund*innen entsteht und der gesamte Überschuss als Produzentenrente angeeignet wird. Dies bedeutet, dass der Verkäufer in der Lage ist, den Reservationspreis jedes*jeder Kund*in genau zu bestimmen und jedes Produkt zu diesem Preis zu verkaufen, wodurch der maximale Gewinn erzielt wird. Der*die Kund*in hingegen hat keinen Überschuss (Konsumentenrente), da er*sie genau den Preis zahlt, den er*sie bereit ist zu zahlen. Der gesamte Überschuss geht somit als Produzentenrente an den*die Verkäufer*in (Scherer & Ross, 1990, S. 489 f.). Da es in der Realität nahezu unmöglich ist, Preisdiskriminierung ersten Grades zu praktizieren, bleibt sie meist ein rein hypothetisches Modell. Es ist

⁵ Transaktionskosten sind die Kosten, die den beteiligten Parteien bei der Aushandlung und Umsetzung einer Vereinbarung entstehen, aber nicht Teil des eigentlichen Kaufpreises sind (Mankiw, 2021, S. 203).

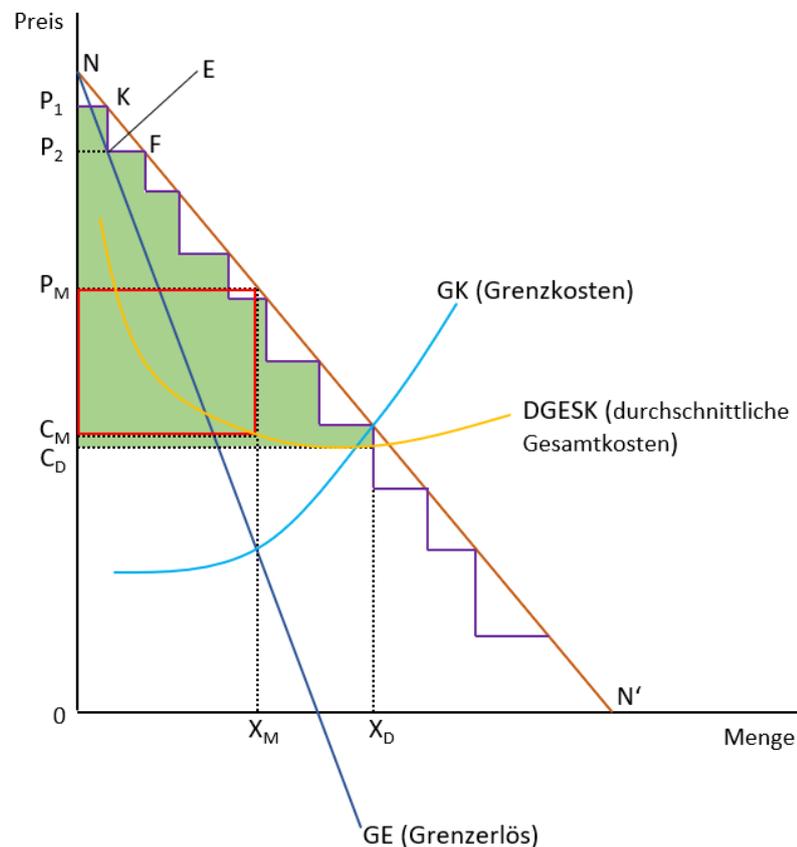
äußerst schwierig und kaum möglich, den genauen Reservierungspreis von einzelnen Kund*innen zu bestimmen. Aus diesem Grund sind praktische Beispiele für Diskriminierung ersten Grades selten. Am ehesten ist es möglich, in dieser Form zu diskriminieren, wenn die finanzielle Situation der eigenen Kund*innen genau bekannt ist. So kommen beispielsweise private Universitäten in den USA mit ihren Studiengebühren einer perfekten Diskriminierung zumindest nahe. Bewerber*innen für Studienplätze bzw. auch deren Eltern müssen gegenüber den Universitäten ihre finanzielle Lage detailliert offenlegen. Neben der finanziellen Situation werden dann auch andere Kriterien, wie z.B. der zu erwartende zukünftige Verdienst der Studierenden oder das Abschneiden bei Eignungstests herangezogen, um gegebenenfalls die Studiengebühren durch die Vergabe von Stipendien, individuellen Nachlässen o.ä. an die Zahlungsfähigkeit bzw. Zahlungsbereitschaft der Bewerber*innen anzupassen (Wiedemann, 2023, S. 94).

Preisdiskriminierung zweiten Grades

Die Diskriminierung *zweiten Grades* ist eine Form der Preisdiskriminierung, bei der ein Unternehmen in der Lage ist, die Nachfrage in verschiedene Segmente einzuteilen und für jedes Segment unterschiedliche Preise zu verlangen. Im Gegensatz zur Preisdiskriminierung ersten Grades, bei der jede*r Kund*in individuell einen Preis erhält, der seinem*ihrem Reservationspreis entspricht, ist die Diskriminierung zweiten Grades weniger genau und teilt die Kund*innen in Gruppen mit ähnlichen Reservationspreisen ein. In diesem Fall hängt der Preis einer Ware oder Dienstleistung häufig von der Abnahmemenge ab. Wenn beispielsweise verschiedene Kund*innen die gleiche Menge eines Gutes kaufen, zahlen sie den gleichen Preis, aber wenn die Menge unterschiedlich ist, kann auch der Stückpreis variieren (z.B. Mengenrabatte). Daher ordnen sich die Kund*innen bei dieser Form der Preisdiskriminierung auch selbst, je nach Zahlungsbereitschaft, einem Segment zu. Preisdiskriminierung zweiten Grades wird aber nicht unbedingt nur über die Quantität der Produkte betrieben. Auch das sogenannte Versioning, bei dem ein grundsätzlich gleiches Produkt in verschiedenen, qualitativ unterschiedlichen Varianten angeboten wird, kann der Preisdiskriminierung zweiten Grades zugeordnet werden. Auch aus ökonomischer Sicht kann Versioning mit der Praxis der Gewährung von Mengenrabatten verglichen werden, da auch hier die Kund*innen ihre Zahlungsbereitschaft selbst bestimmen, indem sie das Produkt in der für sie passenden Variante/Qualität kaufen. Abbildung 3 stellt die Preisdiskriminierung zweiten Grades grafisch dar. Die Standardnachfragekurve ist mit NN' bezeichnet, die Grenzerlöskurve mit NGE und die Grenzkostenkurve mit GK . Ein nicht-

diskriminierendes Unternehmen mit Monopolmacht würde die Grenzkosten mit den Grenzerlösen gleichsetzen, den einheitlichen Preis OP_M für alle Käufer*innen anwenden und dabei OX_M Einheiten verkaufen. Ein Unternehmen, das Preisdiskriminierung praktiziert, kann in der gleichen Situation die Nachfrage in z.B. zehn verschiedene Segmente aufteilen, deren Reihenfolge den absteigenden Reservationspreisen entspricht (siehe Abbildung 3). Es gibt Verbraucher*innen, die bereit sind, P_1K Einheiten zum höchsten erkennbaren Reservationspreis OP_1 zu kaufen, weitere EF -Einheiten aber nur noch zum Preis OP_2 , und so weiter. Der*die Verkäufer*in, der*die für jede Gruppe den ungefähren Reservationspreis verlangt, weitet nun die Produktion aus, bis es keine Segmente mehr gibt, bei denen der Reservationspreis die Grenzkosten übersteigt. Die OX_D -Einheiten werden also (wie in Abbildung 3 dargestellt) zu insgesamt sieben verschiedenen Preisen verkauft. Der Gesamtgewinn ist die abnehmende Spanne des Preises über den durchschnittlichen Stückkosten OC_D für alle verkauften Einheiten (dargestellt durch die grüne Fläche). Dieser Gewinn ist erheblich größer als der Gewinn, der unter einem nicht-diskriminierenden Monopol (rotes Rechteck) erzielt wird (Scherer & Ross, 1990, S. 490; Wiedemann, 2023, S. 95 f.).

Abbildung 3 - Preisdiskriminierung zweiten Grades



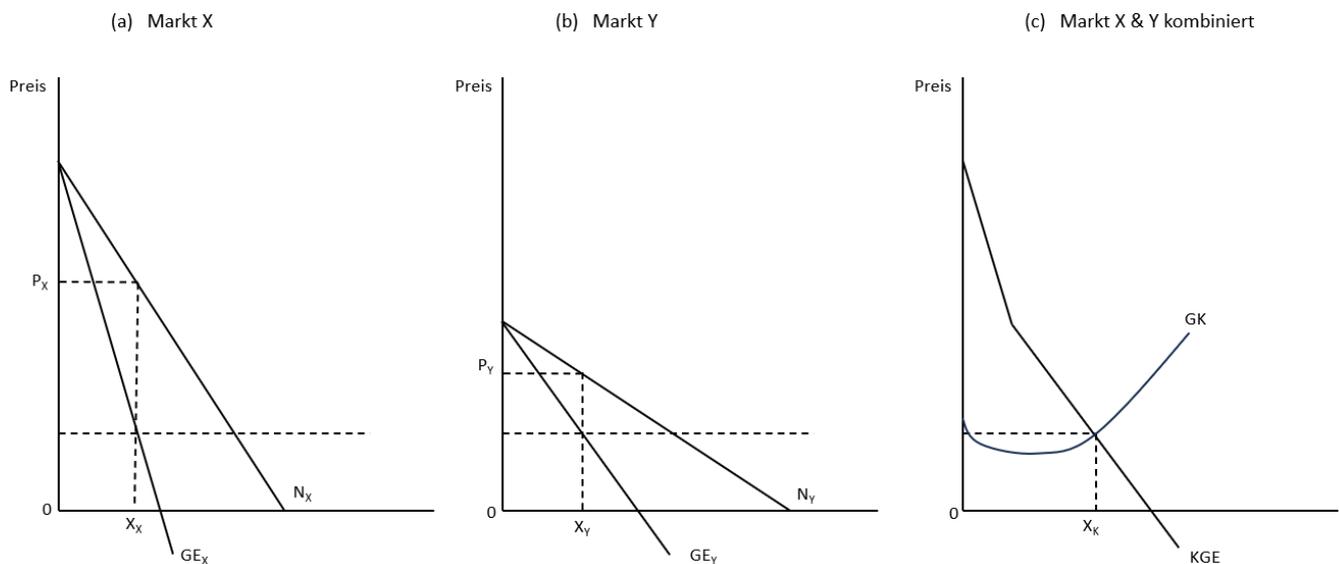
Anmerkung: Eigene Darstellung in Anlehnung an Scherer & Ross (1990, S. 490)

Preisdiskriminierung dritten Grades

Die Diskriminierung *dritten Grades* unterscheidet sich von den anderen Formen der Preisdiskriminierung. Sie ist die in der Praxis am häufigsten anzutreffende Form der Preisdiskriminierung. Bei dieser Art der Diskriminierung teilt der*die Verkäufer*in seine*ihre Kund*innen in zwei oder mehr unabhängige Gruppen ein. Innerhalb dieser Gruppe zahlen alle Personen denselben Preis, unabhängig von der Abnahmemenge eines Gutes. Dabei suchen sich die Kund*innen, im Gegensatz zur Preisdiskriminierung zweiten Grades, nicht den für sie passenden Preis aus. Dieser ist durch die Gruppenzugehörigkeit vorgegeben. Die Gruppenzugehörigkeit wird in der Regel durch verbindende Eigenschaften oder Merkmale der einzelnen Gruppenmitglieder bestimmt und kann auch temporärer Natur sein. Ein klassisches Beispiel sind Rabatte für Studierende und Senior*innen. Hier wird pauschal davon ausgegangen, dass diese beiden Gruppen weniger Geld zur Verfügung haben als z.B. Erwerbstätige. Darauf aufbauend wird angenommen, dass diese Gruppen aufgrund ihrer geringeren finanziellen Mittel einer höheren Preissensibilität unterliegen, weshalb ihnen z.B. im Theater oder Schwimmbad ein Rabatt gewährt wird. Obwohl diese Annahme nicht immer richtig sein muss, da es durchaus auch wohlhabende Pensionist*innen und Studierende gibt. Je grober und pauschaler also eine Gruppe definiert wird, desto stärker weicht der durchschnittlich verlangte Preis vom individuellen Reservationspreis ab. Jede Gruppe verfügt auch über eine eigene Nachfragefunktion. Die Nachfragefunktion gibt an, wie viele Einheiten eines Guts oder Dienstleistung an diese Gruppe zu unterschiedlichen Preisen verkauft werden können. Wenn die Nachfragefunktionen der verschiedenen Gruppen bei gemeinsamen Preisen unterschiedliche Elastizitäten aufweisen, kann der Verkäufer durch unterschiedliche Preise für die verschiedenen Gruppen einen höheren Gewinn erzielen. Angenommen, ein Unternehmen verkauft sein Produkt an zwei verschiedene Gruppen von Kund*innen, Gruppe X und Gruppe Y, und verlangt von beiden Gruppen den gleichen Preis P^* . Die Nachfrageelastizität bei diesem Preis beträgt bei Gruppe X 3,0 und Gruppe Y 6,0, d.h. Gruppe Y reagiert stärker auf Preisänderungen. Aufgrund der unterschiedlichen Elastizitäten kann das Unternehmen mit dem Einheitspreis (P^*) den Gewinn auf diese Weise nicht maximieren. Der Grund dafür ist, dass der Grenzerlös, d.h. der zusätzliche Erlös, der durch den Verkauf einer zusätzlichen Einheit erzielt wird, für die beiden Gruppen unterschiedlich ist. Der Grenzerlös für Gruppe X beträgt die Hälfte des Preises P^* und für Gruppe Y drei Viertel des Preises P^* . Um den Gewinn zu maximieren, sollte das Unternehmen daher einen Teil der Produktion von Gruppe X auf Gruppe Y umverteilen, da der zusätzliche Erlös aus dem Verkauf einer zusätzlichen Einheit an Gruppe Y höher ist als der Verlust aus dem Verkauf einer Einheit

weniger an Gruppe X. Dies kann so lange fortgesetzt werden, bis sich die Preise auf beiden Märkten so verändert haben, dass die Grenzerlöse ausgeglichen sind. Daraus folgt, dass ein Unternehmen, das Diskriminierung dritten Grades praktiziert, den Gewinn maximieren kann, indem es auf dem Markt mit der geringsten Nachfrageelastizität den höchsten Preis und auf dem Markt mit der höchsten Nachfrageelastizität den niedrigsten Preis verlangt (Scherer & Ross, 1990, S. 490 f.; Wiedemann, 2023, S. 99 f.).

Abbildung 4 - Preisdiskriminierung dritten Grades



Anmerkung: Eigene Darstellung in Anlehnung an Scherer & Ross (1990, S. 492)

Abbildung 4 zeigt die traditionelle geometrische Analyse eines Monopols, das eine Diskriminierung dritten Grades auf zwei Märkten, X und Y, ausübt. Die Nachfragekurve (N_X bzw. N_Y) und die Grenzerlöskurve (GE_X bzw. GE_Y) der beiden Märkte sind dabei jeweils in den Diagrammen (a) und (b) dargestellt. Das Ziel ist es nun sicherzustellen, dass der Erlös aus dem Verkauf der letzten Einheit auf Markt X dem Erlös aus dem letzten Verkauf der Einheit auf Markt Y entspricht, d.h. die Grenzerlöse anzugleichen. Um dies zu erreichen werden die Grenzerlöskurven der beiden Märkte horizontal zusammengeführt, wodurch die kombinierte Grenzerlösfunktion KGE in Diagramm (c) entsteht. KGE wird mit den Grenzkosten GK gleichgesetzt, wodurch der gewinnmaximierende Wert OX_K markiert wird. An dem Punkt, an dem GK gleich KGE ist, ist eine horizontale Linie (gestrichelt) zu sehen, deren Schnittpunkt mit den Grenzerlöskurven des jeweiligen Marktes den dort ermittelten Output wiedergibt. Die Preise werden dagegen bestimmt, indem sie aus den entsprechenden

Nachfragefunktionen abgelesen werden, zu denen die gewinnmaximierenden Mengen nachgefragt werden. Somit wird der höhere Preis OP_X im weniger elastischen Markt X und der niedrigere Preis OP_Y im elastischeren Markt Y verlangt (Scherer & Ross, 1990, S. 491).

Die drei Grade der Preisdiskriminierung lassen sich eins zu eins auf die digitale Welt übertragen, auch wenn diese Formen der Preisdiskriminierung durch den Einsatz von Algorithmen und künstlicher Intelligenz immer komplexer und raffinierter werden. So können beispielsweise bei der Preisdiskriminierung ersten Grades unterschiedliche Preise für Flugtickets angeboten werden, je nach Informationen über die finanzielle Situation, den Beruf oder die Reisehistorie der Kund*innen. Ein klassisches Beispiel für die Preisdiskriminierung zweiten Grades sind Softwarepakete, die je nach Funktionsumfang in verschiedenen Versionen (z.B. Basis, Professional, Developer) zu unterschiedlichen Preisen angeboten werden. Als Beispiel für Preisdiskriminierung dritten Grades können die unterschiedlichen Preise für Geschäfts- und Privattarife von Telekommunikationsunternehmen genannt werden (Grochowski et al., 2022, S. 36 f.). Preisdiskriminierung auf der Grundlage von Big Data ist gekennzeichnet durch Preise, die in Echtzeit oder dynamisch angepasst werden, um den größtmöglichen Gewinn zu erzielen und den höchstmöglichen Reservationspreis von den Kund*innen zu erhalten. Mit Hilfe von Algorithmen und Datenanalysen werden unterschiedliche Preise für Produkte oder Dienstleistungen je nach den individuellen Merkmalen der Online-Kund*innen festgelegt. Dies ermöglicht auch eine personalisierte Preisgestaltung im Internet basierend auf Big-Data-Analysen und Algorithmen, die das Verhalten und die Vorlieben der Kund*innen erfassen und auswerten (Townley et al., 2017, S. 683 ff.) .

Belleflamme et al. heben auch die Bedeutung von Datenbroker*innen im Zusammenhang mit Preisdiskriminierung in digitalen Märkten hervor. Diese Unternehmen oder Organisationen sammeln Informationen über Verbraucher*innen, die sie dann verkaufen oder anderen Unternehmen zur Verfügung stellen. Dabei handelt es sich um persönliche Daten, Kauf- und Online-Verhalten, Interessen, Vorlieben und andere relevante Informationen. Unternehmen nutzen diese Databroker*innen als Informationsquelle, um Profile der Verbraucher*innen zu erstellen und dadurch besser zu verstehen, wie sie ihre eigenen Produkte oder Dienstleistungen an unterschiedliche Gruppen von Konsument*innen anpassen können (Belleflamme, Lam & Vergote, 2020, S. 1 ff., S. 8f.).

Preisdiskriminierung in der Praxis

Scherer & Ross (1990, S. 491 ff.) greifen in ihrem Kapitel zur Preisdiskriminierung ein Klassifikationsschema von Fritz Machlup (1955, S. 400–423) auf und modifizieren es, um die in der Realität anzutreffenden Formen der Preisdiskriminierungspraxis zu klassifizieren. Dabei unterscheiden sie drei Hauptformen: Individuelle Diskriminierung, Gruppendifkriminierung und Produktdiskriminierung:

Individuelle Diskriminierung bezeichnet die Ausnutzung der Unterschiede zwischen einzelnen Kund*innen. Ein Beispiel dafür ist das Feilschen, bei dem jedes Geschäft individuell ausgehandelt wird. Dies ist zum Beispiel auf traditionellen Basaren im Nahen Osten oder beim Kauf eines Gebrauchtwagens zu beobachten. Ein anderes Beispiel ist die Praxis des Nachgebens, wenn Verkäufer*innen gegeneinander ausgespielt werden und so gezwungen sind, Abweichungen vom Listenpreis zuzulassen. Auch das Einkommen der Kund*innen spielt bei der individuellen Diskriminierung eine Rolle: Wohlhabende Kund*innen mit unelastischer Nachfrage werden stärker zur Kasse gebeten als weniger wohlhabende, die bei hohen Preisen ihren Konsum einschränken oder einstellen würden. Dies ist häufig bei medizinischen oder juristischen Dienstleistungen der Fall. Schließlich wird auch die Nutzungsdauer bei dieser Form der Diskriminierung berücksichtigt: Kund*innen, die ein Gut oder eine Dienstleistung intensiver nutzen, wird mehr in Rechnung gestellt, auch wenn die Kostenunterschiede kaum spürbar sind. Ein Beispiel hierfür sind die höheren Preise für größere mobile Datenpakete von Telekommunikationsunternehmen (Scherer & Ross, 1990, S. 491 f.).

Unter *Gruppendifkriminierung* wird das Ausnutzen von Unterschieden zwischen verschiedenen Gruppen verstanden. Ein Aspekt davon sind Transportkosten, bei denen Kund*innen, die weiter vom Produktions- oder Lagerstandort entfernt wohnen, höhere Lieferkosten tragen müssen. Ein anderer Aspekt ist die Verdrängung der Konkurrenz, bei der die Preise auf Märkten, die auch von der Konkurrenz bedient werden, systematisch gesenkt werden, um diese zu verdrängen. Eine weitere Praxis der Gruppendifkriminierung besteht darin, überschüssige Waren im Ausland zu reduzierten Preisen anzubieten, um die inländischen Monopolpreise nicht zu drücken. Dabei sind allerdings Anti-Dumping-Regelungen zu beachten (z.B. Verordnung (EU) 2016/1036), dennoch kommt es immer wieder zu Dumping-Vorwürfen. Regionale Unterschiede spielen ebenfalls eine Rolle: In Regionen mit wenig Wettbewerb werden die Preise auf einem hohen Niveau gehalten, im Gegensatz zu Märkten mit starkem Wettbewerb. Die Bindung von Kund*innen ist ein weiterer Faktor: Kund*innen, die viel bei einem Unternehmen

einkaufen, erhalten besondere Rabatte oder Preise. Ein Beispiel hierfür sind Treuepunkteprogramme von Supermärkten. Schließlich spielt auch der Faktor Zeit eine Rolle: Im Einzelhandel ist es üblich, Rabattcoupons an Kundinnen zu verteilen. Wer bereit ist, mehr Zeit für das Sammeln und Einlösen der Coupons aufzuwenden, profitiert eher von Rabatten. Dies gilt insbesondere für Kund*innen mit niedrigem Einkommen, da sie eher bereit sind, Zeit zu investieren, um Geld zu sparen. Die Nachfrageelastizität ist ein weiterer Aspekt der Diskriminierung von Gruppen. Gruppen, die nach Alter, Geschlecht, Beruf usw. unterschieden werden können, haben oft unterschiedliche Reservationspreise oder Nachfrageelastizitäten. Ein Beispiel hierfür sind Flugpreise. Flüge unter der Woche sind in der Regel teurer als Flüge am Wochenende. Der Grund dafür ist, dass bei Geschäftsreisenden, die hauptsächlich unter der Woche fliegen, von einer unelastischen Nachfrage ausgegangen wird. Sie sind daher weniger preissensibel und bereit, höhere Preise zu zahlen. Am Wochenende fliegen dagegen eher Familien, die eine elastische Nachfrage haben. Sie sind preissensitiver und würden bei hohen Preisen ihren Konsum einschränken oder ganz einstellen (Das Europäische Parlament und der Rat der Europäischen Union, 2016a; Scherer & Ross, 1990, S. 492 f.; Wiedemann, 2023, S. 96 f.).

Produktdiskriminierung ist eine Praxis, bei der verschiedene Produkte zu unterschiedlichen Preisen verkauft werden, obwohl sie möglicherweise ähnlich oder sogar identisch sind. Ein Beispiel hierfür ist die Ansprache unterschiedlicher Käuferschichten, bei der Produkte mit einem Premium-Gimmick deutlich teurer verkauft werden als die gleichen Produkte ohne dieses Gimmick, obwohl die Mehrkosten für das Gimmick den höheren Preis nicht rechtfertigen. Unternehmen vertreiben auch physisch homogene Produkte unter verschiedenen Markennamen und verlangen für die bekannteren Marken höhere Preise. Ein weiteres Beispiel der Produktdiskriminierung ist die Reduzierung von Lagerbeständen, bei der zu bestimmten Zeiten Rabatte auf bestimmte Waren angeboten werden, um die Lagerbestände zu reduzieren und den Absatz bei Kund*innen mit geringerer Kaufkraft zu erhöhen. In der Nebensaison werden Dienstleistungen, die bis auf den Zeitraum, in dem sie in Anspruch genommen werden, identisch sind, zu niedrigeren Preisen angeboten. Dies geschieht, um eine gleichmäßigere und rentablere Auslastung zu erreichen und ist ebenfalls eine Form der Produktdiskriminierung. Beispiele hierfür sind Urlaubsorte wie Skigebiete. Bundling ist auch eine Praxis der Produktdiskriminierung, bei der verschiedene Waren und Dienstleistungen zu einem Paket geschnürt und zu einem Preis verkauft werden. Dies geschieht, obwohl die Kund*innen oft nicht alle Komponenten dieses Pakets in Anspruch nehmen wollen. Ein Beispiel hierfür ist der Verkauf von Eintrittskarten für attraktive Veranstaltungen zusammen mit weniger populären

Veranstaltungen im Kulturbereich. Praktiken wie die Forderung nach höheren Transportkosten für wertvollere Güter bei gleichem Transportaufwand oder nach höheren Preis-Kosten-Margen für Güter, die in Bezug auf die Nachfrage unelastisch oder weniger wettbewerbsfähig sind, fallen ebenso in diese Kategorie der Diskriminierung wie die Praxis des Abschöpfens, bei der ein neues (besseres) Produkt zu einem hohen Preis auf den Markt gebracht wird, um den größtmöglichen Umsatz von ungeduldigen Konsument*innen mit hohen Reservationspreisen zu erzielen, und der Preis dann langsam gesenkt wird, um den Markt zu durchdringen (Scherer & Ross, 1990, S. S. 493 f.).

2.2 Technologische Grundlagen für Big-Data-basierte Preisdiskriminierung

Wir leben in einer digitalisierten Welt, in der Unternehmen und Vermarkter*innen alle möglichen Daten über unser Privatleben, unser Einkommen, unsere Vorlieben, unser Ausgabeverhalten, unsere ethnische Zugehörigkeit, unsere Interessen, unsere Hobbys, unseren Gesundheitszustand und eine Vielzahl anderer privater Daten sammeln und im Hinblick auf unser Verhalten, unsere Gewohnheiten, unsere Vorlieben und Abneigungen auswerten. Im Jahr 2022 betrug das Volumen der jährlich generierten oder replizierten Online-Daten 103,66 Zettabyte (= 1,0366e+14 Gigabyte). Waren es 2010 noch 2 Zettabyte, so wird für 2027 ein Wert von 284,3 Zettabyte prognostiziert (Gumbus & Grodzinsky, 2016, S. 118; Statista, 2023).

De Mauro, Greco & Grimaldi (2015, S. 103) definieren Big Data auf der Grundlage ihrer Literaturstudie als Informationen, die sich durch eine so hohe Geschwindigkeit, ein so großes Volumen und eine so große Vielfalt auszeichnen, dass spezielle Technologien und Analysen erforderlich sind, um daraus einen Nutzen zu ziehen. Diese Definition deckt sich mit der von Trübner & Mühlichen (2019, S. 143 f.), die Big Data durch die sogenannten „drei Vs“ charakterisieren: großes Volumen (Volume), hohe Geschwindigkeit (Velocity) und große Vielfalt (Variety) der Daten. Frühere Veröffentlichungen von Big-Data-Definitionen greifen dagegen auf eine negative Abgrenzung zurück, indem sie Big Data als Daten klassifizieren, deren Verarbeitung die Leistungsfähigkeit herkömmlicher Hardware übersteigt (Dittmar, Felden, Finger, Scheuch & Tams, 2016, S. 3).

Durch die Produktion enormer Datenmengen in jüngster Zeit, sei es im Internet durch unzählige Content Creator und Provider oder auch durch die Verbreitung von Sensoren

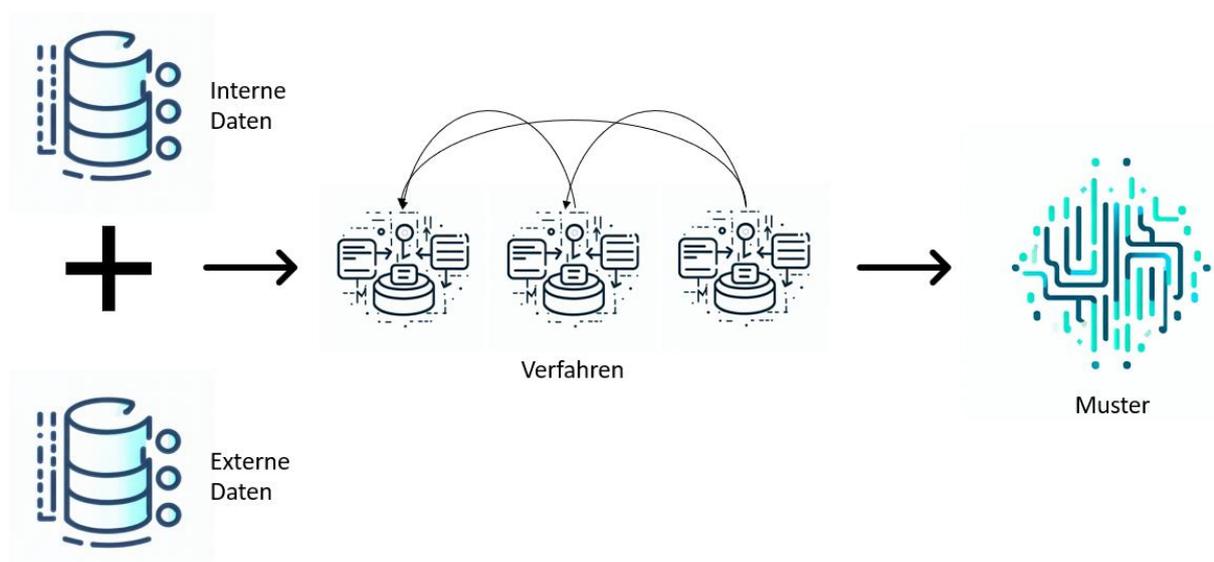
im alltäglichen Leben, ist längst das Zeitalter der miteinander kommunizierenden Dinge (Internet of Things, IoT) angebrochen. Die wachsende Menge an Daten weckt die Erwartung, dass sie wertvolle Informationen enthalten. Diese Informationen könnten beispielsweise für eine personalisierte Ansprache von Kund*innen oder für effizientere Wartungsmaßnahmen genutzt werden, vorausgesetzt, diese riesigen Datenmengen können gezielt analysiert werden. Dies stellt natürlich hohe Anforderungen an eine zielgerichtete Datenverwaltung und -analyse. Infolgedessen gewinnen fortgeschrittene Ansätze und Methoden der Datenanalyse zunehmend an Bedeutung. Diese werden sowohl in der Wissenschaft als auch in der Praxis unter dem Begriff "Business Analytics" oder kurz "Analytics" diskutiert. Ziel dieser Methoden ist es, in großen Datenmengen interessante Muster zu erkennen und Vorhersagen über zukünftige Ereignisse und Entwicklungen zu ermöglichen. Viele dieser Methoden, die unter dem Begriff "Data Mining" zusammengefasst werden, sind bereits seit langem bekannt, wurden aber im Laufe der Jahre weiterentwickelt und verfeinert (Gluchowski, Schieder & Chamoni, 2021, S. 26). Im Gegensatz zur induktiven Statistik geht Data Mining nicht von vorab formulierten Hypothesen aus, sondern nutzt die in den Daten enthaltenen Informationen, um Theorien zu entwickeln. Dieser Ansatz unterstreicht die Bedeutung der datengetriebenen Forschung und ermöglicht es, auf der Grundlage gefundener Muster Hypothesen zu formulieren, die dann mit neuen Daten und bestätigenden Verfahren überprüft werden können. Darüber hinaus kann die bloße Entdeckung von Strukturen von ökonomischem Wert sein, insbesondere bei betrieblichen Anwendungen von Data Mining (Bankhofer & Vogel, 2008, S. 253).

Methoden des Data Mining

Methoden zur Analyse großer Datenmengen können nach ihrem Anwendungszweck klassifiziert werden. Eine grobe Abgrenzung bietet die Unterscheidung zwischen prädiktiven (vorhersagenden) und deskriptiven (beschreibenden) Verfahren. Prädiktive Verfahren können zur Klassifikation oder Regression, deskriptive Verfahren zur Clustering/Segmentierung und Assoziation eingesetzt werden. In einer einfacheren Unterteilung werden deskriptive Verfahren als Verfahren des unüberwachten Lernens ("unsupervised learning") und prädiktive Verfahren als Verfahren des überwachten Lernens ("supervised learning") klassifiziert. Beim unüberwachten Lernen wird nach Mustern oder Ausprägungen in einer oder mehreren (unabhängigen) Variablen gesucht, ohne auf eine erklärende Variable zurückzugreifen. Da hier also keine Prognose für eine

oder mehrere erklärende Variablen durchgeführt wird, entfällt auch ein Vergleich der prognostizierten Werte mit den beobachteten Ausprägungen, es gibt somit auch keine Güte der Prognose, die interpretiert wird. Ziel des unüberwachten Lernens ist es also, Muster oder Strukturen in den Daten zu erkennen, ohne dass dem Algorithmus vorgegeben wird, wonach er suchen soll. Abbildung 5 veranschaulicht diesen Ansatz. Verfahren des unüberwachten Lernens sind die Cluster- und Assoziationsanalyse sowie die Anomalieerkennung (Gluchowski et al., 2021, S. 27).

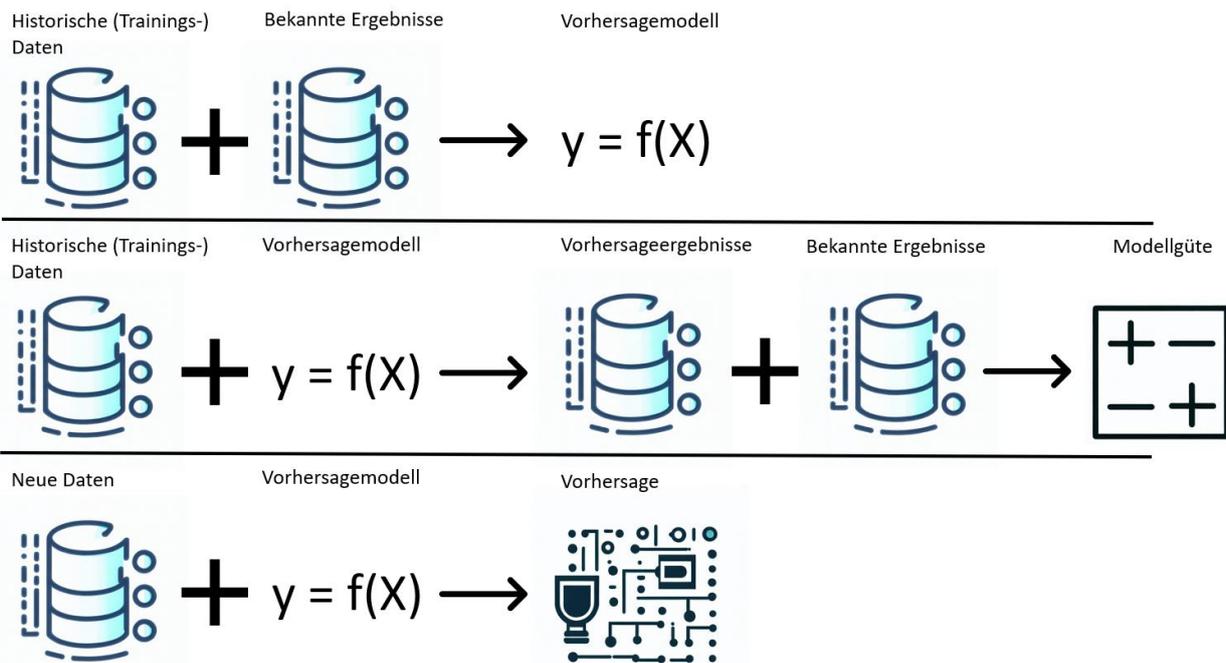
Abbildung 5 - Vorgehensweise bei Verfahren des unüberwachten Lernens



Anmerkung: Eigene Darstellung in Anlehnung an Gluchowski et al. (2021, S. 27). Die Piktogramme wurden mit dem Microsoft Bing Image Creator erstellt.

Im Gegensatz dazu verwenden Verfahren des überwachten Lernens einen vorhandenen Datensatz mit bekannten Ergebnissen und untersuchen den Einfluss einer oder mehrerer (unabhängiger) Eingangsvariablen auf die Ausprägungen einer (in seltenen Fällen auch mehrerer) (abhängiger) Ausgangsvariablen. Das entsprechende Modell „lernt“ also, in welcher Form Ergebnis- bzw. Ausgabewerte aus der gegebenen Kombination von Ausprägungen und Eingangsvariablen bestimmt werden können. Ziel ist es also, einen Algorithmus zu erstellen, der in der Lage ist, auf der Basis des aus den Trainingsdaten gelernten Musters korrekte Vorhersagen oder Klassifikationen für neue, noch nie gesehene Eingabedaten zu treffen. Dieser Ansatz ist in Abbildung 6 dargestellt. Entscheidungsbäume und künstliche neuronale Netze (KNN) können hier als exemplarische Verfahren genannt werden (Gluchowski et al., 2021, S. 27 f.).

Abbildung 6 - Vorgehensweise bei Verfahren des überwachten Lernens



Anmerkung: Eigene Darstellung in Anlehnung an Gluchowski et al.(2021, S. 29). Die Piktogramme wurden mit dem Microsoft Bing Image Creator erstellt.

Empirische Untersuchungen zeigen, dass neben den klassischen statistischen Regressionsmethoden auch Entscheidungsbäume, künstliche neuronale Netze sowie Cluster- und Assoziationsanalysen häufig zur Auswertung umfangreicher Datensätze eingesetzt werden (Gluchowski et al., 2021, S. 28 f.).

Entscheidungsbaumverfahren

Entscheidungsbaumverfahren werden verwendet, um Datenobjekte bestimmten Klassen zuzuordnen. Sie basieren auf historischen Daten, in denen jedes Datenobjekt ein Merkmal hat, das seine Klassenzugehörigkeit angibt. Aus diesen Daten und ihren bekannten Ergebnissen werden Regeln abgeleitet. Diese Regeln werden in Form eines Baumdiagramms dargestellt und können verwendet werden, um neue Datenobjekte zu klassifizieren. Grundsätzlich kann zwischen Klassifikationsbäumen und Regressionsbäumen unterschieden werden. Während Klassifikationsbäume bei nominal skalierten Variablen als abhängige Zielgröße verwendet werden, kommen Regressionsbäume bei quantitativen Variablen zum Einsatz. Da in der Praxis jedoch fortgeschrittene statistische Verfahren für Regressionsmodelle bevorzugt werden, wird im Folgenden der Schwerpunkt auf Klassifikationsbäume gelegt (Bankhofer & Vogel, 2008, S. 273; Gluchowski et al., 2021, S. 30).

Die Erstellung von Klassifikationsbäumen, auch bekannt als Entscheidungsbaum-Lernverfahren, erfolgt in mehreren Schritten. Zunächst werden die Zielvariable (die zu klassifizierende abhängige Variable) und die unabhängigen Variablen identifiziert. Anschließend wird der Datensatz in Trainings- und Testdaten aufgeteilt. Mit den Trainingsdaten wird der Entscheidungsbaum erstellt, mit den Testdaten wird die Genauigkeit der Klassifikation überprüft. Die Trainingsdaten werden schrittweise in immer kleinere Gruppen unterteilt, die in Bezug auf die Zielvariable homogener sind, wobei die Werte der unabhängigen Variablen als Aufteilungskriterien verwendet werden. Jeder Knoten im Baum wird mit einem Homogenitätsmaß bewertet, das angibt, wie homogen die Werte der Zielvariablen in diesem Knoten sind. Ein Knoten gilt als "rein", wenn alle Datenpunkte in diesem Knoten die gleiche Zielvariable haben. Die Aufteilung wird so lange wiederholt, bis die Homogenität in allen Blättern des Baumes einen vorgegebenen Schwellenwert erreicht, d.h. bis sie so homogen wie möglich sind. Dies führt zur Erstellung eines Entscheidungsbaums, der zur Klassifizierung neuer Datenobjekte auf der Grundlage der gelernten Regeln verwendet werden kann. Zur Messung der Homogenität werden spezielle Maße wie die Entropie⁶ oder der Gini-Index⁷ verwendet. Die Entropie bewertet die Homogenität eines Knotens auf der Grundlage der relativen Häufigkeiten der verschiedenen Klassen der Zielvariablen. Ein Knoten ist homogen, wenn die Entropie Null ist, was bedeutet, dass alle Datenpunkte in diesem Knoten zur gleichen Klasse gehören. Die Entropie erreicht ihren höchsten Wert, wenn die Klassen gleich verteilt sind, was bedeutet, dass es keine eindeutige Klassenzugehörigkeit gibt. Nachdem die Berechnung der Homogenität eines Knotens beschrieben wurde, stellt sich die Frage, wie diese Knoten am besten in Teilabschnitte unterteilt werden können. Im Prinzip kann jedes im Datensatz verfügbare Attribut verwendet werden, um den Knoten anhand der Ausprägungen eines dieser Attribute zu unterteilen. Die beste Unterteilung kann anhand des Informationsgewinns erfolgen. Dieser Informationsgewinn wird ermittelt, indem die Entropie des Ausgangsknotens von der gewichteten Summe der Entropien der resultierenden Teilknoten subtrahiert wird. Die Gewichtung berücksichtigt die Anzahl der Datensätze in den Unterknoten im Vergleich zum Ausgangsknoten. Das Attribut mit dem höchsten Informationsgewinn wird für die Unterteilung des Knotens ausgewählt. Dieser Prozess wird für alle Blattknoten im Baum wiederholt. Wenn ein Knoten ein

⁶ Die Entropie kann - einfach ausgedrückt - als ein Maß für die Unordnung in einem System betrachtet werden (Bankhofer & Vogel, 2008, S. 279 ff.; Gluchowski, Schieder & Chamoni, 2021, S. 31).

⁷ Der Gini-Index ist ein Maß für die Wahrscheinlichkeit, dass zwei zufällig ausgewählte Objekte einer Klasse, die sich aus der Aufteilung des Datensatzes nach unabhängigen Variablen ergibt, unterschiedliche Werte für die Zielvariable aufweisen (Bankhofer & Vogel, 2008, S. 276 ff.).

bestimmtes Homogenitätsniveau erreicht, wird er der Klasse zugeordnet, zu der die meisten Datenpunkte in diesem Knoten gehören. Der letzte Schritt bei der Erstellung eines Entscheidungsbaums ist die Überprüfung der Qualität des erstellten Baums anhand von Validierungs- oder Testdaten. Zuvor wurde der Baum mit Trainingsdaten erstellt. Die Qualität des Entscheidungsbaums wird durch die Fehlklassifizierungsrate bewertet, die angibt, wie viele Datensätze das Modell im Verhältnis zur Gesamtzahl der klassifizierten Datensätze falsch klassifiziert hat. Diese Quote ist normalerweise größer als Null, da im Endknoten alle Datensätze derselben Klasse zugeordnet werden, unabhängig von ihrer tatsächlichen Zugehörigkeit. Wenn die Klassifikation der Testdaten ebenfalls den Anforderungen genügt, kann der Entscheidungsbaum zur Klassifikation neuer Datensätze verwendet werden (Bankhofer & Vogel, 2008, S. 273 ff.; Gluchowski et al., 2021, S. 30 ff.).

Zur Veranschaulichung des gesamten Verfahrens wird im Folgenden ein Beispiel in Anlehnung an Bankhofer & Vogel (2008, S. 274 ff.) gegeben.

Ein Unternehmen bietet ein Produkt an, für das es zwei verschiedene Preise gibt, Preis 1 und Preis 2, aber je nach den finanziellen Möglichkeiten der Kund*innen ist nur einer der beiden Preise angemessen. Ohne weitere Informationen liegt die Wahrscheinlichkeit, den Kund*innen den richtigen Preis anzubieten, bei 50%. Ziel ist es nun, aus den vorhandenen Kund*innendaten, die in einem sinnvollen Zusammenhang mit dem Preis stehen müssen, den jeweils richtigen Preis zu ermitteln und damit die Wahrscheinlichkeit zu erhöhen, den Kund*innen das Produkt zum richtigen Preis anzubieten. Die folgende Tabelle 1 zeigt entsprechende Beispieldaten.

Tabelle 1 - Beispieldaten (Entscheidungsbaum)

Nr.	Geschlecht	Alter	Einkommen	Preis
1	Männlich	20	Mittel	1
2	Weiblich	27	Hoch	1
3	Weiblich	48	Hoch	1
4	Männlich	19	Mittel	1
5	Weiblich	52	Mittel	2
6	Männlich	42	Niedrig	2
7	Männlich	61	Mittel	2
8	Weiblich	26	Niedrig	2

Um aus diesen Daten einen Entscheidungsbaum zu konstruieren, werden nun die bedingten Verteilungen des empfohlenen Preises in Abhängigkeit von den verschiedenen

Kund*innenmerkmalen untersucht. Die folgenden Tabellen 2 zeigen die bedingten Verteilungen des empfohlenen Preises bei gegebenem Geschlecht (links), Alter (Mitte) und Einkommen (rechts) der Kund*innen. Beim Alter wird zwischen „ ≤ 30 “ und „ > 30 “ unterschieden.

Tabelle 2 - Bedingte Verteilungen (Entscheidungsbaum)

Nr.	Geschlecht	Preis	Nr.	Alter	Preis	Nr.	Einkommen	Preis
1	Männlich	1	4	19	1	2	Hoch	1
4	Männlich	1	1	20	1	3	Hoch	1
6	Männlich	2	8	26	2	1	Mittel	1
7	Männlich	2	2	27	1	4	Mittel	1
2	Weiblich	1	6	42	2	5	Mittel	2
3	Weiblich	1	3	48	1	7	Mittel	2
5	Weiblich	2	5	52	2	6	Niedrig	2
8	Weiblich	2	7	61	2	8	Niedrig	2

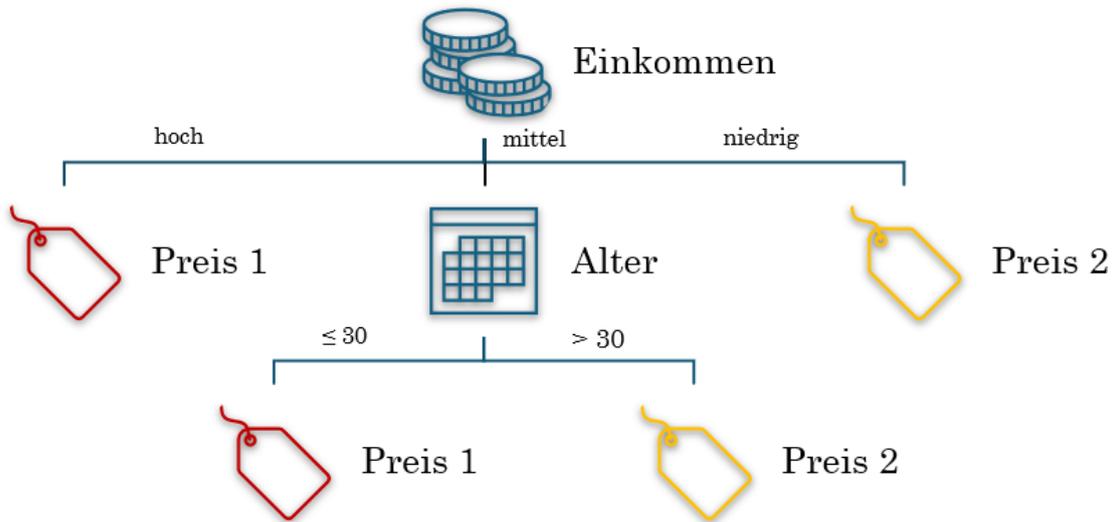
Unter zusätzlicher Berücksichtigung des Alters der Kund*innen in den Fällen, in denen die Kund*innen über ein mittleres Einkommen verfügen, kann der empfohlene Preis eindeutig bestimmt werden, wie aus der folgenden Tabelle 3 ersichtlich ist.

Tabelle 3 - Bestimmung des empfohlenen Preises (Entscheidungsbaum)

Nr.	Einkommen	Alter	Preis
2	Hoch	27	1
3	Hoch	48	1
1	Mittel	20	1
4	Mittel	19	1
5	Mittel	52	2
7	Mittel	61	2
6	Niedrig	42	2
8	Niedrig	26	2

Daraus ergibt sich der in Abbildung 7 dargestellte Entscheidungsbaum

Abbildung 7 - Entscheidungsbaum



Anmerkung: Eigene Darstellung in Anlehnung an Bankhofer & Vogel (2008, S. 275)

Zuerst wird das Merkmal Einkommen berücksichtigt. Ist das Einkommen hoch, wird der höhere Preis (Preis 1) empfohlen. Ist das Einkommen niedrig, wird der niedrigere Preis (Preis 2) empfohlen. Liegt das Einkommen im mittleren Bereich, wird zusätzlich das Alter des*der Kund*in geprüft. Aufgrund dieser Informationen kann auch in diesem Fall der entsprechende Preis angeboten werden.

Anhand dieses einfachen Beispiels kann gezeigt werden, wie wichtig die richtige Auswahl geeigneter Merkmale ist, um die bestmögliche Aufteilung der Objekte im Hinblick auf den Entscheidungsbaum und seine Zielgröße zu erreichen.

Künstliche neuronale Netze

Ein künstliches neuronales Netz (KNN) ist eine Struktur, die aus miteinander verbundenen informationsverarbeitenden Einheiten, den Neuronen, besteht. Diese Neuronen können Informationen austauschen und miteinander kommunizieren. Die Funktionsweise des KNN ist stark an die Informationsverarbeitung im Gehirn von Säugetieren, einschließlich des Menschen, angelehnt. Jedes Neuron im KNN empfängt Eingangssignale (Input) und verarbeitet diese zu Ausgangssignalen (Output) weiter. Im Wesentlichen besteht ein Neuron aus mehreren Komponenten, darunter Eingangssignale, Gewichtungen für diese Signale, einen Schwellenwert und eine Aktivierungsfunktion.

Diese Komponenten werden verwendet, um den Ausgabewert des Neurons auf der Grundlage der Eingabewerte zu berechnen. Zunächst wird also der Gesamtinput (bzw. Netzinput) als Summe der Eingangssignale entsprechend ihrer Bedeutung und Stärke (= Gewichtung) berechnet. Ist dieser Input stark genug, um den Schwellenwert zu überschreiten, wird das Neuron aktiviert und gibt selbst ein Ausgangssignal ab. Wie das Neuron aktiviert wird, hängt von einer Aktivierungsfunktion ab, die verschiedene Formen annehmen kann. Diese Aktivierungsfunktionen können binär, linear oder sigmoid sein. Die erzeugte Ausgabe kann entweder an ein nachgeschaltetes Neuron weitergeleitet oder als Ausgabe des gesamten neuronalen Netzes verwendet werden. Ein KNN kann aus einer oder mehreren Schichten von Neuronen bestehen. Einschichtig ist ein KNN, wenn es nur aus einer Schicht von Neuronen besteht, die jeweils den gesamten Input erhalten und direkt den Output erzeugen. Mehrschichtige KNNs bestehen aus mindestens drei Schichten von Neuronen: einer Eingabeschicht (Input), mindestens einer verdeckten Schicht (Hidden Layer) und einer Ausgabeschicht (Output). Der Unterschied zwischen den beiden KNN-Typen liegt in ihrer Anwendung. Während einschichtige KNNs für eher einfache Aufgaben prädestiniert sind (z.B. Zeichenerkennung), können mehrschichtige KNNs aufgrund ihrer Tiefe und Komplexität für anspruchsvollere Aufgaben und Deep Learning⁸ eingesetzt werden (z.B. Bildverarbeitung, Spracherkennung, etc.). Bei der Datenverarbeitung eines mehrschichtigen, vorwärts gerichteten KNN⁹, wie es häufig in der Klassifikation verwendet wird, werden im ersten Schritt die einzelnen Attributwerte des Datensatzes (= Input) den Neuronen der Eingabeschicht zugewiesen. Diese nehmen die Signale auf und geben sie an die nächste Schicht weiter. In der Regel haben die Neuronen der Input-Schicht keinen Schwellenwert und werden daher bei jedem Input aktiviert. Die Verarbeitung des Inputs erfolgt dann in den Hidden Layers, deren Anzahl vom Aufbau des KNN abhängt. Bei der Klassifikation repräsentiert jedes Neuron der Ausgabeschicht eine Klasse. Wird also ein Neuron der Ausgabeschicht aktiviert, so bedeutet dies, dass der Datensatz zu der entsprechenden Klasse gehört. Neben den starr vorwärts gerichteten neuronalen Netzen gibt es verschiedene Varianten und Alternativen. Beispielsweise gibt es Netze, bei denen Verbindungen zwischen verschiedenen Schichten übersprungen werden, so genannte "shortcut connections". Es gibt auch Netze, bei denen Neuronen innerhalb der gleichen Schicht miteinander verbunden sind. Außerdem gibt es

⁸ Deep Learning ist eine Weiterentwicklung von mehrschichtigen KNNs mit immer tieferen Netzwerkarchitekturen und verbesserter Lernfähigkeit. Ziel ist es, das Verhalten und die Leistungsfähigkeit des menschlichen Gehirns zu simulieren und zu übertreffen (Gluchowski et al., 2021, S. 36 f.; Janiesch, Zschech & Heinrich, 2021, S. 686).

⁹ Bei einem vorwärts gerichteten KNN durchlaufen die Impulse die Struktur vom Input bis zum Output ohne Rückkopplungsschleifen in einer Richtung (Gluchowski et al., 2021, S. 34 f.).

Netze, bei denen die Ausgabe eines Neurons in einem Datensatz als Eingabe für den nächsten Datensatz verwendet wird. Nicht unüblich sind auch Netze mit Rückkopplungen, so genannte "recurrent neural networks" (RNNs). In diesen Netzen gibt es Verbindungen, die zu früheren Zeitpunkten oder Schritten zurückführen, so dass sie eine Art Gedächtnis für frühere Informationen haben. Im Extremfall, wie bei vollständig verbundenen Netzwerken, sind alle Neuronen sowohl vorwärts als auch rückwärts miteinander verbunden. Dies bedeutet, dass Informationen in beide Richtungen fließen können, was für bestimmte Anwendungen nützlich sein kann.

Das Training von KNNs erfolgt häufig mit Hilfe der Back-Propagation-Lernregel, insbesondere bei mehrschichtigen KNNs. Dieser Lernprozess besteht aus drei Phasen:

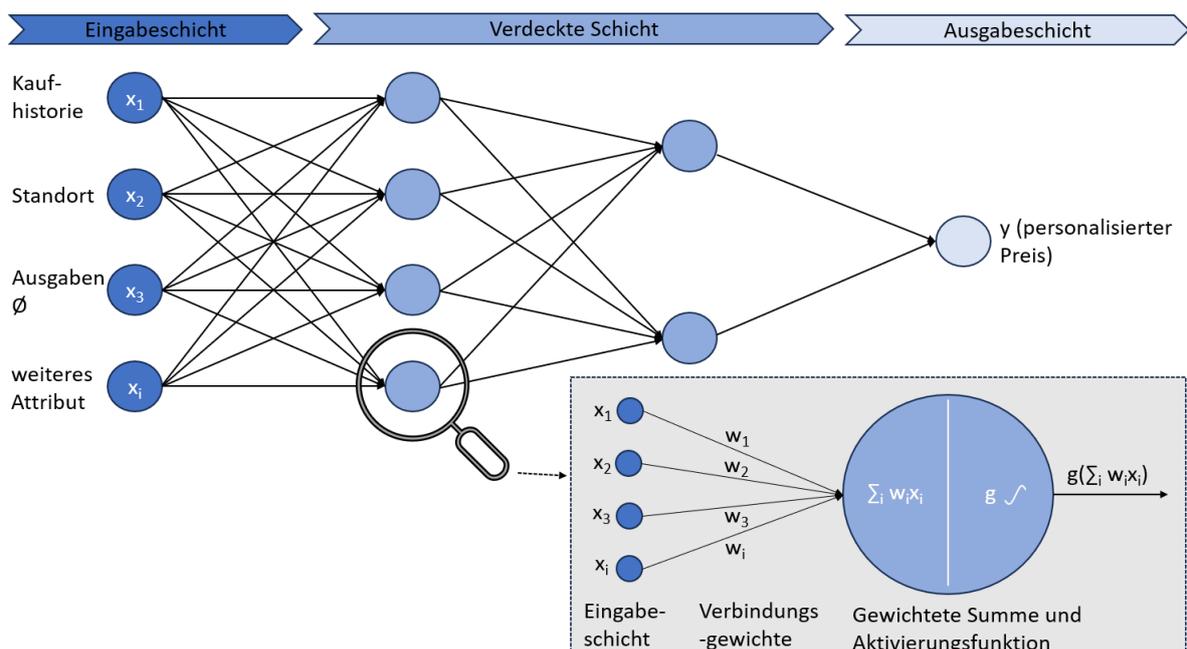
- *Forward Pass*: In dieser Phase werden die Eingangsattribute eines Datensatzes aus der Trainingsmenge in das Netz eingespeist. Die Daten durchlaufen das Netz schichtweise, bis sie die Ausgabeschicht erreichen und einen Ausgabevektor erzeugen.
- *Fehlerberechnung*: In der zweiten Phase wird der Klassifikationsfehler berechnet. Dies geschieht häufig durch die Berechnung der Summe der quadrierten Differenzen zwischen den berechneten Ausgaben und den tatsächlichen Ausgaben für jeden Trainingsdatensatz. Bei Klassifikationsaufgaben kann die Ausgabe mit der tatsächlichen Klassenzugehörigkeit verglichen werden.
- *Backward Pass*: Bei Abweichungen zwischen den berechneten und den tatsächlichen Ausgaben werden in der dritten Phase die Verbindungsgewichte zwischen den Neuronen angepasst. Diese Anpassung erstreckt sich von der Output-Schicht über die Hidden Layers bis hin zur Input-Schicht. Auf diese Weise "lernt" das Netz, indem es die Verbindungsstärken zwischen den Neuronen nach einem vordefinierten Regelwerk anpasst.

Insgesamt handelt es sich um einen iterativen Prozess, bei dem das Netz die Gewichtungen zwischen den Neuronen anpasst, um den Klassifikationsfehler zu minimieren und damit die Fähigkeit des Netzes zur Mustererkennung und Klassifikation zu verbessern. Dieser Lernmechanismus ist ein wesentlicher Bestandteil des Trainings neuronaler Netze für Data Mining und Big Data Analytics (Gluchowski et al., 2021, S. 33 ff.).

In Anlehnung an Nufer & Muth (2022, S. 23 ff.) wird im folgenden Beispiel der Prozess eines KNN im Hinblick auf eine mögliche Anwendung zur Preisdiskriminierung dargestellt.

Angenommen ein Unternehmen verwendet KNNs, um personalisierte Preise für jede*n seiner Kund*innen zu bestimmen. Zunächst werden verschiedene Daten wie z.B. Kaufhistorie, Standort, durchschnittliche Ausgaben etc. der Kund*innen als Eingangswerte (x_i) in die Eingabeschicht des Netzes eingegeben. Diese Eingangswerte werden dann mit einem Verbindungsgewicht (w_i) multipliziert. Diese Gewichte repräsentieren die Bedeutung jeder Eingabe für das Endergebnis und werden im Laufe der Zeit durch den Lernprozess angepasst. Die gewichteten Summen der Eingangswerte ($\sum_i w_i x_i$) werden dann in die verborgene Schicht des Netzes übertragen. Dadurch kann das Netz komplexe Muster in den Daten erkennen. Jedes Neuron in der verdeckten Schicht wendet eine nichtlineare Aktivierungsfunktion (g) auf die gewichtete Summe an. Dadurch werden die Daten weiter transformiert und erhalten zusätzliche Komplexität. Schließlich werden die transformierten Werte an die Ausgabeschicht gesendet, die das Endergebnis erzeugt. In diesem Fall wäre das Ergebnis der personalisierte Preis für den*die Kund*in. Das Netzwerk vergleicht dann das erzeugte Ergebnis mit dem tatsächlichen Wert (z.B. ob der*die Kund*in den Artikel gekauft hat oder nicht) und berechnet den Fehler. Die Verbindungsgewichte werden dann angepasst, um diesen Fehler zu minimieren, und der Prozess wird wiederholt. Durch diesen Prozess kann das neuronale Netz Muster und Beziehungen in den Daten "lernen" und ein Modell erstellen, das zur Vorhersage von Preisen verwendet werden kann. Dies ermöglicht eine personalisierte Preisdiskriminierung, bei der jede*r Kund*in einen individuell angepassten Preis erhält. Abbildung 8 veranschaulicht diesen Prozess.

Abbildung 8 - Künstliches neuronales Netz

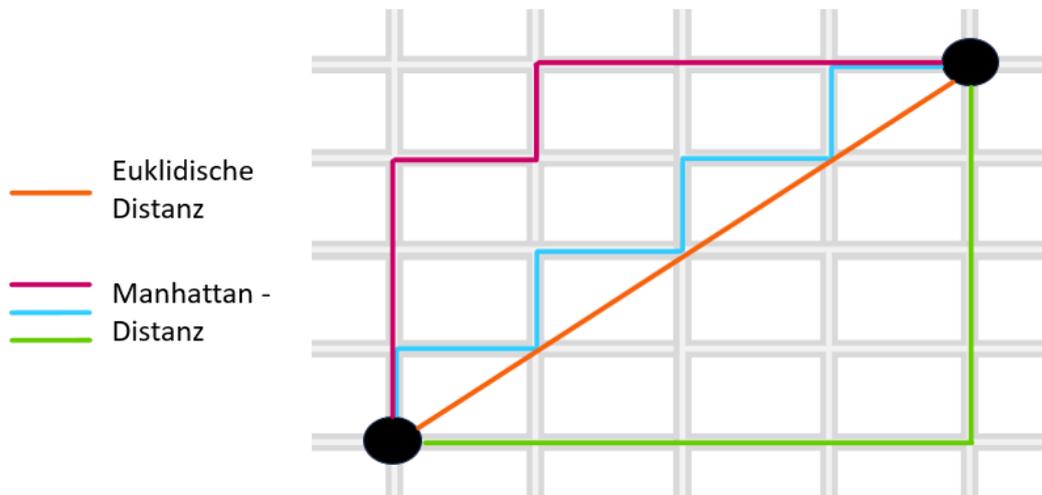


Anmerkung: Eigene Darstellung in Anlehnung an Nufer & Muth (2022, S. 24)

Clusteranalyse

Die Clusteranalyse ist ein Verfahren zur Organisation von Datenobjekten in Segmenten (Cluster), wobei die Objekte innerhalb eines Clusters möglichst ähnlich und die Objekte zwischen den Clustern möglichst verschieden sein sollen. Die Bestimmung der Ähnlichkeit zwischen Datensätzen ist ein wichtiger Schritt in der Clusteranalyse und erfolgt auf Basis der jeweiligen Attributwerte der einzelnen Datenobjekte über die Distanz (Abstand). Je geringer der Abstand zwischen den Objekten ist, desto größer ist die Ähnlichkeit. Daher sollten die Distanzen zwischen den Objekten innerhalb eines Clusters so gering wie möglich und zwischen den Clustern so groß wie möglich sein. Um die Distanz berechnen zu können, ist vorab zu unterscheiden, ob es sich um numerische oder nominale Attribute handelt. Bei numerischen Attributen werden Distanzmaße verwendet, die den absoluten Abstand zwischen Objekten im mehrdimensionalen Raum nutzen, um auf Unähnlichkeiten zu schließen. Bekannte Distanzmaße sind die Euklidische Distanz und die Manhattan-Distanz (auch City-Block-Distanz genannt). Während die Euklidische Distanz den direkten Abstand zwischen zwei Objekten im n-dimensionalen Raum misst, misst die Manhattan-Distanz die Entfernung, die sie im rechten Winkel zurücklegen muss, um von einem Punkt zum anderen zu gelangen. Mit anderen Worten, in einem Raster oder Gitternetz mit vertikalen und horizontalen Linien, ähnlich einem Straßennetz, folgt die Manhattan-Distanz den Straßen, um von einem Punkt zu einem anderen zu gelangen, während die Euklidische Distanz die Luftlinie verwendet, dargestellt in Abbildung 9. Für nominale Merkmale werden Distanzmaße verwendet, die auf der Gleichheit oder Verschiedenheit der Objekte beruhen. Die Abstände werden z.B. durch die Anzahl der nicht übereinstimmenden Merkmalsausprägungen bestimmt. Beispielsweise würden die Objekte A und B mit jeweils unterschiedlichen Ausprägungen in den drei nominalen Attributen Farbe (rot/blau), Größe (groß/klein) und Form (Kreis, Quadrat) eine Distanz von 3 aufweisen. Wäre dagegen ein Attribut identisch, z.B. beide Objekte hätten das Attribut Form = Kreis, und die anderen wären unterschiedlich, so wäre die Distanz zwischen A und B 2 (Bankhofer & Vogel, 2008, S. 165 ff., S. 173; Gluchowski et al., 2021, S. 39 f.).

Abbildung 9 - Euklidische Distanz & Manhattan-Distanz



Anmerkung: Eigene Darstellung

Die Algorithmen, die bei der Clusterbildung verwendet werden, lassen sich in zwei Klassen von Verfahren unterteilen: Die hierarchischen Verfahren und die partitionierenden Verfahren. Die hierarchischen Verfahren lassen sich weiter in divisive und agglomerative Verfahren unterteilen. Divisive Verfahren gehen von einem Cluster aus, der alle Daten umfasst, und unterteilen diesen schrittweise in mehrere Cluster. Die agglomerativen Verfahren, die in der Praxis aus rechentechnischen Gründen eine wesentlich höhere Relevanz besitzen, verfolgen den umgekehrten Ansatz, indem sie jeden Datensatz als eigenen Cluster betrachten und diese schrittweise gruppieren. Zunächst werden dabei die beiden Cluster mit dem geringsten Abstand zu einem neuen Cluster (Partition) zusammengefasst. Dazu wird eine Distanzmatrix erstellt, die die Abstände zwischen allen Clustern enthält. Auf diese Weise kann leicht festgestellt werden, welche beiden Datenpunkte die geringste Distanz zueinander haben. Anschließend wird der Abstand zwischen den resultierenden Clustern berechnet. Es gibt verschiedene Methoden zur Fusionierung von Clustern, darunter Single Linkage (basierend auf dem geringsten Abstand zwischen den Objekten in den Clustern), Complete Linkage (basierend auf dem größten Abstand zwischen den Objekten) und Average Linkage (basierend auf dem Durchschnitt der Abstände zwischen allen Objektpaaren, z.B. nach der Manhattan-Distanz). Beispielsweise werden beim Single Linkage-Verfahren in jedem Iterationsschritt, ausgehend von einzelnen Clustern, zunächst die beiden Cluster mit der geringsten Distanz zueinander gesucht. Diese beiden Cluster werden dann verschmolzen und die Distanz zwischen dem verschmolzenen Cluster und allen anderen Clustern wird

erneut berechnet. Dieser Vorgang wird so lange wiederholt, bis nur noch ein großer Cluster übrigbleibt, der alle Objekte enthält. Der Endzustand, in dem alle Objekte in einem großen Cluster zusammengefasst sind, liefert jedoch keine betriebswirtschaftlich relevanten Informationen. Daher ist es notwendig, auf einen früheren Zustand der Clusterbildung zurückzugreifen. In diesem Fall wird häufig ein Dendrogramm als Visualisierungswerkzeug verwendet, um die verschiedenen Schritte und Fusionen des Clusterbildungsprozesses darzustellen (Bankhofer & Vogel, 2008, S. 197 ff.; Gluchowski et al., 2021, S. 40 f.).

Ein Beispiel in Anlehnung an Gluchowski et al. (2021, S. 41 f.) soll diese Vorgehensweise verdeutlichen.

Ein Online-Shop bietet in seinem Sortiment Artikel aus vier verschiedenen Kategorien an: B = Bücher, F = Filme, M = Musik und S = Spiele. In Tabelle 4 sind Beispieldaten von zwölf Kund*innen (K1, K2, K3, ..., K12) gegeben, deren Kaufverhalten anhand von drei Attributen charakterisiert ist: Artikelanzahl, Durchschnittspreis und gekaufte Artikelgruppen.

Tabelle 4 - Beispieldaten (Clusterbildung)

Kund*innen	Artikelanzahl	Durchschnittspreis	Artikelgruppe
K1	10	15	B, F, S
K2	4	20	B, F, M
K3	12	25	B, S
K4	8	18	B, F, M, S
K5	6	10	B, M, S
K6	3	20	B, F
K7	7	10	B, M
K8	5	15	M, S
K9	15	30	B
K10	6	12	B, F
K11	9	15	B, F
K12	4	18	B, S

Da die obige Tabelle sowohl numerische als auch nominale Attribute enthält, muss ein Distanzmaß verwendet werden, das beide Arten von Attributen berücksichtigt und gleichzeitig normiert. Ein solches Maß ist der Gower-Koeffizient, der eine Möglichkeit bietet, die Ähnlichkeit zwischen Objekten zu quantifizieren. Nach der Zusammenführung von zwei Clustern ergibt sich ein neuer Mittelpunkt des neuen Clusters, der die Mittelwerte der Attribute der im neuen Cluster enthaltenen Objekte repräsentiert. Auf dieser Grundlage können weitere Schritte der Clusteranalyse durchgeführt werden, wie

z.B. die Berechnung der Distanzen zu anderen Clustern oder die Zuordnung neuer Datenobjekte zu bestehenden Clustern. Im vorliegenden Beispiel erfolgt ein Abbruch nach 5 gebildeten Clustern, wie in Tabelle 5 dargestellt.

Tabelle 5 - Ergebnis der Clusterbildung

Cluster	Kund*innen	Interpretation
1	K1, K4	Mehrere Artikel aus allen Kategorien, mittlere Preis (gesamter Unterhaltungsbedarf)
2	K2, K6, K12	Wenige Artikel, vor allem Bücher und Filme (Gelegenheitsbedarf)
3	K3, K9	Viele, teure Bücher und Spiele (Viel-Leser*innen & Viel-Gamer*innen)
4	K5, K7, K8	Mehrere Artikel, aber keine Filme (Musik- und Spieleliebhaber*innen)
5	K10, K11	Moderate Anzahl und moderater Preis für Bücher und Filme

Den Kund*innen aus den einzelnen Clustern könnten nun auf sie zugeschnittene Angebote unterbreitet werden, um eine zielgerichtete Kommunikation zu erreichen und Streueffekte zu vermeiden. Außerdem könnten die Preise je nach Cluster angepasst werden, um die Zahlungsbereitschaft der Kund*innen besser ausnutzen zu können. Dies wäre eine Form der Preisdiskriminierung dritten Grades.

Bei einer großen Anzahl gegebener Datenobjekte ist die Berechnung der Datenmatrix bei hierarchischen Verfahren sehr rechenintensiv. Daher wird in der Praxis häufig auf weniger genaue, aber schnellere Verfahren wie partitionierende Verfahren zurückgegriffen. Im Gegensatz zu hierarchischen Verfahren wird bei partitionierenden Verfahren die Anzahl der Cluster vorgegeben und die Datenobjekte werden schrittweise den Clustern zugeordnet. Dazu wird für jeden Cluster ein Clusterzentrum berechnet, das den Mittelwert der Daten im Cluster darstellt. Die Objekte werden dem Cluster zugeordnet, dessen Zentrum am nächsten liegt. Diese Schritte werden so lange wiederholt, bis sich die Cluster nicht mehr ändern oder ein anderes Abbruchkriterium erreicht wird. Das bekannteste partitionierende Verfahren ist das k-Means Verfahren. Es verwendet die euklidische Distanz, um die Unähnlichkeit von Datenobjekten zu berechnen, weshalb für dieses Verfahren nur quantitative Attribute verwendet werden können. Für das k-Means Verfahren muss zunächst die Anzahl (k) der Cluster festgelegt werden. Anschließend werden die Zentren der einzelnen Cluster zufällig bestimmt und das Datenobjekt mit dem geringsten Abstand zum jeweiligen Zentrum dem entsprechenden Cluster hinzugefügt. Dies wird so lange fortgesetzt, bis alle Objekte einem Cluster zugeordnet sind. Im

nächsten Schritt wird für jeden Cluster ein neues Zentrum berechnet, indem die Mittelwerte der Attributwerte der im Cluster enthaltenen Datenobjekte verwendet werden. Da sich die neuen Zentren nun an anderen Punkten innerhalb der Cluster befinden, beginnt das Verfahren von neuem und die Datenobjekte werden nach und nach, beginnend mit dem nächstgelegenen, den neuen Zentren zugeordnet. Dies wird so lange wiederholt, bis sich in einer Iteration keine Änderungen in der Clusterzusammensetzung mehr ergeben. Das k-Means Verfahren zielt also darauf ab, die Datenpunkte so auf die k Cluster zu verteilen, dass die Summe der Abstände jedes Objektes zum jeweiligen Zentrum minimiert wird (Bankhofer & Vogel, 2008, S. 188 ff.; Gluchowski et al., 2021, S. 43).

Assoziationsanalysen

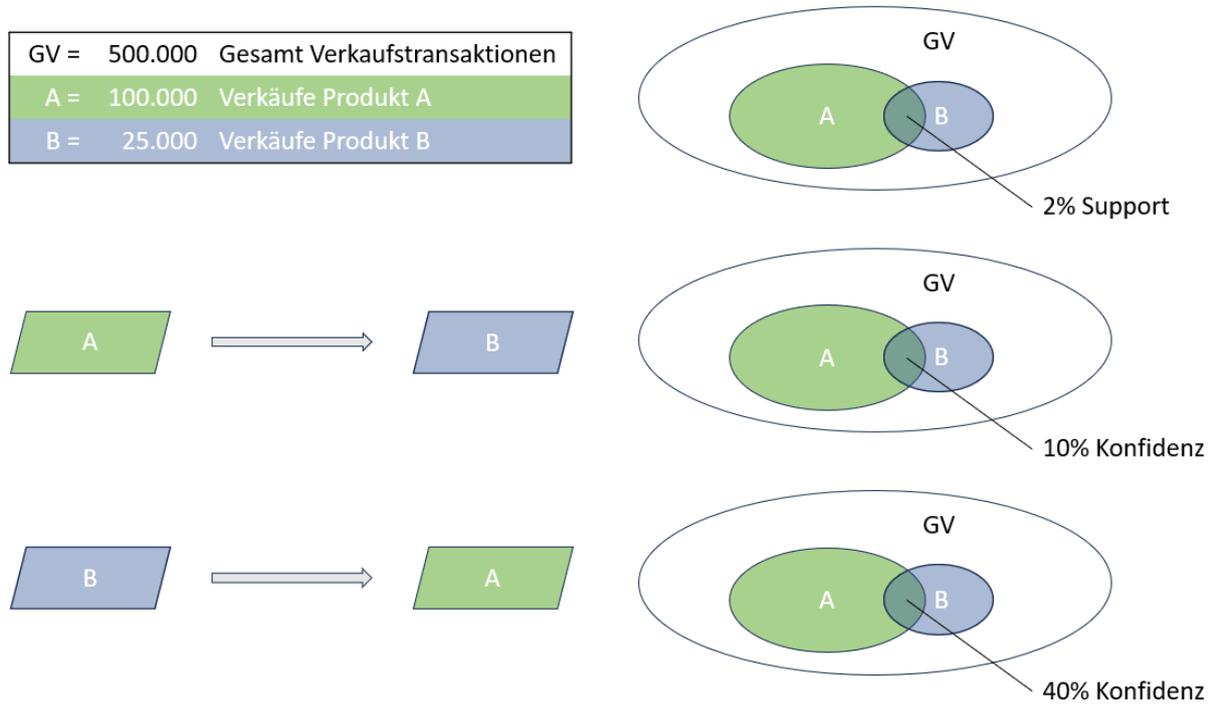
Die Assoziationsanalyse ist ein originäres Verfahren des Data Mining, das dazu dient, Zusammenhänge und Abhängigkeiten in einer Datenbasis aufzudecken. Sie wird häufig in der Warenkorbanalyse, im Bank- und Versicherungswesen und bei der Untersuchung von Wechselwirkungen bei der gleichzeitigen Einnahme verschiedener Medikamente eingesetzt. Die Analyse konzentriert sich auf Ereignisse, die häufig gleichzeitig oder zeitlich versetzt auftreten. Diese Ereignisse werden als parallele und sequenzielle Assoziationen bezeichnet. Ein Beispiel für parallele Ereignisse sind zwei Artikel, die sich in einem Online-Warenkorb befinden und daher fast gleichzeitig gekauft werden. Sequenzielle Assoziationen decken Ereignisse auf, die zeitlich versetzt auftreten, obwohl ein Zusammenhang zwischen ihnen besteht. In der Assoziationsanalyse werden Messgrößen als Items bezeichnet. Diese Items können als Elemente von Mengen oder als einzelne Attributwerte von Datensätzen auftreten. In beiden Fällen wird von Itemsets gesprochen. Ziel der Analyse ist es, aus diesen Itemsets Regeln abzuleiten, die auf Konfidenz und Support basieren. Die Konfidenz gibt an, in welchem Prozentsatz der Fälle Item B auftritt, wenn Item A vorhanden ist. Der Support hingegen gibt an, in welchem Prozentsatz aller Fälle dies der Fall ist. Regeln, die einen Mindestsupport und eine Mindestkonfidenz überschreiten, werden als potenziell interessant betrachtet. Eine Assoziationsregel stellt eine Regel der Form "wenn Item(menge) X, dann Item(menge) Y" dar, wobei die Mengen X und Y disjunkt und echte Teilmengen der Itemmenge I sind. Eine Transaktion erfüllt eine Regel $X \rightarrow Y$ genau dann, wenn alle Items, die durch diese Regel abgebildet werden, auch in der Transaktion vorkommen. Dieser Ansatz ermöglicht es, Muster und Zusammenhänge in großen Datenmengen zu entdecken und zu nutzen. Zur Bewertung einer Assoziationsregel sind mindestens zwei Maße in Form von Support und

Konfidenz erforderlich (Bankhofer & Vogel, 2008, S. 261 f.; Gluchowski et al., 2021, S. 43 f.).

Der Support und die Konfidenz einer Assoziationsregel sollen im Folgenden anhand eines sehr einfachen Zahlenbeispiels in Anlehnung an Bankhofer & Vogel (2008, S. 262 f.) und Gluchowski et al. (2021, S. 44 f.) erläutert werden, das die Preisdiskriminierung betrifft.

Bei insgesamt 500.000 Verkaufstransaktionen in einem Online-Shop wurde 100.000 Mal Produkt A und 25.000 Mal Produkt B gekauft, 10.000 Mal wurden beide Produkte zusammen gekauft. Anhand Abbildung 10 wird deutlich, wie sich der Support für die Kombination der beiden Produkte ergibt. Der Bereich, der den Verkäufen entspricht, in denen Produkt A und Produkt B gemeinsam gekauft wurden, wird dabei ins Verhältnis zur Gesamtzahl aller vorliegenden Verkäufe gesetzt. Dabei ergibt sich in diesem Beispiel ein prozentualer Anteil von 2 % (10.000 von 500.000). Wird nun als mögliche Assoziationsregel die Regel „Wenn Produkt A gekauft wird, dann wird auch Produkt B gekauft“, betrachtet, so ergibt sich eine Konfidenz von 10 %. Anhand des entsprechenden Diagramms wird ersichtlich, dass die Anzahl der Kombinationskäufe von Produkt A und Produkt B jetzt ins Verhältnis zur Gesamtanzahl der Produkt A-Käufe gesetzt wird (10.000 von 100.000). Im Fall einer umgekehrten Assoziationsregel resultiert entsprechend eine Konfidenz von 40 %, da jetzt die Gesamtanzahl der Produkt B-Käufe die Bezugsgröße darstellt (10.000 von 25.000).

Abbildung 10 - Beispiel zu Support und Konfidenz (Assoziationsanalyse)



Anmerkung: Eigene Darstellung in Anlehnung an Bankhofer & Vogel (2008, S. 263)

Damit würde in diesem Beispiel die zweite dargestellte Regel ein Muster in den Daten aufzeigen, das der Online-Shop exemplarisch wie folgt nutzen könnte: Durch die Anordnung von Produkt A in Sichtweite von Produkt B könnte gegebenenfalls eine Zunahme dieser Verbundkäufe erreicht werden. Des Weiteren sollte nicht bei beiden Produkten gleichzeitig ein Sonderangebot gemacht werden, sondern – sofern derartige Angebote überhaupt geplant sind – idealerweise nur bei Produkt B. Diese Angebote könnten dann auch zu einer absoluten Zunahme der Produkt A-Verkäufe führen.

Dieses Beispiel zeigt, wie die Assoziationsanalyse zur Entwicklung von Strategien für die Preisdiskriminierung genutzt werden kann, indem sie Muster im Kaufverhalten der Kund*innen aufdeckt.

2.3 Rechtliche Rahmenbedingungen und ökonomische Auswirkungen

Rechtliche Rahmenbedingungen für Big-Data-basierte Preisdiskriminierung

Gemäß Artikel 21 “Nichtdiskriminierung” der Charta der Grundrechte der Europäischen Union (Europäische Union, 2012a) sind “Diskriminierungen, insbesondere wegen des Geschlechts, der Rasse, der Hautfarbe, der ethnischen oder sozialen Herkunft, der genetischen Merkmale, der Sprache, der Religion oder der Weltanschauung, der politischen oder sonstigen Anschauung, der Zugehörigkeit zu einer nationalen Minderheit, des Vermögens, der Geburt, einer Behinderung, des Alters oder der sexuellen Ausrichtung” verboten. Verboten ist auch “jede Diskriminierung aus Gründen der Staatszugehörigkeit”.

Gemäß Art. 6 Abs. 1 lit. d und Art. 7 Abs.5 lit. c der “Richtlinien zur Auslegung und Anwendung der Richtlinie 2005/29/EG des Europäischen Parlaments und des Rates über unlautere Geschäftspraktiken von Unternehmen gegenüber Verbrauchern im Binnenmarkt” (UGPRL) (Das Europäische Parlament und der Rat der Europäischen Union, 2005) steht es Gewerbetreibenden grundsätzlich frei, die Preise für ihre Produkte festzulegen, solange die Verbraucher*innen ausreichend über die Gesamtkosten oder, falls dies aufgrund der Art des Produkts nicht möglich ist, über die Berechnung der Gesamtkosten informiert werden. Um jedoch einer dadurch möglichen Preisdiskriminierung entgegenzuwirken, gibt es Vorschriften, die eine Preisdiskriminierung verbieten können. In den Leitlinien zur UGPRL (Das Europäische Parlament und der Rat der Europäischen Union, 2021a) wird Preisdiskriminierung als eine Praxis definiert, bei der ein*e Gewerbetreibende*r unterschiedlichen Verbraucher*innen oder Verbraucher*innengruppen unterschiedliche Preise für dieselben Waren oder Dienstleistungen in Rechnung stellt. In diesem Zusammenhang verweisen die Leitlinien auf Artikel 20 der Dienstleistungsrichtlinie (Das Europäische Parlament und der Rat der Europäischen Union, 2006). Dieser enthält ein generelles Verbot der Preisdiskriminierung aufgrund der Staatsangehörigkeit und des Wohnsitzes. Zulässig sind jedoch Unterschiede in den Zugangsbedingungen, die unmittelbar durch objektive Kriterien gerechtfertigt sind. Mehrere sektorspezifische EU-Rechtsvorschriften verbieten ausdrücklich direkte und indirekte Preisdiskriminierung aufgrund der Staatsangehörigkeit oder des Wohnsitzes der Endkund*innen oder aufgrund des Ortes der Niederlassung von Speditionsunternehmen oder von Ticket-Anbieter*innen in der

Union. Diese Verbote gelten beispielsweise für den Luftverkehr¹⁰, den Seeverkehr¹¹, den Schienenverkehr¹² und den Bus- und Bahnverkehr¹³.

Die UGPRL erlaubt es Unternehmen, ihre Preise auf der Grundlage von Online-Überwachung und Profiling zu personalisieren. Gemäß Art. 6 Abs. 1 lit. ea der UGPRL, der durch die Richtlinie (EU) 2019/2161 hinzugefügt wurde, müssen sie die Verbraucher*innen jedoch darüber informieren, ob der Preis auf der Grundlage einer automatisierten Entscheidungsfindung personalisiert wurde. Personalisierte Preise und Angebote können auch mit anderen unlauteren Geschäftspraktiken kombiniert werden. Dies ist beispielsweise der Fall, wenn Gewerbetreibende im Zusammenhang mit datengestützter Personalisierung einen unzulässigen Einfluss auf Verbraucher*innen ausüben, wie in Artikel 8 („Aggressive Geschäftspraktiken“) und Artikel 9 („Belästigung, Nötigung und unzulässige Beeinflussung“) der UGPRL beschrieben. Hinzu kommt, dass Gewerbetreibende, die ihre Preise personalisieren, indem sie personenbezogene Daten von Konsument*innen verwenden, auch die DSGVO (Datenschutz-Grundverordnung) und die Datenschutzrichtlinie für elektronische Kommunikation einhalten müssen.

Artikel 13 der Datenschutzrichtlinie für elektronische Kommunikation (Das Europäische Parlament und der Rat der Europäischen Union, 2009) schreibt vor, dass automatische Anrufsysteme, Faxgeräte oder elektronische Post nur für Zwecke der Direktwerbung verwendet werden dürfen, wenn die Teilnehmer*innen zuvor ihre Einwilligung gegeben haben. Widerspricht eine betroffene Person der Verarbeitung ihrer personenbezogenen Daten zu Direktmarketingzwecken, muss der*die Verantwortliche, gemäß Artikel 21 der DSGVO, die Direktwerbung einstellen. Die Artikel 12 bis 14 der Datenschutz-Grundverordnung enthalten „Informationspflichten in Bezug auf die Verarbeitung personenbezogener Daten“, darunter auch das Recht auf Bereitstellung „aussagekräftiger Informationen über das Bestehen einer automatisierten Entscheidungsfindung“. Gemäß Artikel 22 der DSGVO hat eine „betroffene Person“ das Recht „nicht einer ausschließlich auf einer automatisierten Verarbeitung — einschließlich Profiling — beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder

¹⁰ Vgl. VO (EG) Nr. 1008/2008 Art. 23 Abs. 2 (Das Europäische Parlament und der Rat der Europäischen Union, 2008)

¹¹ Vgl. VO (EG) Nr. 1177/2010 Art. 1 5 Abs. 2 (Das Europäische Parlament und der Rat der Europäischen Union, 2010)

¹² Vgl. VO (EG) Nr. 2021/782 Art. 5 (Das Europäische Parlament und der Rat der Europäischen Union, 2021b)

¹³ Vgl. VO (EG) Nr. 181/2011 Art. 4 Abs. 2 (Das Europäische Parlament und der Rat der Europäischen Union, 2011)

sie in ähnlicher Weise erheblich beeinträchtigt“ (Das Europäische Parlament und der Rat der Europäischen Union, 2016b).

Eine Studie (2022) im Auftrag des “Committee on the Internal Market and Consumer Protection” (IMCO) zur Untersuchung der personalisierten Preisgestaltung unterstreicht die Möglichkeiten der Personalisierung von Preisen für Unternehmen, betont aber auch die Verpflichtung der Händler*innen gemäß Art. 6 Abs. 1 lit. ea der Verbraucherrechterichtlinien, offenzulegen, ob personalisierte Preise auf der Grundlage automatisierter Entscheidungsfindung angewendet werden. Die Studie betont jedoch, dass sich diese Verpflichtung nur auf den Fernabsatz und darauf beschränkt, ob personalisierte Preise angewendet werden, und nicht auf die Art und Weise, in der sie angewendet werden (European Parliament. Directorate General for Internal Policies of the Union., 2022, S. 28 ff.).

Nach Townley et al. ist der rechtliche Rahmen für Big-Data-basierte Preisdiskriminierung durch das EU-Wettbewerbsrecht geregelt. Dies gilt insbesondere für Artikel 102 des Vertrags über die Arbeitsweise der Europäischen Union. Artikel 102 verbietet „die missbräuchliche Ausnutzung einer beherrschenden Stellung auf dem Binnenmarkt oder auf einem wesentlichen Teil desselben durch ein oder mehrere Unternehmen, soweit dies dazu führen kann, den Handel zwischen Mitgliedstaaten zu beeinträchtigen“ (Europäische Union, 2012b). Die Autoren betonen, dass das vorrangige Ziel des EU-Wettbewerbsrechts darin besteht, das Wohl der Verbraucher*innen zu schützen und den Wettbewerb zu fördern. Sie argumentieren, dass Big-Data-basierte Preisdiskriminierung nicht notwendigerweise per se gegen das EU-Wettbewerbsrecht verstößt, da sie je nach den Umständen sowohl positive als auch negative Auswirkungen auf das Wohl der Verbraucher*innen haben kann. Die rechtliche Beurteilung hängt daher von der jeweiligen Marktsituation ab, die die Auswirkungen der Preisdiskriminierung beeinflusst. EU-Verordnungen und -Beschlüsse gelten ab dem Tag ihres Inkrafttretens automatisch in allen EU-Mitgliedstaaten und müssen national umgesetzt werden (Europäische Kommission, 2023; Townley et al., 2017, S. 721 ff.).

Townley et al. betonen auch die Bedeutung des Verbraucher*innenschutzes und der Datensicherheit und plädieren dafür, dass Verbraucher*innen die Wahl haben sollten, ob sie ihre Informationen offenlegen oder schützen möchten. Auf diese Weise können Verbraucher*innen ihre Interessen wahren, sich vor möglicher Ausbeutung durch Unternehmen schützen und sicherstellen, dass ihre Privatsphäre und Sicherheit gewährleistet ist. Sie plädieren also dafür, den Verbraucher*innen mehr Kontrolle und

Wahlmöglichkeiten zu geben, was sie als wichtigen Schutzmechanismus in einer Welt des Online-Handels und der Datenanalyse ansehen (Townley et al., 2017, S. 744 ff.). In der IMCO-Studie gehen die Autoren davon aus, dass sich personalisierte Preise in Zukunft häufen werden und dass solche Praktiken größtenteils zum Nachteil der Verbraucher*innen sind und daher neue Regulierungen erforderlich sind. So schlägt die Studie vor, Preisdiskriminierung ersten Grades zu verbieten, wenn sie zu einem höheren als dem regulären Preis führt. Während (Preis-) Diskriminierungen z.B. aufgrund des Geschlechts oder der ethnischen Herkunft, die unter die Antidiskriminierungsgesetze fallen, generell verboten sind, sind bestimmte andere Kriterien, wie z.B. der Gesundheitszustand oder Angriffspunkte wie z.B. Ängste, die ausgenutzt werden können, von diesen Verboten nicht erfasst und sollten laut der Studie in zukünftige Überlegungen zur Regulierung einbezogen werden. Für die Zukunft schlägt die Studie auch vor, die Informationspflicht darüber, ob personalisierte Preise entstehen, auf die Art und Weise, wie sie entstehen, auszuweiten (European Parliament. Directorate General for Internal Policies of the Union., 2022, S. 12 ff. S. 38 ff.).

In Österreich schützt § 30 Gleichbehandlungsgesetz vor Diskriminierung aufgrund der ethnischen Zugehörigkeit und des Geschlechts. Gemäß §31 und §32 liegt eine unmittelbare Diskriminierung vor, wenn eine Person "beim Zugang zu und bei der Versorgung mit Gütern und Dienstleistungen, die der Öffentlichkeit zur Verfügung stehen [...] eine weniger günstige Behandlung erfährt, als eine andere Person erfährt, erfahren hat oder erfahren würde" (Rechtsinformationssystem des Bundes (RIS), 2023a). Darüber hinaus gibt es auf Länderebene eigene Antidiskriminierungsgesetze, z.B. für Wien (Rechtsinformationssystem des Bundes (RIS), 2023b), die die Diskriminierung von natürlicher Personen aufgrund der „ethnischen Zugehörigkeit, der Religion, der Weltanschauung, einer Behinderung, des Alters, der sexuellen Orientierung, der Geschlechtsidentität und des Geschlechts, insbesondere auch auf Grund von Schwangerschaft und Elternschaft, sowie die Anstiftung einer Person zu solchen Diskriminierungen“ verbieten. Das Bundesgesetz gegen den unlauteren Wettbewerb (UWG) regelt die grundlegenden Materien des österreichischen Wettbewerbsrechts. Das UWG verbietet unlautere Geschäftspraktiken, die das wirtschaftliche Verhalten von Durchschnittsverbraucher*innen wesentlich beeinflussen und damit den Erfordernissen der beruflichen Sorgfalt widersprechen. Weiters verbietet das UWG Geschäftspraktiken, die den Wettbewerb zum Nachteil von Unternehmen nicht nur unerheblich beeinflussen. §1 Abs. 3 UWG definiert unlautere Geschäftspraktiken also solche, die entweder aggressiv oder irreführend sind. Nach §1a UWG gilt eine Geschäftspraktik als aggressiv, „wenn sie

geeignet ist, die Entscheidungs- oder Verhaltensfreiheit des Marktteilnehmers in Bezug auf das Produkt durch Belästigung, Nötigung, oder durch unzulässige Beeinflussung wesentlich zu beeinträchtigen und ihn dazu zu veranlassen, eine geschäftliche Entscheidung zu treffen, die er andernfalls nicht getroffen hätte.“ Nach §2 UWG gilt eine Geschäftspraktik als irreführend, „wenn sie unrichtige Angaben (§39) enthält oder sonst geeignet ist, einen Marktteilnehmer in Bezug auf das Produkt über einen oder mehrere der folgenden Punkte derart zu täuschen, dass dieser dazu veranlasst wird, eine geschäftliche Entscheidung zu treffen, die er andernfalls nicht getroffen hätte“ (Rechtsinformationssystem des Bundes (RIS), 2023c).

Ökonomische Auswirkungen von Big-Data-basierter Preisdiskriminierung

Scherer & Ross (1990) betonen in ihrem Standardwerk zum Thema Preisdiskriminierung, dass das Ziel von Preisdiskriminierung die Steigerung von Gewinn und Wohlfahrt ist. Jedoch heben sie hervor, dass Preisdiskriminierung sowohl positive als auch negative Auswirkungen haben kann. Die zentrale Frage gemäß Scherer & Ross ist, ob Preisdiskriminierung zur Effizienz der Ressourcenallokation beiträgt. In natürlichen Monopolmärkten, in denen es nur eine*n Anbieter*in gibt, führt die Preisdiskriminierung dazu, dass der*die Anbieter*in seine*ihre Produktionskapazität besser auslasten kann. Dies wiederum kann zu niedrigeren Durchschnittskosten führen und die Effizienz steigern. Ein klassisches Beispiel ist die Stromversorgung, bei der Großabnehmer*innen niedrigere Preise zahlen und somit dazu beitragen, dass die Stromerzeugung optimal ausgelastet wird. Preisdiskriminierung führt jedoch nicht immer zu Effizienzsteigerungen. In einigen Fällen kann sie zu einer unnötigen Ressourcenverschwendung führen. In oligopolistischen Märkten können Preiskämpfe auftreten, wenn Anbieter*innen versuchen, Kund*innen mit diskriminierenden Preisen anzulocken. Dies kann zu einer Verschwendung von Ressourcen führen, da die Unternehmen in einen ruinösen Wettbewerb eintreten. Wenn Unternehmen beginnen, diskriminierende Preise anzubieten, um Kund*innen zu gewinnen, kann dies zu einem Verlust an Preisdisziplin führen und letztendlich die Kollusion zwischen den Unternehmen gefährden. Dies kann in weiterer Folge zu weiteren Preisschwankungen führen und auch die Marktstabilität beeinträchtigen. Ein weiterer von Scherr & Ross angesprochener Punkt ist die Einkommensumverteilung, die durch Preisdiskriminierung entsteht. Da Preisdiskriminierung unterschiedliche Preise für verschiedene Gruppen von Kund*innen bedeutet, zahlen einige Verbraucher*innen mehr für dasselbe Gut oder dieselbe Dienstleistung als andere. Ob dies als fair oder unfair angesehen wird, ist

subjektiv und erfordert Werturteile. Die Einkommensumverteilung kann dazu führen, dass einige Verbraucher*innen benachteiligt werden, während andere bevorzugt werden. Scherer & Ross diskutieren verschiedene Ansätze zur Preisbildung und zur Vermeidung von Preisdiskriminierung. Ein Ansatz ist das Basing-Point-System, bei dem ein Produktionsstandort als Referenzpunkt festgelegt wird und alle Preise von diesem Standort aus berechnet werden. Dies kann dazu beitragen, unabhängige Preisinitiativen zu verhindern und die Disziplin des Oligopols aufrechtzuerhalten. Die Frage, ob Preisdiskriminierung positive oder negative Auswirkungen auf die Wohlfahrt hat, lässt sich nicht so einfach beantworten. Es handelt sich um ein komplexes und nuanciertes Thema, das sich nicht einfach verallgemeinern lässt. Die Folgen hängen von den Marktbedingungen, den spezifischen Formen der Diskriminierung und den individuellen Werturteilen ab. Daher bedarf jeder Fall einer differenzierten Analyse, um die Auswirkungen auf Effizienz, Wettbewerb und Einkommensverteilung angemessen beurteilen zu können (Scherer & Ross, 1990, S. 494 ff.)

Auch Steinberg (2020) und Townley et al. (2017) stellen in ihren Arbeiten zu Big-Data-basierter Preisdiskriminierung fest, dass die ökonomischen Auswirkungen von Preisdiskriminierung komplex und situationsabhängig sind und sich sowohl negativ als auch positiv auf Wohlfahrt und Wettbewerb auswirken können. Nach ihrer Analyse hängen die Auswirkungen von Faktoren wie der Marktstruktur, der Nachfrageelastizität und der Art der Preisdiskriminierung ab (Steinberg, 2020, S. 114 f.; Townley et al., 2017, S. 688 ff.)

Townley et al. erläutern, dass die Auswirkungen von Preisdiskriminierung in Märkten mit asymmetrischem Wettbewerb, in denen Preisdifferenzierung beispielsweise auf Markenpräferenzen und Wechselkosten¹⁴ beruht (best-response asymmetry), tendenziell positiver sein können. Hier kann Preisdiskriminierung den Wettbewerb intensivieren und insgesamt zu Vorteilen für die Verbraucher*innen führen, die von differenzierten Angeboten profitieren. In Märkten mit symmetrischem Wettbewerb, in denen Preisdifferenzierung beispielsweise auf Suchkosten¹⁵ basiert (best-response symmetry), kann Preisdiskriminierung dagegen den Wettbewerb eher schwächen. Dies kann zu

¹⁴ Wechselkosten sind die Kosten, die entstehen, wenn ein*e Kund*in von einem*einer Anbieter*in zu einem*einer anderen wechselt. Sie können dabei von physischen Investitionen (z. B. der Kauf von Ausrüstung), über Informationsinvestitionen (z. B. das Erlernen der Produktnutzung), bis hin zu psychologischen Investitionen (z. B. die emotionale Bindung an eine Marke) reichen (Klemperer, 1995, S. 517 f.).

¹⁵ Suchkosten sind die Kosten, die bei der Suche nach Informationen entstehen und sich hauptsächlich in Zeit ausdrücken. Sie sind nicht für alle Konsument*innen gleich, da der Wert von Zeit je nach Einkommen und individuellen Umständen variieren kann (Stigler, 1961, S. 216).

Nachteilen für die Verbraucher*innen führen, da sie möglicherweise nicht in der Lage sind, die besten verfügbaren Angebote zu finden (Townley et al., 2017, S. 688 ff., S. 724).

Die Effektivität von Preisdiskriminierung hängt außerdem von den Informationen ab, die Unternehmen über die Zahlungsbereitschaft der Kund*innen haben. Und zwar von diesen Informationen, die im Internet zu einem großen Teil aus dem beobachteten Verhalten der Konsument*innen mittels Tracking-Tools gewonnen werden. Dies verdeutlicht die Bedeutung von asymmetrischem bzw. symmetrischem Wettbewerb auf die Auswirkungen von Preisdiskriminierung (Steinberg, 2020, S. 100 ff.; Townley et al., 2017, S. 697).

In einem Szenario, in dem ein Unternehmen ein absolutes Monopol hat und Preisdiskriminierung betreibt, wird erwartet, dass die Konsumentenrente im Vergleich zu derjenigen bei einheitlichen Preisen reduziert ist, d.h. die Verbraucher*innen also im Durchschnitt weniger Nutzen haben. Ein Monopolist kann seine Marktmacht ausnutzen und so die Verbraucher*innen individuell ausbeuten, indem er insgesamt höhere Preise verlangt. Townley et al. weisen jedoch darauf hin, dass die Gesamtauswirkungen auf die Konsument*innen von der Art der Nachfragekurve abhängen können. In Märkten mit einer flachen Nachfragekurve, in denen die Verbraucher*innen relativ preiselastisch sind, können Preisänderungen aufgrund von Preisdiskriminierung dazu führen, dass die Verbraucher*innen insgesamt besser abschneiden. Dies ergibt sich daraus, dass Preisdiskriminierung in diesem Fall die Preise für einige Verbraucher*innen senkt, was insgesamt zu einer höheren Konsumentenrente führt. Gemäß Townley et al. gibt es aber nur einen begrenzten theoretischen Zusammenhang zwischen der (ex ante) Marktmacht eines Unternehmens und den daraus resultierenden Auswirkungen auf die Wohlfahrt der Verbraucher*innen. Die Tatsache, dass ein Unternehmen eine marktbeherrschende Stellung innehat, bedeutet nicht zwangsläufig, dass dies zu bestimmten Ergebnissen für die Verbraucher*inne führt (Townley et al., 2017, S. 688 ff.).

Dubé & Misra (2017) beispielsweise untersuchen in ihrer Fallstudie die Auswirkungen von personalisierter Preisdiskriminierung mittels maschinellen Lernens in einem Monopol. Sie kommen zu dem Schluss, dass die Gewinne des untersuchten marktbeherrschenden Unternehmens, in diesem Fall einer großen digitalen Plattform für Human Ressource mit Geschäftskund*innen, um mehr als 10% steigen, wenn es personalisierte Preise anstelle von Einheitspreisen verwendet. Weiterhin stellen sie fest, dass sich auf der Seite der Verbraucher*innen die Gesamtkonsumentenrente reduziert. Eine Einschränkung der Datenverwendung spezifischer Informationen der Verbraucher*innen würde diesen Effekt gemäß den Erkenntnissen der Fallstudie nicht

ausgleichen, sondern sogar noch weiter verstärken. Obwohl die Konsumentenrente insgesamt abnimmt, würde aber die Mehrheit der Verbraucher*innen von niedrigeren Preisen durch personalisierte Preisdiskriminierung im Gegensatz zum Einheitspreis profitieren. Wird neben der Höhe der Gesamtkonsumentenrente auch deren Verteilung betrachtet, so stellen die Autoren fest, dass die Vorteile der Personalisierung durch die Verringerung der Ungleichheit den Verlust der Gesamtkonsumentenrente ausgleichen können. Insbesondere kleine Unternehmen profitieren von diesen allokativen Vorteilen. Dubé und Misra gehen davon aus, dass es für Verbraucher*innen schwierig ist, ihre persönlichen Daten online zu ändern oder zu manipulieren, um von günstigeren Preisen zu profitieren. Obwohl sie nicht speziell untersuchen, wie sich ein anhaltender Widerstand der Verbraucher*innen, der auf Bedenken hinsichtlich der Fairness personalisierter Preise beruht, in den Ergebnissen ihrer Studie widerspiegelt, vermuten sie, dass ein solcher Widerstand im Laufe der Zeit zu einer flexibleren Preisreaktion der Nachfrage führen könnte. Ein solches Verhalten könnte sich nach Ansicht der Autoren negativ auf Märkte auswirken, in denen personalisierte Preise offener angewendet und weniger kritischer betrachtet werden (Dubé & Misra, 2017, S. 35).

Auf Märkten, wo nicht ein einzelnes Unternehmen den Preis diktiert, sondern mehrere Anbieter*innen im Wettbewerb stehen, gestaltet sich die Analyse von Big-Data-basierter Preisdiskriminierung komplex. Nicht jede Form der Preisdiskriminierung wirkt sich gleich aus, und ihre Effekte auf die Verbraucher*innen bzw. die Wohlfahrt können variieren. Preisdiskriminierung, die auf Markenpräferenzen oder Wechselkosten basiert, kann den Wettbewerb zwischen den Anbieter*innen intensivieren. Wenn Unternehmen individuelle Preise anbieten, die auf den Präferenzen und der Loyalität der Kund*innen basieren, kann dies den Wettbewerb zwischen ihnen fördern und zu positiven Ergebnissen für die Verbraucher*innen führen. Andererseits kann Preisdiskriminierung, die auf den Suchkosten der Verbraucher*innen basiert, den Wettbewerb zwischen den Unternehmen behindern, da Verbraucher*innen aufgrund der Kosten und Unsicherheiten bei der Produktsuche weniger geneigt sind, den*die Anbieter*in zu wechseln. Wenn Unternehmen Preise basierend auf dem Suchverhalten von Verbraucher*innen festlegen, kann dies den Einkaufsprozess erheblich erschweren. Angenommen, Verbraucher*innen suchen online nach einem Produkt und stellen fest, dass die Preise variieren, je nachdem, welche Suchmethode sie verwenden oder ob sie eine Website verlassen und später wieder zurückkehren auch könnten manche Online-Shops Rabatte anbieten, um die Verbraucher*innen zum Kauf zu bewegen. Diese ständigen Preisschwankungen, die von Faktoren wie der Suchhistorie abhängen, können nicht nur den Zeitaufwand erhöhen,

sondern auch zu Verwirrung führen, da die Kund*innen nie wirklich sicher sein können, ob sie das beste Angebot erhalten. Diese Praxis kann den Wettbewerb zwischen Unternehmen einschränken, weil es für Verbraucher*innen schwieriger wird, Angebote zu vergleichen. Weniger Wettbewerb bedeutet oft, dass die Preise höher und die Auswahl geringer ist, was letztendlich für die Verbraucher*innen nachteilig ist. Die Komplexität der Kaufentscheidung steigt mit dem Aufwand, der betreiben werden muss, um den besten Preis zu finden. Die Informationsflut und die Komplexität der Preisstruktur können zu Fehlentscheidungen führen. Wenn Verbraucher*innen durch Online-Preisdiskriminierungspraktiken verwirrt sind oder es ihnen schwerfällt, Preise und Produkte effektiv zu vergleichen, führt dies zu nachteiligen Ergebnissen. Kurz gesagt, die Anpassung der Preise an das Suchverhalten kann den Kaufprozess unnötig komplizieren, was zu Verwirrung führt, den Wettbewerb verringert und es schwieriger macht, sicherzustellen, dass der*die Kund*in das beste verfügbare Angebot erhält (Townley et al., 2017, S. 688 ff.)

Belleflamme, Lam & Vergote (2020) untersuchen in ihrer Arbeit Preisdiskriminierung in unvollkommenen Märkten. Als Schlüsselaspekt führen sie das unvollkommene Profiling von Verbraucher*innen durch die Unternehmen an. Das bedeutet, dass die Unternehmen zwar Informationen über die Präferenzen und die Zahlungsbereitschaft der Konsument*innen haben, diese Informationen aber nicht perfekt sind. Als Ergebnis können die Unternehmen verschiedene Gruppen von Verbraucher*innen festlegen, denen sie unterschiedliche Preise anbieten. Dies geschieht aus strategischen Gründen der Gewinnmaximierung und aus Unsicherheit der Unternehmen über die Preisgestaltung der Konkurrenz. Die wirtschaftlichen Folgen dieser Preisdiskriminierung sind vielfältig. Einerseits kann sie für die Unternehmen von Vorteil sein, da sie höhere Preise verlangen und damit höhere Gewinne erzielen können. Auf der anderen Seite können Verbraucher*innen, insbesondere solche mit geringer Zahlungsbereitschaft, unter höheren Preisen leiden. Die Komplexität hinsichtlich der Auswirkungen zeigt sich auch bei den Folgen für die Wohlfahrt der Verbraucher*innen. Obwohl einige Verbraucher*innen höhere Preise zahlen, können andere von niedrigeren Preisen profitieren. Dadurch ist es auch möglich, dass die Gewinne einer Gruppe die Verluste einer anderen Gruppe überwiegen, was zu einem positiven Gesamteffekt auf die Wohlfahrt führt. So zeigen auch Belleflamme et al., dass Preisdiskriminierung nicht zwangsläufig zu Nachteilen für die Verbraucher*innen führen muss. Im Gegenteil, die Möglichkeit zur Preisdiskriminierung kann den Wettbewerb zwischen Unternehmen erhöhen und insgesamt zu niedrigeren Preisen führen. Dies ist besonders relevant, wenn

Unternehmen mit unterschiedlichen Fähigkeiten zum Profiling von Verbraucher*innen miteinander konkurrieren (Belleflamme et al., 2020, S. 1 ff., S. 5ff., S.9 ff.).

Auch Poort & Zuiderveen Borgesius (2019) sehen die Auswirkungen auf die Wohlfahrt differenziert. Preisdiskriminierung kann demnach sowohl für Konsument*innen als auch für Verkäufer*innen von Vorteil sein und den Wohlstand beider Parteien erhöhen. Sie ermöglicht es den Anbieter*innen, ihre Fixkosten zu decken und gleichzeitig eine breite Palette von Kund*innen zu bedienen. Ähnliches gilt für die dynamische Preisgestaltung. Allerdings können solche Preisstrategien für einige Kund*innen höhere Kosten verursachen als einheitliche oder konstante Preise. Sie können daher die Wohlfahrt bestimmter Gruppen von Verbraucher*innen (Konsumentenrente) verringern. Dies ist umso wahrscheinlicher, je ausgeklügelter das Preissystem des Anbietenden ist. Die Wohlfahrtseffekte der Preisdiskriminierung sind demnach nicht eindeutig. Wenn Preisdiskriminierung nicht zu einer signifikanten Marktausweitung führt, kann sie häufig die Gesamtkonsumentenrente zugunsten der Produzentenrente verringern. In einigen Fällen kann Preisdiskriminierung sogar zu einem Nettowohlfahrtsverlust führen, wenn die Produzent*innen gewinnen, die Verbraucher*innen aber mehr verlieren. In einigen Fällen können sogar die anbietenden Unternehmen einen Wohlfahrtsverlust erleiden, wenn der Wettbewerb intensiver wird. Im Allgemeinen muss Preisdiskriminierung, um wohlfahrtssteigernd zu sein, zu einem erheblichen Anstieg der Gesamtproduktion führen, indem Märkte bedient werden, die zuvor nicht bedient wurden. Selbst in diesem Fall ist es wahrscheinlich, dass Verbraucher*innen mit einer hohen Zahlungsbereitschaft unter der Preisdiskriminierung leiden. Je mehr sich die personalisierte Preisgestaltung der Preisdiskriminierung ersten Grades annähert, desto mehr Wohlfahrt wird von den Konsument*innen zu den Produzent*innen verlagert (Poort & Zuiderveen Borgesius, 2019, S. 5).

Eine weitere Marktform, die im Zusammenhang mit Preisdiskriminierung genannt werden kann, ist die monopolistische Konkurrenz. Dabei handelt es sich um eine Marktform zwischen Monopol und vollständigem Wettbewerb. Das Konzept der monopolistischen Konkurrenz geht unter anderem auf Edward Hastings Chamberlin (1933) zurück. Auf einem unvollkommenen Markt gibt es sehr viele kleine Anbieter*innen, die in der Lage sind, den Preis zu variieren. Die Unvollkommenheit entsteht zum einen dadurch, dass die Nachfragenden räumliche, zeitliche oder sachliche Präferenzen haben, zum anderen dadurch, dass die gehandelten Güter zwar vergleichbar sind, sich aber in bestimmten Eigenschaften unterscheiden (Produktdifferenzierung). Dies erlaubt den Verbraucher*innen eine vielfältige Auswahl zwischen substituierbaren

Produkten, die von spezialisierten Unternehmen hergestellt und angeboten werden. Aufgrund der Produktdifferenzierung verfügt jedes anbietende Unternehmen über einen gewissen monopolistischen Spielraum, innerhalb dessen es ihm möglich ist, ähnlich wie in einem Monopol, den Preis oder die Menge festzulegen. In diesem Monopolbereich gilt die Gewinnmaximierungsregel: Grenzerlös gleich Grenzkosten. Verlässt der*die Anbieter*in diesen Bereich, wird der Preis wie beim Polypol durch den Markt festgelegt und der*die Anbieter*in verliert bei höheren Preisen die Nachfragenden an die Konkurrenz (Wied-Nebbeling, 2004, S. 100 ff.). Li und Shuai (2019) untersuchen in ihrer Studie Preisdiskriminierung dritten Grades in einem monopolistisch wettbewerbsorientierten Markt. Ist die Anzahl der Unternehmen in der monopolistischen Konkurrenz festgelegt, und der Markteintritt exogenisiert, führt die Preisdiskriminierung zu höheren Gewinnen für Unternehmen und verringert die Wohlfahrt der Verbraucher*innen im Vergleich zur einheitlichen Preisgestaltung. Wenn der Markteintritt endogenisiert ist, ist die gleichgewichtige Produktvielfalt bei Preisdiskriminierung im Vergleich zum sozialen Optimum immer zu hoch, während sie bei einheitlicher Preisgestaltung zu hoch oder zu niedrig sein kann. Sofern der Markteintritt bei einheitlicher Preisbildung nicht weit unter dem Wohlfahrtsoptimum liegt, führt ein Verbot der Preisdiskriminierung zu einer Verbesserung der Wohlfahrt von Verbraucher*innen und Gesellschaft. (Li & Shuai, 2019, S. 114 ff.).

Dullien (2021) beschreibt in seinem Beitrag, wie die Entwicklung hin zu monopolistischem Wettbewerb auf modernen Märkten, insbesondere im Zusammenhang mit digitalen Unternehmen wie Microsoft, weitreichende ökonomische Auswirkungen hat. Gängige Modelle vollständiger Konkurrenz, die von vollkommenen Märkten ausgehen, werden in Frage gestellt. Stattdessen wird darauf hingewiesen, dass viele relevante Märkte eher monopolistischer Konkurrenz oder globalen Oligopolmärkten unterliegen, in denen Unternehmen von steigenden Skaleneffekten profitieren. Microsoft dient hier als herausragendes Beispiel für ein Unternehmen mit dynamischen Skaleneffekten. Die einmalige Entwicklung von Software wie Windows und Office erfordert zwar anfänglich hohe Kosten, der Verkauf weiterer Kopien verursacht jedoch nur noch geringe Kosten. Diese Skaleneffekte ermöglichen es Microsoft, höhere Gewinne zu erzielen, bessere Löhne zu zahlen und mehr in Forschung und Entwicklung zu investieren. Dullien betont, dass Unternehmen, die von solchen Skaleneffekten profitieren, ihre Marktposition stärken können. Diese Unternehmen können ihren Kund*innen Preise berechnen, die über den Durchschnittskosten liegen, was zu höheren Gewinnen führt. Mit diesen Gewinnen können sie ihre Marktmacht ausbauen und in benachbarte Märkte expandieren. Im

globalen Kontext wird betont, dass Technologieführerschaft und Schlüsselindustrien für die Zukunft entscheidend sind. China strebt aktiv die Technologieführerschaft in wichtigen Zukunftsmärkten an. Der Autor argumentiert, dass, die Wahrscheinlichkeit steigt, dass andere Länder, insbesondere Deutschland und Europa, in diesen Industrien keine Unternehmen mehr haben, wenn China in solchen Schlüsselindustrien die Führung übernimmt. Dies hätte negative Folgen für Einkommen und Wohlstand in den betroffenen Regionen (Dullien, 2021, S. 5 ff.).

Die Diskussion über die Auswirkungen von Preisdiskriminierung in verschiedenen Marktstrukturen sowie die wachsende Macht digitaler Monopole werfen wichtige Fragen bezüglich der Wohlfahrt der Verbraucher*innen und des Wettbewerbs auf. In beiden Diskursen stehen die Interessen der Verbraucher*innen im Mittelpunkt, während die Macht der Unternehmen und ihre Auswirkungen auf den Markt und die Gesellschaft kritisch hinterfragt werden. Die Betrachtung der Auswirkungen von Preisdiskriminierung in monopolistischen Märkten zeigt, dass die Situation komplex ist. Während einige Studien darauf hinweisen, dass Preisdiskriminierung in bestimmten Fällen zu einer höheren Konsumentenrente führen kann, werden auch negative Folgen für die Verbraucher*innen deutlich, insbesondere wenn die Preise aufgrund von personalisierter Preisdiskriminierung steigen und die Gesamtkonsumentenrente reduzieren. Darüber hinaus können Suchkosten und Informationsasymmetrien den Wettbewerb beeinträchtigen und die Verwirrung der Verbraucher*innen verstärken, was zu nachteiligen Ergebnissen führen kann. Diese Herausforderungen werden durch die Dominanz digitaler Monopole weiter verstärkt, wie Hans Jürgen Jakobs aufzeigt. Der deutsche Volkswirt und Wirtschaftsjournalist, beschreibt in seinem Buch „Das Monopol im 21. Jahrhundert“ (2022) die Gefahren, die sich aus Monopolen, Quasi-Monopolen oder sehr konzentrierten Oligopolen für die Wohlfahrt der Konsument*innen ergeben. Das digitale Zeitalter hat eine neue Klasse von Monopolisten hervorgebracht, deren Macht und Einfluss in der Wirtschaftswelt unübersehbar ist. Die "Big Five" - Google, Apple, Facebook, Amazon und Microsoft - haben nicht nur einen wichtigen Platz in der modernen Informationsgesellschaft eingenommen, sondern sind auch zu dominierenden Kräften auf den globalen Märkten geworden. Diese Unternehmen handeln mit einem kostbaren Gut: den Daten der Nutzer*innen. Ihre Macht erstreckt sich auf so unterschiedliche Bereiche wie Energie, Telekommunikation, Handel und Finanzen (Jakobs, 2022, S. 16 ff.).

Die Macht digitaler Monopole wird in Krisenzeiten besonders deutlich, ihre alltäglichen Auswirkungen werden jedoch oft als selbstverständlich hingenommen. Anders als

traditionelle Rohstoffmonopole beziehen digitale Monopole ihre Macht aus der freiwilligen Beteiligung der Nutzer*innen. Kritik an ihrer Monopolstellung wird oft nur von einer Minderheit geäußert, während die Mehrheit der Bevölkerung ihre Dienste ohne direkte Kosten nutzt. Diese Unternehmen sind jedoch keine Wohltäter*innen, sondern hochprofitable Organisationen, die erheblichen Einfluss auf Regierungen, Märkte und Investor*innen ausüben. Ihre Geschäftspraktiken tragen dazu bei, ihre Marktmacht zu festigen und den Wettbewerb zu verzerren. Unternehmen wie Amazon nutzen aggressive Preisstrategien und Preisdifferenzierung (E-Book-Markt, Prime-Programm), um verschiedene Segmente von Kund*innen anzusprechen und ihre Gewinne zu maximieren. Der Aufstieg der digitalen Monopole wurde laut Jakobs durch mehrere Faktoren begünstigt, darunter aggressives Wachstum, exzessive Übernahmen und Netzwerkeffekte. Diese Unternehmen haben einzigartige Ultra-Monopole geschaffen, die es neuen Wettbewerber*innen immer schwerer machen, in den Markt einzutreten und zu konkurrieren. Die Auswirkungen dieser Monopole gehen über rein wirtschaftliche Aspekte hinaus. Sie wirken sich auf die Gesellschaft aus, indem sie Fehlinformationen verbreiten, politische Prozesse beeinflussen und neue Herausforderungen für die psychische Gesundheit und den sozialen Zusammenhalt schaffen. Die Situation wird durch den weltweiten Kampf um Märkte und Wohlstand verschärft, bei dem die Daten-Superkonzerne eine Hauptrolle spielen. Sowohl in den USA ("Big Five") als auch in China (Baidu, Alibaba, Tencent) konkurrieren Unternehmen mit Algorithmen und Künstlicher Intelligenz um die Erschließung wichtiger und ertragreicher Märkte. Während die USA und China in dieser Auseinandersetzung führend sind, hat Europa bisher wenig Einfluss auf dem Gebiet der Datenwirtschaft (Jakobs, 2022, S. 54 ff., S. 67 ff., S. 73 ff., S. 99 ff., S. 266 f.).

Die Herausforderung laut Jakobs besteht darin, sicherzustellen, dass diese Macht verantwortungsbewusst und demokratisch kontrolliert wird, um die Grundprinzipien von Wettbewerb und Freiheit in der digitalen Ära zu wahren. Dazu gehört eine strengere Regulierung, um den Missbrauch von Monopolmacht und Datenschutzverletzungen einzudämmen, sowie eine größere Transparenz und Kontrolle über die Nutzung persönlicher Daten durch Unternehmen. Digitale Monopole führen zu einer Marktdynamik, in der herkömmliche, bereits oligopolisierte Märkte bedroht werden und der Wettbewerb weiter minimiert wird. Diese Konzentration von Marktmacht hat direkte Auswirkungen auf Preise, Wohlstand und technologischen Fortschritt. Jakobs unterstreicht die Notwendigkeit, über Preisbildung und die Zusammenhänge zwischen Oligopolen bzw. Monopolen zu sprechen, insbesondere im Hinblick auf die

Wettbewerbssituation und freie Märkte in Europa. Die von digitalen Monopolen ausgehende Macht wird als globale Gefahr für die Wirtschaftsordnung gesehen, die eine ernsthafte Herausforderung für politische und regulatorische Antworten darstellt (Jakobs, 2022, S. 357 f., S. 359 ff. S. 366 f., S. 369 ff., S. 383, S. 395 f., S. 409 ff.).

Fairness

Ähnlich wie bei der Frage nach den ökonomischen Auswirkungen von Big-Data-basierter Preisdiskriminierung geben Townley et al. auch bei der Frage nach der Fairness keine eindeutige Antwort und weisen auch hier darauf hin, dass mehrere Faktoren darüber entscheiden, ob Preisdiskriminierung als fair oder unfair angesehen werden kann. Unter anderem sind Aspekte wie die subjektive Wahrnehmung oder die Fairness zwischen sozialen Gruppen entscheidend. Verbraucher*innen können Preisdiskriminierung als unfair wahrnehmen, wenn sie den Eindruck haben, dass sie ungleich behandelt werden, oder wenn sie sich getäuscht fühlen, weil sie nicht wissen, dass unterschiedliche Preise angeboten werden. Es kann auch sein, dass bestimmte soziale Gruppen, die im digitalen Umfeld weniger versiert sind, schlechtere Angebote erhalten, was als eine Form sozialer Ungerechtigkeit empfunden werden kann. Die Frage der Fairness ist also ebenso eine differenzierte und muss von Fall zu Fall betrachtet werden (Townley et al., 2017, S. 721). Im Allgemeinen stehen Verbraucher*innen der Preisdiskriminierung eher skeptisch gegenüber und empfinden diese als ungerecht, selbst wenn sie persönlich davon profitieren. Dieses Empfinden wird abgeschwächt, wenn die Preisdiskriminierung an die Zugehörigkeit zu einer Gruppe geknüpft ist (Rabatte für Pensionist*innen, Studierende, etc.) oder wenn die Verbraucher*innen die Preisdiskriminierung durch ihre eigene Entscheidung stärker kontrollieren können, wie dies bei Treueprogrammen von Unternehmen der Fall ist (Konrad & Polk, 2020, S. 797). Auch die IMCO-Studie (European Parliament. Directorate General for Internal Policies of the Union., 2022, S. 20) unterstreicht diese Ergebnisse. Demnach hängt die Einstellung der Konsument*innen zur Preisdiskriminierung stark von der Art und Weise der dafür verwendeten Daten ab. Preisdiskriminierung dritten Grades, d.h. unterschiedliche Preise für unterschiedliche Gruppen, wird im Allgemeinen akzeptiert, da die Verbraucher*innen bereits lange und in der realen Welt mit dieser Praxis konfrontiert sind (z.B. Rabatte für Pensionist*innen, Studierende, etc.). Bei dieser Form der Preisdiskriminierung kennen die Verbraucher*innen im Normalfall die Parameter für die Preisunterschiede, weshalb die

Informationsasymmetrie zwischen Händler*innen und Verbraucher*innen eher gering ist. Ähnliches gilt für die Preisdiskriminierung zweiten Grades. So lässt sich aus den Ergebnissen einer niederländischen Studie (Poort & Zuiderveen Borgesius, 2019, S. 14) ableiten, dass beispielsweise Mengenrabatte als relativ akzeptabel angesehen werden. Preisdiskriminierung ersten Grades hingegen wird nach den Ergebnissen der IMCO-Studie generell abgelehnt. Unterstützt wird diese Aussage wiederum durch die Studie von Poort & Zuiderveen Borgesius (Poort & Zuiderveen Borgesius, 2019, S. 10 f.), in der 80% der Befragten angeben, dass sie Online-Preisdiskriminierung bis zu einem gewissen Grad inakzeptabel und unfair finden. Auch eine Studie aus den USA (Turow, Feldman & Meltzer, 2005) unterstreicht, dass 91 % der Befragten eine negative Einstellung gegenüber personalisierten Preisen in Supermärkten und 64 % eine negative Einstellung gegenüber personalisierten Rabatten haben.

3 Methodik

Im empirischen Teil dieser Arbeit wird eine quantitative standardisierte Beobachtung durchgeführt.

Einleitend werden daher die drei zentralen Gütekriterien in der quantitativen Forschung vorgestellt: Objektivität, Reliabilität und Validität. *Objektivität* bezieht sich auf die Unabhängigkeit der Forschungs- oder Testergebnisse von den durchführenden Personen. Sie umfasst die Durchführungsobjektivität, die Auswertungsobjektivität und die Interpretationsobjektivität. Durchführungsobjektivität kann in der quantitativen Forschung durch Standardisierung erreicht werden, während Interpretationsobjektivität in der sozialwissenschaftlichen Forschung schwer zu erreichen ist, da Interpretationen immer subjektiven Bewertungen unterliegen können. *Reliabilität* gibt an, wie frei ein Test von zufälligen Messfehlern ist und bezeichnet die Reproduzierbarkeit von Messergebnissen. Ein reliabler Test bildet die wahre Merkmalsausprägung des Untersuchungsgegenstandes sehr genau ab. Die Reliabilität einer Messung wird in der Regel bei der Entwicklung des Messinstruments mit Hilfe von Reliabilitätskoeffizienten statistisch geschätzt und inhaltlich interpretiert. Die *Validität* misst, inwieweit ein Messinstrument tatsächlich das Phänomen erfasst, das es zu messen vorgibt. Es wird unterschieden zwischen Inhaltsvalidität (ob die verwendeten Items eine zu messende Eigenschaft umfassend abbilden), Kriteriumsvalidität (misst den Grad der Übereinstimmung zwischen den Messergebnissen und einem relevanten Kriterium außerhalb der Messung) und Konstruktvalidität (prüft, wie gut die Items dem theoretischen Konzept entsprechen, das sie messen sollen). Zwischen diesen drei Gütekriterien besteht eine hierarchische Beziehung: Objektivität ist eine notwendige, aber nicht hinreichende Bedingung für Reliabilität, und Reliabilität ist wiederum eine notwendige, aber nicht hinreichende Bedingung für Validität. Objektive und reliable Messinstrumente sind Grundvoraussetzungen für die Entwicklung valider Messinstrumente (Döring & Bortz, 2016, S. 344; Krebs & Menold, 2019, S. 490 ff.).

3.1 Quantitative standardisierte Beobachtung

Die Beobachtung ist eine Methode der Datenerhebung, bei der Ereignisse, Verhaltensweisen oder Merkmale von Forschenden systematisch erhoben, dokumentiert und ausgewertet werden. Sie wird systematisch geplant und durchgeführt und

unterscheidet sich dadurch von der Alltagsbeobachtung, die durch Willkür gekennzeichnet ist. Beobachtungen können mit Hilfe menschlicher Sinnesorgane und/oder technischer Sensoren durchgeführt werden. Um eine zeitversetzte Analyse und Interpretation der beobachteten Daten zu ermöglichen, werden die Datenerhebungen teilweise oder vollständig medial aufgezeichnet (Döring & Bortz, 2016, S. 323 f.; Thierbach & Petschick, 2019, S. 1165).

In der sozialwissenschaftlichen Methodenliteratur werden verschiedene Formen der Beobachtung anhand unterschiedlicher Dimensionen unterschieden. Diese Unterscheidungen sind hilfreich, um die spezifischen Anwendungsbereiche und Implikationen der einzelnen Beobachtungsformen für die Forschung besser zu verstehen. Zum einen gibt es die Direktheit der Beobachtung, die in direkte und indirekte Beobachtungen unterteilt wird. Direkte Beobachtungen erfolgen unmittelbar durch die Sinnesorgane der Forschenden, während indirekte Beobachtungen auf technisch vermittelten Daten oder der Sammlung von Spuren beruhen. Zum anderen werden wissenschaftliche Beobachtungen nach dem Grad der Strukturierung, der sich auf die Beobachtungssituation bezieht, und dem Grad der Standardisierung, der sich auf das verwendete Beobachtungsinstrument bezieht, unterschieden. Unstrukturierte oder auch unsystematische Beobachtungen erheben visuelle und verbale Daten über den Beobachtungsgegenstand ohne vorherige Beobachtungsrichtlinien. Sie sind somit offen und orientieren sich an den Gegebenheiten des Beobachtungsfeldes. Diese offene Herangehensweise ist vergleichbar mit unstrukturierten Interviews, bei denen auf einen festgelegten Fragenkatalog oder vorgegebene Antwortmöglichkeiten verzichtet wird. Teilstrukturierte Beobachtungen orientieren sich an konkreten Forschungsfragen oder theoretischen Konzepten, ähnlich dem Leitfadeninterview, bei dem die Fragen festgelegt, die Antworten aber offen sind. (Voll-)strukturierte Beobachtungen, auch systematische oder geschlossene Beobachtungen genannt, erheben numerische Messwerte für genau definierte Variablen nach einem vorher festgelegten Beobachtungsschema. Dieser Ansatz entspricht der vollstrukturierten Befragung und verwendet einen standardisierten Erhebungsbogen, ähnlich dem standardisierten Fragebogen für Befragungen. In der Praxis werden unstrukturierte und teilstrukturierte Methoden oft als qualitative Beobachtung klassifiziert, im Gegensatz zur quantitativen Beobachtung, die vollstrukturierte Methoden umfasst. Als weitere Dimension wird die Transparenz der Beobachtung in verdeckte und offene Beobachtungen unterteilt. Bei verdeckten Beobachtungen wissen die Beobachteten nicht, dass sie beobachtet werden, während bei offenen Beobachtungen die Beobachteten wissen, dass sie beobachtet werden. Schließlich

wird die Involviertheit der Beobachterrolle in teilnehmende und nicht-teilnehmende Beobachtung unterschieden. Bei der teilnehmenden Beobachtung nehmen die Forschenden eine aktive Rolle im Feld ein, während bei der nicht-teilnehmenden Beobachtung die Forschenden eine distanzierte Beobachterposition einnehmen. Diese Dimensionen ermöglichen es den Forschenden, die Beobachtungsmethode gezielt nach den Bedürfnissen ihrer Studie und den spezifischen Gegebenheiten des Forschungsfeldes auszuwählen. Die Wahl der Beobachtungsform hat direkte Auswirkungen auf die Art der erhobenen Daten, die Möglichkeiten ihrer Analyse und die Interpretation der Ergebnisse (Döring & Bortz, 2016, S. 328; Thierbach & Petschick, 2019, S. 1166 f.).

Um die Vielfalt und Komplexität von Beobachtungsmethoden in der sozialwissenschaftlichen Forschung vollständig zu erfassen, ist es wichtig, die Unterschiede zwischen standardisierter und nicht-standardisierter Beobachtung zu verstehen. Während die bisherigen Ausführungen ein breites Spektrum an Beobachtungsdimensionen von der Direktheit der Beobachtung bis hin zur Involviertheit der Beobachterrolle aufgezeigt haben, bieten die folgenden Überlegungen eine tiefer gehende Analyse spezifischer standardisierter und nicht-standardisierter Beobachtungsverfahren. Diese Gegenüberstellung ermöglicht es, die jeweiligen Vorteile und Herausforderungen dieser methodischen Ansätze noch genauer zu beleuchten und die Bedeutung der Wahl der Beobachtungsmethode für die Forschungspraxis weiter zu unterstreichen. Bei der standardisierten Beobachtung wird ein vorgegebenes System zur Klassifizierung von Verhaltensweisen oder Phänomenen verwendet, meist in Form von Skalen oder binären Optionen. Die Messgenauigkeit wird durch die Übereinstimmung der Beobachter*innen validiert, was eine zuverlässige Anwendbarkeit des Systems und die Generierung genauer Daten signalisiert. Zuvor wird die Zuverlässigkeit jeder Kategorie innerhalb des Systems in einem Pretest geprüft und dokumentiert, einschließlich der Berechnung von Reliabilitätskoeffizienten. Bei niedriger Beobachterübereinstimmung werden das System und die Schulung der Beobachter*innen überarbeitet, um Unklarheiten zu beseitigen und die Messgenauigkeit zu verbessern. Im Gegensatz dazu basiert die nicht-standardisierte Beobachtung nicht auf einem vordefinierten Kategoriensystem. Sie ist flexibler und weniger strukturiert, was zu einem offeneren Ansatz bei der Datenerhebung führt. Die Messgenauigkeit ist nicht im Voraus definiert oder durch Beobachter*innenübereinstimmung validiert, was zu einer größeren Varianz in der Datenqualität führen kann. Es findet keine formale Reliabilitätsprüfung statt, und die Datenauswertung und -interpretation basiert stärker auf dem Kontext und der Expertise des*der Forschenden. Nicht standardisierte Beobachtungsmethoden sind

anpassungsfähiger an spezifische Forschungskontexte und können nuanciertere Daten liefern, die durch ein standardisiertes System möglicherweise übersehen werden. Standardisierte Beobachtung ermöglicht somit eine systematische, zuverlässige und valide Datenerhebung durch die Anwendung vordefinierter Kriterien und die Sicherstellung einer hohen Beobachter*innenübereinstimmung. Die nicht-standardisierte Beobachtung bietet mehr Flexibilität und Anpassungsfähigkeit an den Forschungskontext, allerdings auf Kosten der standardisierten Messgenauigkeit und der leichteren Vergleichbarkeit der Daten (Döring & Bortz, 2016, S. 346 f.).

Beobachtungsstudien sind ein wesentliches Instrument der wissenschaftlichen Forschung, sie sind jedoch anfällig für eine Reihe von Fehlerquellen, die die Qualität und Validität der erhobenen Daten erheblich beeinträchtigen können. Diese Fehlerquellen können in zwei Hauptkategorien eingeteilt werden: Fehler, die mit der Beobachtungssituation zusammenhängen, und Fehler, die mit dem Verhalten oder den Eigenschaften der beobachtenden Person zusammenhängen. Zu den situationsbedingten Fehlern gehören die Reaktivität und die Künstlichkeit der Beobachtungssituation. Reaktivität tritt auf, wenn das Verhalten der beobachteten Personen durch das Bewusstsein, beobachtet zu werden, beeinflusst wird, was zu nicht repräsentativem Verhalten führen kann. Ebenso kann die Künstlichkeit einer Laborumgebung das natürliche Verhalten der Teilnehmenden verändern und damit die Ergebnisse verfälschen. Beobachterbezogene Fehlerquellen umfassen Wahrnehmungs-, Erinnerungs-, Interpretations- und Reproduktionsfehler. Wahrnehmungsfehler entstehen durch selektive Wahrnehmung, die auf persönlichen Vorannahmen beruht, während Erinnerungsfehler durch die begrenzte Kapazität des menschlichen Gedächtnisses verursacht werden. Interpretationsfehler spiegeln persönliche Urteile oder Vorurteile wider, die die Interpretation der Daten beeinflussen können, und Wiedergabefehler treten auf, wenn korrekt wahrgenommene Beobachtungen bei der Dokumentation verzerrt wiedergegeben werden. Hinzu kommen Fehler wie der Halo-Effekt, bei dem ein Gesamteindruck oder ein markantes Einzelmerkmal die weiteren Urteile der*des Beobachtenden beeinflusst, sowie die Projektion eigener Eigenschaften auf die Beobachteten und die emotionale Beteiligung des Beobachters, die insbesondere bei teilnehmenden Beobachtungen zu Verzerrungen führen können. Fehler durch die Wahl der Beobachtungsform und durch das Beobachtungssystem, das entweder zu viele oder zu wenige Kategorien enthält, sind weitere kritische Punkte, die die Datenqualität beeinträchtigen können. Um diese Fehlerquellen zu minimieren, sind gezielte Gegenmaßnahmen erforderlich. Entscheidend sind die sorgfältige Auswahl der

Beobachter*innen, die Durchführung von Beobachterschulungen, die Praxis der Überprüfung durch Co-Beobachter*innen und die Anpassung der Beobachtungsmethoden und -instrumente an den Forschungskontext (Döring & Bortz, 2016, S. 330 ff.; Gniewosz, 2011, S. 105 f.).

Erklärung der empirischen Methode

In der vorliegenden Arbeit wird eine standardisierte quantitative Beobachtung durchgeführt. Zu diesem Zweck wurde ein Web Scraper entwickelt, der an sieben Tagen in der Woche jeweils dreimal täglich die Preise von acht verschiedenen in Österreich erhältlichen Produkten und Dienstleistungen aufzeichnet. Der Web Scraper simuliert bei der Preisabfrage drei verschiedene User Agents. Jeder User Agent gibt vor, die Preisabfrage von einem anderen Endgerät aus durchzuführen. So wird erreicht, dass die aufgezeichneten Preise von einem Windows PC, einem Apple iPhone und einem Android Smartphone durchgeführt werden. Die Abfrage erfolgt auf zwei Arten: Einmal im normalen Modus des Browsers und einmal im Inkognito-Modus des Browsers. Bei der Abfrage im normalen Modus werden die Cookies der Webseite gespeichert und bei jedem Aufruf mitgegeben. Nach jedem Aufruf werden die Cookies mit den neu erhaltenen überschrieben. Im Inkognito-Modus erfolgt keine Speicherung von Cookies. Die abgefragten Preise werden in Logfiles mit Zeitstempel protokolliert, zusätzlich wird ein Screenshot der aufgerufenen Webseite gespeichert. Um standortbedingte Einflüsse zu eliminieren, werden die Abfragen mit der gleichen IP aus einem Netzwerk durchgeführt.

Ziel ist es, herauszufinden, ob es Preisunterschiede zwischen den User Agents gibt, aber auch, ob es Unterschiede zwischen den jeweiligen User Agents im normalen und Inkognito-Modus gibt.

Um die hier angewandte empirische Methode vor dem Hintergrund der Gütekriterien quantitativer Forschung zu betrachten, ist es unerlässlich, die Implementierung und Funktionsweise des entwickelten Web Scrapers zu verstehen. Dieser automatisierte Ansatz zur Datenerhebung ermöglicht eine detaillierte Bewertung der Objektivität, Reliabilität und Validität der gewonnenen Preisinformationen. Der Web Scraper, der speziell für die Erfassung der Preise von acht verschiedenen Produkten und Dienstleistungen entwickelt wurde, verwendet eine innovative Methodik, indem er drei verschiedene User Agents simuliert und sowohl im normalen als auch im Inkognito-Modus des Browsers agiert. Dieser systematische und standardisierte Ansatz ermöglicht es nicht nur, Preisunterschiede zu identifizieren, die möglicherweise auf die Verwendung

unterschiedlicher Technologien oder Surfmodi zurückzuführen sind, sondern auch, die Datenerhebung ohne menschliches Eingreifen und damit objektiv zu gestalten.

Die *Objektivität* wird durch die Automatisierung und Standardisierung des Erhebungsprozesses gewährleistet. Da der Web Scraper autonom agiert und präzise programmierte Aufgaben ausführt, sind die Ergebnisse unbeeinflusst von menschlicher Voreingenommenheit oder subjektiven Interpretationen. Die gewonnenen Daten - protokolliert in Logfiles mit Zeitstempel und ergänzt durch Screenshots - bieten eine solide Grundlage für eine objektive Auswertung.

Hinsichtlich der *Reliabilität*, also der Zuverlässigkeit der Methode, unterstützt die konsequente Wiederholung der Datenerhebung die Generierung konsistenter und reproduzierbarer Ergebnisse. Durch die dreimal tägliche Abfrage der Preise unter identischen Bedingungen über einen definierten Zeitraum werden mögliche Schwankungen erfasst, was die Messung der tatsächlichen Merkmalsausprägungen der untersuchten Produkte und Dienstleistungen ermöglicht.

Die *Validität* der mit dem Web Scraper erhobenen Daten ist durch die gezielte Ausrichtung auf die Untersuchungsziele hoch. Die Möglichkeit, Preisinformationen unter verschiedenen technischen Bedingungen zu erfassen, spricht direkt die inhaltliche Validität an und stellt sicher, dass das Messinstrument - in diesem Fall der Web Scraper - tatsächlich das misst, was es zu messen vorgibt.

Durch die Anwendung dieser standardisierten quantitativen Beobachtungsmethode wird somit eine hohe Messgenauigkeit erreicht, die auf einer vorher festgelegten und systematisch durchgeführten Beobachtung beruht. Der Einsatz von technischen Sensoren, in diesem Zusammenhang der Web Scraper, gewährleistet eine unabhängige und präzise Datenerhebung.

Trotz der zahlreichen Vorteile dieses methodischen Ansatzes ist es jedoch wichtig, mögliche Beobachtungsfehler und Gegenmaßnahmen zu berücksichtigen. Technische Unzulänglichkeiten, Änderungen am Quellcode der Webseiten oder die Reaktion der beobachteten Webseiten auf die Erkennung von Scraping-Versuchen können die Qualität der erhobenen Daten beeinflussen. Um diese Risiken zu minimieren, sind regelmäßige Überprüfungen und Anpassungen des Scraper-Codes sowie die Validierung der erhobenen Daten durch manuelle Stichproben unerlässlich.

3.2 Das Web-Scraping-Tool: Aufbau und Funktionsweise

Web Scraping basiert auf den im Quellcode von Webseiten gespeicherten Informationen, die durch den direkten Aufruf einer Webadresse (URL) entstehen. Bei dieser Methode werden spezifische Felder oder Datenelemente nicht aus einzelnen Datenbanken, sondern aus Webseiten extrahiert. Der Prozess beginnt auf der Ebene der gerätespezifischen Inhaltsrepräsentation, wobei aus diesen relativ unstrukturierten Daten, die für die Forschungsfrage relevanten Informationen herausgefiltert werden müssen. Erst im Laufe dieses Prozesses entstehen strukturierte Informationen. Ein Nachteil des Web Scraping ist, dass jede Änderung der Benutzeroberfläche auch eine Änderung der Skripte nach sich zieht (Keyling & Jünger, 2016, S. 190 f.). Darüber hinaus ist beim Web Scraping das geltende Recht zu beachten. In Österreich wird dies im Urheberrechtsgesetz (UrhG) geregelt. §42h Abs 1 UrhG erlaubt es „...Texte und Daten in digitaler Form für die wissenschaftliche oder künstlerische Forschung automatisiert auszuwerten und Informationen unter anderem über Muster, Trends und Korrelationen zu gewinnen, wenn er zu dem Werk rechtmäßig Zugang hat. Zu einer solchen Vervielfältigung sind auch einzelne Forscher berechtigt, soweit dies zur Verfolgung nicht kommerzieller Zwecke gerechtfertigt ist.“ (Rechtsinformationssystem des Bundes (RIS), 2024)

Genutzte Technologien und Tools zur Erstellung des Web Scrapers

Für die Entwicklung des Web Scrapers wurden folgenden Technologien und Tools verwendet:

Visual Studio Code (VS Code)

Visual Studio Code, entwickelt von Microsoft, ist ein Quellcode-Editor für Windows, Linux und macOS. Er unterstützt Funktionen wie Debugging, Syntaxhervorhebung, intelligente Codevervollständigung, Snippets, Code-Refactoring und integrierte Versionskontrolle mit Git. Seit seiner Ankündigung im April 2015 und der Freigabe des Quellcodes unter der Lizenz des MIT (Massachusetts Institute of Technology) im November 2015 hat VS Code eine rasche Verbreitung erfahren. Der Editor zeichnet sich durch seine Flexibilität aus, die es den Benutzer*innen ermöglicht, Funktionen durch die Installation von Erweiterungen zu ergänzen (Microsoft, o. J.).

JavaScript

JavaScript ist eine dynamische Programmiersprache, die 1995 von Brendan Eich bei Netscape entwickelt wurde. Ziel war es, Webseiten interaktiv zu gestalten, indem Skripte direkt im Browser des*der Benutzer*in ausgeführt werden können. JavaScript hat sich schnell von einer einfachen clientseitigen Skriptsprache zu einer der wichtigsten Sprachen in der modernen Webentwicklung entwickelt. Die Entwicklung von JavaScript ist eng mit den Standards der European Computer Manufacturers Association (ECMA) verbunden. 1997 wurde der Standard ECMA-262 eingeführt, der die Grundlage für ECMAScript, die Basis von JavaScript, bildet. Dieser Standard wird regelmäßig aktualisiert, um neue Funktionen und Verbesserungen einzuführen, was zu den jährlichen Updates von ECMAScript führt. JavaScript wird nicht nur auf der Client-Seite verwendet, sondern dank Umgebungen wie Node.js auch auf der Server-Seite. Dies hat JavaScript zu einer universellen Sprache gemacht, die sowohl für die Frontend- als auch für die Backend-Entwicklung verwendet wird. Seine Fähigkeiten reichen von einfachen Skripten bis hin zu komplexen Anwendungen mit Event-Handling, asynchroner Programmierung und vielem mehr, unterstützt durch eine Vielzahl von APIs für verschiedene Anwendungsfälle. Aufgrund seiner Vielseitigkeit und ständigen Weiterentwicklung hat sich JavaScript als kritische Komponente bei der Entwicklung interaktiver, dynamischer Webseiten etabliert. Seine Popularität wird auch durch jährliche Umfragen unter Entwickler*innen bestätigt, bei denen JavaScript regelmäßig als eine der am häufigsten verwendeten und beliebtesten Programmiersprachen eingestuft wird (MDN contributors, 2024a; MDN contributors, 2024b; MDN contributors, 2024c; W3Schools, o. J.).

Node.js

Node.js ist eine Webentwicklungstechnologie, die als leistungsfähige JavaScript-Laufzeitumgebung dient und bei Entwickler*innen aufgrund ihrer Fähigkeit, JavaScript-Codes außerhalb eines Webbrowsers auszuführen, sehr beliebt geworden ist. Im Gegensatz zu herkömmlichem JavaScript, das hauptsächlich auf der Client-Seite ausgeführt wird, ermöglicht Node.js die Ausführung von JavaScript auf der Server-Seite, was bisher anderen Programmiersprachen wie Python oder PHP vorbehalten war. Node.js wurde erstmals 2009 von Ryan Dahl mit dem Ziel entwickelt, eine effizientere Methode zur Handhabung von on Ein- und Ausgabeoperationen (I/O-Operationen) zu schaffen. Durch die Verwendung der Google V8 JavaScript Engine, die JavaScript-Codes direkt in Maschinencodes übersetzt, bietet Node.js eine hohe Performance für die Erstellung skalierbarer Webanwendungen. Die asynchrone, ereignisgesteuerte Architektur

ermöglicht es Node.js, eine große Anzahl gleichzeitiger Verbindungen effizient zu verwalten, ohne zusätzliche Threads erstellen zu müssen, was zu einer verbesserten Skalierbarkeit und Ressourceneffizienz führt. Die Einführung von Node.js hat die Art und Weise, wie Webanwendungen entwickelt werden, grundlegend verändert, indem es Entwickler*innen ermöglicht wurde, JavaScript sowohl auf der Client- als auch auf der Serverseite zu verwenden. Dies hat zur Entstehung der Full-Stack JavaScript-Entwicklung geführt. Das riesige und robuste Node.js-Ökosystem, das hauptsächlich über den Node Package Manager (npm) verfügbar ist, bietet eine umfangreiche Sammlung von Bibliotheken und Paketen, die den Entwicklungsprozess vereinfachen. Zu den Meilensteinen in der Geschichte von Node.js zählen die erste stabile Version im Jahr 2010, die rasche Popularität und Akzeptanz durch große Unternehmen wie LinkedIn und Yahoo, die Gründung der Node.js Foundation im Jahr 2015 zur Förderung der nachhaltigen Entwicklung und die breite Akzeptanz in der Unternehmenswelt für den Aufbau skalierbarer und leistungsstarker Anwendungen. Node.js verwendet eine ereignisgesteuerte Architektur mit einer zentralen Komponente, dem Event-Loop, der alle Code-Callback-Funktionen ausführt, die eine Node.js-Anwendung ausmachen. Das Konzept der Single-Threaded-Ausführung in Kombination mit einer nicht-blockierenden I/O-Modellierung ermöglicht es Node.js, Tausende von gleichzeitigen Verbindungen mit einem einzigen Server zu verwalten. Das Herzstück von Node.js ist die V8 JavaScript Engine, die einen Memory Heap und einen Call Stack enthält, die die Ausführung von JavaScript-Code verwalten. Eine weitere wichtige Komponente ist Libuv, eine C-Bibliothek zur Ausführung von I/O-Operationen, die es Node.js ermöglicht, I/O-Anweisungen trotz der Single-Threaded-Natur von JavaScript effizient zu verarbeiten. Die kontinuierliche Weiterentwicklung von Node.js, die von einer aktiven und lebendigen Community weltweit unterstützt wird, sowie regelmäßige neue Features und Verbesserungen sichern Node.js einen Platz als sich ständig weiterentwickelnde Technologie (Faisal, 2023; Mastering Backend, o. J.; Node.js (a), o. J.; Node.js (b), o. J.).

fs in Node.js

Das `fs`-Modul in Node.js bietet eine umfangreiche API für die Interaktion mit dem Dateisystem. Es ermöglicht das Lesen, Schreiben, Modifizieren und Überwachen von Dateien sowie den Zugriff auf deren Metadaten. Zu den grundlegenden Operationen gehören das asynchrone und synchrone Lesen von Dateien, das Schreiben und Anhängen von Inhalten an Dateien sowie das Umbenennen und Löschen von Dateien. Die Interaktion mit dem Dateisystem über das fs-Modul kann über Callbacks, Promises oder die `async/await`-Syntax erfolgen, wobei die Promise-basierten APIs in neueren Node.js-

Versionen eingeführt wurden, um einen moderneren und saubereren Ansatz für asynchrone Operationen zu ermöglichen (De Sousa, 2021; Kinsta, 2023; Node.js (a), o. J.; TutorialsTeacher, o. J.).

Puppeteer

Puppeteer ist eine Node.js-Bibliothek, die eine High-Level-API zur Steuerung von Chrome oder Chromium über das DevTools-Protokoll bereitstellt. Sie läuft standardmäßig im Headless-Modus, kann aber auch für den Vollmodus (Headful) konfiguriert werden. Puppeteer eignet sich für eine Vielzahl von Anwendungsfällen, wie z.B. das Generieren von Screenshots und PDFs von Seiten, das Crawlen von Single-Page-Anwendungen (SPAs) zur Generierung von vorgerenderten Inhalten (SSRs), die Automatisierung von Formulareingaben, UI-Tests, Tastatureingaben und vieles mehr. Es ermöglicht die Erstellung einer automatisierten Testumgebung, die die neuesten JavaScript- und Browser-Features nutzt und kann auch zum Testen von Chrome-Erweiterungen verwendet werden. Verschiedene Konfigurationsoptionen und Umgebungsvariablen stehen zur Verfügung, um das Downloadverhalten anzupassen. Für eine abgespeckte Version ohne Browser-Installation gibt es den „puppeteer-core“, der verwendet wird, wenn eine Verbindung zu einem entfernten Browser hergestellt oder der Browser selbst verwaltet wird. Die Architektur von Puppeteer basiert auf den DevTools-Protokollen zur Verwaltung von Chromium- oder Chrome-Browsern unter Verwendung der von der Node-Bibliothek bereitgestellten APIs. Puppeteer agiert als Node.js-Modul basierend auf dem Projektcode und ermöglicht es den Endbenutzer*innen, den eigentlichen Automatisierungscode in JavaScript zu entwickeln. Puppeteer hat seit seiner Einführung ein zunehmendes Interesse und eine zunehmende Nutzung unter Entwickler*innen und Tester*innen erfahren, obwohl das Interesse als bevorzugtes Testframework im Laufe der Zeit Schwankungen unterworfen war. Dennoch zeigen Umfragen, dass die Bekanntheit und Bereitschaft zur erneuten Nutzung von Puppeteer unter Entwickler*innen von 2019 bis 2021 gestiegen ist. Einer der Hauptvorteile von Puppeteer ist die Kontrolle über das DevTools-Protokoll, was es zu einem leistungsstarken Werkzeug für die Automatisierung von Frontend-Entwicklung und -Tests macht, insbesondere durch die Fokussierung auf die internen Funktionen von Chromium. Die Geschichte von Puppeteer reicht bis in die Zeit zurück, als automatisierte Tests in Webanwendungen nur mit einem automatisierten "Headful" Browser oder einem eigenständigen "Headless" Browser durchgeführt werden konnten. Puppeteer nutzt die Chromium-Browserumgebung, um Tests durchzuführen, und bietet Entwickler*innen die Flexibilität, Headless- oder Vollbrowserversionen zu verwenden, die auf derselben Plattform wie ihre Nutzer*innen laufen. Diese

Anpassungsfähigkeit führte zu einer steigenden Beliebtheit unter Entwickler*innen und zur Integration in populäre Testframeworks wie Mocha (LambdaTest, o. J.; Puppeteer, o. J.).

Der Web Scraper: Aufbau und Quellcode

Das zugrundeliegende Skript des Scrapers ist „scraper.js“, das eine Reihe von Befehlen enthält, die automatisch von einem Computer aus ausgeführt werden können. Im Folgenden werden die einzelnen Befehle des Skripts erläutert. Zum besseren Verständnis ist das Skript in mehrere Quellcodeabschnitte (1 bis 6) unterteilt, leere Zeilen des Skripts werden in den Quellcodes weggelassen.

Quellcode 1 - Import Anweisungen

```
1     import puppeteer from 'puppeteer'
2     import fs from 'fs'
```

Diese beiden Befehle in Zeile 1 und 2 des Quellcode 1 importieren zum einen das Puppeteer-Modul, um automatisierte Tests in einem Browser durchführen zu können, und zum anderen das Node.js Dateisystem-Modul, das Funktionen zum Lesen und Schreiben von Dateien bereitstellt.

Quellcode 2 - Lesen der Konfigurationsdatei

```
5     const config = JSON.parse(fs.readFileSync('./config.json', 'utf-8'))
```

Mit diesem Befehl aus dem Quellcode 2 wird die Datei „config.json“ synchron aus dem Dateisystem gelesen und der JSON-Inhalt in ein JavaScript-Objekt umgewandelt.

Quellcode 3 - Funktion zur Datumsformatierung

```
8     function formatDate(date) {
9         return `${date.getFullYear()}_${date.getMonth() +
            1}_${date.getDate()}_${date.getHours()}_${date.getMinutes()}_${date.g
            etSeconds()}`;
10    }
```

Quellcode 3 beginnt mit der Definition einer Funktion mit dem Namen „formatDate“, die ein Objekt „Date“ als Argument hat und gibt einen formatierten String zurück, der das Datum im Format „Jahr-Monat_Tag_Stunde_Minute_Sekunde“ darstellt. Da in JavaScript die Monate von 0 (= Januar) bis 11 (= Dezember) gezählt werden, wird „getMonth() + 1“ verwendet um die klassische Schreibweise zu erhalten.

Quellcode 4 - Hauptfunktion zum Scrpaen der Preise

```
12  async function scrapePrice({ product, url, xpath, xpathAdd,
13  userAgent, userAgentName }, dateTimeStamp, incognito = false)
14  //const browser = await puppeteer.launch({headless: false})
15  const browser = await puppeteer.launch({headless: 'new'})
16  try {
17    const context = incognito ? await
18    browser.createIncognitoBrowserContext() : browser;
19    const page = await context.newPage();
20    await page.setViewport({ width: 1080, height: 1024 });
21    await page.setUserAgent(userAgent);
22    var cookieExist = false;
23    if (!incognito) {
24      cookieExist = await cookiesExist(product);
25      if (cookieExist) {
26        await loadCookie(page, product);
27      }
28    }
29    await page.goto(url);
30    if (!incognito) {
31      await saveCookie(page, product);
32    }
33    await page.waitForXPath(xpath, { timeout: 60000 });
34
35    var priceAdd = '';
36    if(xpathAdd != null && xpathAdd != '') {
37      await page.waitForXPath(xpathAdd);
38      var priceAddElement = await page.$x(xpathAdd);
39      priceAdd = await page.evaluate(el =>
40      el.textContent, priceAddElement[0]);
41    }
42    const priceElement = await page.$x(xpath);
43    var price = await page.evaluate(el => el.textContent,
44    priceElement[0]);
45    var price = `${price}${priceAdd}`;
46    const logFile = incognito ?
47    `logfile_IM_${formatDate(dateTimeStamp)}.txt` :
48    `logfile_${formatDate(dateTimeStamp)}.txt`;
49    const logMessage = `${incognito ? 'Incognito: ' :
50    ''}Price (User Agent: ${userAgentName}, Product: ${product}):
51    ${price}\n`;
52    fs.appendFile(logFile, logMessage, err => { if (err) throw err;
53    });
54    console.log(logMessage);
55    const filename = `Screenshot_${incognito ? 'Incognito_' :
56    ''}${userAgentName}_${product}_${formatDate(dateTimeStamp)}.pdf`;
57    await page.pdf({ path: filename, printBackground: true });
58    await page.close();
59    await browser.close();
60  } catch (error) {
61    await browser.close();
62    console.error(`Error scraping price for product ${product}:
63    ${error}`);
64  }
```

Quellcode 4 veranschaulicht die Hauptfunktion zum Scrpen der Preise. Zeile 12 definiert eine asynchrone Funktion „scrapePrice“, die ein Objekt mit Produktinformationen und weiteren Parametern annimmt. Zeile 14 startet eine neue Instanz des Browsers im

Headless-Modus, so dass die Browseraufrufe nicht sichtbar sind und im Hintergrund ablaufen. Zu Testzwecken wurde auch der Befehl „const browser = await puppeteer.launch({headless: false})“ in Zeile 13 hinterlegt, um den Aufruf der Browserinstanzen visuell darzustellen. In der finalen Version wurde dieser Befehl jedoch auskommentiert und nur für weitere Tests und Überprüfungen beibehalten. Zeile 16 legt fest, ob ein Inkognito-Browser-Kontext oder der normale Browser-Kontext verwendet werden soll. Zeile 17 öffnet eine neue Seite im angegebenen Kontext. Zeile 18 legt die Größe des Browserfensters für die neue Seite fest. Zeile 19 setzt den User-Agent der Seite auf einen bestimmten Wert. Die Zeilen 20 bis 24 initialisieren eine Variable zur Speicherung der Information, ob Cookies vorhanden sind und überprüfen, ob nicht im Inkognito-Modus gearbeitet wird, und dann, ob Cookies für das Produkt existieren. Lädt diese Cookies gegebenenfalls. Zeile 28 navigiert zu einer angegebenen URL. Zeile 30 und 31 speichert Cookies, wenn nicht im Inkognito-Modus gearbeitet wird. Zeile 34 setzt die Timeout-Zeit auf 60 Sekunden, so lange wird maximal auf das Erscheinen eines bestimmten XPath-Elements auf der aufgerufenen Seite gewartet. Die Zeilen 36 bis 49 prüfen, ob ein zusätzlicher XPath für den Preis existiert, und rufen den zusätzlichen Preis ab, falls vorhanden. In weiterer Folge wird der Preis über den angegebene XPath abgerufen und, falls vorhanden, der Wert des zweiten XPath (xpathAdd) hinzugefügt. Die Angabe des zweiten XPath kann notwendig sein, wenn Euro und Cent nicht im selben XPath der Seite gespeichert sind. Weiters wird der Namen des Logfiles, abhängig vom Inkognito-Status definiert, sowie die Logmeldung mit Preisinformationen. Die Logmeldung wird an die Datei angehängt und wirft einen Fehler aus, wenn ein Fehler auftritt, d.h. sollte ein Problem auftreten, meldet das Skript diesen Fehler sofort. Schließlich sorgt Zeile 51 dafür, dass die Logmeldung an die Konsole ausgegeben wird. Die Zeilen 53 und 54 legen den Namen der PDF-Datei für den Screenshot fest. Erstellen einen PDF-Screenshot der Seite und speichert ihn unter dem angegebenen Pfad. Die Zeilen 56 bis 63 schließen die geöffnete Seite und im Anschluss den Browser. Fehler werden abgefangen und behandelt, indem der Browser geschlossen wird und der Fehler ausgegeben wird.

Quellcode 5 - Cookie-Management

```
65  const saveCookie = async (page, product) => {
66      const cookies = await page.cookies();
67      const cookieJson = JSON.stringify(cookies, null, 2);
68      await fs.writeFileSync(`cookies_${product}.json`, cookieJson);
69  }
71  const loadCookie = async (page, product) => {
72      const cookieJson = await
73      fs.readFileSync(`cookies_${product}.json`);
74      const cookies = JSON.parse(cookieJson);
75      await page.setCookie(...cookies);
76  }
77  const cookiesExist = async (product) => {
78      const path = `cookies_${product}.json`
79  try {
80      if (fs.existsSync(path)){
81          return true;
82      }
83      return false;
84  } catch(err) {
85      console.error(err)
86  }
87  }
```

Quellcode 5 zeigt das Cookie-Management. Die Zeilen 65 bis 69 definieren eine asynchrone Funktion zum Speichern von Cookies. Die Zeilen 71 bis 75 definieren eine asynchrone Funktion, um Cookies aus einer Datei zu laden und auf der Seite zu setzen. Die Zeilen 77 bis 88 definieren eine asynchrone Funktion, die prüft, ob für ein bestimmtes Produkt eine Cookie-Datei existiert.

Quellcode 6 - Hauptfunktion

```
90  async function main() {
91      const dateTimeStamp = new Date();
92      const sleep = (ms) => new Promise(resolve => setTimeout(resolve,
93      ms));
94      for (const productConfig of config) {
95          await scrapePrice(productConfig, dateTimeStamp);
96          await sleep(productConfig.delay);
97      }
99      for (const productConfig of config) {
100         await scrapePrice(productConfig, dateTimeStamp, true);
101         await sleep(productConfig.delay);
102     }
104     console.log('Scraping done!');
105 }
106
107 main().catch(console.error);
```

Quellcode 6 zeigt den letzten Teil des Skripts. Zeile 90 definiert dabei die Hauptfunktion „main“. Zeile 91 erzeugt ein neues Date-Objekt für den Zeitstempel. Die Zeilen 92 bis 102 definieren die Schleifen, die durch die Produktkonfigurationen iterieren und scrapePrice für jedes Produkt sowohl im normalen als auch im Inkognito-Modus aufrufen, mit

Verzögerungen zwischen den Aufrufen. Zeile 104 sorgt dafür, dass auf der Konsole angezeigt wird, dass das Scraping abgeschlossen ist. Zeile 107 ruft „main“ auf und fängt Fehler ab, die in der Konsole ausgegeben werden.

config.json

Der Inhalt der Datei „config.json“ spielt eine zentrale Rolle für die Funktionsweise des Skripts „scrape.js“. Diese Datei dient als Konfigurationsdatei, die festlegt, welche Webseiten das Skript besucht, welche Daten es sammelt und wie es sich dabei verhält. Im Folgenden werden die einzelnen Konfigurationselemente erläutert:

- „product“: Bezeichnet den Namen des Produkts oder der Dienstleistung, das bzw. die auf der Webseite abgefragt wird. Dies hilft bei der Organisation und Zuordnung der gesammelten Daten.
- „url“: Die URL der Webseite, von der Daten gescraped werden sollen. Dies ist die direkte Adresse der Seite, die das Skript besucht.
- „xpath“: Der XPath-Ausdruck, der verwendet wird, um bestimmte Daten aus der Webseite zu extrahieren. Dies ist notwendig, da Webseiten oft komplexe und verschachtelte Strukturen haben und der genaue Ort der benötigten Information angegeben werden muss.
- „xpathAdd“ (optional): Ein zusätzlicher XPath, der verwendet wird, um zusätzliche Informationen zu sammeln. Dies ist notwendig, wenn sich der Preis aus verschiedenen Teilen der Seite zusammensetzt oder wenn zusätzliche Details benötigt werden.
- „userAgentName“: Ein beschreibender Name für den verwendeten User Agent. Dies hilft zu verfolgen und zu organisieren, welcher User Agent (d.h. welches Gerät oder welcher Browser) für den Zugriff verwendet wurde.
- „userAgent“: Der User Agent String, den das Skript verwendet, um die Webseite aufzurufen. Dies beeinflusst, wie die Webseite das anfragende Gerät interpretiert - als Desktop, Smartphone usw.
- „delay“: Eine Verzögerung in Sekunden, die zwischen Scraping-Vorgängen eingehalten wird. Dies wird verwendet, um wie ein menschlicher Benutzer zu erscheinen und somit weniger wahrscheinlich geblockt zu werden.

Der nachfolgende Quellcode 7 zeigt ein Beispiel für einen „config.json“-Eintrag (zur besseren Übersicht gekürzt):

Quellcode 7 - Beispiel-Eintrag der config.json

```
"product": "Mediamarkt",  
"url": "https://www.mediamarkt.at/de/product/...."  
"xpath": "/html/body/div[1]/div[3]/main/div[1]/div[1]/div[2]/...",  
"xpathAdd" :"/html/body/div[1]/div[3]/main/div[1]/div... ",  
"userAgentName": "Safari",  
"userAgent": "Mozilla/5.0 (iPhone; CPU iPhone OS 17_3 like Mac OS X)  
    AppleWebKit/605.1.15 (KHTML, like Gecko) Version/17.2 Mobile/15E148  
    Safari/604.1",  
"delay": 10
```

Die verwendeten User Agents, die die verschiedenen Endgeräte simulieren, wurden der Webseite www.useragentstring.com (UserAgentString.com, o. J.) entnommen.

Da der Scraper spezifische XPath-Elemente verwendet, um die Preise aus einer Webseite zu extrahieren, kann das Fehlen von Preisen in dem erzeugten Logfile immer auf eine Überschreitung der Timeout-Zeit von 60 Sekunden zurückgeführt werden. Dies kann folgende Ursachen haben:

- *Änderung des XPath:* Wenn sich das Layout oder die Struktur einer Webseite ändert, kann der zuvor definierte XPath ungültig werden. Der Scraper wartet auf das hinterlegte XPath-Element, wenn dieses nicht innerhalb der definierten Timeout-Zeit erscheint, kommt es zu einem Timeout-Fehler und die Aktion wird abgebrochen. Dies macht eine Aktualisierung des XPath notwendig.
- *Timeout der Webseite:* Bei Netzwerkproblemen oder Problemen beim Laden der Webseite kann die definierte Timeout-Zeit überschritten werden und der Preis wird deshalb nicht in das Logfile übernommen.
- *Probleme bei der Erreichbarkeit der Webseite:* Sollte die Webseite aufgrund von Wartungsarbeiten, Serverproblemen, Netzwerkausfällen oder ähnlichem nicht erreichbar sein, kann es sein, dass der Scraping-Vorgang nicht durchgeführt werden kann, da auch hier die definierte Timeout-Zeit überschritten wird und somit kein Eintrag in das Logfile erfolgt.

3.3 Datenerhebung

Für die Auswahl der Produkte und Dienstleistungen, deren Preise extrahiert werden und um die Situation speziell in Österreich zu beleuchten wurden fünf E-Commerce-Plattformen aus den zehn umsatzstärksten Webshops in Österreich im Jahr 2021 ausgewählt. Media Markt (238,1 Mio. Umsatz), Universal (140,6 Mio. Umsatz), Otto Versand (126,8 Mio. Umsatz) Shop Apotheke (120,7 Mio. Umsatz) und electronic4you (89,1 Mio. Umsatz) (Brandt, 2022). Darüber hinaus wurde die Liste um drei Unternehmen (Europcar, Ryanair, Urlaubsguru) aus der Reise- bzw. Tourismusbranche erweitert, um unter anderem die These von Konrad & Polk (Konrad & Polk, 2020, S. 796) zu überprüfen, ob Preisdiskriminierung in dieser Branche eher praktiziert wird. Zu diesem Zweck wurden die umsatzstärkste europäische Billigfluggesellschaft (Statista Research Department, 2024), ein Buchungsportal und ein Autovermieter in die Untersuchung einbezogen.

Auf jeder Webseite wurde ein zufälliges Produkt/eine zufällige Dienstleistung ausgewählt und der Preis für den Beobachtungszeitraum in bestimmten Zeitabständen ausgelesen. Die Produkte sind in Tabelle 6 aufgeführt.

Tabelle 6 - Produkte/Dienstleistungen für den Webscraper

Unternehmen	Homepage	Produkt
Electronic4you GmbH	www.electronic4you.at	Samsung Galaxy A54
MediaMarkt Online GmbH	www.mediamarkt.at	De Longhi ECAM 220.21 B Magnifica
Otto GmbH	www.ottoversand.at	Konifera Pavillon „Malaga“
Shop-Apotheke Service B.V.	www.shop-apotheke.at	Dr. Böhm® Magnesium Sport Tabletten (2x60 Stk.)
Universal Versand	www.universal.at	ninebot by Segway E-Scooter
Europcar Österreich, ARAC GmbH	www.europcar.at	Fiat 500 Cabrio (Abholung: Flughafen Wien, 04.11.2024, 10:00 Uhr; Rückgabe: Flughafen Wien, 10.11.2024, 10:00 Uhr)
Ryanair DAC	www.ryanair.com/at/de	Flug: Wien – Lissabon (06.03.2025), 15:10 Uhr bis 17:55 Uhr
Urlaubsguru GmbH	www.urlaubsguru.at	Hotel: Smy Lisboa (Lissabon 06.03.2025 – 10.03.2025, 2 Erwachsene, 1 Zimmer inkl. Flug)

Für die automatisierte Ausführung des Scrapers wurde der Windows-Aufgabenplaner (Task Scheduler) verwendet. Dabei wurde der Task Scheduler so konfiguriert, dass die Konsolenanwendung „cmd.exe“ mit dem Argument „/c node/ ./scraper.js“ verwendet wird. Dazu ist es auch notwendig, im Aufgabenplaner den korrekten Pfad zum entsprechenden Verzeichnis zu hinterlegen, in dem die Datei „scrapers.js“ abgelegt wurde. Das Argument „/c“ teilt „cmd.exe.“ mit, dass es den folgenden Befehl („node ./scraper.js“) ausführen und danach schließen soll. Im Aufgabenplaner wurden zusätzlich drei Trigger gesetzt, die das Skript automatisch jeden Tag zu einer bestimmten Uhrzeit starten. So wurde jeweils ein Trigger für 00:00 Uhr, 08:00 Uhr und 16:00 Uhr gesetzt. Der Aufgabenplaner bietet auch die Möglichkeit, das Skript jederzeit und außerhalb der definierten Trigger per Mausklick zu starten, so dass ein manuelles Auslesen der Preise jederzeit möglich ist. Dies erleichtert die Überprüfung des Skripts, falls Preise in den Logfiles fehlen. In der Konsole kann so nachvollzogen werden, wo im Skript eventuell Timeouts auftreten. Am 04.05.2024 wurden die Trigger für die verbleibenden Tage geändert, um noch weitere Zeitpunkt in die Analyse einzubeziehen, so wurden nach dem Durchlauf um 8:00 Uhr die Trigger jeweils um zwei Stunden vorgezogen (00:00 Uhr → 22:00 Uhr, 08:00 Uhr → 06:00 Uhr, 16:00 Uhr → 14:00 Uhr).

Der Zeitraum für das Scraping wurde vom 04.04.2024 bis zum 11.05.2024 festgelegt. Mindestens dreimal pro Woche wurden die Logfiles durchsucht und auf fehlende Preise überprüft. Wurden solche oder auch fehlende Logfiles gefunden, wurde versucht, die Ursache zu finden und zu beheben. Abschließend wurden die Preise aus den Logfiles in eine einheitliche Excel-Tabelle übertragen, um eine Auswertung zu ermöglichen und die Preise übersichtlich in einer Datei aufzulisten. Dabei wurden auch die aufgetretenen Timeouts vermerkt, zu denen kein Preis extrahiert werden konnte.

Der Einfachheit halber werden in dieser Arbeit die Produkte jeweils mit dem Firmennamen gleichgesetzt.

4 Ergebnisse

Dieses Kapitel gliedert sich in drei Unterkapitel, in denen die Ergebnisse der Datenanalyse dargestellt, diskutiert und vor dem Hintergrund der theoretischen Grundlagen interpretiert werden.

4.1 Auswertung

Während der Datenerhebung vom 04.04.2024 08:00 Uhr (erster Durchlauf) bis zum 11.05.2024 22:00 Uhr (letzter Durchlauf) wurden 111 Scraping-Durchläufe mit dem Scraper durchgeführt. 105 davon waren automatisiert, 6 wurden manuell durchgeführt. Jeder Durchlauf besteht aus 6 Preisabfragen (3 Endgeräte, 2 Browsermodi) für 8 Produkte. In Tabelle 7 sind die Anzahl und die Zeitpunkte der einzelnen Durchgänge im Beobachtungszeitraum aufgeführt.

Aufgrund eines Fehlers im 0-Uhr-Trigger des Aufgabenplaners wurde am 05.04.2024 um Punkt Mitternacht kein Scraping-Lauf durchgeführt, weshalb dieser manuell um 00:15 Uhr nachgeholt wurde. Der nächste manuelle Durchlauf wurde am 09.04.2024 um 16:15 Uhr durchgeführt, da es bei den drei vorherigen Scraping-Durchläufen bei einer Webseite (Urlaubsguru) zu kontinuierlichen Timeouts kam, weshalb der geplante Durchlauf um 16:00 Uhr ebenfalls manuell gestoppt wurde, um das Problem vor dem nächsten Durchlauf zu beheben. Am 10.04.2024 gab es Probleme mit der Internetverbindung, weshalb der Scraping-Durchgang um 16:00 Uhr nicht durchgeführt und um 16:57 Uhr nachgeholt wurde. Am 11.04.2024 gab es ebenfalls Probleme mit dem Internet, weshalb die Scraping-Durchläufe um 08:00 Uhr und 16:00 Uhr nicht durchgeführt wurden. Als dies bei der manuellen Kontrolle festgestellt wurde, wurde ein Durchlauf um 18:24 Uhr gestartet, um sicherzustellen, dass der Scraper ordnungsgemäß funktioniert. Am 12.04.2024 kam es beim Durchlauf um 08:00 Uhr zu Zeitüberschreitungen im Inkognito-Modus einer Webseite (Shop Apotheke), weshalb um 09:17 Uhr ein manueller Durchlauf gestartet wurde, bei dem dieses Problem nicht mehr auftrat. Am 16.04.2024 wurde um 16:55 Uhr ein weiterer manueller Durchlauf gestartet, da sich der XPath eines Preises verändert hat und somit im Skript aktualisiert werden musste. Auch am 06.05.2024 und 07.05.2024 gab es Probleme mit der Netzwerkverbindung, so dass jeweils zwei Scraping-Durchläufe nicht gestartet werden konnten.

Tabelle 7 – Anzahl und Zeitpunkte der Scraping-Durchläufe 04.04.2024 - 11.05.2024

Datum	Scraping-DL	DL 1	DL 2	DL 3	DL 4	manuell
04.Apr	2	08:00	16:00	x	x	x
05.Apr	3	x	08:00	16:00	x	00:15
06.Apr	3	00:00	08:00	16:00	x	x
07.Apr	3	00:00	08:00	16:00	x	x
08.Apr	3	00:00	08:00	16:00	x	x
09.Apr	3	00:00	08:00	x	x	16:15
10.Apr	3	00:00	08:00	x	x	16:57
11.Apr	2	00:00	x	x	x	18:24
12.Apr	4	00:00	08:00	16:00	x	09:17
13.Apr	3	00:00	08:00	16:00	x	x
14.Apr	3	00:00	08:00	16:00	x	x
15.Apr	3	00:00	08:00	16:00	x	x
16.Apr	4	00:00	08:00	16:00	x	16:55
17.Apr	3	00:00	08:00	16:00	x	x
18.Apr	3	00:00	08:00	16:00	x	x
19.Apr	3	00:00	08:00	16:00	x	x
20.Apr	3	00:00	08:00	16:00	x	x
21.Apr	3	00:00	08:00	16:00	x	x
22.Apr	3	00:00	08:00	16:00	x	x
23.Apr	3	00:00	08:00	16:00	x	x
24.Apr	3	00:00	08:00	16:00	x	x
25.Apr	3	00:00	08:00	16:00	x	x
26.Apr	3	00:00	08:00	16:00	x	x
27.Apr	3	00:00	08:00	16:00	x	x
28.Apr	3	00:00	08:00	16:00	x	x
29.Apr	3	00:00	08:00	16:00	x	x
30.Apr	3	00:00	08:00	16:00	x	x
01.Mai	3	00:00	08:00	16:00	x	x
02.Mai	3	00:00	08:00	16:00	x	x
03.Mai	3	00:00	08:00	16:00	x	x
04.Mai	4	00:00	08:00	14:00	22:00	x
05.Mai	3	06:00	14:00	22:00	x	x
06.Mai	1	06:00	x	x	x	x
07.Mai	1	x	x	22:00	x	x
08.Mai	3	06:00	14:00	22:00	x	x
09.Mai	3	06:00	14:00	22:00	x	x
10.Mai	3	06:00	14:00	22:00	x	x
11.Mai	3	06:00	14:00	22:00	x	x

Anmerkung: DL = Durchlauf, x = kein Scraping-DL

Insgesamt wurden somit 5328 Preisabfragen (666 je Produkt) durchgeführt. Dabei kam es in 322 Fällen (6,04%) zu Timeouts beim Aufruf der Webseiten, so dass von einem oder mehreren User Agents kein Preis ausgelesen werden konnte. In der folgenden Tabelle 8

wird die Anzahl der Timeouts den jeweiligen Produkten zugeordnet. Es fällt auf, dass ein Großteil der Timeouts auf den beiden Webseiten von electronic4you und Urlaubsguru auftritt. Während die Timeouts bei Urlaubsguru bis auf einen Zeitraum von 16 Stunden (08.04.2024 16:00 Uhr bis 09.04.2024 08:00 Uhr), in dem die Webseite für den Scraper nicht erreichbar war, immer um 00:00 Uhr bzw. 22:00 Uhr auftraten, trat bei electronic4you ab dem 13.04.2024 (00:00 Uhr) ein Problem auf, das nicht behoben werden konnte und täglich zu zwei Timeouts im normalen Modus des Browsers führte. Während die erste Abfrage im normalen Modus funktionierte, ergaben Überprüfungen, dass der Scraper beim zweiten Aufruf der entsprechenden URL mit dem zweiten Endgerät, aber der gleichen IP, auf eine andere Seite der Homepage weitergeleitet wurde, auf der das interessierende Produkt nicht zu sehen war. Eine manuelle Überprüfung von einem anderen Endgerät aus ergab das gleiche Ergebnis. Offenbar haben die mitgesendeten Cookies dazu geführt, dass die Webseite den Scraper als wiederholten Zugriff erkannte und auf eine andere Seite weiterleitete. Im Inkognito-Modus war dieses Phänomen demnach nicht zu beobachten.

Tabelle 8 - Timeouts Datenerhebung

Produkt	Timeouts in %	Timeout Anzahl		Browsermodus
Shop Apotheke	0,41%	22	0	Normal
			22	Inkognito
Media Markt	0,17%	9	4	Normal
			5	Inkognito
Universal	0,45%	24	12	Normal
			12	Inkognito
Europcar	0,02%	1	1	Normal
			0	Inkognito
electronic4you	3,55%	189	174	Normal
			15	Inkognito
Otto Versand	0,08%	4	1	Normal
			3	Inkognito
Ryanair	0,06%	3	0	Normal
			3	Inkognito
Urlaubsguru	1,31%	70	44	Normal
			26	Inkognito
Gesamt	6,04 %	322	236	Normal
			86	Inkognito

160 Timeouts traten auf, da die Webseite vom Scraper nicht geladen werden konnte und es somit zu Zeitüberschreitungen kam. 30 dieser Timeouts sind darauf zurückzuführen, dass sich jeweils ein XPath eines Produktes geändert hat und korrigiert werden musste. 162 Timeouts sind auf das beschriebene Problem ab dem 13.04. bei electronic4you zurückzuführen.

Um die Unterschiede in den beobachteten Produktpreisen je nach User Agent und Browsermodus interpretieren und verlässliche Aussagen treffen zu können, ist es wichtig, mit präzisen und verlässlichen Daten zu arbeiten. Daher wurde bei der Datenbereinigung ein strenger Ansatz gewählt, bei dem ein gesamter Scraping-Durchlauf eines Produktes gelöscht wurde, sobald auch nur ein Wert in einem Datensatz (gleiches Datum, gleiche Uhrzeit) aufgrund eines Timeouts nicht vorhanden war. Auf diese Weise wurde sichergestellt, dass nur vollständige und vergleichbare Datensätze in die Analyse einfließen. Damit soll die Glaubwürdigkeit und Genauigkeit der Ergebnisse bestmöglich gewährleistet werden. Somit sind alle analysierten Datenpunkte vollständig, was bedeutet, dass alle Vergleiche zwischen den User Agents und Browsermodi auf der gleichen Datenbasis basieren und somit Konsistenz und Reliabilität gegeben ist. Durch das konsequente Löschen unvollständiger Datensätze wird verhindert, dass eine Verzerrung der Daten das Analyseergebnis beeinflusst.

Für die Auswertung der gescrapten Preise wurden die beobachteten Preise in acht CSV-Datensätze aufgeteilt, einer für jedes Produkt/jede Dienstleistung. Jeder Datensatz enthält fünf Spalten (Datum, Uhrzeit, UserAgent, BrowserModus und Preis). Die statistische Auswertung erfolgte in R. Für jeden Datensatz wurden die Mittelwerte der Preise für jeden User Agent und jeden Browsermodus berechnet. Um die Preisentwicklung im Zeitverlauf visuell zu überprüfen, wurden Boxplots und Liniendiagramme erstellt. Zusätzlich wurde eine ANOVA durchgeführt, um zu testen, ob es signifikante Unterschiede in den mittleren Preisen zwischen den verschiedenen User Agents und Browsermodi gibt.

Für die Auswertung der gesammelten Preise der Shop Apotheke Webseite wurden 582 Preise aus 97 Scraping-Durchläufen herangezogen. 14 Durchläufe wurden aufgrund fehlender Daten von der Auswertung ausgeschlossen (Tabelle 9):

Tabelle 9 - Shop Apotheke: Ausgeschlossene Scraping-DL

Produkt	Datum	Uhrzeit
Shop Apotheke	10.Apr	00:00
	10.Apr	08:00
	10.Apr	16:57
	11.Apr	00:00
	12.Apr	08:00
	08.Mai	06:00
	08.Mai	14:00
	08.Mai	22:00
	09.Mai	22:00
	10.Mai	06:00
	10.Mai	14:00
	10.Mai	22:00
	11.Mai	06:00
	11.Mai	22:00

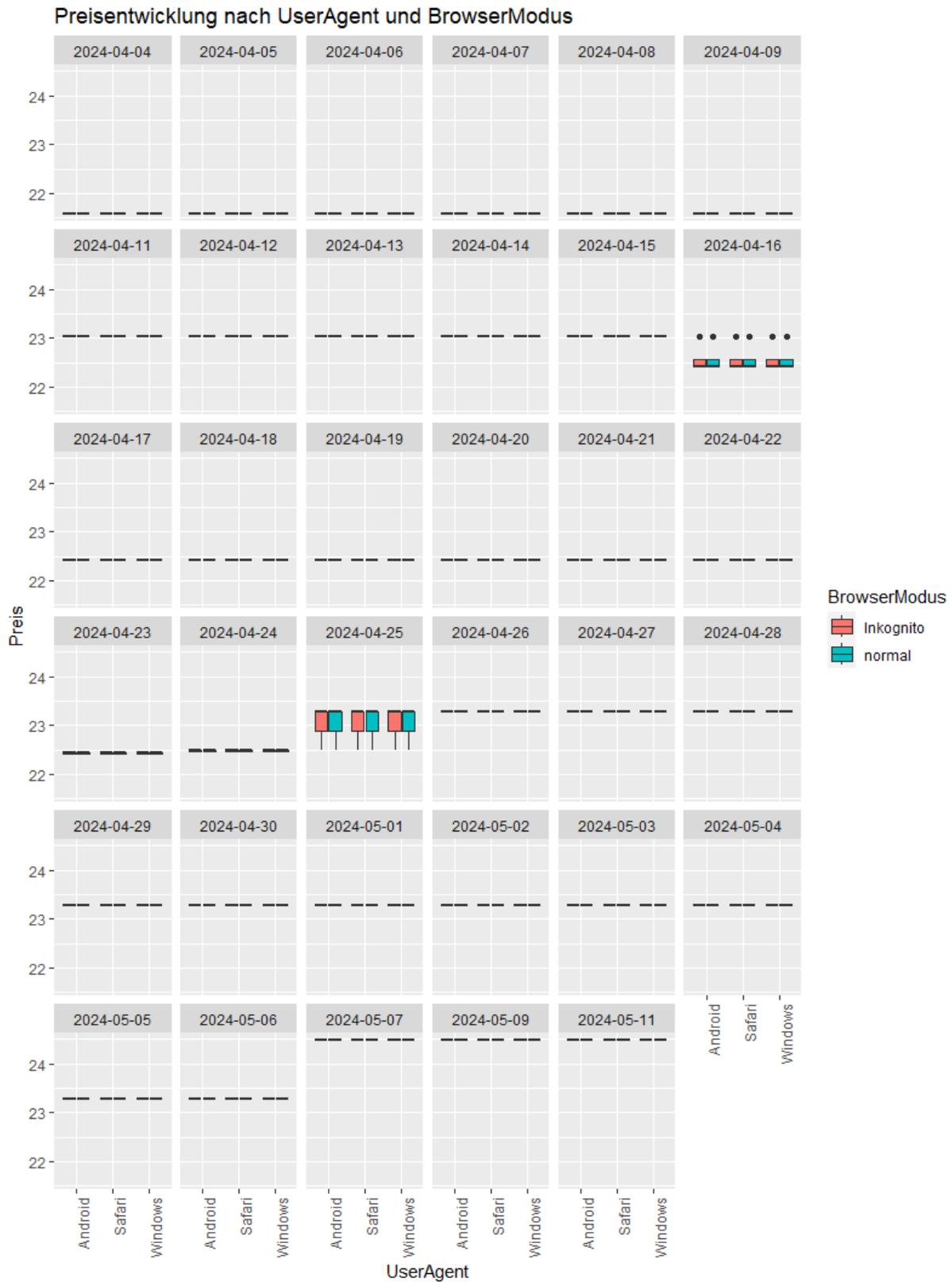
Die Auswertung der Mittelwerte für die Browsermodi und User Agents ergab die in Tabelle 10 dargestellten Werte.

Tabelle 10 - Shop Apotheke: Preismittelwerte

Produkt	User Agent	Browsermodus	Mittelwert Preis (€)
Shop Apotheke	Android	Normal	22,75
	Safari	Normal	22,75
	Windows	Normal	22,75
	Android	Inkognito	22,75
	Safari	Inkognito	22,75
	Windows	Inkognito	22,75

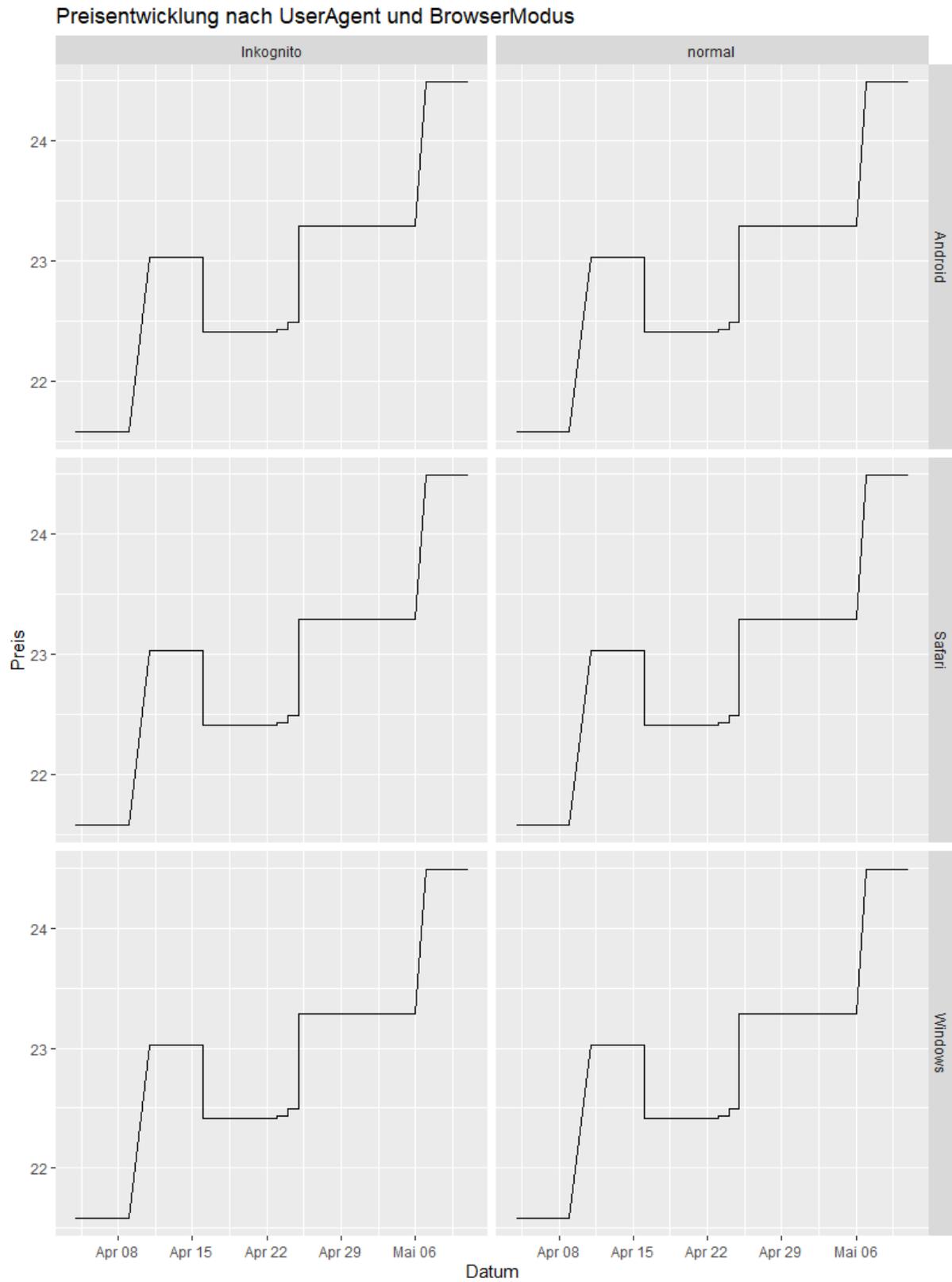
Die Auswertungen der ANOVA ergeben für die abhängige Variable (Preis) in Verbindung mit den unabhängigen Variablen User Agent, Browsermodus und deren Interaktion jeweils einen p-Wert von 1. Die Boxplots (Abbildung 11) und Liniendiagramme (Abbildung 12) der Preise im Zeitverlauf sehen wie folgt aus:

Abbildung 11 - Shop Apotheke: Boxplots



Anmerkung: Erstellt mit R

Abbildung 12 - Shop Apotheke: Liniendiagramme



Anmerkung: Erstellt mit R

Für die Auswertung der gesammelten Preise der electronic4you Webseite wurden 150 Preise aus 25 Scraping-Durchläufen herangezogen. 86 Durchläufe wurden aufgrund fehlender Daten von der Auswertung ausgeschlossen (Tabelle 11):

Tabelle 11 - electronic4you: Ausgeschlossene Scraping-DL

Produkt	Datum	Uhrzeit
electronic4you	10.Apr	00:00
	13.Apr – 11.Mai	alle

Die Auswertung der Mittelwerte für die Browsermodi und User Agents ergab die in Tabelle 12 dargestellten Werte.

Tabelle 12 - electronic4you: Preismittelwerte

Produkt	User Agent	Browsermodus	Mittelwert Preis (€)
electronic4you	Android	Normal	23,80
	Safari	Normal	23,80
	Windows	Normal	23,80
	Android	Inkognito	23,80
	Safari	Inkognito	23,80
	Windows	Inkognito	23,80

Die Auswertungen der ANOVA ergeben für die abhängige Variable (Preis) in Verbindung mit den unabhängigen Variablen User Agent, Browsermodus und deren Interaktion folgende p-Werte: User Agent 0.370; Browsermodus 0.319; Interaktion Browsermodus und User Agent 0.370. Die Boxplots und Liniendiagramme sind im Anhang beigelegt.

Für die Auswertung der gesammelten Preise der Media Markt Webseite wurden 642 Preise aus 107 Scraping-Durchläufen herangezogen. 4 Durchläufe wurden aufgrund fehlender Daten von der Auswertung ausgeschlossen (Tabelle 13):

Tabelle 13 - Media Markt: Ausgeschlossenen Scraping-DL

Produkt	Datum	Uhrzeit
Media Markt	04.Apr	16:00
	12.Apr	08:00
	16.Apr	16:00
	21.Apr	16:00

Die Auswertung der Mittelwerte für die Browsermodi und User Agents ergab die in Tabelle 14 dargestellten Werte.

Tabelle 14 – Media Markt: Preismittelwerte

Produkt	User Agent	Browsermodus	Mittelwert Preis (€)
Media Markt	Android	Normal	343,48
	Safari	Normal	343,48
	Windows	Normal	343,48
	Android	Inkognito	343,48
	Safari	Inkognito	343,48
	Windows	Inkognito	343,48

Die Auswertungen der ANOVA ergeben für die abhängige Variable (Preis) in Verbindung mit den unabhängigen Variablen User Agent, Browsermodus und deren Interaktion jeweils einen p-Wert von 1. Die Boxplots und Liniendiagramme sind im Anhang beigefügt.

Für die Auswertung der gesammelten Preise der Otto Versand Webseite wurden 654 Preise aus 109 Scraping-Durchläufen herangezogen. 2 Durchläufe wurden aufgrund fehlender Daten von der Auswertung ausgeschlossen (Tabelle 15):

Tabelle 15 – Otto Versand: Ausgeschlossenen Scraping-DL

Produkt	Datum	Uhrzeit
Otto Versand	11.Apr	00:00
	20.Apr	00:00

Die Auswertung der Mittelwerte für die Browsermodi und User Agents ergab die in Tabelle 16 dargestellten Werte.

Tabelle 16 – Otto Versand: Preismittelwerte

Produkt	User Agent	Browsermodus	Mittelwert Preis (€)
Otto Versand	Android	Normal	426,60
	Safari	Normal	426,60
	Windows	Normal	426,60
	Android	Inkognito	426,60
	Safari	Inkognito	426,60
	Windows	Inkognito	426,60

Die Auswertungen der ANOVA ergeben für die abhängige Variable (Preis) in Verbindung mit den unabhängigen Variablen User Agent, Browsermodus und deren Interaktion jeweils einen p-Wert von 1. Die Boxplots und Liniendiagramme sind im Anhang beigefügt.

Für die Auswertung der gesammelten Preise der Europcar Webseite wurden 660 Preise aus 110 Scraping-Durchläufen herangezogen. 1 Durchlauf wurden aufgrund fehlender Daten von der Auswertung ausgeschlossen (Tabelle 17):

Tabelle 17 – Europcar: Ausgeschlossenen Scraping-DL

Produkt	Datum	Uhrzeit
Europcar	16.Apr	16:00

Die Auswertung der Mittelwerte für die Browsermodi und User Agents ergab die in Tabelle 18 dargestellten Werte.

Tabelle 18 - Europcar: Preismittelwerte

Produkt	User Agent	Browsermodus	Mittelwert Preis (€)
Europcar	Android	Normal	44,64
	Safari	Normal	44,64
	Windows	Normal	44,64
	Android	Inkognito	44,64
	Safari	Inkognito	44,64
	Windows	Inkognito	44,64

Die Auswertungen der ANOVA ergeben für die abhängige Variable (Preis) in Verbindung mit den unabhängigen Variablen User Agent, Browsermodus und deren Interaktion folgende p-Werte: User Agent 0.368; Browsermodus 0.318; Interaktion Browsermodus und User Agent 0.368. Die Boxplots und Liniendiagramme sind im Anhang beigefügt.

Für die Auswertung der gesammelten Preise der Ryanair Webseite wurden 660 Preise aus 110 Scraping-Durchläufen herangezogen. 1 Durchlauf wurden aufgrund fehlender Daten von der Auswertung ausgeschlossen (Tabelle 19):

Tabelle 19 –Ryanair: Ausgeschlossenen Scraping-DL

Produkt	Datum	Uhrzeit
Ryanair	11.Apr	00:00

Die Auswertung der Mittelwerte für die Browsermodi und User Agents ergab die in Tabelle 20 dargestellten Werte.

Tabelle 20 - Ryanair: Preismittelwerte

Produkt	User Agent	Browsermodus	Mittelwert Preis (€)
Ryanair	Android	Normal	78,99
	Safari	Normal	78,99
	Windows	Normal	78,99
	Android	Inkognito	78,99
	Safari	Inkognito	78,99
	Windows	Inkognito	78,99

Die Auswertungen der ANOVA ergeben für die abhängige Variable (Preis) in Verbindung mit den unabhängigen Variablen User Agent, Browsermodus und deren Interaktion folgende p-Werte: User Agent 0.368; Browsermodus 0.318; Interaktion Browsermodus und User Agent 0.368. Die Boxplots und Liniendiagramme sind im Anhang beigefügt.

Für die Auswertung der gesammelten Preise der Universal Webseite wurden 642 Preise aus 107 Scraping-Durchläufen herangezogen. 4 Durchläufe wurden aufgrund fehlender Daten von der Auswertung ausgeschlossen (Tabelle 21):

Tabelle 21 - Universal: Ausgeschlossenen Scraping-DL

Produkt	Datum	Uhrzeit
Universal	17.Apr	16:00
	03.Mai	16:00
	04.Mai	00:00
	04.Mai	08:00

Die Auswertung der Mittelwerte für die Browsermodi und User Agents ergab die in Tabelle 22 dargestellten Werte.

Tabelle 22 – Universal: Preismittelwerte

Produkt	User Agent	Browsermodus	Mittelwert Preis (€)
Universal	Android	Normal	1.008,76
	Safari	Normal	1.008,76
	Windows	Normal	1.008,76
	Android	Inkognito	1.008,76
	Safari	Inkognito	1.008,76
	Windows	Inkognito	1.008,76

Die Auswertungen der ANOVA ergeben für die abhängige Variable (Preis) in Verbindung mit den unabhängigen Variablen User Agent, Browsermodus und deren Interaktion jeweils einen p-Wert von 1. Die Boxplots und Liniendiagramme sind im Anhang beigefügt.

Für die Auswertung der gesammelten Preise der Urlaubsguru Webseite wurden 516 Preise aus 86 Scraping-Durchläufen herangezogen. 25 Durchläufe wurden aufgrund fehlender Daten von der Auswertung ausgeschlossen (Tabelle 23):

Tabelle 23 - Urlaubsguru: Ausgeschlossenen Scraping-DL

Produkt	Datum	Uhrzeit
Urlaubsguru	08.Apr	16:00
	09.Apr	00:00
	09.Apr	08:00
	11.Apr	00:00
	12.Apr	00:00
	13.Apr	00:00
	14.Apr	00:00
	15.Apr	00:00
	16.Apr	00:00
	17.Apr	00:00
	19.Apr	00:00
	20.Apr	00:00
	21.Apr	00:00
	22.Apr	00:00
	23.Apr	00:00
	24.Apr	00:00
	25.Apr	00:00
	26.Apr	00:00
	28.Apr	00:00
	30.Apr	00:00
	01.Mai	00:00
	03.Mai	00:00
	04.Mai	00:00
	07.Mai	22:00
	08.Mai	22:00

Die Auswertung der Mittelwerte für die Browsermodi und User Agents ergab die in Tabelle 24 dargestellten Werte.

Tabelle 24 – Urlaubsguru: Preismittelwerte

Produkt	User Agent	Browsermodus	Preismittelwert (€)
Urlaubsguru	Android	Normal	733,46
	Safari	Normal	733,67
	Windows	Normal	733,46
	Android	Inkognito	732,26
	Safari	Inkognito	732,26
	Windows	Inkognito	732,26

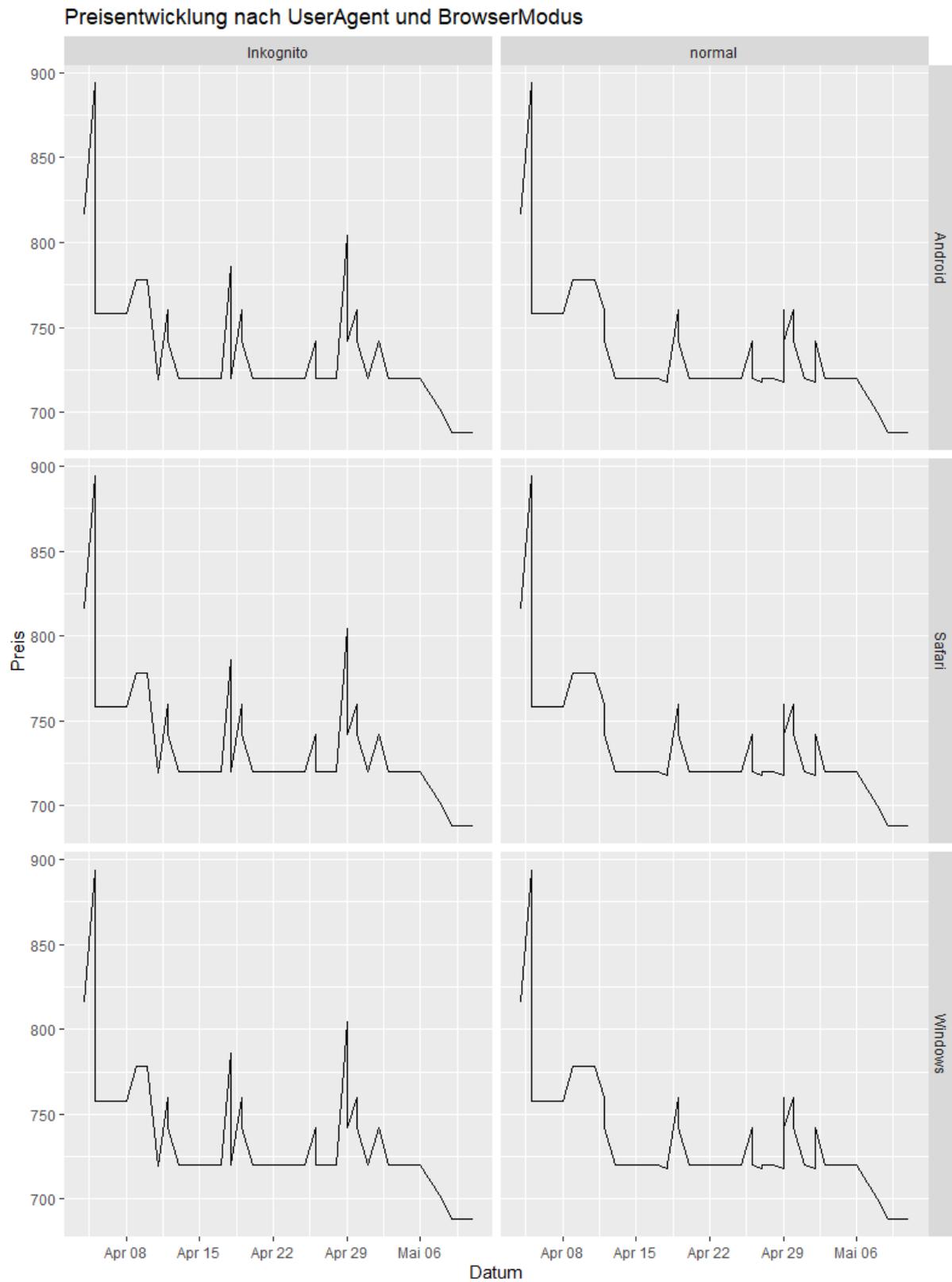
Die Auswertungen der ANOVA ergeben für die abhängige Variable (Preis) in Verbindung mit den unabhängigen Variablen User Agent, Browsermodus und deren Interaktion folgende p-Werte: User Agent 0.999; Browsermodus 0.658; Interaktion Browsermodus und User Agent 0.999. Die Boxplots und Liniendiagramme sind in den Abbildungen 13 und 14 dargestellt.

Abbildung 13 - Urlaubsguru: Boxplots



Anmerkung: Erstellt in R

Abbildung 14 - Urlaubsguru: Liniendiagramme



Anmerkung: Erstellt in R

4.2 Präsentation, Diskussion und Limitation

Die Analyse der quantitativen Beobachtung zeigt, dass es im Beobachtungszeitraum keine signifikanten Preisunterschiede bei den untersuchten Produkten/Dienstleistungen gab. Die unabhängigen Variablen (Browsermodus, User Agent sowie das Zusammenspiel dieser beiden Variablen) hatten somit keinen Einfluss auf die abhängige Variable Preis. Die Mittelwerte der Preise waren bei sieben der acht Produkte/Dienstleistungen identisch. Lediglich auf der Website von Urlaubsguru konnten vereinzelt Preisunterschiede festgestellt werden.

Die in dieser Arbeit aufgestellten Hypothesen lauteten:

- H0 (Nullhypothese): Es gibt keine Unterschiede in den Preisen zu den verschiedenen Zeitpunkten zwischen den verschiedenen User Agents und den verschiedenen Browsermodi.
- H1 (Alternativhypothese): Es gibt Unterschiede in den Preisen zu den verschiedenen Zeitpunkten zwischen den verschiedenen User Agents und den verschiedenen Browsermodi.

Die Ergebnisse der ANOVA stützen die Nullhypothese (H0). Für die abhängige Variable Preis in Verbindung mit den unabhängigen Variablen User Agent, Browsermodus und deren Interaktion ergeben sich jeweils p-Werte, die auf eine fehlende Signifikanz hinweisen ($p\text{-Wert} = > 0,05$). Dies bedeutet, dass die Unterschiede in den durchschnittlichen Preisen statistisch nicht signifikant sind.

Die Mittelwerte der Preise für die verschiedenen User Agents und Browsermodi zeigen keine signifikanten Unterschiede, wie die Darstellung der Preismittelwerte in Kapitel 4.1 verdeutlicht. Dies bestätigt, dass die Preisgestaltung für die untersuchten Produkte unabhängig vom verwendeten Browsermodus oder User Agent konsistent ist.

Die Visualisierung mittels Boxplots und Liniendiagrammen unterstreicht diese Ergebnisse. Für sieben der acht Produkte/Dienstleistungen zeigen die Diagramme identische Verläufe der Preisentwicklung für alle User Agents und Browsermodi im Beobachtungszeitraum. Lediglich bei Urlaubsguru gibt es einige Ausnahmen, die jedoch nicht ausreichen, um eine systematische Preisdiskriminierung nachzuweisen.

Bei der Webseite Urlaubsguru wurden am 11.04.2024 um 18:24 Uhr Preisunterschiede zwischen dem normalen Modus (778 €) und dem Inkognito-Modus (719,34 €) festgestellt.

Ähnliche Unterschiede traten am 18.04.2024 (Normalmodus: 718 €; Inkognito-Modus: 786 €), am 29.04. (Normalmodus: 718 €; Inkognito-Modus: 804,48 €) und am 02.05.2024 auf (Normalmodus: 718 €; Inkognito-Modus: 742 €). Auch am 29.04.2024 um 08:00 Uhr gab es Preisunterschiede, allerdings zeigte sich hier, dass der Preis für den User Agent Safari im Inkognito-Modus identisch mit dem Preis im normalen Modus war (für alle User Agents 760 €), während für die User Agents Windows und Android im Inkognito-Modus ein anderer Preis erhoben wurde (742 €). Wie den Rohdaten zu entnehmen ist, traten bei Urlaubsguru auch an anderen Tagen, insbesondere um 00:00 Uhr, Preisunterschiede vor allem zwischen dem normalen und dem Inkognito-Modus auf, da hier jedoch Timeouts bei einzelnen User Agents auftraten, wurden diese Datensätze aus der Analyse ausgeschlossen. Diese Timeouts könnten auch ein Hinweis darauf sein, dass die Preisunterschiede auf regelmäßige Preisaktualisierungen (dynamische Preisgestaltung) der Webseite um ca. 00:00 Uhr zurückzuführen sind. So kann es sein, dass die Webseite kurz nach 00:00 Uhr aktualisiert wird, so dass im Normalmodus noch der alte Preis angezeigt wird, im Inkognito-Modus jedoch der neue Preis, da der Web Scraper die Preise nicht gleichzeitig, sondern nacheinander abfragt. Die Zeitüberschreitungen um Mitternacht können daher auch darauf zurückzuführen sein, dass die Webseite während dieser Aktualisierung nicht erreichbar ist.

Insgesamt zeigen die Ergebnisse, dass im Untersuchungszeitraum und für die untersuchten Produkte/Dienstleistungen keine systematische Preisdiskriminierung aufgrund des Browsermodus und des User Agents nachgewiesen werden konnte. Dies deutet darauf hin, dass die untersuchten E-Commerce-Plattformen keine derartigen Praktiken auf der Grundlage des verwendeten Endgeräts anwenden. Die spezifischen Beobachtungen bei Urlaubsguru könnten auf eine dynamische Preisgestaltung oder technische Schwierigkeiten beim Scraping hindeuten, lassen aber ebenfalls keine Preisdiskriminierung erkennen.

Eine Einschränkung dieser Studie ist der relativ kurze Beobachtungszeitraum. Aufgrund begrenzter Ressourcen und Zeit war es nicht möglich, eine längerfristige Datenerhebung durchzuführen. Ein längerer Beobachtungszeitraum könnte helfen, saisonale oder kurzfristige Schwankungen in der Preisgestaltung zu identifizieren und zu analysieren.

Die Studie konzentrierte sich auf die Untersuchung der Variablen User Agent und Browsermodus. Die Konzentration auf diese beiden Variablen ist eine bewusste

Entscheidung, die auch auf theoretischen Überlegungen und pragmatischen Erwägungen der technischen Machbarkeit beruht. Dieser Ansatz wurde auch bewusst gewählt, um andere einfließende Variablen wie den geografischen Standort, das Surfverhalten oder demografische Daten auszuschließen, obwohl diese Variablen für die Preisgestaltung und eine mögliche Preisdiskriminierung relevant sein können.

Die Entwicklung und Implementierung des Web-Scraping-Tools erforderte eine steile Lernkurve, da zu Beginn keine Vorkenntnisse in den Bereichen Programmierung und Web Scraping vorhanden waren. Das Wissen und die Herangehensweise für die Entwicklung des Web Scrapers mussten daher erst erworben werden. Trotz der erfolgreichen Implementierung traten technische Herausforderungen wie Timeouts und Änderungen im Quellcode der Webseiten auf, die die Datenerhebung beeinträchtigten. Aufgrund dieser Herausforderungen mussten die Daten vor der Auswertung bereinigt werden, was sich auf die Qualität und Vollständigkeit der Daten auswirkte.

Die verfügbaren personellen und finanziellen Ressourcen waren begrenzt, was die Möglichkeiten für eine umfassendere Datenerhebung und -analyse einschränkte.

4.3 Interpretation

Die Ergebnisse der quantitativen Studie zeigen, dass es im Beobachtungszeitraum keine signifikanten Preisunterschiede für die untersuchten Produkte und Dienstleistungen gab. Dies deutet darauf hin, dass die Preisbildung im Internet für die untersuchten Fälle unabhängig vom Browsermodus und User Agent erfolgt. Die untersuchten E-Commerce-Plattformen scheinen keine Preisdiskriminierung aufgrund dieser Parameter vorzunehmen. Dies könnte auf verschiedene Faktoren zurückzuführen sein:

- **Technologische Grenzen:** Die technologische Infrastruktur und die Datenanalyse der Händler sind möglicherweise noch nicht ausgereift genug, um eine feingranulare Form der Preisdiskriminierung zu unterstützen. Der Einsatz von Algorithmen und Big Data erfordert, wie in Kapitel 2.2 beschrieben erhebliche Investitionen und technisches Fachwissen, das eventuell nicht von allen Unternehmen aufgebracht werden kann.
- **Regulatorische und ethische Bedenken:** Unternehmen könnten sich aus regulatorischen oder ethischen Gründen gegen Preisdiskriminierung entscheiden. In vielen Ländern gibt es strenge Datenschutzgesetze, die den Umgang mit

personenbezogenen Daten regeln. Zudem könnten Unternehmen negative Reaktionen der Verbraucher*innen fürchten, die Preisdiskriminierung als unfair empfinden könnten. Diese ethischen Bedenken wurden in Kapitel 2.3 diskutiert, wobei die potenziellen negativen Auswirkungen auf das Vertrauen der Kund*innen hervorgehoben wurden.

- **Ökonomische Auswirkungen:** Preisdiskriminierung könnte auch wirtschaftliche Risiken mit sich bringen, wie den Verlust des Vertrauens der Kund*innen oder rechtliche Konsequenzen. Unternehmen könnten daher den potenziellen wirtschaftlichen Nutzen, wie in Kapitel 2.3 dargelegt, gegen die Risiken abwägen und sich gegen eine solche Praxis entscheiden.

Die Interpretation der Ergebnisse deutet darauf hin, dass die untersuchten E-Commerce-Plattformen möglicherweise keine Preisdiskriminierung anwenden, weil ihnen entweder die technischen Möglichkeiten fehlen oder weil sie die wirtschaftlichen und ethischen Risiken zu hoch einschätzen. Es ist auch möglich, dass Unternehmen bewusst auf Preisdiskriminierung verzichten, um ihre Marktposition und ihren Ruf zu schützen. Dies könnte insbesondere in wettbewerbsintensiven Märkten der Fall sein, wo das Vertrauen und die Loyalität der Kund*innen entscheidend sind (Elegido, 2011, S. 634 f., S. 637, S. 639; Sun & Tang, 2021, S. 2, S. 10).

Die erhobenen Daten beschränken sich auf einen bestimmten Zeitraum und eine begrenzte Anzahl von Produkten und Plattformen in Österreich. Ein längerer Beobachtungszeitraum und eine breitere Datenbasis könnten zu anderen Ergebnissen führen. Saisonale Effekte oder Sonderaktionen von Produkten und Dienstleistungen wurden nicht berücksichtigt, waren aber für die Zielsetzung auch nicht relevant und wären nur dann von Interesse gewesen, wenn aktionsbedingte Preise nur für einzelne Variablenkategorien beobachtet worden wären.

5 Zusammenfassung und Schlussfolgerung

Die vorliegende Arbeit untersucht das Phänomen der Big-Data-basierten Preisdiskriminierung in der digitalen Welt, insbesondere im Online-Handel. Ziel der Arbeit ist es zu untersuchen, inwieweit Preisdiskriminierung im Online-Handel praktiziert wird und welche Faktoren einen Einfluss darauf haben. Dazu wurde als Grundlage folgende Forschungsfrage formuliert: *Wie kommen unterschiedliche Preise für ein und dasselbe Produkt für Verbraucher*innen im Internet zustande und was sind die ökonomischen Auswirkungen davon?*

Die Arbeit beginnt mit einer Einführung in das Thema Preisdiskriminierung in der digitalen Welt, um die Relevanz des Themas aufzuzeigen. So werden historische Bezüge erläutert, wo und in welcher Branche Online-Preissetzungsmechaniken erstmals beobachtet wurden (Spann et al., 2005, S. 53 f.) und dass sich diese Praxis in den letzten Jahren, aufgrund der zunehmenden Datenmengen (Big Data) und der technischen Möglichkeiten zu deren Verarbeitung, immer weiter verbreitet hat (Chen et al., 2016, S. 1339 f.). Ein Praxisbeispiel und Erkenntnisse aus wissenschaftlichen Arbeiten und Studien zum Thema Online-Preisdiskriminierung helfen, die Relevanz des Themas zu unterstreichen.

Die allgemeinen theoretischen Grundlagen der Preisdiskriminierung werden in Kapitel 2 erläutert, einschließlich der Definition von Preisdiskriminierung und der Abgrenzung zu verwandten Begriffen wie Preisstreuung und dynamische Preisgestaltung, der notwendigen Voraussetzungen für Preisdiskriminierung, der verschiedenen Grade von Preisdiskriminierung und praktischer Beispiele. Dabei wird unter anderem auf die Grundlagenliteratur von Scherer & Ross (1990) zum Thema Preisdiskriminierung zurückgegriffen. Die Arbeit konzentriert sich anschließend auf die technologischen Grundlagen der Preisdiskriminierung auf der Basis von Big Data. Big Data wird dabei definiert als Informationen, die sich durch hohe Geschwindigkeit, großes Volumen und große Vielfalt auszeichnen (De Mauro et al., 2015, S. 103; Trübner & Mühlichen, 2019, S. 143 f.). Technologische Fortschritte ermöglichen eine fein abgestufte Form der Preisdiskriminierung, die auf der Analyse großer Datenmengen basiert (Data Mining). Unternehmen können Algorithmen und maschinelles Lernen einsetzen, um komplexe Muster in den Daten zu erkennen und personalisierte Preise für ihre Kund*innen festzulegen. Es werden vier Data-Mining-Methoden vorgestellt, mit denen es theoretisch möglich ist, nützliche Informationen aus vorhandenen Daten zu extrahieren und Preisdiskriminierung in Echtzeit durchzuführen. Dieser theoretische Prozess wird

anhand von einfachen Beispielen veranschaulicht. Im nachfolgenden Abschnitt werden die rechtlichen Bestimmungen und ökonomischen Implikationen dargestellt, die bei der Anwendung von Preisdiskriminierung zu berücksichtigen sind. Letztere werden auch im Zusammenhang mit verschiedenen Marktformen beleuchtet. Für Österreich sind dabei die rechtlichen Rahmenbedingungen der EU, wie z.B. die Verbraucherrechtlinien (UGPRL), oder nationale Gesetze, wie z.B. das Gleichbehandlungsgesetz zu beachten. (Das Europäische Parlament und der Rat der Europäischen Union, 2005; Rechtsinformationssystem des Bundes (RIS), 2023a). Die ökonomischen Auswirkungen von Preisdiskriminierung sind vielfältig und komplex (Poort & Zuiderveen Borgesius, 2019, S. 5; Scherer & Ross, 1990, S. 494 ff.; Steinberg, 2020, S. 114 f.; Townley et al., 2017, S. 688 f.). Die Unternehmen müssen sorgfältig abwägen, ob die potenziellen Gewinne die möglichen Risiken und Kosten überwiegen. Zu den wirtschaftlichen Risiken gehören der Verlust des Vertrauens der Kund*innen, mögliche rechtliche Konsequenzen und negative Reaktionen der Verbraucher*innen, die Preisdiskriminierung als unfair empfinden könnten (European Parliament. Directorate General for Internal Policies of the Union., 2022, S. 20; Konrad & Polk, 2020, S. 797; Poort & Zuiderveen Borgesius, 2019, S. 10 f.; Sun & Tang, 2021, S. 2; Townley et al., 2017, S. 721; Turow et al., 2005). Diese negativen Reaktionen können sich auf das Image und die Marktposition eines Unternehmens auswirken. Preisdiskriminierung kann jedoch auch positive wirtschaftliche Auswirkungen haben, indem sie Unternehmen in die Lage versetzt, ihren Umsatz zu maximieren, indem sie ihre Preise an die Zahlungsbereitschaft der Kund*innen anpasst. Dies kann zu einer effizienteren Preisgestaltung und einer besseren Ausschöpfung von Marktsegmenten führen. Preisdiskriminierung kann auch dazu beitragen, dass Unternehmen in einen Wettbewerb eintreten, von dem die Verbraucher*innen profitieren, da die Preise insgesamt sinken (Belleflamme et al., 2020, S. 1 ff., S. 5ff., S. 9ff.; Townley et al., 2017, S. 688 ff.). Darüber hinaus wird die Bedeutung digitaler Monopole wie Google, Apple, Amazon, Microsoft, Tencent, Baidu, Alibaba und Facebook beleuchtet. Diese Unternehmen haben durch ihre Marktmacht und den Zugang zu großen Datenmengen einen erheblichen Einfluss auf die Preisgestaltung und den Wettbewerb auf dem digitalen Markt (Jakobs, 2022, S. 54 ff., S. 67 ff., S. 73 ff., S. 99 ff., S. 266 f.).

Die Methodik des empirischen Teils wird einleitend in Kapitel 3 beschrieben. Für die vorliegende Arbeit wurde eine standardisierte quantitative Beobachtung durchgeführt, bei der die Preise von acht Produkten/Dienstleistungen auf verschiedenen E-Commerce-Seiten in Österreich über einen Zeitraum von ca. fünf Wochen regelmäßig erhoben wurden. Dabei wurden drei verschiedene Userprofile mittels Web Scraping verwendet.

Diese Userprofile wurden durch die verwendeten Endgeräte definiert und die Beobachtungen wurden sowohl im Normalmodus als auch im Inkognito-Modus des Browsers durchgeführt, um mögliche Unterschiede in der Preisgestaltung zu identifizieren. Die Beobachtung wurde mit einem speziell entwickelten Web Scraper durchgeführt. Abschließend wird erläutert, welche E-Commerce-Plattformen für die Beobachtung herangezogen wurden und wie die Datenerhebung geplant und durchgeführt wurde.

Kapitel 4 beschreibt die Auswertung, Präsentation, Limitation und Interpretation der Beobachtung und der dabei erhobenen Daten. Die Datenerhebung erfolgte über einen Zeitraum von ca. 5 Wochen und umfasste 111 Scraping-Durchläufe. Es wird erläutert, welche Daten für die statistische Analyse verwendet wurden, welche ausgeschlossen wurden und wo Probleme bei der Datenerhebung auftraten. In der Darstellung der Ergebnisse der quantitativen Beobachtung werden diese diskutiert und die Grenzen der Studie aufgezeigt. Die Analyse zeigt, dass es im Beobachtungszeitraum keine signifikanten Preisunterschiede bei den untersuchten Produkten und Dienstleistungen gab. Die unabhängigen Variablen (Browsermodus, User Agent sowie deren Interaktion) hatten keinen Einfluss auf die abhängige Variable Preis. Die Mittelwerte der Preise waren bei sieben der acht Produkte/Dienstleistungen identisch. Lediglich auf der Webseite von Urlaubsguru konnten vereinzelt Preisunterschiede festgestellt werden. Diese Unterschiede traten jedoch nicht regelmäßig auf und reichten nicht aus, um eine systematische Preisdiskriminierung nachzuweisen. Die untersuchten E-Commerce-Plattformen scheinen keine Preisdiskriminierung aufgrund dieser Parameter vorzunehmen. Mögliche Gründe für den Verzicht auf Preisdiskriminierung werden ebenfalls genannt.

Die Forschungsfrage dieser Arbeit lautet: *„Wie kommen unterschiedliche Preise für ein und dasselbe Produkt für Verbraucher*innen im Internet zustande und was sind die ökonomischen Auswirkungen davon?“* Im theoretischen Teil der Arbeit wurden die Mechanismen erläutert, die zu unterschiedlichen Preisen im Internet führen können. Im empirischen Teil der Arbeit wurde eine quantitative standardisierte Beobachtung durchgeführt, bei der die Preise von acht Produkten auf acht E-Commerce-Plattformen über einen Zeitraum von fünf Wochen mit drei verschiedenen User Agents (Windows, Safari, Android) und in zwei verschiedenen Browsermodi (normal und inkognito) erfasst wurden. Die Auswertung der gesammelten Daten ergab jedoch keine signifikanten Preisunterschiede aufgrund des verwendeten User Agents oder des Browsermodus. Dies

deutet darauf hin, dass die Preisgestaltung für die untersuchten Produkte und Dienstleistungen im Internet unabhängig von diesen Faktoren erfolgt. Es wurde festgestellt, dass die Preise für die gleichen Produkte über den Beobachtungszeitraum hinweg konsistent waren, unabhängig davon, welcher User Agent oder Browsermodus verwendet wurde. Die ökonomischen Auswirkungen dieser Befunde sind, dass Verbraucher*innen im Allgemeinen erwarten können, dass sie für die gleichen Produkte auf den untersuchten Plattformen die gleichen Preise zahlen, unabhängig von ihrem User Agent oder Browsermodus. Es ist jedoch wichtig zu beachten, dass diese Ergebnisse auf der spezifischen Stichprobe von Produkten und Plattformen basieren, die in dieser Studie untersucht wurden, und dass sie möglicherweise nicht auf andere Produkte, Plattformen oder Zeiträume verallgemeinert werden können. Weitere Forschung ist erforderlich, um diese Ergebnisse zu bestätigen und zu erweitern und um andere Faktoren zu untersuchen, die die Preisgestaltung im Internet beeinflussen können.

Aus der vorliegenden Arbeit lassen sich einige wichtige Schlussfolgerungen ziehen. Erstens zeigt die empirische Untersuchung, dass im österreichischen Online-Handel im Beobachtungszeitraum keine signifikante Preisdiskriminierung für die untersuchten Plattformen, von denen fünf zu den Umsatzstärksten in Österreich zählen (Brandt, 2022), festgestellt werden konnte. Dies deutet darauf hin, dass derartige Praktiken in diesem Marktsegment noch nicht weit verbreitet sind. Trotz der begrenzten empirischen Evidenz besteht jedoch theoretisch ein erhebliches Potenzial für Preisdiskriminierung im Online-Handel. Die technischen Möglichkeiten der Datenerhebung und -analyse ermöglichen eine individuelle Preisgestaltung auf der Grundlage des Nutzerverhaltens, demographischer Merkmale und anderer personalisierter Daten. Zum anderen könnten regulatorische Vorgaben und ethische Bedenken eine wichtige Rolle dabei spielen, dass Preisdiskriminierung bislang noch nicht umfassend eingesetzt wird. Datenschutzgesetze und Bedenken hinsichtlich einer fairen Preisgestaltung sind wichtige Hindernisse. Darüber hinaus müssen Händler*innen aus ökonomischen Gründen sorgfältig abwägen, ob die potenziellen Gewinne durch individualisierte Preise die Risiken eines Vertrauensverlustes und negativer Reaktionen der Kund*innen rechtfertigen. Drittens zeigt die Analyse, dass mit fortschreitender technologischer Entwicklung und zunehmender Datenverfügbarkeit die Bedingungen für Preisdiskriminierung im Online-Handel günstiger werden könnten. Unternehmen könnten zunehmend in der Lage sein, Preisdiskriminierung effizient und im Einklang mit regulatorischen und ethischen Standards umzusetzen. Schließlich wird deutlich, dass weitere Untersuchungen erforderlich sind, um ein umfassenderes Bild zu erhalten. Insbesondere die Analyse der

Auswirkungen von Preisdiskriminierung auf Verbraucher*innen und Märkte verdient zusätzliche Aufmerksamkeit. Weiters zeigt sich, dass sowohl die Debatte über Preisdiskriminierung als auch die Diskussion über digitale Monopole ähnliche Schlussfolgerungen aufweisen. Darüber hinaus lassen sich aus der Arbeit wichtige Empfehlungen für Verbraucher*innen ableiten. Verbraucher*innen sollten sich der Praxis der Preisdiskriminierung bewusst sein und verstehen, wie ihre Daten im Internet genutzt werden können, um personalisierte Preise festzulegen. Bildung und Aufklärung zu diesen Themen können helfen, informierte Kaufentscheidungen zu treffen. Außerdem sollten sich Verbraucher*innen über die Maßnahmen bewusst sein, die sie ergreifen können, um ihre Online-Privatsphäre zu schützen. Dazu gehören beispielsweise die Verwendung von VPNs (Virtual Private Network), Inkognito-Modi in Webbrowsern, das regelmäßige Löschen von Cookies und die Beschränkung der Menge an persönlichen Informationen, die sie online weitergeben. Es wird empfohlen, Preisvergleichsseiten zu nutzen und Produkte auf verschiedenen Plattformen zu prüfen, um sicherzustellen, dass sie den besten verfügbaren Preis erhalten. Transparenz seitens der Anbieter*innen kann hier unterstützend wirken. Des Weiteren bedarf es einer ausgewogenen Regulierung, die den Wettbewerb schützt, die Wohlfahrt der Verbraucher*innen gewährleistet und sicherstellt, dass die Macht der Unternehmen verantwortungsbewusst und demokratisch kontrolliert wird. Nur so kann eine faire und wettbewerbsfähige digitale Wirtschaft gewährleistet werden, die allen zugutekommt.

Basierend auf den Ergebnissen dieser Arbeit lassen sich mehrere Vorschläge für zukünftige Forschung im Bereich der Preisdiskriminierung im Online-Handel ableiten. Erstens wäre es sinnvoll, ähnliche empirische Untersuchungen in verschiedenen geografischen Regionen und Marktsegmenten durchzuführen, um zu prüfen, ob die beobachteten Ergebnisse auch in anderen Kontexten gültig sind. Dies könnte zu einem umfassenderen Verständnis der Verbreitung und Wirkungsweise von Online-Preisdiskriminierung beitragen. Darüber hinaus könnten zukünftige Studien die langfristigen Auswirkungen von Preisdiskriminierung auf das Kaufverhalten der Verbraucher*innen analysieren. Insbesondere wäre es interessant zu untersuchen, ob und wie sich die Wahrnehmung von Fairness und das Vertrauen in Online-Händler verändern, wenn Verbraucher*innen erkennen, dass Preise individuell angepasst werden. Solche Untersuchungen könnten wertvolle Erkenntnisse darüber liefern, wie Unternehmen Preisdiskriminierung einsetzen können, ohne das Vertrauen ihrer Kund*innen zu verlieren. Ein weiterer wichtiger Forschungsbereich betrifft die technischen und ethischen Aspekte der Datenerhebung und -verarbeitung. Zukünftige Studien könnten

sich darauf konzentrieren, wie Algorithmen zur Vorhersage des Verbraucher*innen- und Kaufverhaltens verbessert werden können, um eine präzisere und fairere Preisgestaltung zu ermöglichen. Gleichzeitig sollten die ethischen Implikationen solcher Technologien genauer untersucht werden, um sicherzustellen, dass der Datenschutz und die Rechte der Verbraucher*innen gewahrt bleiben. Es wäre auch sinnvoll, den regulatorischen Rahmen für Preisdiskriminierung genauer zu untersuchen. Eine vergleichende Analyse der rechtlichen Regelungen in verschiedenen Ländern könnte aufzeigen, wie unterschiedliche Regelungen die Praxis der Preisdiskriminierung beeinflussen und welche Regulierungsmaßnahmen am effektivsten sind, um die Verbraucher*innen zu schützen und gleichzeitig die Innovationsfähigkeit der Unternehmen zu erhalten. Schließlich könnte zukünftige Forschung auch die ökonomischen Auswirkungen von Preisdiskriminierung auf Wettbewerb und Marktdynamik untersuchen. Es wäre interessant zu analysieren, ob und wie Preisdiskriminierung zu einer Verschiebung von Marktanteilen zwischen Unternehmen führt und welche langfristigen Auswirkungen dies auf die Marktstruktur hat. Auch die Rolle der künstlichen Intelligenz im Zusammenhang mit Preisdiskriminierung ist sicherlich für zukünftige Forschungsarbeiten von Interesse.

6 Literaturverzeichnis

- Ahlheim, Michael/Rose, Manfred (1992). Das Konzept der Konsumentenrente (Dupuit, Marshall). In *Messung individueller Wohlfahrt* (S. 34–39). Berlin, Heidelberg: Springer Berlin Heidelberg. Online: http://link.springer.com/10.1007/978-3-642-58160-1_4 [Abruf am 29.05.2024].
- Azzolina, Stefano/Razza, Manuel/Sartiano, Kevin/Weitschek, Emanuel (2021). Price Discrimination in the Online Airline Market: An Empirical Study. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(6), 2282–2303.
- Bankhofer, Udo/Vogel, Jürgen (2008). *Datenanalyse und Statistik*. Wiesbaden: Gabler. Online: <http://link.springer.com/10.1007/978-3-8349-9654-1> [Abruf am 16.09.2023].
- Belleflamme, Paul/Lam, Wing Man Wynne/Vergote, Wouter (2020). Competitive Imperfect Price Discrimination and Market Power. *Marketing Science*, 39(5), 996–1015.
- van Boom, Willem H./van der Rest, Jean-Pierre I./van den Bos, Kees/Dechesne, Mark (2020). Consumers Beware: Online Personalized Pricing in Action! How the Framing of a Mandated Discriminatory Pricing Disclosure Influences Intention to Purchase. *Social Justice Research*, 33(3), 331–351.
- Brandt, Mathias (2022). Die Top 10 der Online-Shops in Österreich. Statista. Online: <https://de.statista.com/infografik/716/die-top-10-online-shops-in-oesterreich-nach-umsatz/> [Abruf am 15.02.2024].
- Chamberlin, Edward Hastings (1933). *The Theory of Monopolistic Competition*. Cambridge: Harvard University Press.
- Chen, Le/Mislove, Alan/Wilson, Christo (2016). An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace. *Proceedings of the 25th International Conference on World Wide Web. WWW '16: 25th International World Wide Web Conference, Montréal Québec Canada: International World Wide Web Conferences Steering Committee*. Online: <https://dl.acm.org/doi/10.1145/2872427.2883089> [Abruf am 22.04.2023].
- Das Europäische Parlament und der Rat der Europäischen Union (2005). Richtlinie 2005/29/EG des Europäischen Parlaments und des Rates vom 11. Mai 2005 über unlautere Geschäftspraktiken im binnenmarktinternen Geschäftsverkehr zwischen Unternehmen und Verbrauchern und zur Änderung der Richtlinie 84/450/EWG des Rates, der Richtlinien 97/7/EG, 98/27/EG und 2002/65/EG des Europäischen Parlaments und des Rates sowie der Verordnung (EG) Nr. 2006/2004 des Europäischen Parlaments und des Rates. eur-lex.europa.eu. Online: <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=celex%3A32005L0029> [Abruf am 13.10.2023].
- Das Europäische Parlament und der Rat der Europäischen Union (2006). Richtlinie 2006/123/EG des Europäischen Parlaments und des Rates vom 12. Dezember 2006 über Dienstleistungen im Binnenmarkt. eur-lex.europa.eu. Online: <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=celex%3A32006L0123> [Abruf am 13.10.2023].

- Das Europäische Parlament und der Rat der Europäischen Union (2008). Verordnung (EG) Nr. 1008/2008 des Europäischen Parlaments und des Rates vom 24. September 2008 über gemeinsame Vorschriften für die Durchführung von Luftverkehrsdiensten in der Gemeinschaft. eur-lex.europa.eu. Online: <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=celex%3A32008R1008> [Abruf am 13.10.2023].
- Das Europäische Parlament und der Rat der Europäischen Union (2009). Richtlinie 2002/58/EG des Europäischen Parlaments und des Rates vom 12. Juli 2002 über die Verarbeitung personenbezogener Daten und den Schutz der Privatsphäre in der elektronischen Kommunikation (Datenschutzrichtlinie für elektronische Kommunikation). eur-lex.europa.eu. Online: <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A02002L0058-20091219> [Abruf am 13.10.2023].
- Das Europäische Parlament und der Rat der Europäischen Union (2010). Verordnung (EU) Nr. 1177/2010 des Europäischen Parlaments und des Rates vom 24. November 2010 über die Fahrgastrechte im See- und Binnenschiffsverkehr und zur Änderung der Verordnung (EG) Nr. 2006/2004. eur-lex.europa.eu. Online: <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32010R1177> [Abruf am 13.10.2023].
- Das Europäische Parlament und der Rat der Europäischen Union (2011). Verordnung (EU) Nr. 181/2011 des Europäischen Parlaments und des Rates vom 16. Februar 2011 über die Fahrgastrechte im Kraftomnibusverkehr und zur Änderung der Verordnung (EG) Nr. 2006/2004. eur-lex.europa.eu. Online: <https://eur-lex.europa.eu/legal-content/DE/ALL/?uri=celex%3A32011R0181> [Abruf am 13.10.2023].
- Das Europäische Parlament und der Rat der Europäischen Union (2016a). Verordnung (EU) 2016/1036 des Europäischen Parlaments und des Rates vom 8. Juni 2016 über den Schutz gegen gedumpte Einfuhren aus nicht zur Europäischen Union gehörenden Ländern (kodifizierter Text). eur-lex.europa.eu. Online: <https://eur-lex.europa.eu/eli/reg/2016/1036/oj?locale=de> [Abruf am 06.09.2023].
- Das Europäische Parlament und der Rat der Europäischen Union (2016b). Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung). eur-lex.europa.eu. Online: <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A02016R0679-20160504> [Abruf am 13.10.2023].
- Das Europäische Parlament und der Rat der Europäischen Union (2018). Verordnung (EU) 2018/302 des Europäischen Parlaments und des Rates vom 28. Februar 2018 über Maßnahmen gegen ungerechtfertigtes Geoblocking und andere Formen der Diskriminierung aufgrund der Staatsangehörigkeit, des Wohnsitzes oder des Ortes der Niederlassung des Kunden innerhalb des Binnenmarkts und zur Änderung der Verordnungen (EG) Nr. 2006/2004 und (EU) 2017/2394 sowie der Richtlinie 2009/22/EG. eur-lex.europa.eu. Online: <https://eur-lex.europa.eu/eli/reg/2018/302/oj> [Abruf am 21.06.2023].
- Das Europäische Parlament und der Rat der Europäischen Union (2021a). Leitlinien zur Auslegung und Anwendung der Richtlinie 2005/29/EG des Europäischen Parlaments und des Rates über unlautere Geschäftspraktiken von Unternehmen

- gegenüber Verbrauchern im Binnenmarkt. eur-lex.europa.eu. Online: <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52021XC1229%2805%29> [Abruf am 13.10.2023].
- Das Europäische Parlament und der Rat der Europäischen Union (2021b). Verordnung (EU) 2021/782 des Europäischen Parlaments und des Rates vom 29. April 2021 über die Rechte und Pflichten der Fahrgäste im Eisenbahnverkehr (Neufassung). eur-lex.europa.eu. Online: <https://eur-lex.europa.eu/eli/reg/2021/782/oj?locale=de> [Abruf am 13.10.2023].
- De Mauro, Andrea/Greco, Marco/Grimaldi, Michele (2015). What is big data? A consensual definition and a review of key research topics. INTERNATIONAL CONFERENCE ON INTEGRATED INFORMATION (IC-ININFO 2014): Proceedings of the 4th International Conference on Integrated Information, Madrid, Spain: o.V. Online: <https://pubs.aip.org/aip/acp/article/1644/1/97-104/883119> [Abruf am 08.09.2023].
- De Sousa, Brian (2021). File processing in Node.js: A comprehensive guide. LogRocket Frontend Analytics. Online: <https://blog.logrocket.com/file-processing-node-js-comprehensive-guide/> [Abruf am 12.04.2024].
- Dittmar, Carsten/Felden, Carsten/Finger, Ralf/Scheuch, Rolf/Tams, Lars (2016). Big Data - Ein Überblick. Heidelberg: dpunkt.verlag.
- Döring, Nicola/Bortz, Jürgen (2016). Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. Berlin, Heidelberg: Springer Berlin Heidelberg. Online: <https://link.springer.com/10.1007/978-3-642-41089-5> [Abruf am 10.04.2024].
- Dubé, Jean-Pierre/Misra, Sanjog (2017). Personalized Pricing and Consumer Welfare. SSRN Electronic Journal. Online: <https://www.ssrn.com/abstract=2992257> [Abruf am 22.04.2023].
- Dullien, Sebastian (2021). Nach der Corona-Krise: Die nächste Phase der (De-)Globalisierung und die Rolle der Industriepolitik. IMK Policy Brief, No. 100. Hans-Böckler-Stiftung, Institut für Makroökonomie und Konjunkturforschung (IMK). Düsseldorf. Online: <http://hdl.handle.net/10419/237915> [Abruf am 20.12.2023].
- Elegido, Juan M. (2011). The Ethics of Price Discrimination. Business Ethics Quarterly, 21(4), 633–660.
- Europäische Kommission (2023). Anwendung des EU-Rechts. commission.europa.eu. Online: https://commission.europa.eu/law/application-eu-law/implementing-eu-law_de [Abruf am 13.10.2023].
- Europäische Union (2012a). Charter of Fundamental Rights of the European Union. eur-lex.europa.eu. Online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT> [Abruf am 13.10.2023].
- Europäische Union (2012b). Vertrag über die Arbeitsweise der Europäischen Union. eur-lex.europa.eu. Online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012E%2FTXT> [Abruf am 18.10.2023].
- European Parliament. Directorate General for Internal Policies of the Union. (2022). Personalised pricing. LU: Publications Office. Online: <https://data.europa.eu/doi/10.2861/127035> [Abruf am 22.04.2023].

- Faisal (2023). Node.js: Beyond the Basics - An In-Depth Exploration Through its History and Architecture. Online: <https://thefaisal.dev/nodejs-beyond-the-basics> [Abruf am 12.04.2024].
- Gluchowski, Peter/Schieder, Christian/Chamoni, Peter (2021). Methoden des Data Mining für Big Data Analytics. In Sara D'Onofrio/Andreas Meier (Hrsg.). *Big Data Analytics* Edition HMD. (S. 25–48). Wiesbaden: Springer Fachmedien Wiesbaden. Online: https://link.springer.com/10.1007/978-3-658-32236-6_2 [Abruf am 08.09.2023].
- Gniewosz, Burkhard (2011). Beobachtung. In Heinz Reinders/Hartmut Ditton/Cornelia Gräsel/Burkhard Gniewosz (Hrsg.). *Empirische Bildungsforschung* (S. 99–107). Wiesbaden: VS Verlag für Sozialwissenschaften. Online: http://link.springer.com/10.1007/978-3-531-93015-2_8 [Abruf am 09.06.2023].
- Grochowski, Mateusz/Jablonowska, Agnieszka/Lagioia, Francesca/Sartor, Giovanni (2022). Algorithmic Price Discrimination and Consumer Protection. *Technology and Regulation*, 36-47 Pages.
- Gumbus, Andra/Grodzinsky, Frances (2016). Era of big data: danger of descrimination. *ACM SIGCAS Computers and Society*, 45(3), 118–125.
- Hindermann, Christoph Michael (2018). Price Discrimination in Online Retail. Online: <http://hdl.handle.net/10419/181294> [Abruf am 09.06.2023].
- Jakobs, Hans-Jürgen (2022). *Das Monopol im 21. Jahrhundert: Wie private Unternehmen und staatliche Konzerne unseren Wohlstand zerstören* (1. Auflage). München: Deutsche Verlags-Anstalt.
- Janiesch, Christian/Zschech, Patrick/Heinrich, Kai (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685–695.
- Kalka, Regine/Krämer, Andreas (Hrsg.) (2020). *Preiskommunikation: strategische Herausforderungen und innovative Anwendungsfelder*. Wiesbaden [Heidelberg]: Springer Gabler.
- Keyling, Till/Jünger, Jakob (2016). Observing Online Content. In Gerhard Vowe/Philipp Henn (Hrsg.). *Political communication in the online world: theoretical approaches and research designs* Routledge research in political communication. (S. 183–200). New York, NY: Routledge.
- Kinsta (2023). Understanding the Node.js File System Module (FS). Online: <https://kinsta.com/knowledgebase/nodejs-fs/> [Abruf am 12.04.2024].
- Klemperer, P. (1995). Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade. *The Review of Economic Studies*, 62(4), 515–539.
- Kollmann, Tobias (Hrsg.) (2020). *Handbuch Digitale Wirtschaft*. Wiesbaden: Springer Fachmedien Wiesbaden. Online: <http://link.springer.com/10.1007/978-3-658-17291-6> [Abruf am 22.04.2023].
- Konrad, Vera/Polk, Andreas (2020). Big Data und Preisdiskriminierung. *Wirtschaftsdienst*, 100(10), 793–798.

- Krebs, Dagmar/Menold, Natalja (2019). Gütekriterien quantitativer Sozialforschung. In Nina Baur/Jörg Blasius (Hrsg.). *Handbuch Methoden der empirischen Sozialforschung* (S. 489–504). Wiesbaden: Springer Fachmedien Wiesbaden. Online: http://link.springer.com/10.1007/978-3-658-21308-4_34 [Abruf am 10.04.2024].
- LambdaTest (o. J.). Puppeteer Tutorial: Complete Guide to Puppeteer Testing. Online: <https://www.lambdatest.com/puppeteer> [Abruf am 12.04.2024].
- Li, Youping/Shuai, Jie (2019). Monopolistic competition, price discrimination and welfare. *Economics Letters*, 174, 114–117.
- Machlup, Fritz (1955). Characteristics and Types of Price Discrimination. In *Business Concentration and Price Policy* National Bureau of Economic Research. (S. 397–440). Princeton: Princeton University Press.
- Mankiw, Nicholas Gregory (2021). *Principles of economics (Ninth edition)*. Boston, MA: Cengage.
- Mastering Backend (o. J.). Introduction to Node.js. Online: <https://masteringbackend.com/hubs/nodejs-essentials/overview-of-node-js> [Abruf am 12.04.2024].
- MDN contributors (2024a). JavaScript. Online: <https://developer.mozilla.org/en-US/docs/Learn/JavaScript> [Abruf am 12.04.2024].
- MDN contributors (2024b). JavaScript — Dynamic client-side scripting. Online: <https://developer.mozilla.org/en-US/docs/Learn/JavaScript> [Abruf am 12.04.2024].
- MDN contributors (2024c). What is JavaScript? Online: https://developer.mozilla.org/en-US/docs/Learn/JavaScript/First_steps/What_is_JavaScript [Abruf am 12.04.2024].
- Microsoft (o. J.). Visual Studio Code - Code Editing. Redefined. Online: <https://code.visualstudio.com/> [Abruf am 12.04.2024].
- Narahari, Y./Raju, C.V.L./Ravikumar, K./Shah, Sourabh (2005). Dynamic pricing models for electronic business. *Sadhana*, 30(2–3), 231–256.
- Neubert, Michael (2022). A Systematic Literature Review of Dynamic Pricing Strategies. *International Business Research*, 15(4), 1.
- Node.js (a) (o. J.). Node.js v21.7.3 documentation. Online: <https://nodejs.org/api/fs.html> [Abruf am 12.04.2024].
- Node.js (b) (o. J.). Introduction to Node.js. Online: <https://nodejs.org/en/learn/getting-started/introduction-to-nodejs> [Abruf am 12.04.2024].
- Nufer, Gerd/Muth, Manuel (2022). Der Einsatz Künstlicher Neuronaler Netze zur Analyse von Markenattributen hinsichtlich der Markenbewertung im Sportmarketing. Online: https://zs.thulb.uni-jena.de/receive/jportal_jparticle_01244733 [Abruf am 22.09.2023].
- O. V. (2005). Preisdifferenzierung. In *Wettbewerbsökonomie* Springer-Lehrbuch. (S. 205–244). Berlin/Heidelberg: Springer-Verlag. Online: http://link.springer.com/10.1007/3-540-26342-X_10 [Abruf am 29.08.2023].

- Pigou, Arthur Cecil (1932). *The Economics of Welfare* 4. Aufl. London: MACMILLAN AND CO.
- Poort, Joost/Zuiderveen Borgesius, Frederik J. (2019). Does everyone have a price? Understanding people's attitude towards online and offline price discrimination. *Internet Policy Review*, 8(1). Online: <https://policyreview.info/node/1383> [Abruf am 24.10.2023].
- Puppeteer (o. J.). Introduction | Puppeteer. Online: <https://pptr.dev/category/introduction> [Abruf am 12.04.2024].
- Rechtsinformationssystem des Bundes (RIS) (2023a). Bundesgesetz über die Gleichbehandlung (Gleichbehandlungsgesetz – GIBG). [ris.bka.gv.at](https://www.ris.bka.gv.at). Online: <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20003395> [Abruf am 13.10.2023].
- Rechtsinformationssystem des Bundes (RIS) (2023b). Gesetz zur Bekämpfung von Diskriminierung (Wiener Antidiskriminierungsgesetz). [ris.bka.gv.at](https://www.ris.bka.gv.at). Online: <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=LrW&Gesetzesnummer=20000171> [Abruf am 13.10.2023].
- Rechtsinformationssystem des Bundes (RIS) (2023c). Bundesgesetz gegen den unlauteren Wettbewerb 1984 – UWG. [ris.bka.gv.at](https://www.ris.bka.gv.at). Online: <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10002665> [Abruf am 13.10.2023].
- Rechtsinformationssystem des Bundes (RIS) (2024). Gesamte Rechtsvorschrift für Urheberrechtsgesetz. [ris.bka.gv.at](https://www.ris.bka.gv.at). Online: <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10001848> [Abruf am 10.04.2024].
- Scherer, F.M./Ross, David (1990). *Industrial market structure and economic performance* (3rd ed). Boston: Houghton Mifflin.
- Spann, Martin/Klein, Joachim/Makhlouf, Karim/Bernhardt, Martin (2005). Interaktive Preismaßnahmen bei Low-Cost-Fluglinien. In Günter Fandel/Hans Botho von Portatius (Hrsg.). *Revenue Management* (S. 53–78). Wiesbaden: Gabler Verlag. Online: http://link.springer.com/10.1007/978-3-663-11304-1_3 [Abruf am 22.04.2023].
- Spann, Martin/Skiera, Bernd (2020). Dynamische Preisgestaltung in der digitalisierten Welt. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung*, 72(3), 321–342.
- Statista (2023). Daten - Volumen der weltweit generierten Daten bis 2027. Statista. Online: <https://de.statista.com/statistik/daten/studie/267974/umfrage/prognose-zum-weltweit-generierten-datenvolumen/> [Abruf am 08.09.2023].
- Statista Research Department (2024). Umsatz ausgewählter Low Cost Carrier weltweit im Jahr 2022-2023. Statista. Online: <https://de.statista.com/statistik/daten/studie/262245/umfrage/umsatz-der-low-cost-carrier/#:~:text=Southwest%20Airlines%20aus%20den%20USA,rund%2012%2C4%20Milliarden%20Euro.> [Abruf am 15.02.2024].

- Steinberg, Etye (2020). Big Data and Personalized Pricing. *Business Ethics Quarterly*, 30(1), 97–117.
- Stigler, George J. (1961). The Economics of Information. *Journal of Political Economy*, 69(3), 213–225.
- Sun, Luping/Tang, Yanfei (2021). Data-Driven Discrimination, Perceived Fairness, and Consumer Trust—The Perspective of Consumer Attribution. *Frontiers in Psychology*, 12, 748765.
- Thierbach, Cornelia/Petschick, Grit (2019). Beobachtung. In Nina Baur/Jörg Blasius (Hrsg.). *Handbuch Methoden der empirischen Sozialforschung* (S. 1165–1181). Wiesbaden: Springer Fachmedien Wiesbaden. Online: http://link.springer.com/10.1007/978-3-658-21308-4_84 [Abruf am 10.04.2024].
- Townley, Christopher/Morrison, Eric/Yeung, Karen (2017). Big Data and Personalized Price Discrimination in EU Competition Law. *Yearbook of European Law*, 36, 683–748.
- Trübner, Miriam/Mühlichen, Andreas (2019). Big Data. In Nina Baur/Jörg Blasius (Hrsg.). *Handbuch Methoden der empirischen Sozialforschung* (S. 143–158). Wiesbaden: Springer Fachmedien Wiesbaden. Online: http://link.springer.com/10.1007/978-3-658-21308-4_9 [Abruf am 10.04.2024].
- Turow, Joseph/Feldman, Lauren/Meltzer, Kimberly (2005). Open to Exploitation: America's Shoppers Online and Offline. A Report from the Annenberg Public Policy Center of the University of Pennsylvania. Online: <https://repository.upenn.edu/handle/20.500.14332/2000>.
- TutorialsTeacher (o. J.). Node.js File System. Online: <https://www.tutorialsteacher.com/nodejs/nodejs-file-system> [Abruf am 12.04.2024].
- UserAgentString.com (o. J.). List of User Agent Strings. User Agent String.Com. Online: <https://www.useragentstring.com/pages/useragentstring.php> [Abruf am 18.03.2024].
- W3Schools (o. J.). JavaScript History. Online: https://www.w3schools.com/js/js_history.asp [Abruf am 12.04.2024].
- Webster, Jane/Watson, Richard T. (2002). Analyzing the Past To Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2), 13–23.
- Wiedemann, Klaus (2023). *Rechtliche Implikationen Profiling-basierter Personalisierung*. S.l.: Springer-Verlag Berlin AN.
- Wied-Nebbeling, Susanne (2004). Monopolistische Konkurrenz. In *Preistheorie und Industrieökonomik* Springer-Lehrbuch. (S. 100–124). Berlin, Heidelberg: Springer Berlin Heidelberg. Online: http://link.springer.com/10.1007/978-3-662-09437-2_3 [Abruf am 07.12.2023].
- Zhao, Bo (2017). Web Scraping. In Laurie A. Schintler/Connie L. McNeely (Hrsg.). *Encyclopedia of Big Data* (S. 1–3). Cham: Springer International Publishing. Online: http://link.springer.com/10.1007/978-3-319-32001-4_483-1 [Abruf am 03.06.2023].

7 Abbildungsverzeichnis

Abbildung 1 - Preisabfrage Sixt: normaler Modus (Bing Browser)	3
Abbildung 2 - Preisabfrage Sixt: Inkognito-Modus (Bing Browser)	3
Abbildung 3 - Preisdiskriminierung zweiten Grades.....	14
Abbildung 4 - Preisdiskriminierung dritten Grades.....	16
Abbildung 5 - Vorgehensweise bei Verfahren des unüberwachten Lernens	22
Abbildung 6 - Vorgehensweise bei Verfahren des überwachten Lernens	23
Abbildung 7 - Entscheidungsbaum	27
Abbildung 8 - Künstliches neuronales Netz.....	30
Abbildung 9 - Euklidische Distanz & Manhattan-Distanz.....	32
Abbildung 10 - Beispiel zu Support und Konfidenz (Assoziationsanalyse).....	37
Abbildung 11 - Shop Apotheke: Boxplots	77
Abbildung 12 - Shop Apotheke: Liniendiagramme	78
Abbildung 13 - Urlaubsguru: Boxplots.....	85
Abbildung 14 - Urlaubsguru: Liniendiagramme.....	86
Abbildung 15 - Electronic4you: Boxplots.....	110
Abbildung 16 - Electronic4you: Liniendiagramme	111
Abbildung 17 - Europcar: Boxplots.....	112
Abbildung 18 - Europcar: Liniendiagramme.....	113
Abbildung 19 - Media Markt: Boxplots	114
Abbildung 20 - Media Markt: Liniendiagramme	115
Abbildung 21 - Otto Versand: Boxplots.....	116

Abbildung 22 - Otto Versand: Liniendiagramme.....	117
Abbildung 23 - Ryanair: Boxplots.....	118
Abbildung 24 - Ryanair: Liniendiagramme.....	119
Abbildung 25 - Universal: Boxplots.....	120
Abbildung 26 - Universal: Liniendiagramme.....	121

8 Tabellenverzeichnis

Tabelle 1 - Beispieldaten (Entscheidungsbaum).....	25
Tabelle 2 - Bedingte Verteilungen (Entscheidungsbaum)	26
Tabelle 3 - Bestimmung des empfohlenen Preises (Entscheidungsbaum)	26
Tabelle 4 - Beispieldaten (Clusterbildung).....	33
Tabelle 5 - Ergebnis der Clusterbildung	34
Tabelle 6 - Produkte/Dienstleistungen für den Webscraper	70
Tabelle 7 – Anzahl und Zeitpunkte der Scraping-Durchläufe 04.04.2024 - 11.05.2024 ...	73
Tabelle 8 - Timeouts Datenerhebung	74
Tabelle 9 - Shop Apotheke: Ausgeschlossene Scraping-DL	76
Tabelle 10 - Shop Apotheke: Preismittelwerte.....	76
Tabelle 11 - electronic4you: Ausgeschlossene Scraping-DL	79
Tabelle 12 - electronic4you: Preismittelwerte.....	79
Tabelle 13 - Media Markt: Ausgeschlossenen Scraping-DL	79
Tabelle 14 – Media Markt: Preismittelwerte	80
Tabelle 15 – Otto Versand: Ausgeschlossenen Scraping-DL	80
Tabelle 16 – Otto Versand: Preismittelwerte.....	80
Tabelle 17 – Europcar: Ausgeschlossenen Scraping-DL.....	81
Tabelle 18 - Europcar: Preismittelwerte	81
Tabelle 19 –Ryanair: Ausgeschlossenen Scraping-DL.....	81
Tabelle 20 - Ryanair: Preismittelwerte	82
Tabelle 21 - Universal: Ausgeschlossenen Scraping-DL.....	82

Tabelle 22 – Universal: Preismittelwerte	82
Tabelle 23 - Urlaubsguru: Ausgeschlossenen Scraping-DL.....	83
Tabelle 24 – Urlaubsguru: Preismittelwerte.....	84

9 Quellcodeverzeichnis

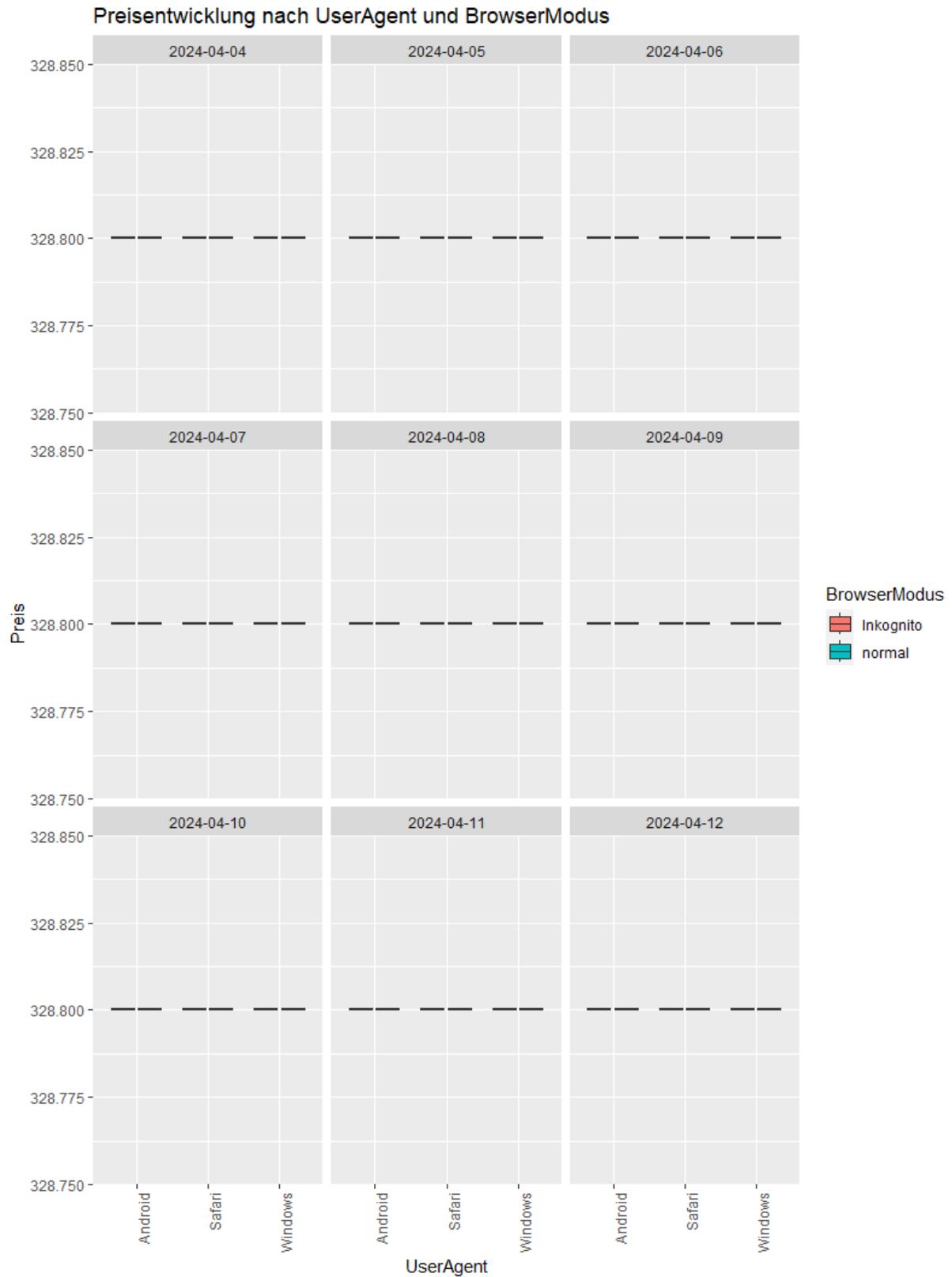
Quellcode 1 - Import Anweisungen	63
Quellcode 2 - Lesen der Konfigurationsdatei	63
Quellcode 3 - Funktion zur Datumsformatierung.....	63
Quellcode 4 - Hauptfunktion zum Scrpaen der Preise.....	64
Quellcode 5 - Cookie-Management	66
Quellcode 6 - Hauptfunktion	66
Quellcode 7 - Beispiel-Eintrag der congig.json.....	68

10 Abkürzungsverzeichnis

ACP	Algorithmic Consumer Price Discrimination
API	Application Programming Interface
DSGVO	Datenschutz-Grundverordnung
ECMA	European Computer Manufacturers Association
EU	Europäische Union
IMCO	Committee of the Internal Market and Consumer Protection
IOT	Internet of Things
KNN	Künstliche neuronale Netze
NPM	Node Package Manager
PDF	Portable Document Format
PHP	PHP: Hypertext Preprocessor
RNN	Recurrent neural network
SPA	Singel-page application
SSR	Server Side Rendering
UGPRL	Richtlinie über unlautere Geschäftspraktiken (EU)
UI	User Interface
UrhG	Urheberrechtsgesetz
URL	Uniform Resource Locator
UWG	Bundesgesetz gegen den unlauteren Wettbewerb
VPN	Virtual Private Network
VS Code	Visual Studio Code

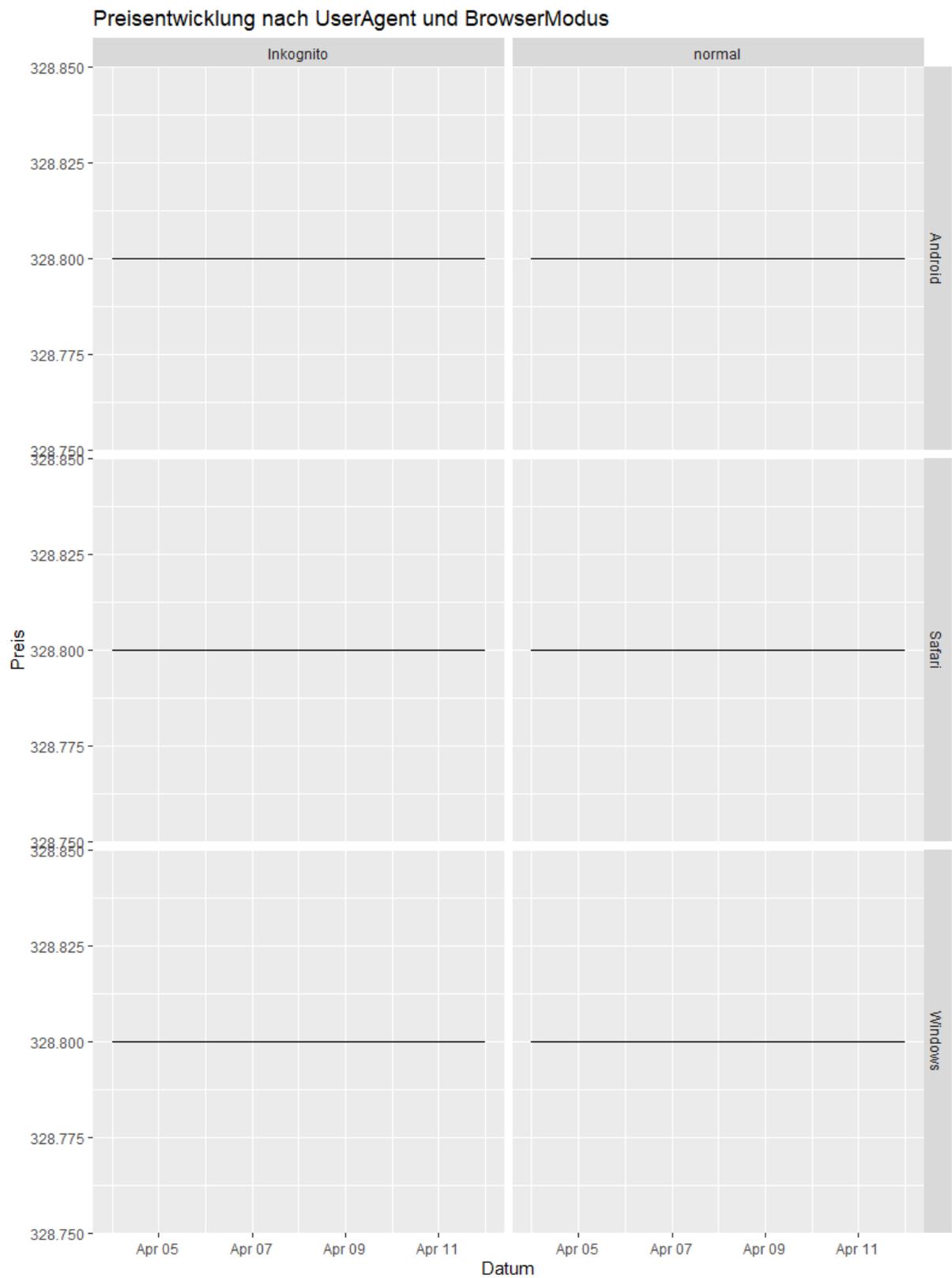
Anhang A

Abbildung 15 - Electronic4you: Boxplots



Anmerkung: Erstellt in R

Abbildung 16 - Electronic4you: Liniendiagramme



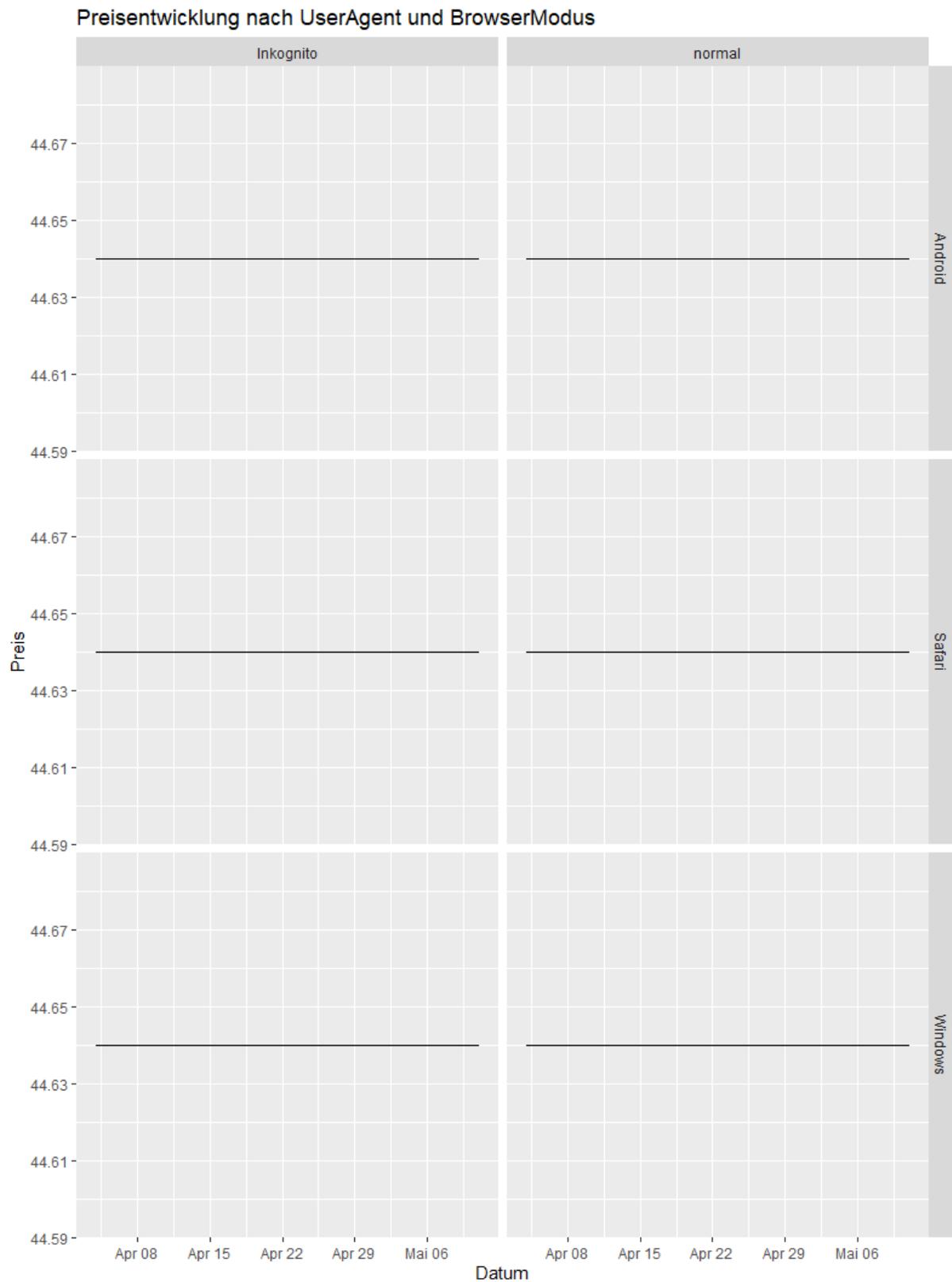
Anmerkung: Erstellt in R

Abbildung 17 - Europcar: Boxplots



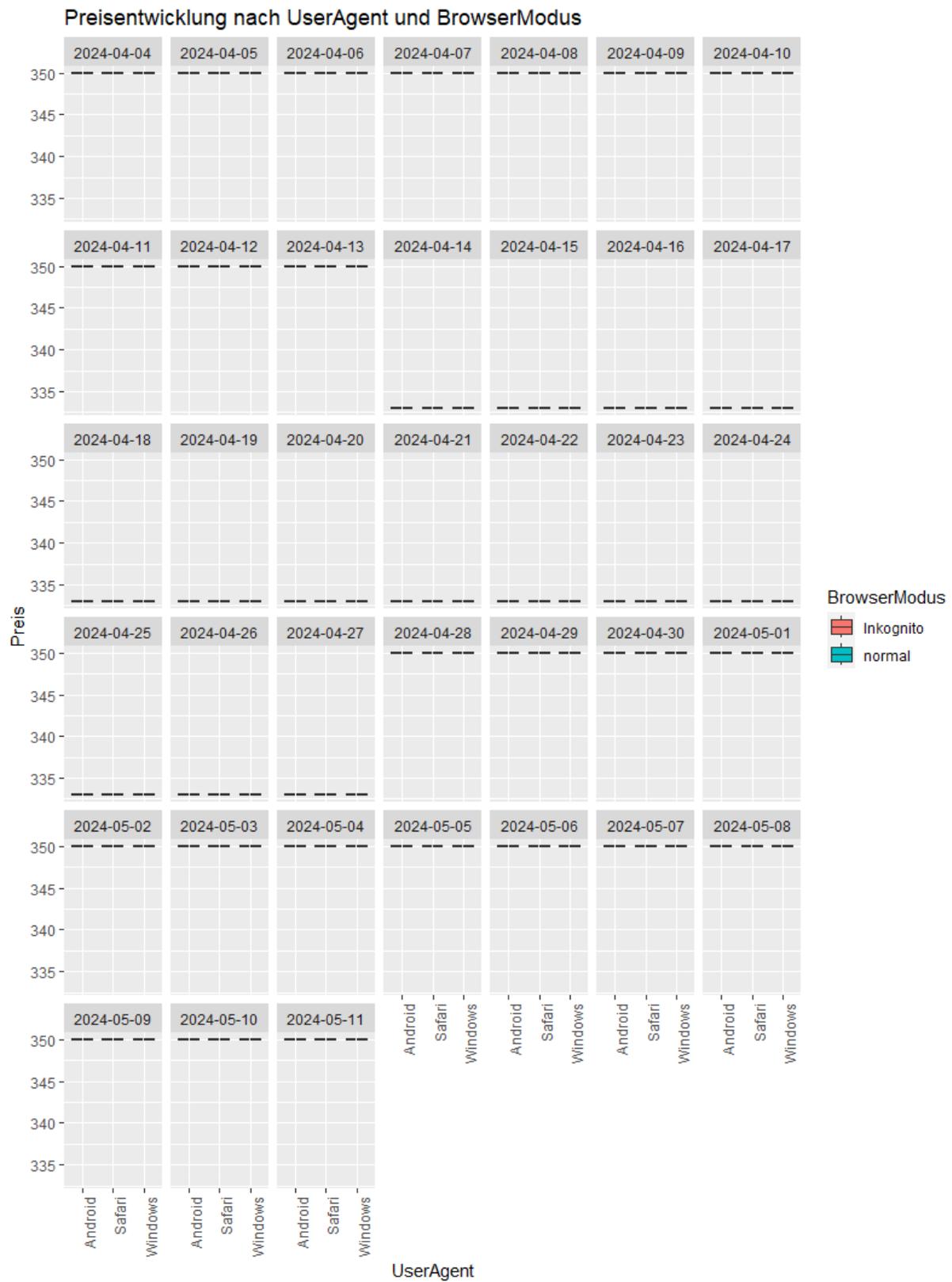
Anmerkung: Erstellt in R

Abbildung 18 - Europcar: Liniendiagramme



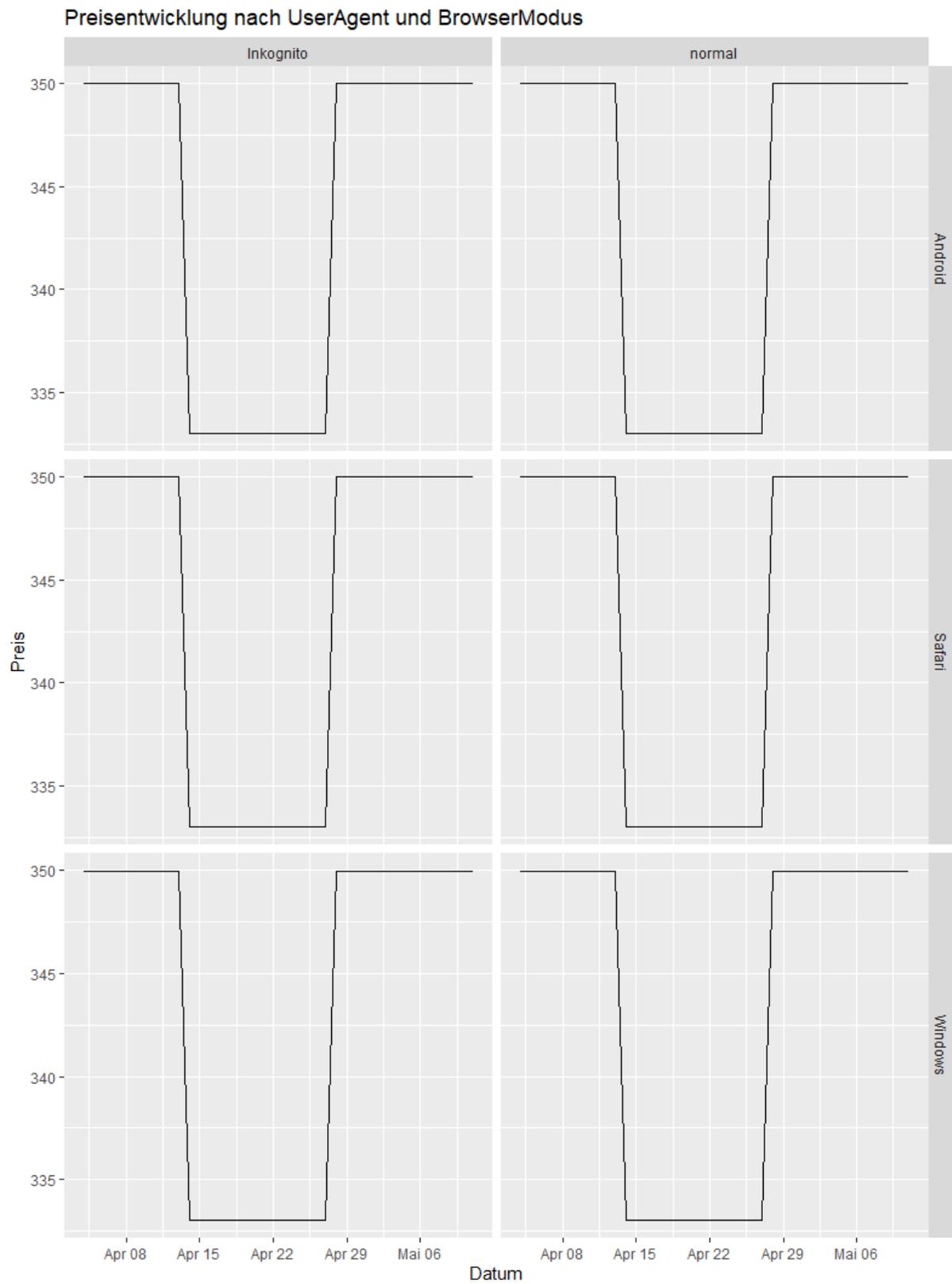
Anmerkung: Erstellt in R

Abbildung 19 - Media Markt: Boxplots



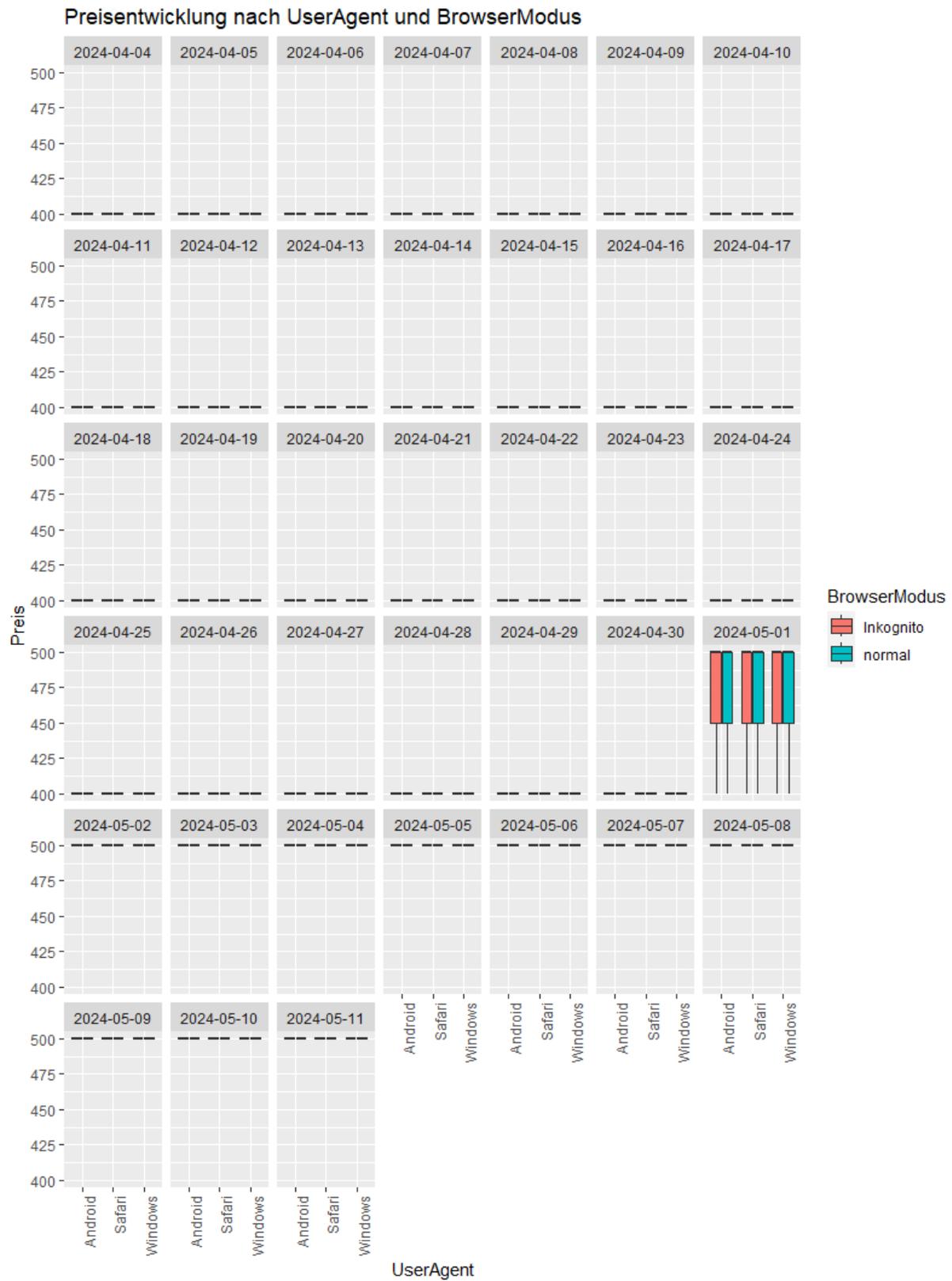
Anmerkung: Erstellt in R

Abbildung 20 - Media Markt: Liniendiagramme



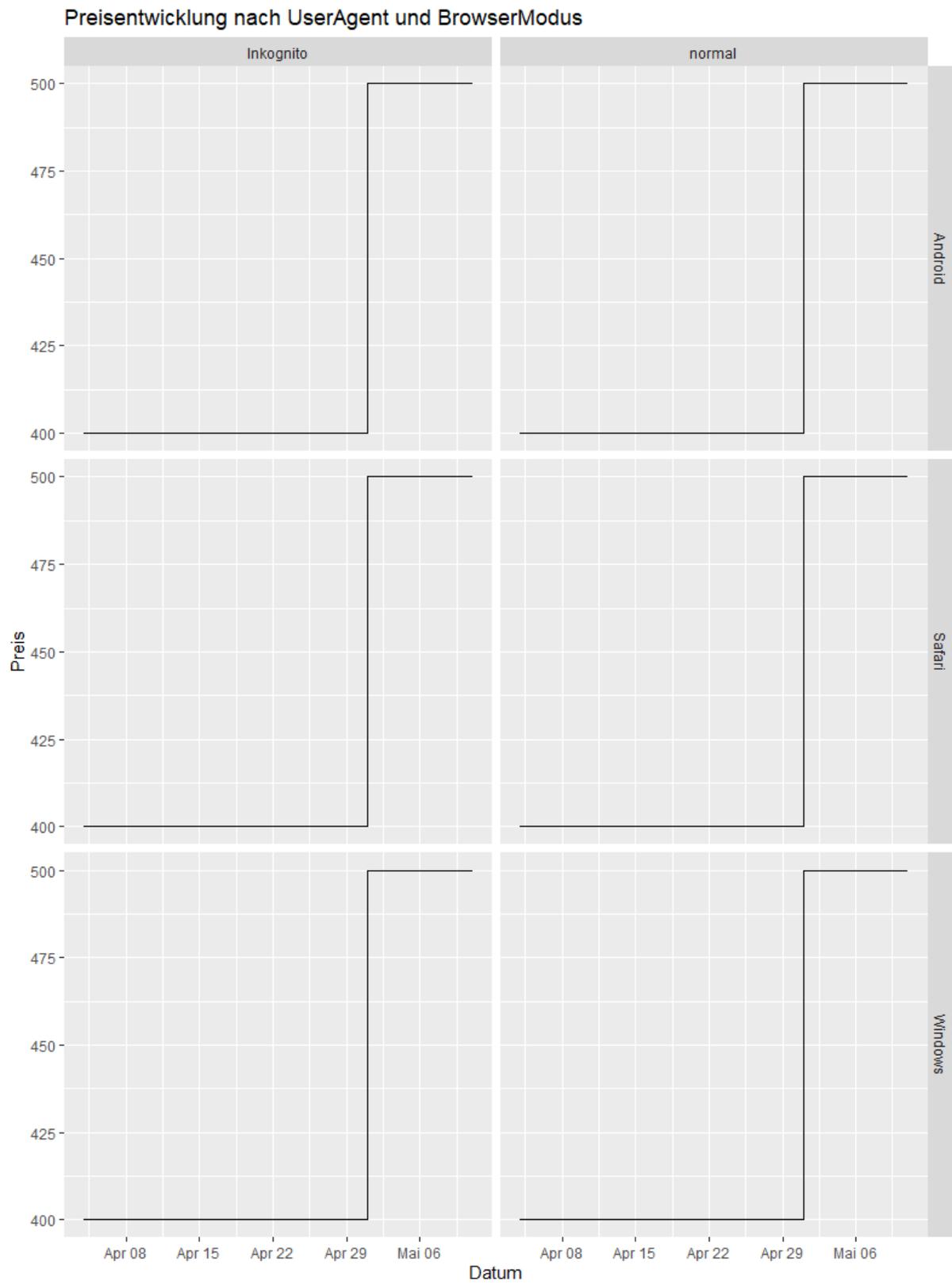
Anmerkung: Erstellt in R

Abbildung 21 - Otto Versand: Boxplots



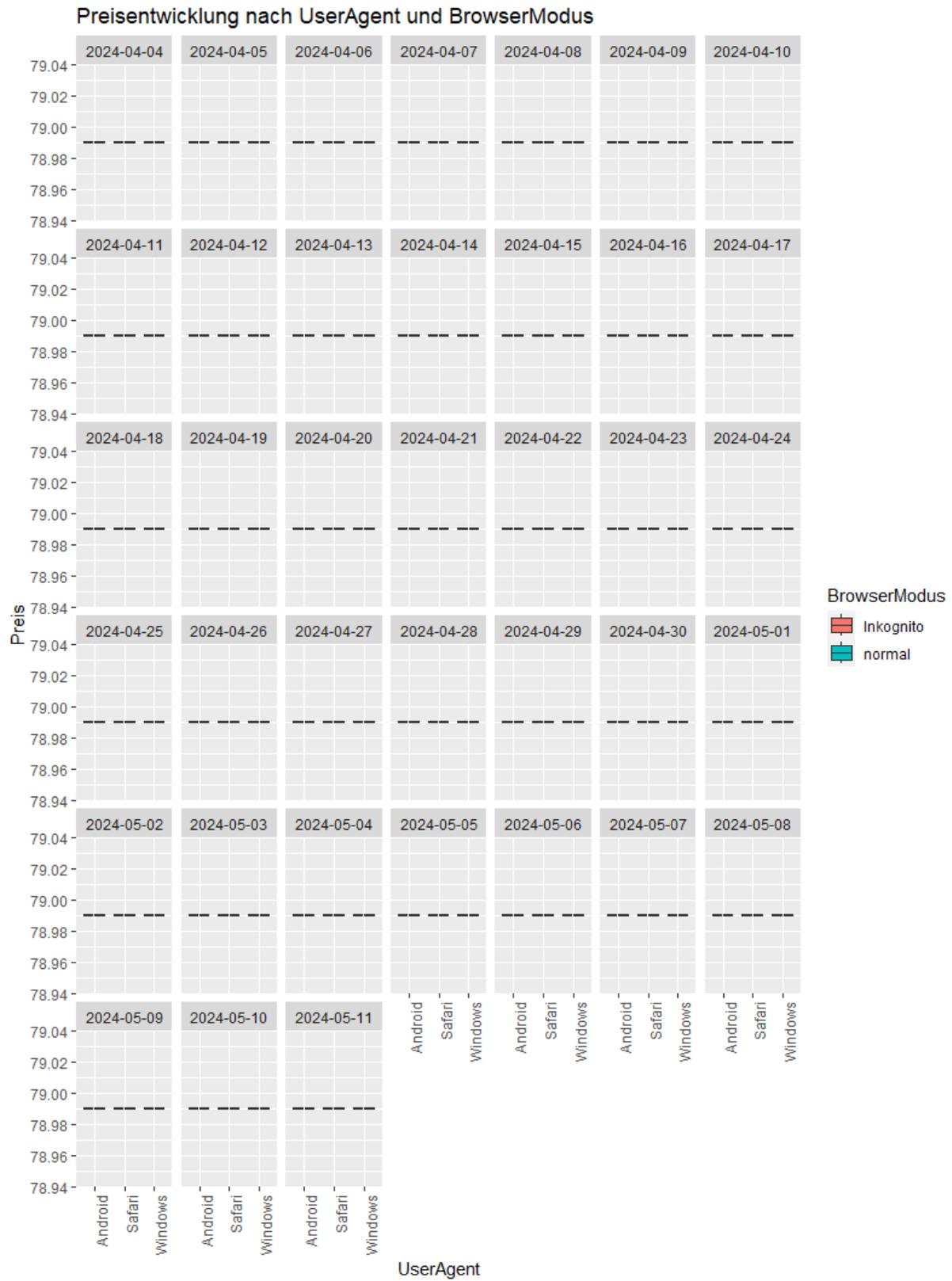
Anmerkung: Erstellt in R

Abbildung 22 - Otto Versand: Liniendiagramme



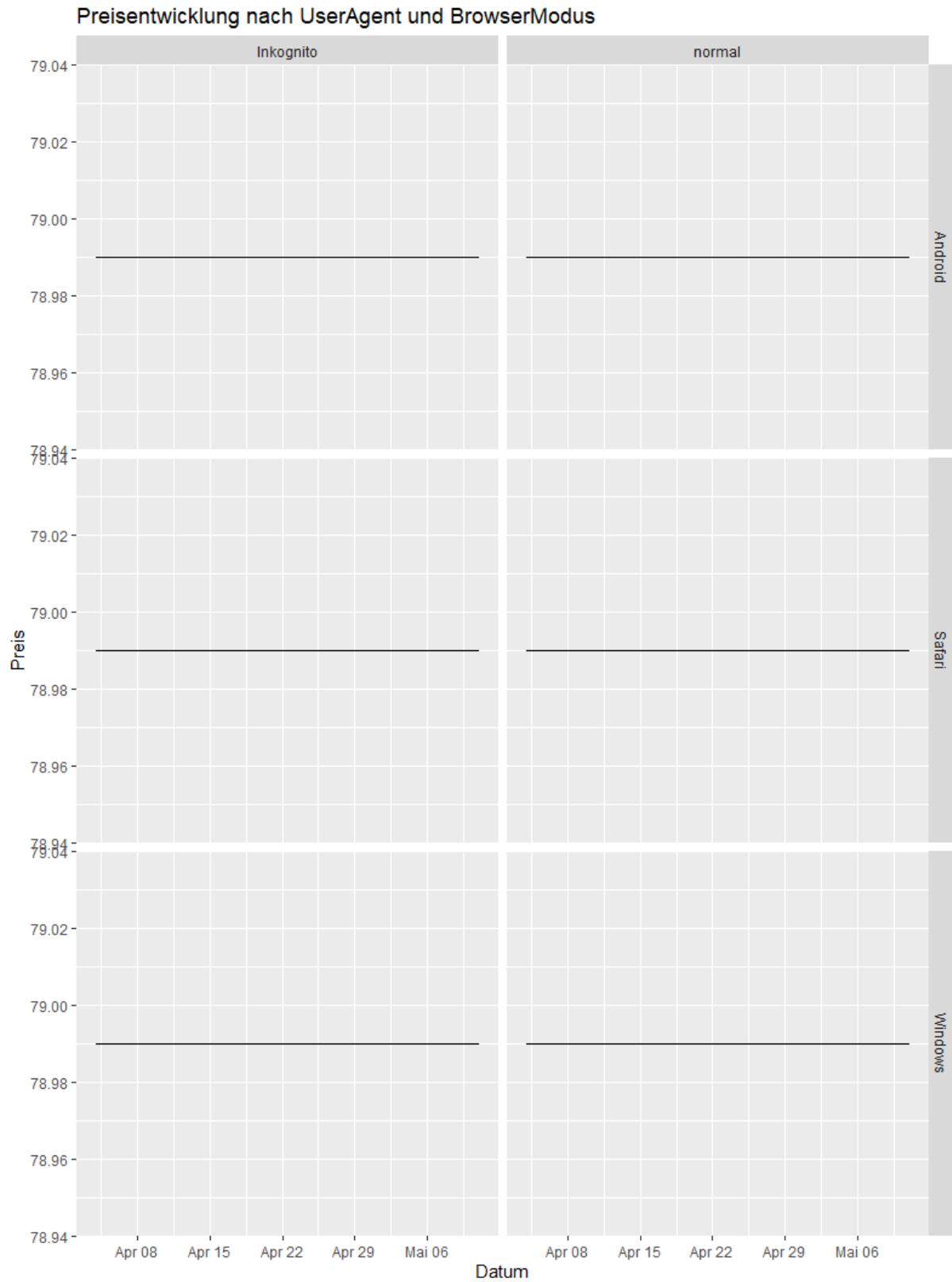
Anmerkung: Erstellt in R

Abbildung 23 - Ryanair: Boxplots



Anmerkung: Erstellt in R

Abbildung 24 - Ryanair: Liniendiagramme



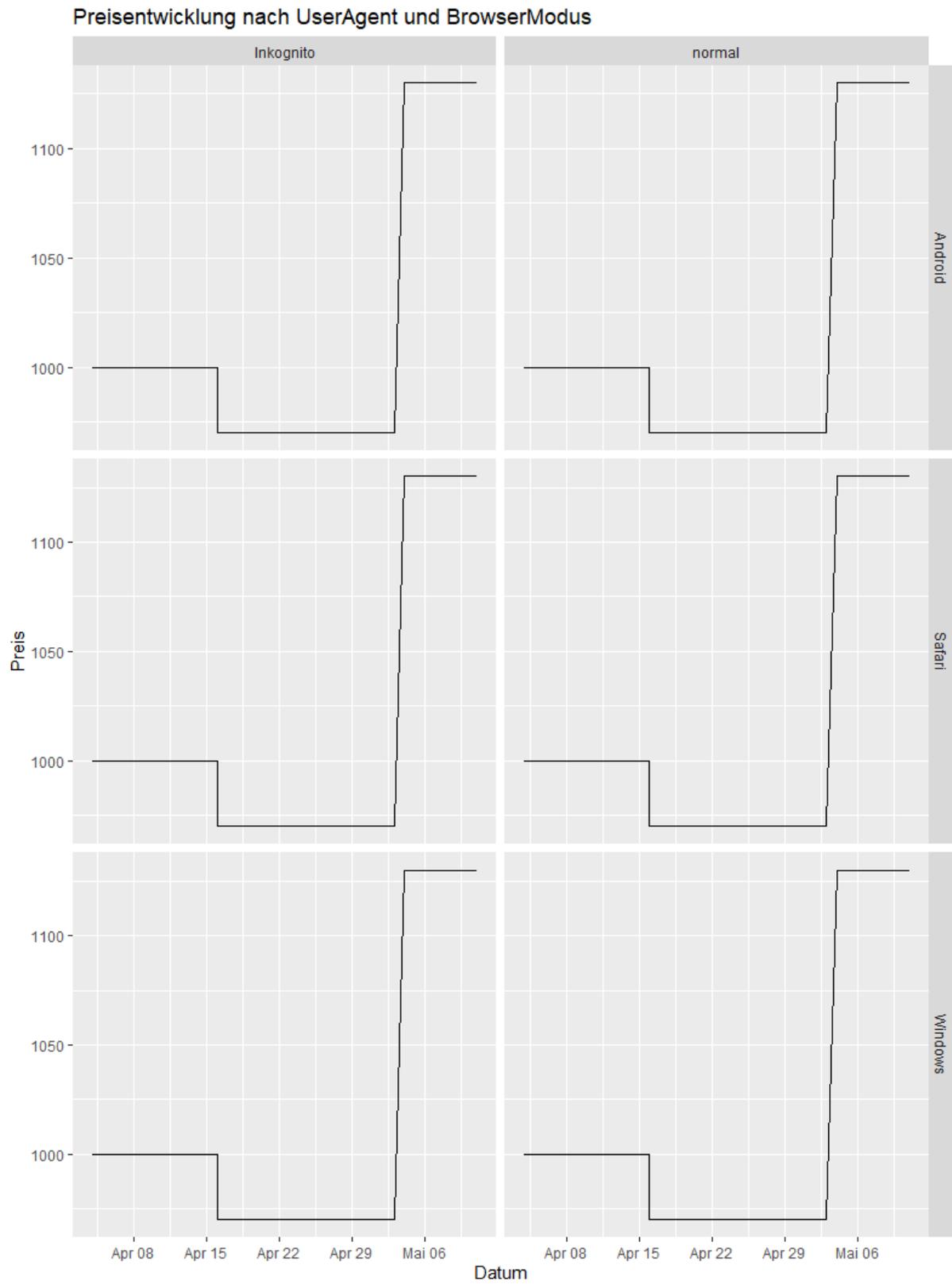
Anmerkung: Erstellt in R

Abbildung 25 - Universal: Boxplots



Anmerkung: Erstellt in R

Abbildung 26 - Universal: Liniendiagramme



Anmerkung: Erstellt in R