

# **Wallstreetbets**

## **Statistischer Zusammenhang von Unternehmenserwähnungen auf reddit.com/r/Wallstreetbets und deren Aktienkurs**

### **Bachelorarbeit**

eingereicht von: **Ing. Johannes Hatter**  
Matrikelnummer: 51905460

im Fachhochschul-Bachelorstudiengang Wirtschaftsinformatik (0470)  
der Ferdinand Porsche FernFH

zur Erlangung des akademischen Grades eines  
**Bachelor of Arts in Business**

Betreuung und Beurteilung: Dr. Valentin Hofstätter

Wiener Neustadt, Mai 2022

# Ehrenwörtliche Erklärung

Ich versichere hiermit,

1. dass ich die vorliegende Bachelorarbeit 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 Bachelorarbeit bisher weder im Inland noch im Ausland in irgendeiner Form als Prüfungsarbeit zur Beurteilung vorgelegt oder veröffentlicht habe.

Ebreichsdorf, 18.05.2022

---

Unterschrift

## **Creative Commons Lizenz**

Das Urheberrecht der vorliegenden Arbeit liegt bei Ing. Johannes Hatter. Sofern nicht anders angegeben, sind die Inhalte unter einer Creative Commons „Namensnennung - Nicht-kommerziell - Weitergabe unter gleichen Bedingungen 4.0 International Lizenz“ (CC BY-NC-SA 4.0) lizenziert.

Die Rechte an zitierten Abbildungen liegen bei den in der jeweiligen Quellenangabe genannten Urheber\*innen.

Die Kapitel 2 bis 4 der vorliegenden Bachelorarbeit wurden im Rahmen der Lehrveranstaltung „Bachelor Seminar 1“ eingereicht und am 26.01.2022 als Bachelorarbeit 1 angenommen.
--

## **Kurzzusammenfassung:**

Wallstreetbets - Statistischer Zusammenhang von Unternehmenserwähnungen auf [reddit.com/r/Wallstreetbets](https://www.reddit.com/r/Wallstreetbets) und deren Aktienkurs

Der Short Squeeze der GameStop-Aktie 2021 hat weit über die Grenzen der Finanzwelt für Aufsehen gesorgt. Von der breiten Medienlandschaft wurde von einem märchenhaften Sieg der Kleinanleger, organisiert über die Internetplattform [www.reddit.com/r/wallstreetbets](https://www.reddit.com/r/wallstreetbets), gegenüber den großen Hedgefonds berichtet. Neben der GameStop-Aktie werden auf der oberhalb angeführten Plattform auch weitere Unternehmensaktien in Zusammenhang mit potentiellen Short Squeezes diskutiert. Die gegenständliche Arbeit untersucht, ob tatsächlich ein statistischer Zusammenhang zwischen der Anzahl der Erwähnungen einer Unternehmensaktie auf Reddit und deren Aktienkurs nachweisbar ist. Dazu werden Textmining- und Sentimentanalyse-Verfahren eingesetzt, um die gewünschten Informationen von [www.reddit.com/r/wallstreetbets](https://www.reddit.com/r/wallstreetbets) zu extrahieren. Neben der Anzahl der Erwähnungen wird auch der jeweilige Kontext der Unternehmenserwähnung untersucht (positiv, neutral, negativ). Diese so gewonnenen Daten werden anschließend den jeweiligen Aktienkursen gegenübergestellt und mittels Kreuzkorrelation mit Zeitversatz auf mögliche Zusammenhänge untersucht.

Die Ergebnisse der Untersuchung zeigen, dass eine geringe Korrelation zwischen den Unternehmenserwähnungen ausgewählter Firmen auf der Internetplattform [www.reddit.com/r/wallstreetbets](https://www.reddit.com/r/wallstreetbets) und deren täglichem Aktienschlusskurs besteht.

## **Schlagwörter:**

Reddit; wallstreetbets; Statistik; Kreuzkorrelation; Spearman;

## **Abstract:**

Wallstreetbets – Statistical correlation between the number of company mentions on [reddit.com/r/Wallstreetbets](https://www.reddit.com/r/Wallstreetbets) and its stock price

The GameStop Short Squeeze in 2021 raised attention exceeding the boundaries of finance specific news. Broadcast companies reported a fairy-tale like win from the small investors, organized through the internet platform [www.reddit.com/r/wallstreetbets](https://www.reddit.com/r/wallstreetbets) over the big hedge funds. In addition to the GameStop stock (GME), other stocks are in discussion concerning short squeezing on Reddit. The goal of this paper is to examine if a statistical correlation does exist between the number of mentions of a company's stock per day on Reddit and its actual stock development. To do so, text mining and sentiment detection methods are used to extract the needed information form the comments of [www.reddit.com/r/wallstreetbets](https://www.reddit.com/r/wallstreetbets). In addition to the number of mentions per day, the context of the company mentioning is considered (positive, neutral, negative). Afterwards the generated data is compared to the historical stock data of the associated company. Cross corelation is performed to investigate possible nexuses between stock development and Reddit-activity.

The results of the study show, that only slightly significant correlations between the mentioned time series are provable.

## **Keywords:**

Reddit; wallstreetbets; statistics; cross corellation; spearman;

# Inhaltsverzeichnis

<b>GLOSSAR</b>	<b>1</b>
<b>1. EINLEITUNG</b>	<b>2</b>
1.1 Ziel der Arbeit	3
1.2 Vorgehensweise	4
1.3 Aufbau der Bachelorarbeit	5
<b>2. WIRTSCHAFTLICHE ENTWICKLUNG VON GAMESTOP UND VERGLEICHSUNTERNEHMEN</b>	<b>7</b>
<b>3. AUSGANGSLAGE DER LITERATUR</b>	<b>9</b>
<b>3.1 Ausgewählte Kapitalmarktgrundlagen</b>	<b>9</b>
3.1.1 Leerverkäufe (Short-Selling)	9
3.1.2 Short Squeeze	10
3.1.3 Neo-Broker	11
<b>3.2 Textmining und Sentimentanalyse</b>	<b>11</b>
3.2.1 Das Saarbrücker Pipelinemodell	12
3.2.1 Textmining Tools	12
<b>3.3 Mathematische Ermittlung statistischer Zusammenhänge</b>	<b>15</b>
3.3.1 Pearson-Korrelationskoeffizient	19
3.3.2 Kreuzkorrelation	21
3.3.3 Autokorrelation	22
<b>4. WIRTSCHAFTSPSYCHOLOGISCHE EINFLUSSFAKTOREN</b>	<b>24</b>
4.1 Boom der Kleinanlegerschaft	24
4.2 Trading Apps & beeinflussende Design Patterns	24
4.3 Steigende Risikobereitschaft von Privatanlegern	27
4.4 Social Media & Herdenverhalten	27
<b>5. BERECHNUNGSMODELL ZUR STATISTISCHEN ZUSAMMENHANGSERMITTLUNG</b>	<b>30</b>
5.1 Dateneinholung	30

5.1.1	Kommentarexport www.reddit.com	30
5.1.2	Einholung Finanzmarktdaten	35
5.2	Datenaufbereitung	36
5.3	Berechnungsmodell	37
6.	<b>ERSTELLUNG KURSPAARE &amp; VISUALISIERUNG</b>	<b>43</b>
7.	<b>ANWENDUNG DES BERECHNUNGSMODELLS</b>	<b>51</b>
7.1	Normalverteilung	51
7.2	Spearman-Korrelationskoeffizient	54
7.3	Autokorrelation	58
7.4	Kreuzkorrelation	59
8.	<b>SCHLUSSFOLGERUNGEN AUS DEN ERGEBNISSEN</b>	<b>64</b>
9.	<b>CONCLUSIO</b>	<b>68</b>
	<b>LITERATURVERZEICHNIS</b>	<b>71</b>
10.	<b>ANHANG</b>	<b>77</b>
10.1	Schiefe & Kurtosis	77
10.2	Shapiro-Wilk-Test	77
10.3	Spearman Rangkorrelationskoeffizient	78
10.3.1	Spearman Rangkorrelationskoeffizient Sentimentdaten	80
10.4	Autokorrelation	81
10.5	Kreuzkorrelation	82
10.6	AMC Entertainment Holdings, Inc. (AMC)	84
10.7	Virgin Galactic Holdings, Inc. (SPCE)	87
10.8	Tootsie Roll Industries, Inc. (TR)	90
10.9	Palantir Technologies Inc. (PLTR)	93
10.10	Nokia Oyj (NOK)	96
10.11	The Macerich Company (MAC)	99
10.12	American Airlines Group Inc. (AAL)	102
10.13	Express, Inc. (EXPR)	105

<b>10.14</b>	<b>Eastman Kodak Company (KODK)</b>	<b>108</b>
<b>10.15</b>	<b>BlackBerry Limited (BB)</b>	<b>111</b>
<b>10.16</b>	<b>Bed Bath &amp; Beyond Inc. (BBBY)</b>	<b>114</b>
<b>10.17</b>	<b>GameStop Inc. (GME)</b>	<b>117</b>

## Glossar

<b>Begriff</b>	<b>Bedeutung</b>
API	Application Programming Interface. Ein Satz von Befehlen, Funktionen, Protokollen und Objekten, die Programmierer verwenden können, um eine Software zu erstellen oder mit einem externen System zu interagieren.
Frontpage	Die erste dargestellte Seite auf einer Webseite, im aktuellen Fall eines Subreddits
Gamification	Verharmlosung von Finanzanlagevorgängen. Der Begriff wird eingesetzt, wenn der Handel am Finanzmarkt als Glücksspiel dargestellt oder verkauft wird.
Lag	Verzögerung eines Signals A zu Signal B
Lead	Vorlauf eines Signals A zu Signal B
Leverage	Hebelwirkung. Mit Hilfe von Leverage kann der mögliche Gewinn im guten Fall, um ein Vielfaches vergrößert werden
Library	Eine Sammlung von vorbereiteten Methoden und Funktionen in einer Programmiersprache
Neo-Broker	Online Broker mit günstigen Angeboten für Kleinanleger
Polarität	Ausprägung des Inhalts eines Texts – z.B.: positiv, neutral, negativ
Submission	Auch als Thread bezeichnet. Teilbereich eines Subreddits
Subreddit	Spezifisches, inhaltlich zusammenhängendes Forum auf der Internetplattform <a href="http://www.reddit.com">www.reddit.com</a>

# 1. Einleitung

Anfang des Jahres 2021 kam es an der New Yorker Börse (NYSE) zu einem Ereignis, welches in seiner Ausprägung und Form in der modernen Finanzwirtschaft völlig neu und noch nie dagewesen war.

Die Aktie der seit Jahren wirtschaftlich strauchelnden US-amerikanischen Einzelhandelskette GameStop erfuhr im Januar 2021 einen Kursanstieg von 20 US-Dollar auf zwischenzeitlich 480 US-Dollar. (Thorbecke 2021).

Da im Vorlauf des Kursanstiegs der GameStop-Aktie (GME) mehrere große Hedgefonds mit Leerverkäufen auf einen Fall der Aktie gesetzt hatten, führte der rasante Kursanstieg für manche von ihnen zu Verlusten in Milliardenhöhe (Thorbecke 2021).

Dieser sogenannte Short Squeeze sorgte auch weit über die Grenzen der Finanzwelt für Aufsehen und wurde in der breiten medialen Berichterstattung intensiv erörtert. In diesem Zusammenhang wurde von den großen Medienverlegern weltweit die heroische Geschichte eines modernen Robin Hood-Märchens erzählt, in der das kollektive Handeln von Kleinanlegern große Hedgefonds, wie etwa Melvin Capital, in die Knie gezwungen haben soll. Durch das gezielte Akquirieren große Mengen der GME-Aktie im Verbund soll dies gelungen sein. Organisiert über die Internetplattform [www.reddit.com](http://www.reddit.com), insbesondere über den Subreddit /r/Wallstreetbets, wurde der Short Squeeze eingeleitet. Eine detaillierte wissenschaftliche Untersuchung, ob diese Behauptungen tatsächlich der Realität entsprechen, wurde dabei von den Zeitschriften nicht durchgeführt. Vorhandene Literatur zeigt, dass es in dem gegenständlichen Fall erkennbare Zusammenhänge zwischen Internetaktivitäten und dem Kurs der GameStop-Aktie gibt. So zeigt eine aufgestellte Untersuchung von Evangelos, Eleftheria und Polydoros aus dem Jahr 2021, dass im Zeitraum des enormen Kursanstiegs von GME ein nachweisbarer statistischer Zusammenhang zwischen Google-Suchanfragen und der Entwicklung des Aktienkurses besteht (Evangelos, Eleftheria und Polydoros 2021).

In der Vergangenheit wurde zudem nachgewiesen, dass Aktivitäten auf Social Media wie etwa Twitter, den Aktienkurs ausgewählter Unternehmen beeinflussen können. Als plakatives Beispiel kann hier das Unternehmen Tesla mit Ihrem CEO Elon Musk erwähnt werden. Dieser nutzt gezielte Tweets zur Einflussnahme auf den Aktienkurs seiner Firma. So werden etwa jüngste Unternehmenserfolge und Investments über diesen Kanal der breiten Öffentlichkeit kundgetan. Eine von Bollen, Mao und Zeng 2010 durchgeführte Studie hat gezeigt, dass insbesondere Tweets den Aktienkurs bestimmter Unternehmen beeinflussen können (Bollen, Mao und Zeng 2010).

Auch im Zusammenhang mit der gegenständlichen Untersuchung des Aktienkurses der GME-Aktie und verwandter Aktiengruppen hat ein Tweet von Elon Musk das Herdenverhalten der Kleinanleger und dadurch die darauffolgende Kursentwicklung stark beeinflusst. Mehr zur Beeinflussbarkeit von unerfahrenen Anlegern und deren Herdenverhalten wird im Folgekapitel 4 detaillierter ausgeführt.

Die im Juni 2021 angestellte Arbeit von Anand und Pathak, vom indischen Institut für Management, zeigt zudem einen statistischen Zusammenhang zwischen den Ereignissen an der NYSE und den Reddit-Aktivitäten im Zeitraum von 01. Dezember 2020 bis 31. April 2021 (Anand und Pathak 2021). Es wird allerdings nicht beleuchtet, ob dieser Zusammenhang für weitere Unternehmen repräsentativ sein könnte, oder der Fall GameStop ein einmaliges Ereignis bleibt. Auch wird die Frage, ob in den Short Squeeze etwaige Großanleger, unabhängig von der Reddit-Community, involviert waren, außen vorgelassen.

Neben der GameStop-Aktie werden Aktien weiterer Unternehmen im Zusammenhang mit möglichen Short Squeezes im Forum auf [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets) geführt und wiederkehrend diskutiert. Auch diese Unternehmen weisen ungewöhnliche (nicht dem Marktwert entsprechende) Kursverläufe auf. Bei diesen Firmen handelt es sich um ein Konvolut von 22 Unternehmen, welche ebenfalls stark leerverkauft waren, beispielsweise den US-amerikanischen Kinobetreiber AMC-Entertainment.

## 1.1 Ziel der Arbeit

Im Rahmen dieser Arbeit wird beantwortet, ob die Erkenntnisse von Anand und Pathak, 2021, hinsichtlich des GME-Short Squeezes reproduzierbar sind. Weiterführend wird untersucht, ob ein statistischer Zusammenhang zwischen der Anzahl der Erwähnungen ausgewählter auf [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets) zum Short Squeeze aufgerufener Unternehmen, und dem Aktienkurs ebendieser Unternehmen, nachweisbar ist.

Konkret lautet die Forschungsfrage: Besteht im nachfolgend definierten Untersuchungszeitraum ein statistischer Zusammenhang zwischen der Erwähnungsanzahl bestimmter, zum Short Squeeze ausgewählter Unternehmensaktien im Forum [www.reddit.com/r/wallstreetbets](http://www.reddit.com/r/wallstreetbets) und dem Aktienkurs ebendieser Unternehmen?

Die angenommene Hypothese ist hierbei, dass zwischen der Anzahl der Erwähnungen einzelner, ausgewählter, zum Short Squeeze auserkorenen Unternehmensaktien im Forum [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets) und dem Aktienkurs ebendieser Unternehmensaktien, ein statistischer Zusammenhang besteht.

Die Entwicklung eines Prognosemodells zur etwaigen Frühzeiterkennung von ähnlichen Marktentwicklungen ist nicht Teil der gegenständlichen Arbeit.

## 1.2 Vorgehensweise

Als Betrachtungszeitraum für die Untersuchung wird Juni 2020 – der Start der Entwicklungen rund um GME - bis Juni 2021 gewählt. Aus dem Forum [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets) werden dazu sämtliche Kommentare im Betrachtungszeitraum exportiert und mittels Textmining und -analyseverfahren auf mehrere Gesichtspunkte, wie die Anzahl der Unternehmenserwähnungen und dem Kontext (positiv, neutral, negativ), in welchem diese Aktien erwähnt wurden, tagesfein untersucht (Sentimentanalyse).

Anschließend werden die so gewonnenen Kommentarkurse den täglichen Aktienkursen gegenübergestellt, und eine Überprüfung von statistischen Zusammenhängen durchgeführt.

Im ersten Schritt wird dabei mittels Pearson-Korrelationskoeffizient geprüft, ob im Betrachtungszeitraum zwischen den gegenständlichen Stichproben - Anzahl der Unternehmenserwähnungen ausgewählter Unternehmen und deren Aktienkursen – generell ein linearer Zusammenhang besteht. Vertiefend wird anschließend auf die normalisierte Kreuzkorrelation unter Berücksichtigung eines Zeitversatzes zurückgegriffen. Dabei wird dezidiert ermittelt, ob eine Korrelation zwischen der Zeitreihe der täglichen Unternehmenserwähnungen in den Kommentaren unter [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets) und dem Aktienkurs dieses Unternehmens nachweisbar ist. Insbesondere wird dabei im Falle einer Korrelation untersucht, mit welchem Lag (Zeitversatz) jene auftritt. Weiterführend wird anhand der Autokorrelation untersucht, ob und mit welchem Versatz die täglichen Aktien- und Kommentarkurse mit sich selbst korrelieren. Um Falschinterpretationen von Scheinkorrelationen zu vermeiden, kommen Signifikanztests und unterstützend Sentimentanalysen der einzelnen Kommentare zum Einsatz. Zusätzlich werden vor den Korrelationsprüfungen Untersuchungen der Stichprobenverteilungen mittels Histogrammen, Formmaßen und Shapiro-Wilk-Test durchgeführt, um die korrekten Berechnungswerkzeuge auszuwählen.

Weiterführend werden die volkswirtschaftlichen und wirtschaftspsychologischen Faktoren, unter anderem das Herdenverhalten und die Risikobereitschaft bei unerfahreneren Kleinanlegern, welche auch das Eintreten und das Wiederkehren von solchen volkswirtschaftlichen Phänomenen begünstigen könnten, beleuchtet.

Im Rahmen der wirtschaftspsychologischen Aufarbeitung der Umstände, die das Eintreten des GME-Kursanstiegs gefördert haben, werden keine neuen Untersuchungen durchgeführt, sondern vorhandene wissenschaftliche Arbeiten zusammengetragen, entsprechende Einflussfaktoren kumuliert und in den Kontext der gegenständlichen Forschungsfrage gesetzt.

Vor dem Hintergrund dieser Arbeit wird dabei insbesondere auch der Einfluss, den bestimmte Design Patterns von Trading Plattformen auf das Anlageverhalten von unerfahrenen Tradern nehmen können (Gamification) (Chaudhry und Kulkarni 2021), und die Organisation und Informationsgewinnung von Kleinanlegern über Social Media Plattformen beleuchtet (Yarovaya et al. 2021).

Zudem wird dargestellt, wie sich ebendiese Informationsgewinnung auch seit dem Start der COVID-19-Pandemie entwickelt und verändert hat.

Darauf aufbauend wird kurz auf das im Jänner 2021 zunehmend starke mediale Interesse am gegenständlichen Fall und die mögliche Befeuerung des dadurch sich entwickelnden Hypes eingegangen.

Aufgrund des sich zunehmend verändernden Gefüge des Kapitalmarkts, hier sei im aktuellen Fall die steigende Anzahl der Kleinanleger, sowie der Einfluss von Social Media auf das Anlageverhalten einer breiten Personengruppe erwähnt, soll die gegenständliche Arbeit dazu dienen, für Kleinanleger, Unternehmen aber auch Finanzintermediäre und Aufsichtsorgane die umfangreiche Beeinflussung von Marktgegebenheiten durch die gezielte Organisation von Kleinanlegern darzustellen.

Sie soll als Ausgangspunkt für Überlegungen zur Stabilisierung oder Revolutionierung von Sektoren, sowie Regularien des Kapitalmarkts dienlich sein. Es wird dargestellt, dass der Finanzmarkt des 21. Jahrhunderts bisher noch nie dagewesene Dynamiken entwickeln kann.

Persönlich haben mich die Ereignisse rund um den GameStop-Short Squeeze fasziniert, da ich selbst gelegentlich im Forum [www.reddit.com/r/Wallstreetbets](https://www.reddit.com/r/Wallstreetbets) unterwegs bin. Ich möchte daher herausfinden, ob ein statistischer Zusammenhang zwischen Aktienkurs und Forenaktivität für die weiteren zum Short Squeeze aufgerufenen Unternehmen nachweisbar ist.

### **1.3 Aufbau der Bachelorarbeit**

Im folgenden Kapitel wird die wirtschaftliche Entwicklung von GameStop seit der Jahrtausendwende beschrieben, um ein besseres Bild über den exorbitanten Kursanstieg darzustellen. Darauffolgend werden in den nächsten Abschnitten die volkswirtschaftlichen Einflussfaktoren, welche eine Begünstigung der Ereignisse mit sich gebracht haben könnten, dargelegt.

Es folgt eine Darstellung der Möglichkeiten zur Textanalyse und welche Verfahren für die Bearbeitung des gegenständlichen Untersuchungsinhalts zweckdienlich sind.

Die Aufarbeitung der derzeitigen wissenschaftlichen Grundlagen schließt mit der Beschreibung der mathematischen Parameter zum Nachweis statistischer Zusammenhänge. Insbesondere liegt hier der Fokus auf dem Pearson-Korrelationskoeffizienten und der Kreuzkorrelation.

In dem darauffolgenden Kapitel werden die volkswirtschaftlichen und wirtschaftspsychologischen Faktoren, die das Eintreten solcher Marktphänomene begünstigen können, anhand einer Reihe von durchgeführten Studien detailliert untersucht, korreliert und Schlussfolgerungen aufgestellt.

Anschließend wird theoretisch erläutert, welcher Lösungsweg zur weiteren Beantwortung der Forschungsfrage gewählt wird. Dabei wird im Detail auf die durchzuführenden Arbeitsschritte und Analyseverfahren eingegangen.

Im zweiten Block der Arbeit wird der zuvor entwickelte Lösungsweg für die Aktienkurse ausgewählter Unternehmen im Betrachtungszeitraum angewandt. Die Untersuchungsergebnisse werden analysiert, gegen die Hypothese geprüft und zusammengefasst. Erkenntnisse und mögliche Folgemaßnahmen werden gelistet.

Abschließend folgt eine Conclusio, sowie eine Handlungsempfehlung.

## **2. Wirtschaftliche Entwicklung von GameStop und Vergleichsunternehmen**

Um die Brisanz der dramatischen Kursentwicklung und das darauffolgende Medieninteresse besser zu verstehen, wird im folgenden Kapitel ein kurzer Überblick über die vergangenen 20 Geschäftsjahre des Unternehmens GameStop gegeben.

GameStop ist eine börsennotierte, US-amerikanische Einzelhandelskette für Computerspiele und Unterhaltungssoftware. Zusätzlich tritt sie auch als Verleger des Computerspielmagazins „Game Informer“, sowie als Eigentümer diverser Tochtergesellschaften auf. Das Unternehmen betreibt weltweit mehr als 7500 Filialen, davon 1200 in Europa (United States Securities and Exchange Commission 2016).

Im Jahr 2002 wurde GameStop an die Börse gebracht. Durch eine kluge Fusions- und Erwerbsstrategie von Mitbewerbern erlebte Gamestop in den Jahren 2004 bis 2016 einen wirtschaftlichen Aufschwung.

Seit 2016 fiel der Börsenwert des Unternehmens sukzessiv (Bary 2017). Ausschlaggebend dafür war der branchenübergreifend wachsende Online Handel. Bereits 2012 mussten mehrere hunderte Filialen geschlossen werden, 2017 folgten weitere. 2018 wurde mit 673 Millionen US-Dollar der bis dato größte Verlust der Firmengeschichte geschrieben. Seither wird ein Verkauf des Unternehmens an einen Investor nicht ausgeschlossen. Nach erneuten umsatzbedingten Schließungen im Jahr 2019 erzielte das Unternehmen im zweiten Quartal 2019 einen Verlust von 400 Millionen US-Dollar (Wikipedia 2021).

Durch den steigenden Konsum von Computerspielen profitierte ab Herbst 2020 auch GameStop, obwohl ein Großteil der Filialen geschlossen bleiben musste. Anfang 2021 stieg der Kurs der Unternehmensaktie von 20 US-Dollar auf über 480 US-Dollar am 28.01.2021. Als Grund gilt der Einstieg des Investors Ryan Cohen und einer Vielzahl von Privatanlegern nach Diskussionen auf Twitter und dem Subreddit „Wallstreetbets“.

Diese Entwicklung wurde zum Problem für Hedgefonds, die zuvor Leerverkäufe von GME-Aktien getätigt hatten – weit mehr als Aktien vorhanden waren -, da auf fallende Kurse und daraus zu erzielende Gewinne gesetzt wurde.

Im Zusammentreffen mit der kurstreibenden Spekulation der Privatanleger wurden die Leerverkäufer jedoch genötigt, ihre Positionen mit Rückkäufen zu jedem zustande kommenden Preis aufzulösen, was zu Milliardenverlusten für die Hedgefonds führte. Es kam zu einem Short Squeeze (Thorbecke 2021). Die Marktkapitalisierung des Unternehmens GameStop stieg so zwischenzeitlich auf über 20 Milliarden US-Dollar.

Während des Erfolgs mit der GameStop-Aktie (aus Sicht der Reddit-User) wurden weitere Unternehmen mit einer hohen Shortquote identifiziert und von der Userschaft zum Investment aufgerufen.

Jene Unternehmen verzeichnen eine ähnliche wirtschaftliche Entwicklung wie GameStop. Auch sie haben zu Jahresbeginn 2021 starke Kursanstiege erfahren. Es handelt sich dabei um die folgenden Firmen, die im Rahmen der weiteren Arbeit, gemeinsam mit GameStop, auf statistische Zusammenhänge zwischen der Forenaktivität und dem Aktienkurs untersucht werden:

- AMC Entertainment Holdings, Inc. (AMC)
- AMC Networks Inc. (AMCX)
- American Airlines Group Inc. (AAL)
- BB Liquidating Inc. (OTC Pink: BLIAQ)
- Bed Bath & Beyond Inc. (BBBY)
- BlackBerry Limited (BB)
- Build-A-Bear Workshop, Inc. (BBW)
- Eastman Kodak Company (KODK)
- Express, Inc. (EXPR)
- Fossil Group, Inc. (FOSL)
- Genius Brands International (GNUS)
- iRobot Corporation (IRBT)
- Koss Corporation (KOSS)
- Ligand Pharmaceuticals Incorporated (LGND)
- The Macerich Company (MAC)
- Naked Brand Group (NAKD)
- National Beverage Corp. (FIZZ)
- Nokia Oyj (NOK)
- Palantir Technologies Inc. (PLTR)
- Siebert Financial (SIEB)
- Tootsie Roll Industries, Inc. (TR)
- Virgin Galactic Holdings, Inc. (SPCE)

### **3. Ausgangslage der Literatur**

Um das aufgetretene Phänomen GameStop 2021 und weitere auf [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets) diskutierte Unternehmen hinsichtlich eines statistischen Zusammenhangs zwischen Erwähnungsanzahl und Aktienkurs zu untersuchen, werden in den folgenden Kapiteln Auszüge aus volkswirtschaftlichen, mathematischen und informationstechnischen Themengebieten ausgeführt. Dies dient dazu die Grundlagen und den aufbauenden Stand aus Wissenschaft und Technik, zum gewählten Lösungsweg angeführt in Kapitel 5, darzulegen.

Begonnen wird einleitend mit einem Überblick zu den benötigten Kapitalmarktgrundlagen.

#### **3.1 Ausgewählte Kapitalmarktgrundlagen**

##### **3.1.1 Leerverkäufe (Short-Selling)**

Unter einem Leerverkauf (engl. Short Sale) wird der Verkauf von Wertpapieren, Waren, Devisen, Optionen oder Futures im Kassa- oder Termingeschäft verstanden, die sich nicht im Eigentümern des Verkäufers (Leihgut) befinden. Diese Geschäftspraktik wird dann angewandt, wenn der Verkäufer davon ausgeht, dass der Wert des Verkaufsguts fällt und er jenes Gut später zu einem günstigeren Zeitpunkt zurückkaufen kann. Aus der Differenz zwischen Verkaufs- und Kaufpreis erzielt er dadurch Profit (Differenzgeschäft). Man spricht in so einem Fall von einer Short-Position (Cambridge Dictionary 2021).

Es wird beim Leerverkauf zwischen gedeckten und ungedeckten Leerverkäufen unterschieden. Bei gedeckten Leerverkäufen verfügt der Verkäufer über das geliehene Verkaufsgut der Transaktion. Im Falle von ungedeckten Leerverkäufen („Naked Short Sale“) ist der Verkäufer dagegen zum Zeitpunkt des Verkaufs weder im Besitz des Guts, noch hat er es geliehen (Whitebox 2020).

Traditionell wird die Geschäftspraktik der Leerverkäufe von Hedgefonds angewandt – so auch in den Fällen der „Wallstreetbets“. Die Leerverkaufsquote im Falle von GameStop betrug im Januar 2021 140%, was bedeutet, dass 140% des Handelsvolumens leerverkauft waren (Burnette 2021). Die durchschnittliche Shortquote von Unternehmen beträgt nur rund 5% (Angel 2021).

Leerverkäufe bergen jedoch ein hohes Risiko. Fallen die Kurse nicht, sondern steigen sie, streicht der Leerverkäufer einen Verlust ein. Dieser Verlust kann – mit steigendem Kurs des Guts – theoretisch ins Unendliche steigen.

### 3.1.2 Short Squeeze

Ein Short Squeeze tritt nun ein, wenn ein Basiswert, beispielhaft eine Aktie, im Kurs stark ansteigt, was die Marktteilnehmer, die auf fallende Kurse gesetzt haben (Short Positionen) dazu zwingt, ihre Shortpositionen zurückzukaufen um ihre Verluste zu minimieren.

Marktbedingungen die zum Eintreten eines solchen Ereignisses führen, sind etwa grundlegende Unternehmensveränderungen oder unerwartete Erfolgsmeldungen, die die Hoffnung auf einen Kursaufschwung schüren (Evangelos 2021). Dies tritt jedoch auch nur dann ein, wenn genügend Leverage (Markteinfluss) der entsprechenden Nachfragenden vorhanden ist, um den Aktienkurs demgemäß anzukurbeln (Burnette 2021).

Das Eintreten von Short Squeezes ist nichts Neues. Allen, Litov und Mei dokumentierten, dass es bereits im frühen 18. und 19. Jahrhundert regelmäßig zu Squeezes und Corners<sup>1</sup> gekommen ist. Ein klassisches Beispiel ist etwa der Harlem Corner aus dem Jahr 1863 (Allen, Litov und Mei 2006).

Der umfänglich größte und aktuellste Short Squeeze, vor den „Wallstreetbets“, war der Fall Volkswagen-Porsche im Jahr 2008. Damals stieg der Preis der Volkswagen-Aktie von EUR 210,- am 24.10.2008 auf EUR 1005,- am 28.10.2008 und machte Volkswagen kurzzeitig zum wertvollsten Unternehmen weltweit, basierend auf der Marktkapitalisierung (Allen et. al 2017). Der damalige Short Squeeze der VW-Aktie wurde von Porsche, welches sich als Shareholder von VW in einer katastrophalen Wirtschaftslage befand, initiiert, um deren eigene Liquidität zu erhöhen (Allen et. al 2017).

---

<sup>1</sup>Extremform eines Squeezes, bei dem die Käuferseite nahezu die komplette Kontrolle über alle Basiswerte inne hat

### **3.1.3 Neo-Broker**

Unter Neo-Brokern versteht man eine neue Form von Online Brokern, die seit einigen Jahren am Markt vertreten sind. Von der Gruppe der Neo-Broker wird eine differenzierte Geschäftsstrategie im Vergleich zu konventionellen Bank-Brokern verfolgt (Gäbler 2021).

Das Ziel der Neo-Broker wie etwa Trade Republic oder Robin Hood ist es, die Marktzugänglichkeit und -informationen für jedermann zur Verfügung zu stellen, unabhängig von möglichen Investmentvolumen. Meist erfolgt diese Bereitstellung des Marktzugangs in Form von Mobile Apps oder Webauftritten. Um dieses Ziel zu erreichen wird auf schlanke Unternehmensstrukturen und simple Anwenderoberflächen gesetzt, die dadurch gleichzeitig wartungsarm und kostenschonend sind. Durch die Verschlankung des Investmentprozesses und die einfache Strukturierung der Anwendungen gelingt es Neo-Brokern einen breiten Userkreis anzusprechen. Insbesondere die junge Investmentzielgruppe ab 18 Jahren wird damit erreicht. Erfahrene Anleger stoßen jedoch an die Grenzen der Investmentmöglichkeiten von Neo-Brokern, da durch die Simplifizierung und das Prinzip der Kostendeckelung bestimmte Teile des Kapitalmarkts über die Neo-Broker nicht zugänglich sind (Gäbler 2021).

## **3.2 Textmining und Sentimentanalyse**

Unter Textmining oder Natural Language Processing (NLP) wird eine Subkategorie des Data Mining verstanden, welche sich mit der maschinenbasierten Analyse und Informationsgewinnung aus großen Textmengen befasst. Die Eingliederung im Wissenschaftsfeld des Data Minings kann wie folgt getroffen werden.

Data Mining sucht relevante Information in Daten. Text Mining sucht relevante Information in Sprachdaten. Sentimentanalyse oder Opinion Mining versucht, Meinungsäußerungen (Information) in Newsgroups und Foren automatisch zu erkennen und zu klassifizieren, um damit letztlich Wissen über Meinungen zu extrahieren.

Jene Methoden werden im Social Media Bereich zur Analyse von Kommentarzusammenhängen oder zur Gewinnung von neuen Informationen aus umfangreichen Bibliotheken von Schriftstücken angewandt. Generell spricht man im deutschsprachigen Raum daher von Computerlinguistik (CL) oder linguistischer Datenverarbeitung (LDV). Das Ziel dabei ist die Polarität einer Aussage (positiv, neutral, negativ) und deren Subjektivität maschinenbasiert aus Texten zu extrahieren.

Sentimentanalyseverfahren werden auch in der gegenständlichen Untersuchung eingesetzt, um positive oder negative Polaritäten in Kommentaren zu erkennen und deren eventuelle Auswirkung auf den Aktienkurs richtig zu kategorisieren.

Um natürliche Sprache in Form von Text- und Sprachdaten mit Hilfe eines Computers algorithmisch verarbeiten zu können, bedarf es einer Reihe von Arbeitsschritten, die sequentiell durchlaufen werden müssen. Erst dann ist es möglich, korrekte Schlüsse aus dem vorliegenden Datenmaterial zu ziehen. Bei den durchzuführenden Schritten spricht man üblicherweise von dem Saarbrücker Pipelinemodell.

### **3.2.1 Das Saarbrücker Pipelinemodell**

Falls der Text in Schallinformation vorliegt, so muss dieser mittels Spracherkennung erst in Text umgewandelt werden. Im nächsten Schritt wird die Buchstabenkette in Wörter, Sätze etc. segmentiert. Hierbei ist beim Einsatz für Online Foren weiters die Entfernung von etwaigen Hyperlinks, Sonderzeichen oder Emojis zu achten. Dieser Schritt wird als Tokenisierung bezeichnet.

Weiterführend wird eine Morphologische Analyse durchgeführt. Personalformen und Fallmarkierungen werden dabei analysiert und Wörter in ihre Grundform zurückgeführt. Beispielhaft wird aus dem Begriff „trank“ wie in „Er trank seinen Tee.“, durch morphologische Analyse die Grundform „trinken“ gebildet. Anschließend erfolgt eine syntaktische und semantische Analyse jedes Satzes bzw. deren Satzteile. Diese Schritte umfassen potentiell eine Vielzahl verschiedener Einzelschritte.

Darauffolgend kann im Bedarfsfall auch eine Dialog- und Diskursanalyse implementiert werden, in welcher die Beziehungen zwischen aufeinander folgenden Sätzen erkannt werden (z.B.: Frage und Antwort).

Das Saarbrücker Pipelinemodell ist so konzipiert, dass nicht immer sämtliche Verfahren der Computerlinguistik die gesamte Kette durchlaufen müssen. Die zunehmende Verwendung von maschinellen Lernverfahren hat zu der Einsicht geführt, dass auf jeder Analyseebene statistische Regelmäßigkeiten existieren, welche zur Sprachmodellierung eingesetzt werden können (Kolb 2011).

### **3.2.1 Textmining Tools**

Um Natural Language Processing (Textmining) durchzuführen, müssen die Algorithmen zur Abwicklung der einzelnen Arbeitsschritte des Saarbrücker Pipelinemodells nicht neu entwickelt werden, sondern es kann auf eine breite Palette von bereits vorhandenen technischen Tools und Systemen zurückgegriffen werden. Im Folgenden wird nur ein kleiner Bruchteil der Möglichkeiten von bestehenden Textmining-Systemen vorgestellt und angeführt. Es wird weiterführend dargelegt, weshalb genau die gewählte Methode für die Beantwortung der gegenständlichen Fragestellung eingesetzt wird.

Vor allem die beiden Programmiersprachen R und Python bieten kostenlos umfangreiche Bibliotheken zur Durchführung von computerlinguistischen

Untersuchungen an. Zudem stellen beide Sprachen Libraries zum Export von Kommentaren aus der Internetplattform [www.reddit.com](http://www.reddit.com) zur Verfügung.

Im aktuellen Anwendungsfall wird aufgrund der umfangreicheren Möglichkeiten zum Reddit-Kommentarexport und der NLP-Funktionalitäten die Programmiersprache Python in der Version 3.10. eingesetzt (Reitz und Schuster 2016a).

Um die geforderten textuellen Analysen durchführen zu können, bedarf es einer Reihe von bereits entwickelten Bibliotheken, wovon die wichtigsten dabei nachfolgend kurz beschrieben werden.

### Pandas

Die Pandas Bibliothek stellt Funktionen und Datenstrukturen zur Manipulation von numerischen Tabellen, Texten und Zeit-Serien zur Verfügung. Sie wird gegenständlich verwendet um den Text gemäß dem Saarbrücker Pipelinemodell aufzubereiten und zu tokenisieren (Kolb 2011). Weiters bietet Pandas umfangreiche Funktionen zur Analyse von Zeitreihen.

### NLTK

Die Abkürzung NLTK steht für Natural Language Toolkit und diese Library stellt ebendiese Funktionen für den User dar. Mithilfe der von der Stanford University entwickelten Bibliothek ist es möglich, morphologische Analysen, sowie syntaktische und semantische Auswertungen inklusiver aller Teilschritte durchzuführen. Das Paket bietet diese Funktionalität allerdings nur für die englische Sprache an, was für die weitere Lösungsfindung der gegenständlichen Bachelorarbeit jedoch irrelevant ist (Reitz und Schuster 2016b).

### TextBlob

Aufgrund der umfangreichen Möglichkeiten von NLTK wurde mit Textblob eine Bibliothek entwickelt, die eine simple API darstellt, welche auf der Library NLTK aufbaut und es ermöglicht, komplizierte Textanalysemethoden mithilfe von vorbereiteten Funktionen, einfach durchzuführen. Textblob wird im Rahmen der Lösungsfindung zur Anzahlermittlung der einzelnen einschlägigen Kommentare (enthält zum Beispiel die Erwähnung von AMC) und zur Sentimentanalyse eingesetzt (Textblob 2021).

### PRAW

PRAW steht für Python Reddit API Wrapper und ermöglicht es, wie es der Name verspricht, Informationen direkt von der Webplattform [www.reddit.com](http://www.reddit.com) zu extrahieren. Mit dieser Bibliothek ist es möglich nahezu alle einsehbaren Informationen der Redditplattform abzurufen und automatisiert weiterzuverarbeiten. Der große Vorteil gegenüber etwa vergleichbaren Bibliotheken bei anderen Programmiersprachen wie R besteht bei PRAW darin, dass die Funktionen in Einklang mit den Reddit-Coding-

Richtlinien entwickelt wurden, und so der User nicht auf spezifische Regelungen wie zum Beispiel die maximal erlaubte Abfrageanzahl von Kommentaren pro Minute, Rücksicht nehmen muss.

Mithilfe der PRAW-Bibliothek werden im Rahmen dieser Forschungsarbeit die Kommentare vom Subreddit „Wallstreetbets“ extrahiert und in eine weiter verarbeitbare Form gebracht (PRAW 2021).

Neben der weiter oben bereits beschriebenen Pandas-Library kommen zur Berechnung etwaiger statistischer Zusammenhänge die folgenden Python Programmbibliotheken zum Einsatz:

### NumPy

Hierbei handelt es sich um eine Bibliothek, die dem Anwender umfassende mathematische Berechnungen für Datenfelder (Arrays) zur Verfügung stellt. Mithilfe von NumPy können Operationen wie das Wurzelziehen bis hin zu einfachen Anwendungen der deskriptiven Statistik durchgeführt werden (NumPy 2021).

### Statsmodels

Statsmodels ist ein Python-Modul, das eine Vielzahl von Klassen und Funktionen zur Bewertung und Einschätzung statistischer Modelle, Durchführung von statistischen Tests und vertiefter Datenanalyse bereitstellt. (Statsmodels 2021).

### Matplotlib

Die Matplotlib-Library wird im Rahmen der gegenständlichen Untersuchung eingesetzt, um die gewonnenen Erkenntnisse grafisch zu visualisieren. Mithilfe dieser Bibliothek ist es möglich mathematische Darstellungen aller Art anzufertigen (Matplotlib 2021).

### 3.3 Mathematische Ermittlung statistischer Zusammenhänge

#### Normalverteilung

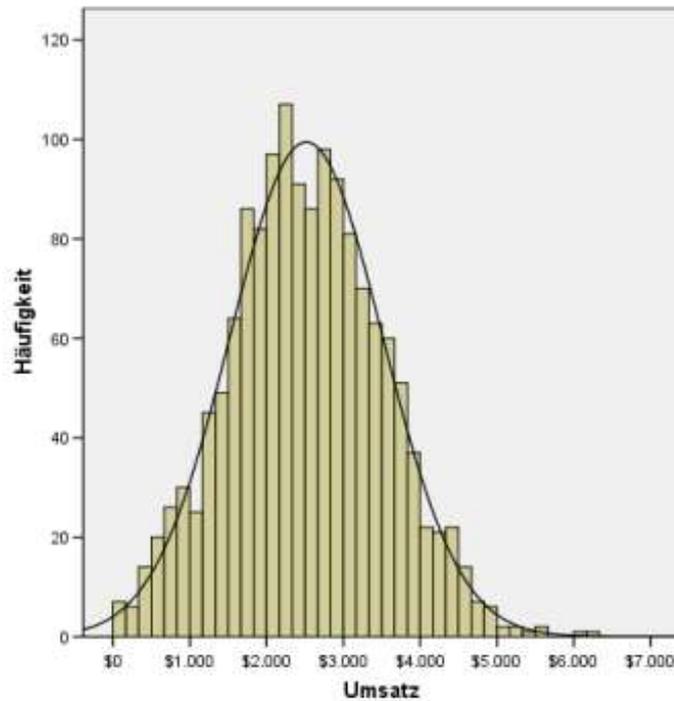
Viele Sachverhalte aus Natur- und Sozialwissenschaften folgen dem Prinzip der Normalverteilung. Jenes besagt, dass der Großteil der Werte in einer zufällig zusammengestellten Stichprobe dem Durchschnitt der Einzelwerte entspricht und nur wenige Werte viel kleiner oder viel größer als dieser Durchschnitt sind. Eine Normalverteilung wird in der Statistik dann eingesetzt, wenn die tatsächliche Verteilungsfunktion einer Stichprobe unbekannt ist (Staudinger 2021).

Hintergrund des hohen Stellenwerts der Normalverteilung nimmt der zentrale Grenzwertsatz an, welcher besagt, dass der Mittelwert einer Stichprobe aus der Grundgesamtheit näherungsweise dem Mittelwert der Grundgesamtheit selbst entspricht. Für vertiefende Informationen sei auf die zugrunde liegende Literatur verwiesen (Staudinger 2021).

Aufgrund der weiten Verbreitung der Normalverteilung wird daher in der gegenständlichen Untersuchung angenommen, dass auch die Unternehmenserwähnungen in den Kommentaren auf der Plattform [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets), sowie der Aktienkurs der jeweiligen Unternehmen im Betrachtungszeitraum, normalverteilte Daten darstellen.

#### Histogramm

Um diese Annahme zu verifizieren und sicherzustellen, dass es sich bei den gegenständlichen Stichproben um normalverteilte Daten handelt, kann im ersten Schritt ein Histogramm der vorliegenden Werte angefertigt werden. Anhand des Histogramms ist es sodann möglich zu beurteilen, ob eine Normalverteilung bzw. eine annähernde Normalverteilung vorliegt (Ebermann 2010a). Diese Erkenntnis aus der grafischen Beurteilung stellt die Basis für alle weiteren Berechnungsschritte dar. Unterhalb ist ein Beispielhistogramm, welches eine deutliche Normalverteilung darstellt, abgebildet.



**Abbildung 1: Beispielhistogramm Normalverteilung (Ebermann 2010a)**

### Formmaße zur Beurteilung einer Normalverteilung

Da anhand des Histogramms nur ein erster grafischer Eindruck über die Stichprobenverteilung gewonnen werden kann, wird im folgenden Kapitel auf die Formmaße Schiefe und Kurtosis eingegangen.

Diese beiden Werkzeuge werden eingesetzt um objektiv zu überprüfen, ob und wie sehr die gegenständlichen Daten der Stichprobe von einer Normalverteilung abweichen.

### Schiefe

Die Schiefe ist eine statistische Kennzahl, welche die Ausprägung der Asymmetrie einer Verteilung angibt. Sie zeigt an, ob und wie stark eine Verteilung nach links oder rechts geneigt ist. Generell wird jede nicht symmetrische Verteilung als schief bezeichnet.

Die Kennzahl wird anhand der folgenden Formel geschätzt:

$$Sch = \frac{\sum_{m=1}^n (x_m - \bar{x})^3}{n * s_x^3}$$

**Abbildung 2: Formel Schiefe (Blanz 2021)**

$x_m$  stellt dabei den jeweiligen Wert der Stichprobe,  $\bar{x}$  den Mittelwert,  $n$  die Anzahl der Werte in der Stichprobe und  $s$  die Standardabweichung dar. Eine Verteilung mit der Schiefe von 0 stellt eine perfekt symmetrische Verteilung dar. Verteilungen mit einem Schiefewert größer 0 werden linkssteil, Werte kleiner 0 als rechtssteil bezeichnet. Je näher eine Verteilung an 0 liegt, desto symmetrischer ist sie (Blanz 2021).

### Kurtosis

Das zweite gegenständlich anzuwendende Formmaß wird Kurtosis oder auch Exzess bzw. Wölbung bezeichnet. Die Kennzahl der Kurtosis gibt die Steilheit einer Verteilung an und drückt dadurch auch die Abweichung des Verlaufs einer Verteilung vom Verlauf einer Normalverteilung aus.

Es wird hierbei zwischen breitgipfliger (platykurtischer) und schmalgipfliger (leptokurtischer) Ausprägung unterschieden. Die Berechnung erfolgt anhand der untenstehenden Formel:

$$Ku = \frac{\sum_{m=1}^n (x_m - \bar{x})^4}{n * s_x^4} - 3$$

**Abbildung 3: Formel Kurtosis (Blanz 2021)**

Wie auch bei der Schiefe stellt  $x_m$  dabei den jeweiligen Wert der Stichprobe,  $\bar{x}$  den Mittelwert,  $n$  die Anzahl der Werte innerhalb der Stichprobe und  $s$  die Standardabweichung dar. Ein Kurtosiswert von 0 bedeutet gegenständlich, dass die Verteilung einer Normalverteilung entspricht. Verteilungen mit einem Kurtosiswert größer 0 werden schmalgipflig, Werte kleiner 0 als breitgipflig bezeichnet.

### Shapiro-Wilk-Test

Der Shapiro-Wilk-Test stellt eine weitere Möglichkeit zur Untersuchung einer Stichprobe auf das Vorliegen einer Normalverteilung dar. Es handelt sich hierbei um einen statistischen Signifikanztest. Als Nullhypothese wird dabei angenommen, dass die Grundgesamtheit normalverteilt ist (Shapiro und Wilk 1965).

Der von Shapiro und Wilk entwickelte Normalverteilungstest liefert eine Kennzahl, die die Ausprägung der Verteilung ausdrückt. Neben grafischen Prüfungen anhand von Histogrammen, welche zu einem gewissen Grad von der Subjektivität des Betrachters abhängen, wird so eine objektive Betrachtung der Stichprobe ermöglicht.

Zur Durchführung werden initial die untenstehende Nullhypothese  $H_0$  und die Alternativhypothese  $H_1$  aufgestellt.

$$H_0: F = F_0 \text{ und } H_1: F \neq F_0$$

**Abbildung 4: Hypothese und Alternativhypothese Shapiro-Wilk-Test (Shapiro und Wilk 1965)**

Zusätzlich wird das Signifikanzniveau, üblicherweise  $\alpha=5\%$ , gewählt.

Im nächsten Schritt werden die Ordnungsstatistiken der Stichprobe erstellt. Dazu werden die Werte nach aufsteigender Größe sortiert und jedem Wert ein Rang zugeordnet. Anschließend erfolgt die Berechnung der Teststatistik ( $W$ ) anhand der folgenden Formel.

$$W = \frac{b^2}{(n-1)s_x^2}$$

**Abbildung 5: Teststatistik ( $W$ ) Shapiro-Wilk-Test**

$b^2$  steht hierbei für den unter Vorliegen einer Normalverteilung erwarteten Wert der Varianz der Stichprobe. Zur Ermittlung der erwarteten Varianz  $b^2$  werden die zuvor angefertigten Ordnungsstatistiken herangezogen. Davon werden paarweise die Differenzen zwischen dem größten und dem kleinsten Wert, dem zweitgrößten und dem zweitkleinsten Wert usw. gebildet. Bei ungerader Werteanzahl bleibt so der mittlere Wert übrig. Die ermittelten Differenzen werden mit den von Shapiro-Wilk vorgegebenen Koeffizienten multipliziert, über alle Paare addiert und die daraus entstandene Summe abschließend quadriert.

Der Wert  $s_x^2$  steht für die nicht korrigierte Stichprobenvarianz. (Shapiro und Wilk 1965). Der Wert der Teststatistik  $W$  wird folglich einem kritischen Wert  $Wk$  für einen gegebenen Stichprobenumfang gegenübergestellt und mit dem zuvor gewählten Signifikanzniveau verglichen. Ist der Wert  $W$  größer als  $Wk$  wird die Nullhypothese nicht abgelehnt und es wird angenommen, dass eine Normalverteilung vorliegt. Die kritischen Werte  $Wk$  liegen in Abhängigkeit vom Stichprobenumfang und des Signifikanzniveaus tabelliert vor.

Die Teststatistik  $W$  kann wie ein Korrelationskoeffizient, der Werte zwischen 0 und 1 annehmen kann, interpretiert werden. Je näher die Teststatistik an 1 liegt, desto weniger Abweichungen zeigt die tatsächliche Varianz von der hypothetischen, unter Annahme des Vorliegens einer Normalverteilung. Gibt es jedoch signifikante statistische Abweichungen und  $W$  ist kleiner als  $Wk$ , so wird die Nullhypothese

abgelehnt und die Alternativhypothese angenommen – es liegt keine Normalverteilung vor (Shapiro und Wilk 1965).

Der Shapiro-Wilk-Test hat, verglichen mit anderen bekannten Normalverteilungstests, eine hohe statistische Teststärke – höher auch als der oft eingesetzte Kolmogorov-Smirnov-Test (Razali und Wah 2011), weswegen er in der gegenständlichen Arbeit zum Einsatz kommt.

Die zuvor beschriebenen mathematischen Operationen werden im gegenständlichen Fall mithilfe der Python-NumPy-Bibliothek berechnet.

### 3.3.1 Pearson-Korrelationskoeffizient

Der Pearson-Korrelationskoeffizient wird zum Nachweis eines linearen Zusammenhangs zwischen zwei Zufallsvariablen eingesetzt. Im ersten Berechnungsschritt werden dazu die Kovarianzen der beiden Mittelwerte der betrachteten Zufallsvariablen gebildet.

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

#### Abbildung 6: Formel Kovarianz (Staudinger 2021)

Da die auf diesem Weg ermittelten Kovarianzen jedoch von der Dimension der beiden Größen abhängig sind (Beispiel: Angabe der Größe in m oder cm), müssen jene normiert werden um ein dimensionsloses Maß zu erhalten. Dies wird mittels des Korrelationskoeffizienten nach Pearson erreicht. Dabei werden die Kovarianzen durch die jeweilige Standardabweichung dividiert. Der so generierte Quotient stellt den Korrelationskoeffizienten dar.

Der ermittelte Korrelationskoeffizient kann einen Wert zwischen -1 und +1 annehmen und drückt dadurch die Ausprägung der Korrelation aus. Die unterschiedlichen Ausprägungsformen sind in der Tabelle/Abbildung unterhalb dargestellt: (Staudinger 2021).

Korrelationskoeffizient nach Pearson	Bedeutung
$r = -1$	vollständige lineare Abhängigkeit
$-1 < r \leq -0.8$	starker negativer linearer Zusammenhang
$-0.8 < r \leq -0.6$	mäßig starker negativer linearer Zusammenhang
$-0.6 < r \leq -0.4$	mittlerer negativer linearer Zusammenhang
$-0.4 < r \leq -0.2$	geringer negativer linearer Zusammenhang
$-0.2 < r < 0$	sehr schwache Korrelation
$r = 0$	Lineare Unabhängigkeit
$0 < r < 0.2$	sehr schwache Korrelation
$0.2 \leq r < 0.4$	geringer positiver linearer Zusammenhang
$0.4 \leq r < 0.6$	mittlerer positiver linearer Zusammenhang
$0.6 \leq r < 0.8$	mäßig starker positiver linearer Zusammenhang
$0.8 \leq r < 1$	starker positiver linearer Zusammenhang
$r = 1$	vollständige lineare Abhängigkeit

Abbildung 7: Ausprägungsformen des linearen Zusammenhangs (Staudinger 2021)

### Signifikanzprüfung

Um folglich nachzuweisen, dass es sich dabei um eine signifikante Korrelation handelt, muss eine Signifikanzprüfung des Korrelationskoeffizienten ( $r$ ) durchgeführt werden.

Dazu wird der zugehörige p-Wert (Wahrscheinlichkeit) aus dem t-Wert (Teststatistik) des Korrelationskoeffizienten berechnet. Die Teststatistik ( $t$ ) wird hierbei mit der folgenden Formel berechnet

$$t = \frac{r \times \sqrt{n - 2}}{\sqrt{1 - r^2}}$$

Abbildung 8: Formel t-Teststatistik (Walther 2020)

Im Falle eines negativen Korrelationskoeffizienten ( $r$ ) muss der Absolutbetrag des Korrelationskoeffizienten für die Berechnung der Teststatistik eingesetzt werden. Der Wert  $n$  stellt im Rahmen der Formel die Anzahl der Wertepaare bei der Ermittlung des Korrelationskoeffizienten dar (Ebermann 2010b).

Die Ermittlung des p-Wertes kann nachfolgend auf Basis der zweiseitigen T-Verteilung erfolgen. Im Rahmen der gegenständlichen Arbeit wird für die Berechnung der Werte auf die Python-NumPy-Bibliothek und die vordefinierten Formelmethode zurückgegriffen.

Ist die mittels zweiseitiger T-Verteilung ermittelte Wahrscheinlichkeit  $p$  geringer als das zuvor gewählte Signifikanzniveau  $\alpha$  (z.B. 5%), so kann von einer signifikanten Korrelation zwischen den zuvor untersuchten Zufallsvariablen ausgegangen werden. (Ebermann 2010b).

Ein Faktor, der im Rahmen der oberhalb dargestellten statistischen Betrachtungen außen vor gelassen wird, ist der zeitliche. Um zu beurteilen ob die Anzahl der Unternehmenserwähnungen auf [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets) pro Tag mit dem tatsächlichen Aktienkurs dieses Unternehmens korreliert, benötigt es die Betrachtung der beiden Zeitreihen mittels Auto- als auch Kreuzkorrelation.

### 3.3.2 Kreuzkorrelation

Die Methode der Kreuzkorrelation wird angewandt um Korrelationen zwischen zwei Signalen zu erkennen. Als Signal in der aktuellen Beschreibung wird eine Zeitreihe aufgezeichneter Daten verstanden (Bundesverband Geothermie 2020).

Gegenständlich stellen die beiden Signale einerseits die Kommentaranzahl (Signal 1) und andererseits den Aktienkurs (Signal 2) des jeweiligen Unternehmens dar.

Wie auch beim Pearson-Korrelationskoeffizienten, ist bei der Kreuzkorrelation auf die Dimensionen der beiden Signale zu achten. Sind diese unterschiedlich - zum Beispiel weist Signal A einen ähnlichen Verlauf auf wie Signal B, allerdings sind die Werte von Signal A zweimal kleiner - so wird mit einem nicht-normalisierten Kreuzkorrelationsverfahren die Korrelation nicht entdeckt. Aus diesem Grund muss auch hier eine Normalisierung der Kreuzkorrelation eingesetzt werden. Dadurch wird es möglich, die Korrelation zweier Signale mit unterschiedlichen Amplituden zu erfassen.

Da die Kommentaranzahl pro Tag deutlich höher als der tatsächliche Aktienkurs in US-Dollar sein wird, ist im gegenständlichen Fall eine Normalisierung der Kreuzkorrelation von Nöten.

Identisch zum Pearson-Korrelationskoeffizienten gilt auch hier – je höher der Wert, desto höher die Korrelation. Wenn beide Signale ident sind, beträgt die Korrelation den Maximalwert 1. Sollten die Signale vollständig entgegengesetzt sein, wird der Mindestwert -1 ausgewiesen.

### Time Shift

Oftmals ist es weniger interessant herauszufinden, ob zwei Signale generell korrelieren, sondern mit welchem Zeitversatz (Time Shift) die beiden zusammenwirken.

Das Anwenden eines Zeitversatzes auf die normalisierte Kreuzkorrelation führt somit zu einer „normalisierten Kreuzkorrelation mit einem Zeitversatz von X“. In der aktuellen Untersuchung könnte man zum Beispiel davon ausgehen, dass sich die Kommentare nicht umgehend, sondern mit 24 Stunden Zeitversatz auf den Aktienkurs auswirken.

Mithilfe einer zeitversetzten Kreuzkorrelation können „Lags“ (Verzögerungen) oder „Leads“ (Vorläufer) zwischen zwei oder mehreren Signalen identifiziert werden. Um zu ermitteln ob zwei Signale mit einem Zeitversatz korrelieren, müssen alle möglichen zeitlichen Verzögerungen geprüft werden. Die Programmiersprache Python stellt hierfür mit statsmodels eine Bibliothek zur Verfügung, die diese umfangreichen Berechnungen durchführen kann (Statsmodels 2021). Durch Annahme sinnvoller Lag bzw. Lead-Grenzen kann die Berechnung weiter beschleunigt werden (z.B. die Kommentaranzahl am 18.12.2020 wird sich unwahrscheinlich mit einem Lag von 365 Tagen auf den 17.12.2021 auswirken).

### **3.3.3 Autokorrelation**

Unter Autokorrelation versteht man den Vergleich einer Zeitreihe von Daten, zu verschiedenen Zeitpunkten, mit sich selbst. Es wird beispielsweise dazu eingesetzt um sich wiederholende Muster oder Saisonalitäten zu erkennen (Stocker 2021). Im gegenständlichen Fall wird die Methode der Autokorrelation angewandt, da angenommen werden kann, dass der gestrige Aktienschlusskurs Einfluss auf den heutigen Aktienschlusskurs haben wird, oder umgekehrt ausgedrückt, der heutige Aktienkurs bis zu einem gewissen Grad vom gestrigen abhängig ist.

Eine Kennzahl die zur generellen Detektion einer solchen Autokorrelation 1. Ordnung eingesetzt werden kann, ist der Durbin-Watson-Koeffizient. Bildet man die Differenzen zwischen den Residuen und ihrem Vorperiodenwert, und dividiert man anschließend die damit berechenbare Quadratsumme der Differenzen, durch die Residuenquadratsumme, so erhält man die dazu erforderliche Teststatistik (Stocker 2021).

Diese kann Werte zwischen 0 und 4 annehmen. Je näher der Wert dabei an 2 liegt, desto geringer ist das Ausmaß der Autokorrelation. Werte zwischen 1,5 und 2,5 werden als akzeptabel betrachtet, Werte unter 1 oder über 3 deuten auf klare Autokorrelation hin.

Wie auch bei der Kreuzkorrelation ist es jedoch oftmals interessanter zu erheben, mit welchem Zeitversatz ein Signal mit sich selbst korreliert (Time Shift). Ident zur

Kreuzkorrelation stellt Python auch für die Autokorrelation umfangreiche Möglichkeiten zur Detektion von Saisonalitäten und Mustern zur Verfügung (NumPy 2021).

#### Zusammenfassung mathematische Grundlagen

Zur Nachweisführung etwaiger Korrelationen in der gegenständlichen Untersuchung werden die erhobenen Daten initial auf das Vorhandensein einer Normalverteilung überprüft. Dadurch wird sichergestellt, dass die oberhalb beschriebenen statistischen Methoden auch anwendbar und zielführend sind. Es werden Histogramme angefertigt, sowie Schiefe und Kurtosis berechnet. Abschließend erfolgt ein Shapiro-Wilk-Test zur objektiven Nachweiserbringung des Vorliegens einer Normalverteilung.

Erst danach erfolgen die eigentlichen Prüfungen auf etwaig vorhandene Korrelationen. Im ersten Schritt wird hierbei mittels Pearson-Korrelationskoeffizient abgetastet, ob lineare Zusammenhänge zwischen den täglichen Unternehmenserwähnungen auf [www.reddit.com/r/wallstreetbets](http://www.reddit.com/r/wallstreetbets) und deren Aktienschlusskursen bestehen. Diese Berechnungen werden auch für die Tagesgewinne und -verluste des Aktienkurses gegenüber der mittels Sentimentanalyse ermittelten positiv und negativ gepolten Kommentare pro Tag, je Einzelunternehmen und aller Betrachtungsunternehmen aggregiert, analysiert. Die ermittelten Ergebnisse werden mittels zweiseitiger T-Verteilung signifikanzgeprüft.

Im zweiten Schritt erfolgt die vertiefte Betrachtung der Daten unter Berücksichtigung des zeitlichen Faktors mittels Auto- und Kreuzkorrelation. Hierbei wird untersucht, inwiefern die Aktien- bzw. Kommentarkurse mit sich selbst (Autokorrelation) und wie sehr die beiden Signale zueinander (Kreuzkorrelation) korrelieren. Dabei werden auch mögliche Saisonalitäten und Zeitversätze untersucht.

In Kapitel 3 wurden die literarischen Grundlagen zum Verständnis der weiteren Untersuchungen, der in [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets)-Kommentaren erwähnten Unternehmen in Zusammenhang mit potentiellen Short Squeezes beschrieben. Unter Punkt 3.1 wurde auf ausgewählte Kapitalmarktgrundlagen eingegangen und unter Punkt 3.2 die einzusetzenden Tools und Libraries zur Prüfung der Hypothese beschrieben. Abschließend fanden sich in Kapitel 3.3 die anzuwendenden statistischen Berechnungsmethoden zum Korrelationsnachweis allgemein dargestellt.

## **4. Wirtschaftspsychologische Einflussfaktoren**

In dem vorangegangenen Kapitel 3.1 wurde das Eintreten und die Möglichkeiten eines Short Squeezes generell erläutert. Darin wurde beschrieben, dass eine entsprechende Marktmacht vorhanden sein muss, um ein solches Szenario herbeizuführen.

Das Ziel von Kleinanlegern ist grundsätzlich die Erwirtschaftung von Profit für die eigene Tasche. Sie haben dabei im Vergleich zu institutionellen Investoren einen um ein Vielfaches kleineren Anteil an Investitionskapital, was dem individuellen Kleinanleger kein Leverage zur Beeinflussung etwaiger Basiswerte gewährleistet (Burnette 2021).

Was die Fälle von „Wallstreetbets“ jedoch so interessant und völlig neuartig macht, ist der Fakt, dass sich eben diese Kleinanleger nun zur Steigerung des gemeinsamen Leverage bündeln. Welche Faktoren diesen Zusammenschluss begünstigen, wird in den folgenden Unterkapiteln erläutert.

### **4.1 Boom der Kleinanlegerschaft**

In den letzten Jahren sind neue Entwicklungen am Kapitalmarkt bemerkbar. Die Anzahl von Neo- bzw. Internet-Brokern wie Trade-Republic oder Robin Hood ist merklich gestiegen (Riedl 2021). Mit dem Ausbruch der Corona-Virus Pandemie Anfang 2020 hat sich dieser Trend weiter verstärkt. Im ersten Quartal 2020 wurden allein in den USA über eine Million neue Accounts bei diversen Online Brokern angelegt (Burnette 2021). Zusätzlich bot der noch nie dagewesene Absturz des Aktienmarkts zum Beginn der COVID-19-Pandemie einen idealen Einstiegspunkt für neue Anleger. Kleinanleger waren in den Monaten Juli und August für 25% aller Trades in den Vereinigten Staaten verantwortlich (McCrank 2021). Einer Studie aus dem Jahr 2020 zufolge wurden damals bereits 77% der Gesamtmarktkapitalisierung von Aktien in den USA direkt oder indirekt über Investment- oder Rentenfonds von Privatanleger gehalten (Mackintosh 2020).

Insbesondere kommissionsfreie Broker-Apps wie Robin Hood stellen bei den Anlegern der Wallstreetbets den Markteinstiegspunkt der Wahl dar.

### **4.2 Trading Apps & beeinflussende Design Patterns**

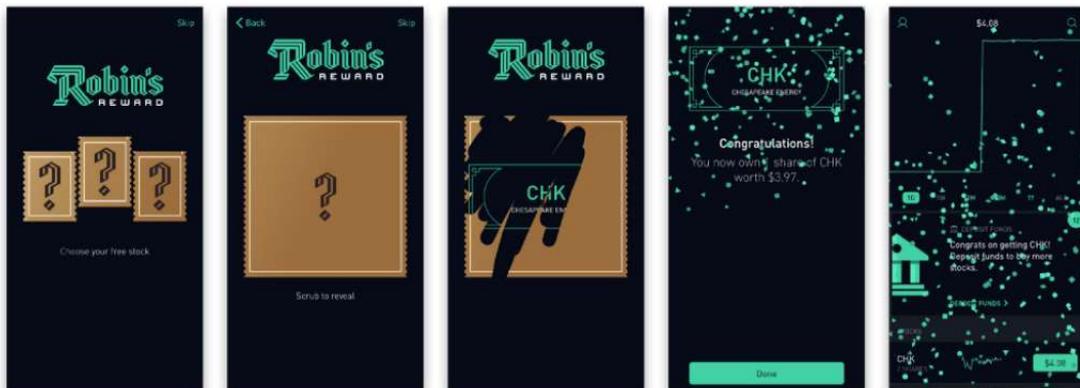
Die vorherrschende „App Culture“ wurde durch die Coronapandemie weiter verstärkt und eine der Branchen, die einen immensen Anstieg an neuen Mobilanwendungen verzeichnen konnte, war die Finanzbranche (Malhotra 2020). Die durch den Lockdown gewonnene Zeit, etwa durch Home Office Möglichkeiten, gaben Kleinanlegern mehr Zeit sich mit dem Kapitalmarkt zu befassen (Anand 2020). Im Angesicht von möglichen

Lohnkürzungen haben die simplen und teilweise kommissionsfreien Angebote zum Zuverdienst durch Trading Apps einen Silberstreif am Horizont dargestellt (Malhotra 2020). Um eine hohe Akzeptanz zu erreichen, wurden von den Trading Apps fünf, von der Literatur identifizierte Schlüsselfunktionalitäten, perfekt umgesetzt:

- Einfache Bedienbarkeit
- App Performance und Geschwindigkeit
- User Interface Design
- Investment Analyse und Informationen in einfacher Form
- Sicherheits- und Privatsphäre Mechanismen

Zusammen mit dem Faktor der meist wegfallenden Brokergebühren, der vor allem für die junge Anlegerschaft einen entscheidenden Punkt darstellt, führen diese Funktionalitäten zur hohen Akzeptanz von diversen mobilen Trading Apps. (Malhotra 2020).

Ein weiterer, mit der simplen Anwendungsgestaltung einhergehender Aspekt, ist die Gamification des Investmentprozesses. Unter Gamification wird (von engl. "game": "Spiel") die Übertragung von spieltypischen Elementen und Vorgängen in spielfremde Zusammenhänge verstanden (Bendel 2013). Spezifischer formuliert kann man behaupten, dass Trading Apps technische und soziale Interaktionsebenen darstellen, die das Investmentverhalten gezielt beeinflussen können. Bestimmte Designs, Architekturen und Features von Investment-Apps können die User dazu bewegen weniger rationale sondern eher instinktive Entscheidungen und Investments durchzuführen (Jameson et. al. 2014). Ein Beispiel hierfür stellt etwa die „Golden Ticket“ Funktion von Robin Hood, dargestellt in Abbildung 1, dar. Robin Hood bietet seinen Usern eine gratis Aktie im Gegenzug zur Anmeldung auf der Plattform oder Weiterempfehlung an einen Freund an. Dabei wird eine Rubbelbildanimation eingesetzt, die den Anwender in den Gedanken überführen könnte, beim Anlegen handle es sich um eine Form des Glücksspiels (Wursthorn und Choi 2020).



**Abbildung 9: Robin Hood Golden Ticket Funktion (Chaudhry und Kulkarni 2021)**

Festlegungen bei der Konzeption des User Interfaces, wie zum Beispiel welche Inhalte zuerst sichtbar sind, welche Farben gewählt werden oder welche Einstellung als Default-Einstellungen gesetzt sind, beeinflussen die heuristische Entscheidungsfindung von Anwendern (Lockton 2012). Zudem wurde in Studien festgestellt, wie gezielte Designfestlegungen, Nachrichten oder Texte überzeugender wirken lassen können (Chaudhry und Kulkarni 2021). Besonders ist dabei, auch in Zusammenhang mit der Designwahl von Robin Hood (siehe Abbildung 1), die Erkenntnis, dass vor allem dunkle Designpatterns die heuristische Situationsbeurteilung so beeinflussen, dass instinktivere Entscheidungen getroffen werden (Chaudhry und Kulkarni 2021). In einer von Tanaka und Kawabata durchgeführten Studie wurde weiterführend festgestellt, wie die Vorhersehbarkeit von Anwendungsschritten in Userinterfaces dazu führen kann, welche Wetten von Anwendern platziert werden (Tanaka und Kawabata 2020).

Zusätzlich wurde in einer gemeinsamen Forschungskoooperation der Universitäten St. Gallen, Paderborn und New York entdeckt, dass Neo-Broker durch gezielte E-Mail oder Pop-Up-Benachrichtigungen zum Investieren in bestimmte Basiswerte erreichen, dass der Handel mit diesen Basiswerten auch tatsächlich signifikant zunimmt (Arnold 2021).

Die dadurch hervorgerufene Änderung in der Anlegerdemografie und die drastische Verringerung der Einstiegsbarrieren zum Handel am Kapitalmarkt durch Neobroker, vor allem in App-Format, stellen einflussreiche Faktoren dar, die die Möglichkeiten von Ereignissen wie etwa den GameStop Short Squeeze begünstigen. Der Handelsanteil von Privatanlegern hat sich ausreichend vergrößert, um zu großem Teil mitverantwortlich für die steigenden Aktienkurse und die hohen Optionsaktivitäten zu sein (Burnette 2021).

### 4.3 Steigende Risikobereitschaft von Privatanlegern

Weiterführend kann ein signifikanter Wechsel der Anlagestrategie von Privatanlegern von Investitionen in wachsende Unternehmen, zu Unternehmen mit geringerer Profitabilität und höherer Volatilität festgestellt werden (Burnette 2021).

Eine mögliche Erklärung der Veränderung der Anlagestrategie ist, dass Privatanleger überwiegend Männer darstellen und jene generell risikofreudiger agieren als Frauen. Eine Studie aus dem Jahr 1998 zeigt, dass Männer im Vergleich zu Frauen eine deutlich höhere Risikobereitschaft aufweisen, um in volatile Anlageformen wie Aktien zu investieren (Jianakoplos und Bernasek, 1998). Zudem wurde in einer von 1991 bis 1997 angestellten Untersuchung von 35.000 Privatanlegern festgestellt, dass männliche Marktteilnehmer bis zu 45% öfter am Markt aktiv werden, und kaufen oder verkaufen, als Frauen (Barber und Odean 1998).

Einen weiteren demografischen Aspekt stellt die durch die Verringerung der Markteintrittsbarrieren wie Trading Apps deutlich jüngere Anlegerschaft dar. Die oberhalb erwähnte Studie hat zudem festgestellt, dass jüngere Investoren eher volatilere Portfolios besitzen als ältere.

Jüngere Anleger tendieren zudem eher zu risikoreicheren Transaktionen und diese Risikobereitschaft nimmt erst mit zunehmendem Alter ab (Jianakoplos und Bernasek, 1998).

### 4.4 Social Media & Herdenverhalten

Hinzukommt wie diese neue Anlegerschaft ihre Finanzmarktdaten einholt. Einen elementaren Teil bei der Entscheidungsbildung nehmen hier die sozialen Medien und damit einhergehende emotionale Befindlichkeiten einzelner Individuen ein. Die Literatur zeigt, dass insbesondere bei Investitionsentscheidungen die emotionale Komponente eine Rolle spielt (Dowling und Lucey 2003).

In Zeiten der „App Culture“ beeinflussen Social Media Kanäle eben diese emotionale Meinungsbildung. Mit [www.reddit.com](http://www.reddit.com) und den „Wallstreetsbets“ ist dazu das Paradebeispiel zur Meinungsbildung über Social Media gefunden. Über [www.reddit.com](http://www.reddit.com) erfolgte schnell die Informationsweiterverteilung zu etwaigen Short Squeeze-Aktivitäten auf Twitter. Die kombinierte Userschaft von Reddit und Twitter erlaubte es den Kleinanlegern, im Verbund, einen entsprechenden Leverage-Effekt zu erzielen und so den GME-Preis in die Höhe zu treiben.<sup>2</sup>

Hierbei sei weiterführend erwähnt, dass gerade Twitter im vergangenen Jahrzehnt einen beachtlichen Einfluss auf den Aktienmarkt ausgeübt hat. Eine Studie hat gezeigt, dass „Twitter-Daten den Dow Jones Industrial Average mit einer Genauigkeit von 87,6% vorhersagen“ (Burnette 2021). Diese Vorhersagegenauigkeit demonstriert die

Marktmacht, welche Privatanlegern über Social Media-Kanäle zu Teil wird. Diese Information und das darauf aufbauende Herdenverhalten der Kleinanleger zeigt, wie

<sup>2</sup>Twitter verfügt über 192 Millionen (Lin 2021), Reddit über 52 Millionen tägliche User (Affde 2021)

Ereignisse wie der GameStop-Short Squeeze möglich werden. Als Beispiel für den Twittereinfluss kann ein von Elon Musk veröffentlichter Tweet aus dem Jahr 2018 dienen. Er twitterte der Aktienpreis von Tesla sei zu hoch, was dazu führte, dass der Aktienkurs umgehend um 10% sank (Shead 2021). Die unterhalb dargestellte Abbildung 10, zeigt weitere direkte Zusammenhänge zwischen dem Tesla-Aktienkurs und den Tweets des Unternehmensinhabers aus dem Jahr 2018.



Abbildung 10: Tesla Aktienkurs in Zusammenhang mit Elon Musk (Dounis 2020)

Im Jahr 2016 wurde veröffentlicht, dass der Einfluss von Social Media auf den Aktienmarkt „signifikant größer“ als jener der herkömmlicher Medienkanäle ist (Jiao, Veiga und Walther 2016).

Betrachtet man die Hintergründe zur Entstehung des Herdenverhaltens wird deutlich, welchen Einfluss Social Media gerade im Finanzsektor ausüben kann. Insbesondere zeigt sich dies bei jungen Kleinanlegern, welchen in der Regel das erforderliche Hintergrundwissen zu rationalen Investmenthandlungen am Kapitalmarkt fehlt.

Der Begriff des Herdenverhaltens stammt aus der Tierwelt. Im Falle von Unsicherheit oder Nahrungssuche, wird durch Zusammenschluss einzelner Individuen, die

Sicherheit für die gesamte Gruppe erhöht. In der Menschenwelt kann dieses Verhalten im Rahmen von Modeerscheinungen, Demonstrationen oder eben am Kapitalmarkt erkannt werden. Menschen neigen insbesondere in unsicheren Entscheidungssituationen dazu, nicht rational auf eigene Erfahrungsschätze zurückzugreifen, sondern sich an der Masse zu orientieren und „mitzuschwimmen“ (Fabio 2015). Aufgrund des fehlenden Finanzwissens trifft nun vor allem die neue junge Anlegerschicht, wie im Falle der Wallstreetbets, Entscheidungen nicht selbstständig, sondern sie orientiert sich an ihrem Umfeld (Fabio 2015). Im digitalen Zeitalter stellen die Kanäle von Social Media, im Finanzbereich besonders Reddit und Twitter, ebendieses Umfeld dar. Just dort wurde dann auch das Short Squeezing von GME instinktiv beworben und dazu aufgerufen. Ein weiterer fördernder Faktor des Herdenverhaltens ist der ‚sharing-the-blame-effect‘, der besagt, dass bei der Annahme von Entscheidungen anderer, die Verantwortung einer Fehlentscheidung mit allen Entscheidungsträgern geteilt wird. Die Fehlentscheidung der Masse stellt dabei eine emotional akzeptablere Situation, als das alleinige Verfehlen dar (Fabio 2015). Dieses Verhaltensmuster wurde von John Maynard Keynes treffend festgehalten: *„Worldly wisdom teaches that it is better for reputation to fail conventionally than to succeed unconventionally“*.

In Zusammenhang mit der COVID-19-Pandemie und dem Wegfall des persönlichen Face-to-Face-Austauschs durch Ausgangsbeschränkungen, fand ein noch stärkerer Umschwung zur Informationsverteilung und -einholung, sowie eine Verstärkung dieses Phänomens über Social Media Kanäle statt (Yarovaya et al. 2021). Konsequenzen eines solchen Herdenverhaltens spiegeln sich folglich in stark schwankenden Kursen und volatilen Kapitalströmen wieder, wie es etwa im Fall GameStop eingetreten ist (Fabio 2015).

Der in der bisherigen Literatur eingesetzte Begriff der „Informationsansteckung“, der Contagion-Effekt, wurde durch Social Media und dies insbesondere während der COVID-19-Pandemie nicht nur verstärkt, sondern auch beschleunigt (Yarovaya et al. 2021).

Neben den sozialen Medien hat aber auch das Internet als allgemeine Informationsquelle einen nicht unerheblichen Einfluss auf den Aktienmarkt. Eine Untersuchung hat gezeigt, dass ein statistischer Zusammenhang zwischen den Google-Suchanfragen nach bestimmten Aktien und deren Kurs nachweisbar ist (Evangelos, Eleftheria und Polydoros 2021).

## 5. Berechnungsmodell zur statistischen Zusammenhangsermittlung

Aufgrund der zuvor beschriebenen wirtschaftspsychologischen Faktoren scheint es naheliegend, dass es einen statistischen Zusammenhang im Sinne der angenommenen Hypothese geben könnte.

Im folgenden Kapitel wird daher der gewählte Lösungsansatz zur Nachweisführung eines statischen Zusammenhangs zwischen der Anzahl der Erwähnungen ausgewählter auf [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets) zum Short-Squeeze aufgerufener Unternehmen (siehe Kap.2) und dem Aktienkurs ebendieser Unternehmen dargelegt.

Die Hypothese lautet dabei, dass zwischen der Anzahl der Erwähnungen einzelner, ausgewählter Unternehmen und dem Aktienkurs dieser Firmen eine statistische Korrelation besteht.

Dazu wird beginnend geschildert wie die entsprechenden Kommentare von der Plattform [www.reddit.com](http://www.reddit.com) exportiert, die Finanzmarktdaten eingeholt und ebendiese aufbereitet werden. Anschließend wird im Detail auf die gewählte Berechnungsmethode eingegangen. Darauf folgt eine Zusammenfassung des Kapitels, sowie ein Ausblick auf weitere mögliche Untersuchungen.

### 5.1 Dateneinholung

#### 5.1.1 Kommentarexport [www.reddit.com](http://www.reddit.com)

Nachfolgend werden die einzelnen Arbeitsschritte zum automatisierten Download ausgewählter Inhalte von der Webseite [www.reddit.com](http://www.reddit.com) beschrieben.

##### Automatisierter Zugriff auf [www.reddit.com](http://www.reddit.com)

Um Informationen von der Internetplattform [www.reddit.com](http://www.reddit.com) zu exportieren, ist es erforderlich ein Benutzerkonto anzulegen, welches es einem erlaubt, auf die Inhalte der Webseite zuzugreifen. Jene können zwar auch ohne Account abgerufen werden, um erweiterte Funktionen wie Schnittstellen zu eigenen IT-Anwendungen freizuschalten, ist ein Benutzerkonto jedoch unumgänglich.

Nach erfolgreicher Benutzeranlage kann unter dem Teilbereich */Preferences/Apps* der Plattform eine eigene App erzeugt werden. Diese App muss angelegt werden um später automatisiert, über das zu erstellende Python-Skript, auf die Inhalte der Plattform zugreifen zu können. Im Hintergrund wird auf der Plattform eine neue Client-Instanz mit eigenem Zugangsschlüssel (Secret) generiert, über die dann später die Master-Datenbank (die gesammelten Webseiteninhalte) eingesehen werden kann.

## Installation Python Reddit API Wrapper

Die Python Bibliothek Python Reddit API Wrapper (PRAW) kann über mehrere Wege installiert werden. Im gegenständlichen Fall wird dies über die Windows PowerShell über das „pip“-Kommando ausgeführt. Eingesetzt wird PRAW für die Python Version 3.10. Auf die generellen Schritte zum Setup von Python auf Windows-Betriebssystemen wird hier nicht weiter eingegangen. Hier sei auf weiterführende Literatur verwiesen (Reitz, Kenneth und Tanya Schuster. 2016a).

## Export der Kommentare

Damit automatisiert auf die Kommentare von [www.reddit.com/r/wallstreetbets](http://www.reddit.com/r/wallstreetbets) zugegriffen werden kann, muss im ersten Schritt im Python-Skript eine neue „Reddit-Instanz“ angelegt werden. Über diese Instanz wird die Verbindung zur Plattform Reddit hergestellt. Um diese Verbindung aufzubauen, müssen der Reddit-Instanz die zuvor generierten Anmeldeinformationen - die Client-Instanz und der Zugangsschlüssel - sowie Benutzername und Passwort des Reddit-Accounts mitgegeben werden.

Nach erfolgreicher Verbindung mit der Plattform Reddit, können im weiteren Schritt nun die eigentlichen Aktionen zum Kommentarexport gesetzt werden (PRAW 2021).

Dazu wird nachfolgend kurz der generelle Aufbau der Plattform erläutert. [www.reddit.com](http://www.reddit.com) besteht aus mehreren Subreddits, die sich jeweils einem thematisch abgegrenzten Gebiet widmen. Im aktuellen Fall widmet sich etwa der Subreddit „Wallstreetbets“ allen erdenklichen Themen zur Finanzwirtschaft. Diese Subreddits sind weiter in einzelne Threads (auch Submissions genannt) eingeteilt. Jene Threads können wiederum in Subthreads unterteilt sein.

Um nun auch tatsächlich alle Kommentare abgreifen zu können muss anhand iterativer Vorgehensweisen, die gesamte Kommentarbaumstruktur durchlaufen und analysiert werden.

Dazu wird initial der Subreddit gewählt aus welchem die Kommentare exportiert werden sollen (siehe Code 1).

```
subreddit = reddit.subreddit(`wallstreetbets`)
```

### **Code 1: Export subreddit**

*reddit* stellt dabei die Reddit-Instanz dar. Im weiteren Schritt muss ausgewählt werden, welchen Bereich des Subreddit man untersuchen möchte.

Hier gibt es die Unterscheidung zwischen „hot“ – aktuell „heiße“ Nachrichten und Informationen, „new“ – neueste, „rising“- stark und schnell an Interaktion zunehmende Posts (Threads), „top“ – die meistgelikten Threads in einer freigewählten Zeitspanne, oder „controversial“- Posts mit vielen Up- aber auch vielen Downvotes zugleich.

Eine von Long, Lucey und Yaroyava durchgeführte Studie hat ergeben, dass im gegenständlichen Subreddit, lange Threads mit einer Vielzahl von Kommentaren und

hoher Reichweite das Userverhalten beeinflussen können, während bei kurzen Threads keine Auswirkungen auf die breite Masse feststellbar waren (Long, Lucey und Yarovaya. 2021).

Um eine vollumfängliche Untersuchung zu gewährleisten werden im aktuellen Fall neben dem Bereich „hot“ auch die weiteren Bereiche des Subreddits ausgewertet. Da es sich bei diesem Bereich jedoch um die „Frontpage“, die erste und am meisten interagierte Seite, im Subreddit „Wallstreetbets“ handelt, wird davon ausgegangen, dass die meisten relevanten Kommentare aus diesem Bereich exportiert werden. Hier sind die Thread-Highlights mit hohen Kommentaranzahlen und großer Reichweite gebündelt abrufbar, und somit jene Informationen verfügbar, die die breite Masse erreichen

Über die Übergabe eines Exportlimits oder Zeitgrenzen kann die Anzahl der exportierten Threads eingeschränkt werden. Im aktuellen Fall sind nur die Threads von 01.06.2020 bis 01.06.2021 interessant und der Export wird daher auf diesen Zeitraum beschränkt. Aufgrund einer API-Änderung bei der PRAW-Library ist es seit November 2021 nicht mehr möglich, mehr als 1000 der letzten Submissions von der Plattform zu exportieren (PRAW 2021). Im gegenständlichen Fall ist es über das PRAW-API aufgrund der Vielzahl an Submissions im Subreddit „Wallstreetbets“ daher nur mehr möglich circa die letzten drei Kalendertage zu exportieren.

Um diese Limitierung zu umgehen wird im Rahmen der Untersuchung auf die Pushshift-Library und das psaw-API zurückgegriffen. Pushshift stellt hierbei eine unabhängige Datenbank aller Informationen von [www.reddit.com](http://www.reddit.com) dar, welche laufend mit der Plattform synchronisiert werden (Pushshift 2021).

Neben den Submissions zu allen Subreddits können dort alle zugehörigen Kommentare abgerufen werden. Ein weiterer Vorteil der Pushshift-Library besteht darin, dass die exportierten Daten anhand der Submission-ID ebenfalls mit der PRAW-Library weiterverarbeitet werden können und so die volle Funktionspalette des PRAW-API zur Analyse bereitsteht.

Kommentare auf reddit.com sind wie oberhalb erwähnt in einer Baumstruktur untergliedert. Über die Funktion `.comments` können nun einzelne Kommentare zu den Threads in `subreddit` abgefragt werden. Dabei werden jedoch nur die Hauptkommentare (First-Level-Kommentare) zu einem Thread exportiert. Etwaige Antworten auf dieses Hauptkommentar werden nicht berücksichtigt. In einem nächsten Schritt müssen deshalb auch die Antworten zu den Hauptkommentaren (Second-Level-Kommentare) und nachfolgend die Antworten zu den Antworten (Third-Level-Kommentare) usw. exportiert werden.

Dazu müssen iterativ zuerst der Thread und danach First-Level und jeweils die zugehörigen Second- und Third-Level-Kommentare abgefragt werden um alle relevanten Kommentarinformationen zu exportieren. Da aufgrund dieser Struktur die

Kommentarebenen theoretisch beliebig tief werden können z.B.: Fifth-Level oder gar Tenth-Level-Kommentar, stellt PRAW mit der Funktion `.list()` ein Tool bereit, dass je Hauptkommentar alle Kommentarebenen durchläuft und so sicherstellt, dass jedes Kommentar zu einer Submission exportiert werden kann (PRAW 2021).

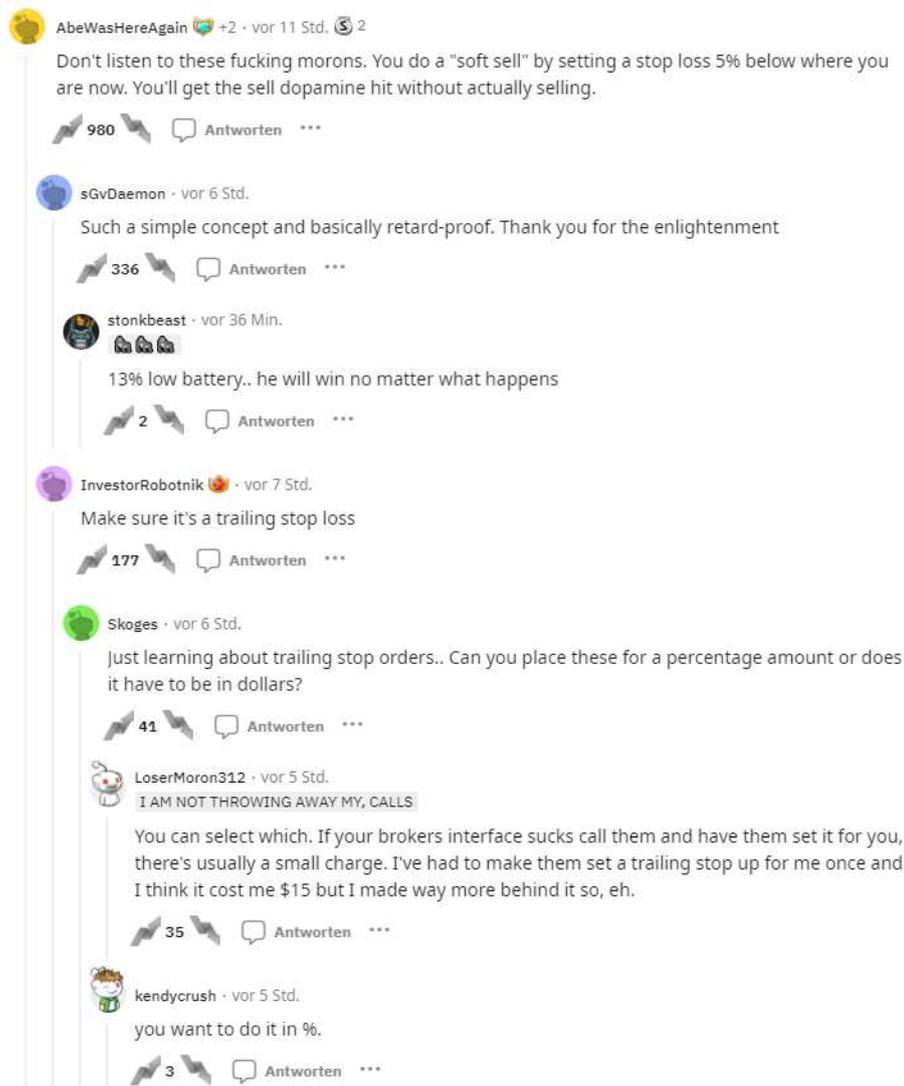


Abbildung 11: reddit Kommentarstruktur (Reddit 2021)

Um nicht alle Kommentare zu jedem diskutierten Thema im Subreddit „Wallstreetbets“ downzuloaden, werden bestimmte Suchbegriffe definiert und nur Threads und Kommentare mit einem Bezug zu diesen Begriffen exportiert. Bezug wird hierbei als Erwähnung des Suchbegriffes in einem Top-Level-Kommentar oder bei Lower-Level-Kommentaren deren Zugehörigkeit zu einem Top-Level-Kommentar mit einem enthaltenen Suchbegriff definiert.

Die folgenden Suchbegriffe werden hierbei festgelegt um die exportierten Informationen zu begrenzen:

<b>Unternehmen</b>	<b>Suchbegriff 1 (Aktienname)</b>	<b>Suchbegriff 2 (Aktienkurzzeichen)</b>	<b>Suchbegriff 3 (Reddit-Spezifika)</b>
AMC Entertainment Holdings, Inc.	AMC Ent	AMC	
AMC Networks Inc.	AMC Net	AMCX	
Virgin Galactic Holdings, Inc.	Virgin	SPCE	
Tootsie Roll Industries, Inc.	Tootsie	TR	
Siebert Financial	Siebert	SIEB	
Palantir Technologies Inc.	Palantir	PLTR	
Nokia Oyj	Nokia	NOK	
National Beverage Corp.	National Beverage	FIZZ	
Naked Brand Group	Naked Brand	NAKD	
The Macerich Company	Macerich	MAC	
Ligand Pharmaceuticals Incorporated	Ligand	LGND	
Koss Corporation	Koss	KOSS	
American Airlines Group Inc.	American Air	AAL	
iRobot Corporation	iRobot	IRBT	
Genius Brands International	Genius Brands	GNUS	
Fossil Group, Inc.	Fossil	FOSL	
Express, Inc.	Express	EXPR	

Eastman Kodak Company	Kodak	KODK	
Build-A-Bear Workshop, Inc.	Build-A-Bear	BBW	Build a Bear
BlackBerry Limited	BlackBerry	BB	Black Berry
Bed Bath & Beyond Inc.	Bed Bath & Beyond	BBBY	Bed Bath and Beyond
BB Liquidating Inc.	BB Liquid	BLIAQ	Blockbuster
GameStop Inc.	GameStop	GME	

**Abbildung 12: Reddit Suchbegriffe**

Dabei wurde besonders darauf geachtet, dass bei der Definition der Suchbegriffe aus den Unternehmensnamen eindeutige Bezüge zur gegenständlichen Analyse hergestellt werden können und der fälschliche Export von nicht geeigneten Kommentaren vermieden wird. So wird etwa beim Unternehmen Genius Brands International nicht der Suchbegriff „Genius“ alleinstehend benutzt, um hier nicht Kommentare zu exportieren in denen das Wort „genius“ womöglich anerkennend zwischen zwei Usern eingesetzt wird (Beispiel: „congratulations you genius“).

Im Rahmen der Suche nach den passenden Kommentaren ist zudem darauf zu achten, dass die Groß-/Kleinschreibung einzelner Kommentare für die Suche unerheblich ist.

Die so exportierten Kommentare werden je Suchbegriff in zwei separate Textdateien geschrieben, welche zu einem späteren Zeitpunkt für die Nutzung im Berechnungsmodell relevant werden. Der Aufbau der Textdateien und die Analyse ebendieser werden in Kapitel 5.2 näher beschrieben.

### 5.1.2 Einholung Finanzmarktdaten

Die in der gegenständlichen Untersuchung verwendeten Finanzmarktdaten stammen von [www.ariva.de](http://www.ariva.de). Je Unternehmen werden hierbei für den Betrachtungszeitraum von 01.06.2020 bis 01.06.2021 folgende Informationen tagesfein in tabellarischer (als .csv) und grafischer Form extrahiert (Ariva 2021).

Datum	Startkurs	Höchstkurs	Tiefstkurs	Schlusskurs	Stücke	Volumen
-------	-----------	------------	------------	-------------	--------	---------

**Abbildung 13: [www.ariva.de](http://www.ariva.de) - Export Finanzmarktdaten Inhalte**

An Tagen an jenen die Börsen geschlossen waren und kein Aktienhandel stattfand (Wochenenden und Feiertage) werden für die Gegenüberstellung des Kommentar- und

Aktienkurses im Rahmen des nachfolgenden Berechnungsmodells folgende Parameter festgelegt:

Aktienchlusskurs: An handelsfreien Tagen wird als Schlusskurs der letztgültige Kurs des Vortags herangezogen. Sollte am Vortag die Börse ebenfalls geschlossen gewesen sein, so wird der letzte Tag mit gültigem Schlusskurs angenommen.

Gewinn/Verlust: An handelsfreien Tagen wird der eventuelle Gewinn respektive der Verlust einer Unternehmensaktie mit 0 festgelegt.

Diese getroffenen Modellparameter sind für die spätere Berechnung der Korrelationsexistenzen in linearer als auch zeitreihenbasierter Form erforderlich.

## 5.2 Datenaufbereitung

Um die gewünschten Nachweisführungen herzuleiten, müssen die Finanzmarktdaten sowie die Daten der Plattform [www.reddit.com](http://www.reddit.com) in geeignete Form gebracht und analysiert werden.

### Finanzmarktdaten

Aus den gewonnenen Finanzmarktdaten werden im weiteren Verlauf der Analyse, das Datum, der Schlusskurs und die Differenz aus Start- und Schlusskurs benötigt. Alle weiteren Parameter werden nicht weiter berücksichtigt.

### Reddit-Daten

In Kapitel 5.1.1 wurde ausgeführt, dass aus den gewonnenen Kommentaren des Subreddits „Wallstreetbets“ je zwei Dateien pro Unternehmen generiert werden sollen. Ziel ist es, im Rahmen des Kommentarexports bereits alle relevanten Informationen aus den Kommentaren zu extrahieren, damit diese Daten kein zweites Mal transformiert werden müssen (siehe dazu Saarbrückener Pipelinemodell in Kapitel 3.2.1).

Es wird daher bereits vor dem Schreiben der Kommentarinformationen in die beiden Dateien eine Sentiment Analyse des jeweiligen Kommentars durchgeführt.

Dazu wird die unter Kapitel 3.2 beschriebene Textblob-Library eingesetzt. Der Kommentartext wird nach dem Saarbrückener Pipelinemodell durch diese Library aufbereitet und etwaige Störfaktoren wie zum Beispiel Sonderzeichen oder Emoticons aus den Kommentaren entfernt. Die anschließend durchgeführte Sentimentanalyse gibt für jeden Kommentar einen Wert der Polarität der Nachricht aus (positiv, negativ, neutral). Dazu wird die bereits in der Library Textblob vorhandene Funktion `blob.sentiment.polarity` eingesetzt wobei `blob.sentiment` hier den Text und `.polarity` die entsprechende Funktion darstellt. Diese Funktion gibt für den Kommentartext den Polaritätswert in einem Bereich von -1 bis +1 zurück. -1 stellt hierbei ein maximal negatives Kommentar und +1 ein maximal positives Kommentar

dar (Textblob 2021). Für die weitere Bewertung der Polarität (py) einzelner Kommentare werden die folgenden Wertgrenzen für positiv, negativ und neutral gewählt:

Positiv:  $py > 0.5$

Neutral:  $-0.5 < py < 0.5$

Negativ:  $py < -0.5$

In der ersten Datei – weiterführend als Kommentardatei bezeichnet – werden daher je Kommentar die folgenden Informationen gespeichert.

Thread	Datum des Verfassens	Kommentartext	Sentimentanalysewert (py)
--------	----------------------	---------------	---------------------------

**Abbildung 14: Kommentardatei Aufbau**

In der zweiten Datei, der Kursdatei, werden pro Tag und Unternehmen die untenstehenden Informationen gespeichert:

Datum	Anzahl Erwähnungen Total	Anzahl positiver Erwähnungen	Anzahl neutraler Erwähnungen	Anzahl negativer Erwähnungen	Differenz Erwähnungen zum Vortag
-------	--------------------------	------------------------------	------------------------------	------------------------------	----------------------------------

**Abbildung 15: Kursdatei Aufbau**

Aus der Information „Anzahl Erwähnungen Total“ pro Tag in der aggregierten Kursdatei ergibt sich der Kommentarkurs des jeweiligen Unternehmens.

Mithilfe der Matplotlib-Library von Python können die so generierten Kommentarkurse visualisiert und den Aktienkursen grafisch gegenübergestellt werden.

Der auf diesem Weg kreierte Unternehmenskommentarkurs kann weiterführend im Berechnungsmodell als Gegenstück zum Aktienkurs eingesetzt werden. Das Ziel des nachfolgenden Berechnungsmodells ist es sodann, etwaige statistische Korrelationen zwischen den auf diesem Wege erstellten Unternehmenskommentarkursen und deren Aktienkursen zu ermitteln. Anhand der Information „Differenz Erwähnungen zum Vortag“ wird die Möglichkeit geschaffen, eine Gegenüberstellung der Kommentaranzahl und etwaigen Gewinnen oder Verlusten der Aktie darzustellen. Die gewonnenen Sentimentanalysedaten sollen dabei zur Bestätigung oder Widerlegung von Scheinkorrelationen dienen.

### 5.3 Berechnungsmodell

Das gewählte Modell zum Nachweis statistischer Zusammenhänge ist stufenweise aufgebaut und gliedert sich in mehrere Schritte. Begonnen wird mit der Ermittlung von linearen Zusammenhängen zwischen der täglichen Kommentaranzahl mit Unternehmenserwähnungen (folglich als Kommentarkurs bezeichnet) und dem Aktienkurs der Unternehmen. Nachfolgend werden die Gewinne bzw. Verluste der

Aktien, der Zunahme bzw. Abnahme der Unternehmenserwähnungen in Kommentaren auf [www.reddit.com/r/wallstreetbets](http://www.reddit.com/r/wallstreetbets) zum Vortag, gegenüberstellt. Die gewonnenen Erkenntnisse werden mittels Signifikanzprüfungen weiter untersucht.

Anschließend erfolgt der Vergleich der jeweiligen Zeitreihen des Kommentar- und des Aktienkurses mittels Kreuzkorrelationsverfahren, um festzustellen ob zwischen diesen Zeitreihen eine Korrelation existiert. In einer letzten Stufe wird untersucht, ob eine Wirkungsverzögerung zwischen den Zeitreihen (Zeitversatz) besteht.

Zur vertieften Validierung finden in jeder Berechnungsstufe Sentimentanalyseverfahren der Kommentare statt, um etwaige Scheinkorrelationen bestmöglich ausschließen zu können.

### Normalverteilungsprüfung

Im ersten Schritt zur Bestimmung etwaiger statistischer Zusammenhänge zwischen den Kommentaren der Plattform [www.reddit.com](http://www.reddit.com) und den jeweiligen Unternehmensaktienkursen wird die zuvor unter Kapitel 3.3 aufgestellte Hypothese, dass es sich bei den beiden Zeitreihen von Daten um normalverteilte Informationen handelt, geprüft. Dies wird initial durch eine grafische Prüfung umgesetzt. Dazu werden die exportierten Finanzmarktdaten, sowie die Kommentarkurse mithilfe der matplotlib-Library in Graphen visualisiert. Zusätzlich werden die generierten Daten einer Schiefe-, sowie Kurtosisbeurteilung unterzogen.

Abschließend erfolgt der Shapiro-Wilk-Test zur statistischen Prüfung des Vorliegens einer Normalverteilung je Unternehmen und Zeitreihe. Bestätigt sich die Hypothese der normalverteilten Informationen wird wie ab Kapitel 3.3.1 beschrieben, mit den Korrelationsprüfungen begonnen. Im Falle einer Widerlegung der Hypothese kommt anstelle des Pearson-Korrelationskoeffizienten weiterführend ein Rangkorrelationskoeffizient zur Zusammenhangsermittlung zum Einsatz.

### Korrelationskoeffizient – Nachweis linearer Zusammenhang

Nachfolgend wird zur Bestimmung möglicher statischer Zusammenhänge zwischen Aktienkurs der ausgewählten Unternehmen und der täglichen Erwähnungsanzahl ebendieser Unternehmen auf der Onlineplattform [www.reddit.com/r/wallstreetbets](http://www.reddit.com/r/wallstreetbets) initial auf lineare Merkmalszusammenhänge zwischen den jeweiligen bivariaten Tagesdaten geprüft.

Dabei soll untersucht werden, ob die tägliche Kommentaranzahl im Betrachtungszeitraum in einem linearen Wirkungszusammenhang zum zugehörigen Unternehmensaktienkurs steht. Wie aus der Hypothese ersichtlich ist, wird hierbei angenommen, dass eine positive Korrelation zwischen diesen Größen besteht.

Der Nachweis eines linearen Zusammenhangs wird anhand der Korrelationsrechnung geführt. Im ersten Schritt wird dazu die Kovarianz aus den Mittelwerten der Unternehmenserwähnungen auf Reddit pro Tag, über den Betrachtungszeitraum, und

dem jeweiligen Tagesschlusskursen der betroffenen Unternehmensaktie im Betrachtungszeitraum ermittelt.

Da die Kovarianzwerte von der Dimension der beiden Größen abhängig sind, erfolgt in einem zweiten Schritt die Ermittlung des Pearson-Korrelationskoeffizienten, um die Kovarianzen zu normieren und ein dimensionsloses Maß zu ermitteln.

Um das Missinterpretieren von möglichen Scheinkorrelationen auszuschließen wird im nächsten Schritt auch der Pearson Korrelationskoeffizient zwischen der Differenz der Kommentarerwähnungen der einzelnen Unternehmen zum Vortag und dem täglichen Gewinn bzw. Verlust des jeweiligen Aktienkurses ermittelt.

Hier werden als Wertepaare pro Tag jeweils die Kommentardifferenz und der Tagesgewinn bzw. -verlust eines Unternehmens gegenübergestellt. Der Vergleich von Gewinnen bzw. Verlusten stellt ein in der Praxis gängiges Instrument zur Äquivalenzprüfung von Aktienkursen und Indizes untereinander dar (Reider 2020).

Besteht auch hier eine Korrelation zwischen den beiden Größen, ist eine Scheinkorrelation zu einem gewissen Grad ausschließbar (Ebermann 2010b).

#### T-Test –Signifikanzprüfung

Um folglich nachzuweisen, dass es sich hierbei um signifikante Korrelationen handelt, wird eine Hypothesenprüfung des Korrelationskoeffizienten durchgeführt. Dazu wird die in Kapitel 3.3 erwähnte Methodik angewandt.

Die auf diesem Weg ermittelten Ergebnisse werden einer ersten Interpretation unterzogen und Schlussfolgerungen zu den Berechnungsergebnissen aufgestellt.

Damit die Sicherheit der getroffenen Aussagen weiter erhöht wird, kommen folglich Sentimentanalysedaten zum Einsatz.

Dabei werden Tagen mit Kursgewinnen, die Anzahl von Unternehmenserwähnungen in einem positiven Kommentarkontext auf [www.reddit.com/r/wallstreetbets](http://www.reddit.com/r/wallstreetbets) gegenübergestellt. Umgekehrt erfolgt dieses Vorgehen auch für den Aktienkurs/ Tage mit Kursverlusten, welchen die Anzahl der negativen Erwähnungen gegenübergestellt wird. Weiters wird die Anzahl der Unternehmenserwähnungen mit positiver sowie negativer Polarität den Tagesgewinnen bzw. den Tagesverlusten der Unternehmensaktien gegenübergestellt. Diese Vergleichsgrößen wurden gewählt um etwaige Zusammenhänge zwischen positiven und negativen Kommentarentwicklungen gegenüber den Tagesgewinnen oder -verlusten zu untersuchen.

Betrachtet werden hierbei bei den Aktienkursen nur die Schlusskurse und nicht Tageshöchst- oder -tiefstwerte.

Die Gegenüberstellung und Korrelationsprüfung wird anhand des Pearson- bzw. bei Nichtvorliegen normalverteilter Daten mittels Spearman-Korrelationskoeffizienten durchgeführt.

Ein Vergleich dieser Größen ist aufgrund von ermittelten Ergebnissen aus einer Untersuchung von 58 Million Social Media Nachrichten in China naheliegend. Diese Studie deutet darauf hin, dass Aktiengewinne, insbesondere untertägig sehr stark von der positiven Polarität von Chatnachrichten bei Amateurinvestoren ausgehen (Dong und Gil-Bazo 2020).

Zusätzlich werden diese Sentimentdaten auch empirisch anhand grafischer Visualisierungen überprüft, um zeitliche Auswirkungen und Zusammenhänge zu erkennen, da dies ebenfalls Aufschluss auf Marktgegebenheiten liefern könnte (Long, Lucey und Yarovaya. 2021).

Alle oberhalb beschriebenen Berechnungsschritte werden in Python über die NumPy- und Statsmodels-Bibliothek umgesetzt. Die anschließende Visualisierung der Ergebnisse erfolgt über die Library Matplotlib.

### Autokorrelation

Weiterführend wird das Vorliegen etwaiger Autokorrelationen untersucht - ob und mit welchem Zeitversatz (Timeshift) die täglichen Aktien- und Kommentarkurse mit sich selbst korrelieren.

Initial wird im Rahmen dieser Untersuchung der unter Kapitel 3.3.3 beschriebene Durbin Watson Test je Kommentar- und Aktienkurs der Unternehmen durchgeführt und die ermittelten Ergebnisse interpretiert, um festzustellen ob überhaupt ein autokorrelierendes Verhalten der Daten vorliegt. Tritt dieser Fall ein, wird in einem weiteren Schritt die Autokorrelationsfunktion berechnet, um etwaige Saisonalitäten der Aktienkurse, sowie der Kommentarkurse festzustellen.

Die mathematischen Rechenoperationen werden mithilfe der Python API „statsmodels“ durchgeführt, interpretiert und bei Bedarf über die Library Matplotlib visualisiert.

### Normalisierte Kreuzkorrelation – TimeShift

Ein Faktor der im Rahmen der oberhalb beschriebenen Berechnungen, abgesehen vom generellen Untersuchungszeitraum und der grafischen Betrachtung, außer Acht gelassen wird, ist die zeitliche Anordnung der Daten innerhalb der Zeitreihen zueinander. Mithilfe der Zusammenhangsprüfung nach Pearson wird eine eventuelle, lineare Korrelation im Betrachtungszeitraum festgestellt, jedoch nicht, wie sich jene der beiden Zeitreihen zueinander verhält und mit welchem etwaigen Lag (Zeitversatz) eine Größe A mit der Größe B korreliert. Dies gilt sowohl für die Gegenüberstellung von

Kommentarkurs und Aktienkurs der ausgewählten Unternehmen, sowie deren Gewinn- und Verlustgegenüberstellung. Auch die weiterführenden Autokorrelationstests geben nur Auskunft über die Korrelation der Zeitreihe mit sich selbst, sowie etwaigen zugehörigen Saisonalitäten.

Um den zuvor geschilderten Faktor in die Beweisführung mitaufzunehmen, wird folglich eine normalisierte Kreuzkorrelationsberechnung mit Zeitversatz (=Time-Shift) durchgeführt. Gemäß der Beschreibung unter Punkt 3.3 stellen der Kommentarkurs und der Aktienkurs, sowie die Kommentarentwicklung/Tag und die Kursentwicklung/Tag unsere Signale A und B dar. Diese Signale (Zeitreihen) werden dazu nach der unter Punkt 3.3 beschriebene Methodik untersucht.

Auch hier wird methodisch in den selben Ausprägungen wie bei der Ermittlung linearer Zusammenhänge vorgegangen. Zuerst werden der Kommentar- und der Aktienkurs des jeweiligen Unternehmens gegenübergestellt. Anschließend erfolgt eine Gegenüberstellung der Kursentwicklung (Gewinn/Verlust) und der Kommentarentwicklung (Anstieg/Abfall). Diese Zeitreihen werden mittels normalisierter Kreuzkorrelation untersucht.

Die Untersuchung der Kurs- und Kommentarentwicklung soll auch hier einer Missinterpretation von Scheinkorrelationen entgegenwirken.

Im nächsten Schritt wird die normalisierte Kreuzkorrelation noch um die Zeitversatzermittlung erweitert, da es naheliegend scheint, dass sich hohe Kommentaranzahlen mit Unternehmenserwähnungen und stark steigende Kurse wohl erst mit einem gewissen Zeitversatz gegenseitig beeinflussen. Aufgrund der vorliegenden Datenbasis wird dieser Zeitversatz (Time-Shift) tagesfein untersucht.

Abschließend werden zur Stützung der Erkenntnisse und Widerlegung von Scheinkorrelationen erneut die Polaritäten der Einzelkommentare ausgewertet und mittels Kreuzkorrelation inklusive Zeitversatzprüfung untersucht. Diese Analyse folgt dabei jener Systematik, die bereits unter dem Unterkapitel „Korrelationskoeffizient – Nachweis linearer Zusammenhang“ für die lineare Korrelationsprüfung beschrieben wurde.

Die so ermittelten Erkenntnisse werden anschließend grafisch mittels der Matplotlib-Library in Python aufbereitet und final interpretiert. Es erfolgt eine Gegenüberstellung der Auswertungsergebnisse aus Pearson-Korrelationskoeffizient und normalisierter Kreuzkorrelation mit Zeitversatz.

Abschließend wird ein Ausblick über mögliche Folgeaktivitäten gegeben.

### Mögliche Folgeaktivitäten

Weiterführend, aber nicht mehr im Rahmen dieser Arbeit behandelt, könnte aufbauend auf etwaig ermittelten Korrelationen ein Granger-Kausalitätstest durchgeführt und so die Basis für ein Reddit-Aktienkurs-Prognosemodell geschaffen werden. Theoretisch wäre denkbar, dass anhand der Unternehmenserwähnungen an einem Kalendertag die Aktienkursentwicklung dieses Unternehmens in der Zukunft prognostiziert werden könnte. Zusätzlich könnte verstärkt auf die Sentimentanalysedaten eingegangen werden, um etwa Prognosen aus den Polaritäten zu ermitteln

## 6. Erstellung Kurspaare & Visualisierung

Im nachfolgenden Kapitel wird ein Überblick über die exportierten Unternehmensdaten gegeben. Dabei werden zuerst die allgemeinen Fakten zum Kommentarexport von der Plattform [www.reddit.com/r/wallstreetbets](http://www.reddit.com/r/wallstreetbets) tabellarisch dargestellt und nachfolgend die von [www.ariva.de](http://www.ariva.de) exportierten Aktienkurse den ermittelten Kommentarkursen graphisch gegenübergestellt.

### Kommentarexportzahlen

Im Rahmen des Kommentarexports wurden sämtliche Kommentare aller Submissions im Subreddit „Wallstreetbets“, im Untersuchungszeitraum von 01.06.2020 00:00:01 Uhr bis 01.06.2021 00:00:00 Uhr, auf die unter Kapitel 5.1.1 erwähnten Suchbegriffe geprüft und im Falle, dass ein gesuchter Begriff enthalten war, einer Sentimentanalyse unterzogen. Auf diesem Wege wurden mehr als 5.76 Millionen Kommentare untersucht und davon über 690.000 Kommentare mit einem Bezug zu den definierten Suchbegriffen mittels Sentimentanalyseverfahren betrachtet. Die so überprüften Kommentarsummen werden in der nachfolgenden Tabelle dargestellt.

Als problematisch stellte sich hierbei die volatile Verfügbarkeit des pushshift-Servers heraus, weshalb die Exportroutinen mehrmals über längere Zeiträume nicht durchführbar waren (Serverausfälle) oder neu gestartet werden mussten. Durch Parallelisierung der Export- und Analyseaktivitäten auf zwei Auswerterechner konnte dieser Umstand zumindest geringfügig abgefedert werden. Weiters wurde aufgrund der sporadisch auftretenden Serverausfälle die Sample-Size des Exports von einem Jahresexport auf Monatsexporte der Kommentare reduziert. Die beiden kommentarreichsten Monate rund um das Gamestop Short Squeeze Ereignis Januar und Februar 2021 (2.6 Mio. Kommentare) wurden wegen der deutlich höheren Kommentaranzahl in diesem Zeitraum jeweils auf zwei Teile (Monatsanfang – Monatsmitte und Monatsmitte bis Monatsende) aufgeteilt.

Tabelle 1: Exportinformationen [www.reddit.com/r/wallstreetbets](http://www.reddit.com/r/wallstreetbets)

Unternehmen (Aktienkürzel)	Kommentar betrachtungszeitraum 01.06.2020 – 01.06.2021			
	Anzahl Kommentarerwähnungen	Davon positiv	Davon neutral	Davon negativ
AAL	3.360	203	3.068	89
AMC	91.946	7.018	81.806	3.143
AMCX	326	9	299	21
BB	64.000	4.575	52.638	2.339
BBBY	2.569	152	2.316	91
BBW	341	38	295	8
BLIAQ	274	19	249	6
EXPR	4.433	239	4.055	137
FIZZ	509	17	484	8
FOSL	502	17	476	9
GME	322.288	22.660	289.128	10.499
GNUS	907	69	814	24
IRBT	104	11	92	1
KODK	2.642	149	2.391	102
KOSS	838	38	792	8
LGND	122	5	112	5
MAC	16.647	784	15.139	390
NAKD	503	22	476	5
NOK	39.933	2.577	36.318	1.308
PLTR	64.218	4.841	56.935	2.442
SIEB	44	1	41	2
SPCE	8.212	561	7.383	258
TR	66.628	4.854	59.247	2.527
Summe	691.346			

Die oberhalb dargestellte Tabelle zeigt, dass etwa 12% aller Kommentare im Betrachtungszeitraum, im Subreddit „Wallstreetbets“, Erwähnungen der Untersuchungsunternehmen enthalten.

### Bipolare Kommentare

Ein Phänomen welches im Rahmen des Kommentarexports festgestellt wurde, und anhand der oberhalb dargestellten Tabelle ebenfalls erkennbar ist, ist der hohe Anteil an neutralen Kommentaren, gemessen an den Gesamtkomentaren total und je Einzelunternehmen. Die Polaritätseingliederung der Kommentare wurde anhand der Anwendungsempfehlung der Textblob-Library (Textblob 2021), wie unter Kapitel 3.2.1 beschrieben, umgesetzt. Durch die dort festgelegte Definition wird eine hohe Anzahl von Kommentaren als neutral eingestuft. Über alle Unternehmen hinweg beträgt jener in der gegenständlichen Untersuchung 88,9%.

Dieses Auftreten ist auf zwei Hauptfaktoren zurückzuführen.

1. Den Auswertungsalgorithmus der Textblob-Library und die dort empfohlenen Polaritätsgrenzen sowie
2. die Art der Formulierung von Kommentaren auf der Plattform [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets)

Ad 1) Der Algorithmus der Textblob Library ist so konzipiert, dass der Text anhand des Saarbrückener Pipelinemodells analysiert wird, positive sowie negative Textelemente gewichtet und jene für eine finale Gesamtbeurteilung der Polarität gegenüber gestellt werden.

Ad 2) Wird nun ein Kommentar so formuliert, dass es einerseits einen negativen Textteil wie etwa eine Beschimpfung oder Beschwerde enthält, diesen aber im weiteren textuellen Verlauf in eine positive Botschaft umwandelt, kann dies dazu führen, dass der Kommentar als gesamtes von Textblob als neutral eingestuft wird (Textblob 2021). Als repräsentatives Beispiel sei folgendes vulgäres Kommentar dargestellt:

„**F\*\*\* wallstreet and their fonds!!** lets go GME!!, ... making money big time, just bought again, how're you guys doin?“. (Reddit 2021)

Die rote (negative) und grüne (positive) Hinterlegung stellen diesen Umstand exemplarisch dar. Menschlich interpretiert würde das Kommentar aufgrund der positiven Wendung und des Geldgewinns als überwiegend positiv eingestuft werden. Im Rahmen der faktenbasierten maschinellen Auswertung erfolgt die Einstufung in solchen Fällen in den Bereich „neutral“.

### Kommentarkurse

Die oberhalb dargestellten Summeninformationen basieren auf tagesfeinen Erwähnungsdaten je Unternehmen, anhand welcher aufbauend die entsprechenden Kommentarkurse erstellt und weiter untersucht wurden. Beispielhaft ist ein Auszug einer solchen Datei unterhalb dargestellt.

Datum	Gesamt <u>KommAnz</u>	positiv	neutral	negativ	Differenz Vortag
20210531	354	31	315	8	-231
20210530	585	56	520	9	-244
20210529	829	87	707	35	-728
20210528	1557	162	1350	45	32
20210527	1525	142	1328	55	798
20210526	727	49	648	30	493
20210525	234	19	204	11	56
20210524	178	17	153	8	108
20210523	70	4	64	2	-26
20210522	96	5	88	3	-76
20210521	172	13	156	3	39
20210520	133	17	113	3	-81
20210519	214	25	184	5	-240
20210518	454	37	404	13	192
20210517	262	30	226	6	185
20210516	77	10	64	3	-78
20210515	155	15	131	9	-291
20210514	446	36	395	15	-85

**Abbildung 16: AMC Kursdatei Auszug Auswertungsmonat Mai 2021**

Im Rahmen der Datenaufbereitung hat sich gezeigt, dass die zur Auswertung geplanten untenstehenden Unternehmen im Verhältnis zu anderen Unternehmen seltener im Untersuchungszeitraum in Kommentaren erwähnt werden, weshalb sie für die weitere Berechnung als nicht relevant eingestuft und daher nicht weiter berücksichtigt wurden. Als Ausschlussgrenze wurde eine Mindestanzahl von 1000 Kommentarerwähnungen im Untersuchungszeitraum definiert.

**Tabelle 2: Nicht weiter berücksichtigte Unternehmen**

<b>Nicht weiter berücksichtigte Unternehmen</b>	<b>Aktienkürzel</b>	<b>Anzahl Kommentarerwähnungen total</b>
AMC Networks Inc.	AMCX	326
Build-A-Bear Workshop, Inc.	BBW	341
BB Liquidating Inc.	BLIAQ	274
National Beverage Corp.	FIZZ	509
Fossil Group, Inc.	FOSL	502
Genius Brands International	GNUS	907
iRobot Corporation	IRBT	104
Koss Corporation	KOSS	838
Ligand Pharmaceuticals Incorporated	LGND	122
Naked Brand Group	NAKD	503
Siebert Financial	SIEB	44

Neben den Kursdateien wurden wie unter Kapitel 5.2 beschrieben, jeweils je Exportzeitraum (Monat bzw. halbes Monat) eine Exportdateiversion, sowie eine daraus aggregierte Fileversion für den Gesamtauswertungszeitraum der Kommentardatei erstellt. Die Anfertigung dieser Dateien erfolgte, um weitere, später stattfindende Untersuchungen zu ermöglichen, sowie zu etwaigen Beweis- und Widerlegungszwecken. Weiterführend wurden darin auch die Polaritäten der einzelnen Kommentare, welche im Rahmen des Kommentarexports, nach dem Saarbrückener Pipelinemodell erhoben wurden, hinterlegt. Das Speichern der o.a. Informationen soll dazu dienen, nachgelagerte Sentimentanalysen (siehe Teilabschnitt „Mögliche Folgeaktivitäten“ in Kapitel 5.3) durchführen zu können. Unterhalb ist exemplarisch eine solche Kommentartextdatei dargestellt.

Option %,1622406964.0,20210530,why exercise? amc probably wont exist when your great great grand son is ready for his first lambo.,0.51'  
 Option %,1622478242.0,20210531,i can't say if you should or shouldn't, but i've got \$50,000 in amc and i'm holding, if that helps.,0.0"  
 Option %,1622477783.0,20210531,thats whats up. good play!!! this ape still learning. im 3 for 5 on my option trades (amc+, mro, ccl). keep  
 Option %,1622419675.0,20210531,i maxed out a heloc on amc . this is absolutely a yolo for me. i also hold shares.,0.2'  
 Option %,1622496364.0,20210531,you pretty much answered your own question by saying you have a rh account. gme has exponentially more ups  
 Option %,1622408153.0,20210530,i bought more calls after reading possible bankruptcy. i know america will never abandon the movies as a pa  
 Option %,1622496418.0,20210531,thats literally how nothing works. how has amc changed its business model so it wont end Polartät it was p  
 Thread %,1622448337.0,20210531, you assume only amc can keep the tradition going? \n\ndo you think they are the only ones that know  
 Option %,162246499.0,20210531,amc with blockbuster.,0.0'  
 Option %,162243477.0,20210531, you assume only amc can keep the tradition going? \n\ndo you think they are the only ones that know  
 Option %,1622479679.0,20210531,i hate the movies, and so does a lot of other americans, and the quality of cinemas are going down the shit  
 Option %,1622443390.0,20210531,he bought 3 lottery tickets. \nwhen the amc numbers are greater than his ticket numbers, he makes money.  
 Option %,1622435542.0,20210531,amc is in the best position to, most brick and mortar- they have \$\$\$\nyou want to start a theater and go up  
 Option %,1622413709.0,20210530,id buy cinemark too if i could. \n\nbut amc has an army of ,0.0'  
 Option %,1622499912.0,20210531,the idea that there has been some cultural or societal awakening that is going to have people flocking fore  
 Option %,1622470417.0,20210531,youre spot on dude. what the heck is even loews cineplex? 10/10 im watching a movie at amc. comfy seats, ab  
 Option %,1622464487.0,20210531,literally watched amc buy out every movie theater within 50-100 miles around me. the biggest one was a driv  
 Option %,1622488841.0,20210531,galaxy, cinemark, and a plethora of smaller local chains. \nimo amc has some of the worst maintained theate

Abbildung 17: AMC Kommentardatei Auszug Auswertungsmonat Mai 2021<sup>1</sup>

Nachfolgend werden für einzelne ausgewählte Unternehmen die beiden Zeitreihen Kommentarkurs und Aktienkurs, sowie deren Verlust- und Gewinnkurse, der Differenz der täglichen Kommentarerwähnungen graphisch gegenübergestellt. Eine Visualisierung aller Unternehmensgrafiken in größerer Darstellung kann dem Anhang entnommen werden.

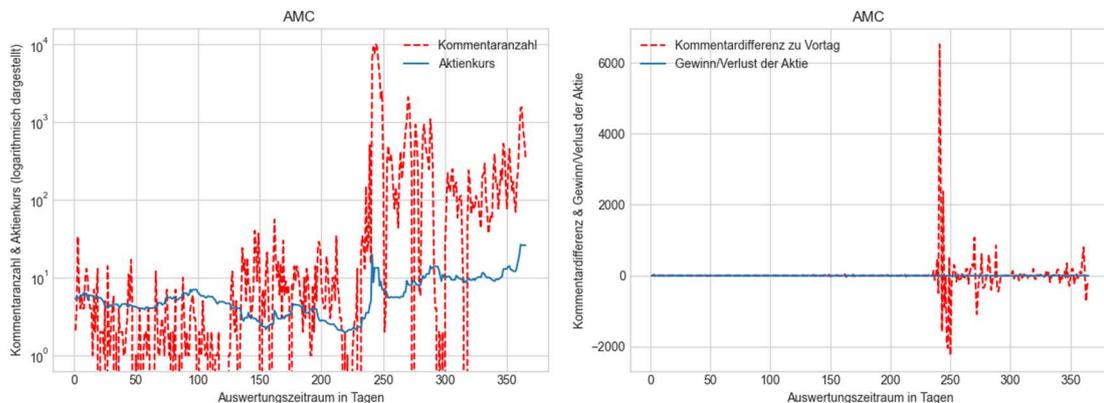


Abbildung 18: Kursverläufe AMC

<sup>1</sup>Abbildung wurde aus Gründen der besseren Lesbarkeit am rechten Ausschnitttrand zugeschnitten

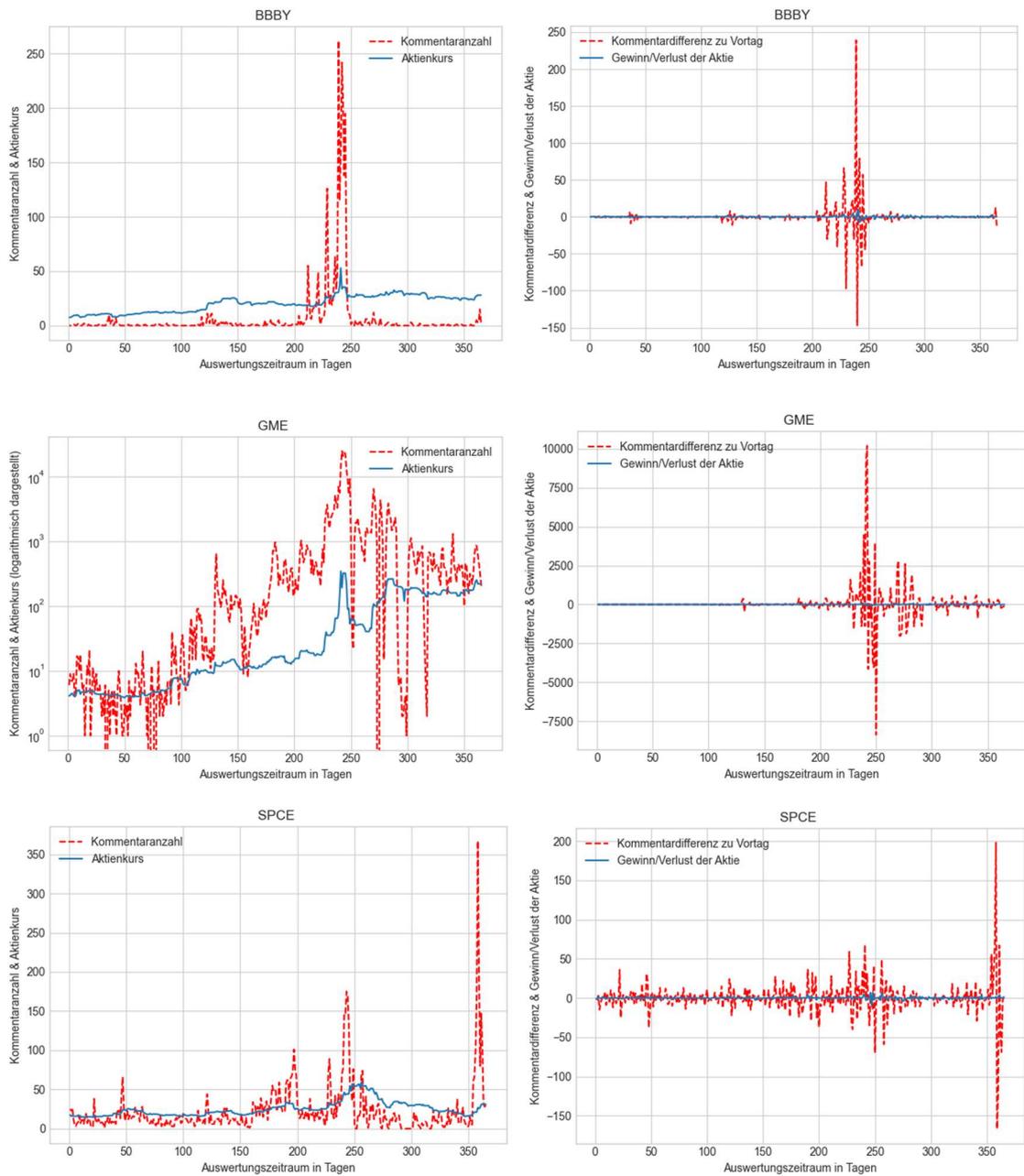


Abbildung 8: Kursverläufe BBBY, GME, SPCE

An dieser Stelle sei weiterführend auf die unter Kapitel 5.1.2 getroffenen Modellparameter hinsichtlich der Behandlung von marktfreien Tagen (Wochenenden und Feiertagen) hingewiesen.

#### Graphische Interpretation

Im gegenständlichen Kapitel wurden die erhobenen Daten kurz präsentiert und auszugsweise graphisch visualisiert.

Die oberhalb dargestellten Abbildungen, sowie jene des Anhangs lassen für ausgewählte Unternehmen eine subjektive Vermutung, über die Möglichkeit eines vorliegenden statistischen Zusammenhangs zwischen dem jeweiligen Kommentar- und Aktienkurs zu (Long, Lucey und Yarovaya. 2021). Es ist erkennbar, dass im Falle einer eklatanten Aktienschlusskursänderung auch der Austausch zu dem Unternehmen auf der Internetplattform [www.reddit.com/r/wallstreetbets](http://www.reddit.com/r/wallstreetbets) zunimmt oder umgekehrt betrachtet, die erhöhte Forenaktivität etwaige Einflüsse auf den Aktienschlusskurs des entsprechenden Unternehmens haben könnte. Um dieses subjektive Empfinden objektiv zu untersuchen wird folglich das Berechnungsmodell umgesetzt. Im Folgekapitel 7 wird die Anwendung und die Ergebnisse des unter Abschnitt 5 beschriebenen Berechnungsmodells dargelegt und nachfolgend eine objektive Beurteilung der Daten präsentiert.

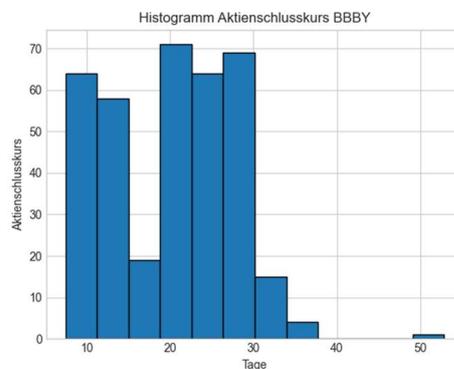
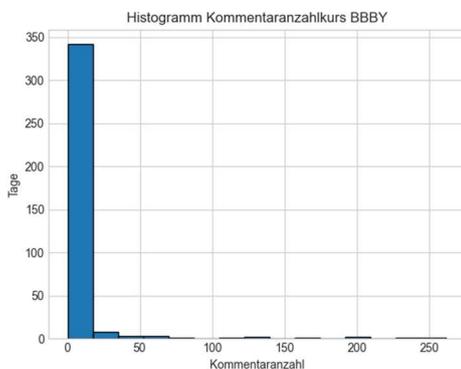
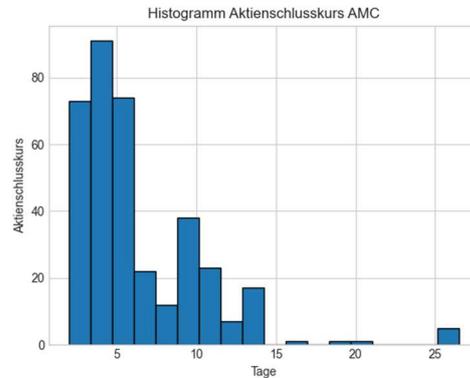
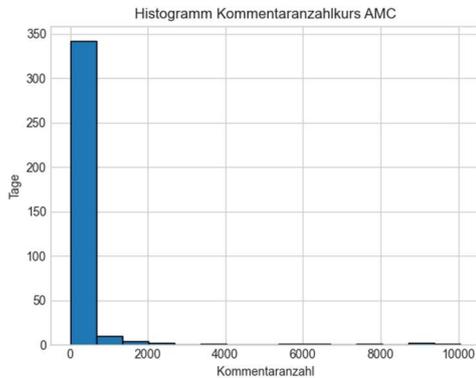
## 7. Anwendung des Berechnungsmodells

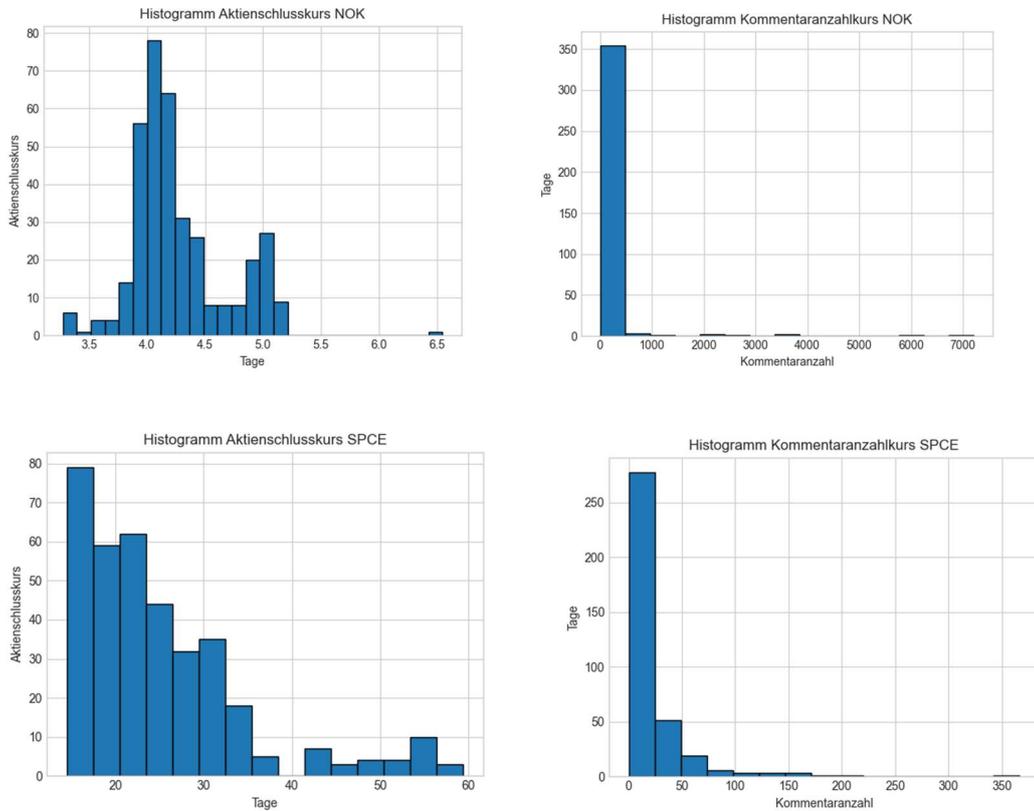
Im gegenständlichen Abschnitt der Untersuchung werden die Ergebnisse der unter Kapitel 5 beschriebenen Berechnungsmethoden präsentiert und etwaige Hindernisse und Probleme in der Anwendung des Berechnungsmodells, sowie Gegenmaßnahmen zur Herbeiführung der vorliegenden statistischen Erkenntnisse beschrieben.

Es werden dabei nicht die Daten zu allen 23 Untersuchungsunternehmen präsentiert, sondern nur Auszugsunternehmen dargestellt bzw. graphisch visualisiert. Dies soll zur besseren Nachvollziehbarkeit der Vorgehensweise dienen und eine Überladung der Textpassagen mit Grafiken und Datentabellen verhindern. Die gesammelten Auswertungen je Einzelunternehmen können dem Anhang entnommen werden.

### 7.1 Normalverteilung

Nachfolgend werden für einzelne ausgewählte Unternehmen die Daten der beiden Zeitreihen Kommentarkurs und Aktienkurs, sowie deren Verlust- und Gewinnkurse in Histogrammen graphisch dargestellt. Eine Visualisierung aller Unternehmensgrafiken findet sich im Anhang.





**Abbildung 20: Histogramme AMC, BBY, NOK, SPCE**

Die Histogramme zeigen, dass sich bei graphischer Beurteilung die Annahme des Vorliegens normalverteilter Daten nicht bestätigt. Neben den vorliegenden auszugsweise dargestellten Unternehmen oberhalb, ist dies auch für alle weiteren Firmen der Untersuchung der Fall.

Um die subjektive Einschätzung einer objektiven Prüfung zu unterziehen wurden Schiefe- und Kurtosisbeurteilungen durchgeführt. Die Ergebnisse für die Auszugsunternehmen finden sich unterhalb.

**Tabelle 3: Auszug Schiefe & Kurtosisberechnung**

<b>Auszug Schiefe und Kurtosisberechnung</b>				
Unternehmen (Aktienkürzel)	Aktienchlusskurs		Kommentaranzahlkurs	
	Schiefe	Kurtosis	Schiefe	Kurtosis
AMC	2,1	6,6	6,9	50,9
BBBY	0,1	-0,4	6,4	44,0
NOK	0,9	1,8	8,7	84,4
SPCE	1,6	2,4	5,0	37,4

Alle weiteren Ergebnisse für die Betrachtungsunternehmen können dem Anhang entnommen werden.

Abschließend wurde zur Normalverteilungsprüfung je Unternehmen und Kurs der Shapiro-Wilk-Test durchgeführt. Für die Auszugsunternehmen finden sich die Erkenntnisse in der nachstehenden Tabelle. Alle weiteren sind dem Anhang zu entnehmen.

**Tabelle 4: Auszug Shapiro-Wilk-Test**

<b>Auszug Shapiro-Wilk-Test (<math>\alpha=5\%</math>)</b>				
Unternehmen (Aktienkürzel)	Aktienchlusskurs		Kommentaranzahlkurs	
	Teststatistik	p-Wert	Teststatistik	p-Wert
AMC	0.8061434031	1.3617218496e-20	0.224773109	3.1577621005e-36
BBBY	0.937645435333252	3.038139184674549e-11	0.2500004768371582	8.447272157183888e-36
NOK	0.9108852744102478	7.458085798245362e-14	0.1637866497039795	3.244994701108379e-37
SPCE	0.827286422252655	1.687847693500704e-19	0.5387645363807678	7.958761802750808e-30

Die oberhalb dargestellten Daten, als auch die Berechnungsergebnisse des Anhangs zeigen, dass es sich bei den Kommentar- als auch Aktienkursen im gegenständlichen Betrachtungszeitraum, nicht um normalverteilte Daten handelt. Die Umsetzung der weiteren Schritte des Berechnungsmodells können somit wie nicht wie vorgesehen

durchgeführt werden. Anstelle des beschriebenen Pearson-Korrelationskoeffizienten, für den Nachweis etwaiger linearer Zusammenhänge, muss nun ein Rangkorrelationskoeffizient eingesetzt werden. Hierbei wird auf den Spearman-Korrelationskoeffizienten zurückgegriffen (Staudinger 2021).

## 7.2 Spearman-Korrelationskoeffizient

Da sich die Hypothese der normalverteilten Daten in der gegenständlichen Untersuchung, wie im vorangegangenen Abschnitt geschildert, nicht bestätigt hat, wurden die weiteren linearen Korrelationsprüfungen, nicht wie unter Kapitel 3.3 beschrieben, mit dem Pearson-Korrelationskoeffizienten sondern mit dem Spearman-Rangkorrelationskoeffizienten durchgeführt.

Zur Ermittlung des Spearman-Korrelationskoeffizienten wurde auf die Python-Library Numpy und das Funktionskonvolut *spearmanr()* zurückgegriffen.

Dabei wird die folgende Formel angewandt:

$$r_s = \frac{\sum_{i=1}^n (r_{si} - r_x^-)(r_{yi} - r_y^-)}{\sqrt{\sum_{i=1}^n (r_{si} - r_x^-)^2} * \sqrt{\sum_{i=1}^n (r_{yi} - r_y^-)^2}}$$

Abbildung 21: Formel Spearman-Korrelationskoeffizient (NumPy 2021)

Die Ausprägungsformen des Spearman-Rangkorrelationskoeffizienten folgen jenem des Pearson-Korrelationskoeffizienten. Die Ergebnisse können daher ebenfalls nach der unter Kapitel 3.3 angeführten Systematik interpretiert werden.

Tabelle 5: Auszug Spearman-Rangkorrelationskoeffizient

Spearman-Rangkorrelationskoeffizient ( $\alpha=5\%$ )				
Unternehmen (Aktienkürzel)	Aktienschlusskurs & Kommentaranzahlkurs		Gewinn/Verluste Aktienkurs & Differenz Unternehmenserwähnungen zum Vortag	
	Korrelation	p-Wert	Korrelation	p-Wert
AMC	0.457646489	2.7034493444e-20	0.04492494994	0.392120520909
BBBY	0.2229085405	1.7214815191e-05	0.06312835612	0.228926400897
GME	0.7595852711	8.4459816857e-70	0.15457912461	0.00306724814
PLTR	0.7995045902	2.2930550247e-82	-0.07037875730	0.17971392368

Die gegenständliche Untersuchung zeigt, dass zwischen dem täglichen Aktienschlusskurs und dem Kommentaranzahlkurs der ausgewählten Unternehmen lineare statistische Zusammenhänge bestehen. Interessant ist hierbei, dass jene Zusammenhänge bei intensiver diskutierten Unternehmen (höhe Erwähnungsanzahlen) stärker ausgeprägt sind als bei Firmen, die seltener erwähnt wurden. Beispielhaft können hier GameStop, Blackberry oder Palantir angeführt werden. Der Korrelationskoeffizient beträgt hierbei 0,75 (GME), 0,67 (BB) und 0,79 (PLTR).

Die Ergebnisse zeigen nicht nur das Bestehen einer linearen Korrelation, sondern auch, dass jene für den Kommentaranzahlkurs und den zugehörigen Aktienschlusskurs signifikant sind.

Weiterführend liefert die Berechnung jedoch auch die Erkenntnis, dass zwischen der Differenz der Unternehmenserwähnungen auf [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets) zum Vortag (Kommentaranstieg oder -abfall) und den Tagesgewinnen bzw. Tagesverlusten des Tagesaktienschlusskurses (Kursanstieg oder -abfall) kein signifikanter linearer Zusammenhang besteht. Am Unternehmen GameStop wurde zwar als einzige Ausnahme eine signifikante lineare Korrelation festgestellt, welche mit einem Wert von 0,15 jedoch sehr schwach ausgeprägt ist. Es kann daher nicht angenommen werden, dass im Falle eines exorbitanten Kommentaranzahlanstiegs eine entsprechende Reaktion am Aktienmarkt folgen wird oder umgekehrt.

#### Zweiseitiger T-Test – Signifikanzprüfung

Zur weiterführenden Signifikanzprüfung wurde anhand der Python-Library NumPy ein zweiseitiger T-Test durchgeführt. Es wurde dazu das Funktionskonvolut `ttest_ind()` genutzt. Als Signifikanzniveau wurde ebenfalls  $\alpha=5\%$  gewählt. Die Ergebnisse der T-Tests stützen die Erkenntnisse des Spearman-Rangkorrelationskoeffizienten.

Anhand des Berechnungsmodells konnte sodann nicht nur nachgewiesen werden, dass zumindest im Untersuchungszeitraum ein linearer statistischer Zusammenhang zwischen Kommentarkurs und Aktienkurs besteht, sondern auch dass die ermittelten Ergebnisse signifikant sind.

### Sentimentanalyse positive und negative Kommentare

Aufbauend auf die oberhalb dargestellten Erkenntnisse, kamen in dieser Phase des Berechnungsmodells erstmals auch die aus den Sentimentanalysen gewonnenen Daten zum Einsatz. Zur weiteren Stützung der Hypothese wurden die folgenden Wertepaare auf das Vorliegen linearer Zusammenhänge untersucht.

- Szenario 1: Die Anzahl der Kommentare positiver Polarität wurden den Aktienschlusskursgewinnen und -verlusten gegenübergestellt  
Dabei wurde geprüft, ob zwischen der Anzahl positiv gepolter Kommentare und den Aktienschlusskursgewinnen bzw. den -verlusten ein linearer statistischer Zusammenhang erkennbar ist.
- Szenario 2: Die Anzahl der Kommentare negativer Polarität wurden den Aktienschlusskursgewinnen und -verlusten gegenübergestellt  
Der Systematik des vorangegangenen Tests folgend wurde hierbei geprüft, ob ein Zusammenhang zwischen der Anzahl negativer Kommentare und den Aktienschlusskursgewinnen bzw. den -verlusten existiert.
- Szenario 3: Im Rahmen dieses Szenarios wurde untersucht, ob an Tagen mit positiv gepolten Kommentaren und Tagen mit Kursgewinnen eine lineare Korrelation zwischen der Anzahl der positiven Kommentare und dem tatsächlichen Kursgewinn nachweisbar ist. Dabei wurden für Tage ohne positive Kommentare oder an Tagen mit Kursverlusten die Kommentaranzahl mit 0 festgelegt. Dies war erforderlich um eine lineare Zusammenhangsermittlung zu ermöglichen.
- Szenario 4: Identisch der Systematik aus Szenario 3, wurde im Zuge des Szenarios 4 dies auch für Tage mit negativ gepolten Kommentaren und Kursverlusten durchgeführt. Dabei wurde für Tage ohne negative Kommentare oder an Tagen mit Kursgewinnen die Kommentaranzahl mit 0 festgelegt.

Die Ergebnisse der vier Untersuchungsszenarien können auszugsweise der nachfolgenden Tabelle entnommen werden.

Tabelle 6: Auszug Spearman Rangkorrelationskoeffizient Sentimentdaten

<b>Auszug Spearman-Rangkorrelationskoeffizient für Sentimentinformationen (<math>\alpha=5\%</math>)</b>								
Aktie	Szenario 1		Szenario 2		Szenario 3		Szenario 4	
	Korr.	p-Wert	Korr.	p-Wert	Korr.	p-Wert	Korr.	p-Wert
<b>GME*</b>	0.0094	0.8582	-0.0038	0.9420	<b>0.1336*</b>	<b>0.0106*</b>	<b>-0.1236*</b>	<b>0.0182*</b>
<b>KODK*</b>	- <b>0.1345*</b>	<b>0.0101*</b>	- <b>0.1077*</b>	<b>0.0398*</b>	0.0286	0.5860	<b>-0.17441*</b>	<b>0.0008*</b>
<b>NOK*</b>	0.0285	0.5871	- <b>0.1058*</b>	<b>0.0433*</b>	0.0981	0.0613	<b>-0.1377*</b>	<b>0.0084*</b>
<b>PLTR*</b>	-0.0997	0.0568	- <b>0.1398*</b>	<b>0.0075*</b>	<b>0.3041*</b>	<b>2.996e-09*</b>	<b>-0.4080*</b>	<b>4.533e-16*</b>
<b>SPCE*</b>	<b>0.1096*</b>	<b>0.0365*</b>	0.0415	0.4298	<b>0.1929*</b>	<b>0.0002*</b>	-0.0445	0.3962

\*... *signifikante Ergebnisse anhand des gewählten Signifikanzniveaus*

Die errechneten Werte aus den oberhalb geschilderten vier Szenarien liefern divergierende Ergebnisse, welche nicht pauschal beurteilt werden können.

Auffällig ist jedoch, dass die mittels Sentimentanalyse gewonnenen Daten mit Ausnahme der Unternehmen Kodak (KODK), Nokia (NOK) Palantir (PLTR) und Virgin Galactic (SPCE), anhand fehlender signifikanter Zusammenhänge die Erkenntnis aus den vorangegangenen Spearmankorrelationsberechnungen bestätigen, dass im Untersuchungszeitraum zwischen den Kommentaren und Aktienkursgewinnen bzw. -verlusten nur in Ausnahmefällen ein signifikanter, linearer, statistischer Zusammenhang besteht. Dabei ist die Polarität der Kommentare belanglos. Selbst bei den zuvor angeführten Unternehmen mit vorhandener signifikanter linearer Korrelation ist jene mit einem durchschnittlichen Wert von 0.1 aber sehr gering ausgeprägt.

Für das Unternehmen Palantir, welches mit 30.09.2020 innerhalb des Untersuchungszeitraums den Börsengang vornahm, konnte zudem festgestellt werden, dass zwischen Tagen mit positiven Kommentaren und Tagen mit Aktienschlusskursgewinnen ein geringer positiver linearer Zusammenhang besteht. Gegenübergestellt konnte weiterführend ein mittlerer negativer linearer Zusammenhang zwischen Tagen mit steigenden negativen Kommentarzahlen und Tagen mit Aktienschlusskursverlusten verzeichnet werden.

Auffallend ist weiters, dass das Unternehmen Kodak im Untersuchungszeitraum signifikante lineare Zusammenhänge zwischen den positiven und negativen Kommentaren und Aktienkursgewinnen und -verlusten verzeichnet.

Die Ergebnisse aller Einzelberechnungen je Unternehmen, sowie graphische Visualisierungen davon können dem Anhang entnommen werden.

## 7.3 Autokorrelation

### Durbin-Watson-Test

In der nächsten Phase des Berechnungsmodells wurden der Aktienkurs und der Kommentarkurs jedes Unternehmens auf das Vorliegen einer Autokorrelation untersucht. Insbesondere das Konstrukt des Aktienkurses legt die Annahme nahe, dass jener mit einer bestimmten Saisonalität mit sich selbst korreliert. Zur Bestimmung der Existenz von Autokorrelationen wurde aufgrund der Teststärke (Razali und Wah 2011) gemäß Kapitel 5.3 der Durbin-Watson-Test je Kurs durchgeführt. Die Berechnung wurde mithilfe der Python-Library „statsmodels“ und deren Funktionskonvolut `stats.durbin_watson()` umgesetzt.

Nach Systematik des Durbin-Watson-Tests konnten bei den Aktienkursen klare Autokorrelationen nachgewiesen werden. Die Testergebnisse lagen im Bereich von 0.0004 bis 0.04. Die detaillierten Erkenntnisse können dem Anhang entnommen werden. Neben den Aktienkursen konnte auf diese Weise auch festgestellt werden, dass die Kommentaranzahlkurse mit Werten zwischen 0.2 und 0.9 sehr stark autokorreliert sind (Statsmodels 2021).

### Saisonalitätsuntersuchung

Da die Durbin-Watson-Werte eindeutig das Vorhandensein von Autokorrelationen der einzelnen Aktien -und Kommentaranzahlkurse dokumentieren, wurde im weiteren Verlauf, die Autokorrelationsfunktion mit Time-Shift berechnet, um Saisonalitäten der Autokorrelation zu erkennen (Stocker 2021). Die Berechnung wurde mithilfe der Python-Library „statsmodels“ und deren Funktionskonvolut `tsa.stattools.acf()` durchgeführt. Dabei wurde sowohl bei den Aktien- als auch den Kommentaranzahlkursen deutlich, dass im zeitlichen Naheverhältnis eine hohe Autokorrelation vorliegt – der tägliche Aktienschlusskurs und. Kommentaranzahlkurs somit maßgeblich von den Vortageskursen abhängig ist. Diese vorhandene Autokorrelation nimmt mit jedem weiteren Vortag ab. Zur Verdeutlichung dieser Aussage finden sich unterhalb die Durbin-Watson-Werte, sowie Auszüge der Autokorrelationsuntersuchung für die Unternehmen AMC, BB und MAC im Betrachtungszeitraum. Alle weiteren Ergebnisse können dem Anhang entnommen werden.

Tabelle 7: Auszug Autokorrelationsberechnung

Unternehmen (Aktienkürzel)	Durbin Watson Test		Autokorrelationswert Vortag	
	Aktienkurs	Kommentaranzahlkurs	Aktienkurs	Kommentaranzahlkurs
AMC	0,0263	0,1986	0,9214	0,8953
BB	0,0092	0,1875	0,9650	0,9000
MAC	0,0023	0,2067	0,9748	0,8227

## 7.4 Kreuzkorrelation

Im abschließenden Schritt des Berechnungsmodells erfolgte eine normalisierte Kreuzkorrelationsbeurteilung mit Zeitversatz. Die Berechnung wurde mithilfe der Python-Library „statsmodels“ und deren Funktionsgruppe `tsa.stattools.ccf()` realisiert. Das Ziel hierbei war es, herauszufinden mit welchem Zeitversatz der Kommentarkurs mit dem Aktienkurs womöglich korreliert und wie sich dessen Ausprägung gestaltet. Im Falle einer nachweisbaren Ausprägung der Korrelation wäre die Hypothese bestätigt und die Nachweisführung für das Vorhandenseins eines statistischen Zusammenhangs zwischen den Aktivitäten auf [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets) und dem Aktienkurs der jeweiligen Unternehmen belegt und abgeschlossen.

### Sentimentanalyse Polaritätsinformationen

Neben der Untersuchung der allgemeinen Kommentar- und Aktienschlusskurse wurde, im Sinne der bereits bei der linearen Zusammenhangsprüfung angewandten Systematik, auch in der Kreuzkorrelationsphase auf Sentimentanalysedaten zurückgegriffen. Dabei wurden ebenfalls die bereits unter Kapitel 7.2 beschriebenen vier Szenarien betrachtet.

Im Rahmen der Untersuchung stellte sich heraus, dass über alle verbliebenen Unternehmen hinweg behauptet werden kann, es liegt ein mittels normalisierter Kreuzkorrelationsverfahren ermittelter statistischer Zusammenhang im Untersuchungszeitraum vor. Der Zeitversatz (Time-Shift), an welchem die Korrelation am höchsten ist, variiert jedoch zwischen den einzelnen Unternehmen. Die stärkste Ausprägung einer Korrelation wurde für AMC in der Höhe von 0.8 mit einem Zeitversatz von 119 Tagen zwischen Kommentarkurs und Aktienschlusskurs festgestellt. Hierbei ist jedoch klar festzuhalten, dass etwaige Korrelationen mit einem Time-Shift von über 10 Kalendertagen aufgrund der Funktionsweise der Finanzmärkte nicht als relevante statistische Zusammenhänge, sondern als auftretende Zufallserscheinungen, basierend auf der zugrundeliegenden Datenbasis, zu interpretieren sind.

Alle weiteren Peak-Werte der Kreuzkorrelation finden sich für die Betrachtung des Kommentaranzahlkurses gegenüber dem Aktienschlusskurs im Bereich von 0.38 bis 0.65. Der Time-Shift der beiden Größen variiert zwischen 0 Tagen - Zusammenhang von Kommentaranzahl und Aktienschlusskurs noch am selben Tag - und 119 Tagen Verzögerung, was wiederum im Kontext etwaiger, an die Arbeit anschließender Kausalitätsprüfungen, detaillierter zu betrachten wäre.

Hervorzuheben ist jedoch, dass die Peak-Werte der Korrelation bei fünf der Untersuchungsunternehmen mit einer mäßigen bis stärkeren Korrelationsausprägung von 0.38 bis 0.65 und einem Time-Shift von bis zu sieben Kalendertagen auftreten. Daraus ist zu schließen, dass im unmittelbaren zeitlichen Naheverhältnis eine Korrelation des Kommentarkurses mit dem Aktienschlusskurs besteht, dieser in gewisser Weise einen Vorlaufindikator des Aktienschlusskurses darstellen kann.

Insbesondere bei den Firmen BB, BBBY, KODK und MAC ist dies innerhalb eines Time-Shifts von zwei Tagen zu beobachten.

Vergleichbare Ergebnisse wurden auch für die Untersuchung der Zeitreihen der Differenz der Kommentarerwähnungen zum Vortag des jeweiligen Unternehmens und dessen Tagesgewinnen und -verlusten festgestellt. Hier konnte bei sechs der verbliebenen Firmen festgestellt werden, dass innerhalb eines Time-Shifts von zwei Tagen die höchsten Korrelationswerte von 0.38 bis 0.81 vorliegen. Speziell die Unternehmen Kodak mit einem Wert von 0.75 und Nokia mit 0.81 weisen eine stark ausgeprägte Korrelation zwischen den täglichen Kommentardifferenzen und den Tagesgewinn/-verlusten der Einzelunternehmen aus.

Ident zur Vorgehensweise der linearen Korrelationsprüfung mittels Korrelationskoeffizient nach Spearman wurde anschließend die normalisierte Kreuzkorrelation mit Zeitversatz für die ermittelten vier Polaritätsszenarien angewandt.

Im Rahmen der Untersuchung des Szenarios 1 – Zusammenhang positiver Kommentare gegenüber den Aktiengewinnen bzw. – verlusten konnten nur für bestimmte angewandte Zeitverschiebungen (lags) der beiden Zeitreihen geringe statistische Zusammenhänge detektiert werden. Weiterführend war über die Unternehmen hinweg anhand der ermittelten Daten keine Systematik erkennbar, ob die unterschiedlichen lags der Kreuzkorrelation unternehmensübergreifend einem bestimmten Muster folgen. Jene Zeitversätze mit den größten Korrelationswerten (0.07 bis 0.46) bewegten sich zwischen 0 bis 361 Tagen. Auch hier sei nochmals auf die bereits zuvor ausgeführten möglichen Zufallserscheinungen von Korrelationen bei Zeitversätzen von über 10 Kalendertagen verwiesen. Jene Ergebnisse sind auf die Datengrundlage sowie deren Aufbereitung zurückzuführen und nicht als tatsächliche statistische Zusammenhänge zu werten.

Ein ähnliches Bild ergaben auch Untersuchungen des Szenario 2 – Zusammenhang negativer Kommentare gegenüber den Aktiengewinnen bzw. – verlusten. Hier sind an

weit gestreuten Lagpunkten die höchsten Kreuzkorrelationswerte festgestellt worden, was keine konkreten Schlüsse auf etwaige Systematiken der Korrelationen zulässt. Einzig das Unternehmen Kodak sticht auch hier mit einer mäßig ausgeprägten Korrelation von 0.46 und einem Time-Shift von 0 hervor.

Interessante Erkenntnisse lieferte die Analyse des Polaritätsszenarios 3 – Korrelation von Tagen mit positivem Kommentarkurs gegenüber Gewinntagen des Aktienschlusskurses.

Die Berechnungsergebnisse zeigen für dieses Szenario bei der Hälfte der untersuchten Unternehmen eine geringe bis mittlere Korrelation zwischen den beiden Zeitreihen ohne Lag - noch am selben Kalendertag. Am stärksten ausgeprägt ist neben Nokia mit einem Höchstwert von 0.53 auch hier das Unternehmen Kodak mit 0.5. Für drei weitere Firmen waren geringe Korrelationen mit einem Lag von bis zu 60 Tagen feststellbar.

Das abschließende Polaritätsszenario 4 – als Gegenstück zu Szenario 3 – Tage mit negativen Kommentarkursen gegenüber Tagen mit Kursverlusten lieferte keine weiteren aussagekräftigen Ergebnisse. Für alle Unternehmen konnten sehr geringe Korrelationen mit stark variierenden Lags von 78 bis 361 Kalendertagen detektiert werden.

Weiterführend wurde geprüft, ob die Entwicklung des Aktienschlusskurses in einem statistischen Zusammenhang zur Kommentaranzahl (dem Kommentaranzahlkurs) im Subreddit „Wallstreetbets“ steht. Dabei wurden zwei weitere Zeitreihenpaare mittels normalisierter Kreuzkorrelation untersucht:

1. Der Aktienschlusskurses gegenüber dem Kommentaranzahlkurs
2. Gewinne und Verluste des Aktienschlusskurses gegenüber der Kommentardifferenz des Unternehmens zu den Vortagen

In beiden oberhalb geschilderten Fällen konnte für mindestens die Hälfte der Betrachtungsunternehmen im Untersuchungszeitraum eine geringe Korrelation, mit einem Höchstwert spätestens am Folgetag, festgestellt werden. Dies lässt den Schluss zu, dass ein geringer statistischen Zusammenhang zwischen den Aktienschlusskursen und den Kommentaranzahlkursen des Folgetags vorliegt. Die davon betroffenen Unternehmen lauten AMC, BB, BBBY, EXPR, GME und KODK.

Im Falle des zweiten oberhalb geschilderten Zeitreihenpaares tritt dieser Zusammenhang auch für die Unternehmen NOK und PLTR auf. Daraus lässt sich ableiten, dass zwischen den Gewinnen- bzw. Verlusten einer Aktie und der Anzahl der Kommentarveränderung (Anstieg/Abfall der Kommentaranzahl) spätestens am Folgetag ein geringer statistischer Zusammenhang besteht.

Die Detaillergebnisse der zuvor beschriebenen Berechnungen können auszugsweise der Tabelle unterhalb entnommen werden. Die vollständigen Daten sind im beiliegenden Anhang zu finden.

**Tabelle 8: Auszug Kreuzkorrelation Kommentaranzahlkurs Aktienschlusskurs**

Unternehmen (Aktienkürzel)	Kommentaranzahlkurs zu Aktienschlusskurs		Differenz Unternehmenserwähnungen zum Vortag zu Gewinn/Verluste Aktienkurs	
	Peak	Peak-Wert	Peak	Peak-Wert
BBBY	2	0,340	2	0,488
KODK	0	0,652	0	0,746
SPCE	7	0,376	339	0,173
TR	63	0,579	60	0,287

**Tabelle 9: Auszug Kreuzkorrelation Sentimentdaten**

Unternehmen (Aktienkürzel)	Positive Kommentare zu Gewinn bzw. Verlust Aktienschlusskurs		Negative Kommentare zu Gewinn bzw. Verlust Aktienschlusskurs		Tage mit pos. Kommentarkurs vs. Gewinntage Aktienschlusskurs		Tage mit neg. Kommentarkurs vs. Verlusttage Aktienschlusskurs	
	Peak	Peak- Wert	Peak	Peak- Wert	Peak	Peak- Wert	Peak	Peak- Wert
BBBY	12	0,239	12	0,130	12	0,413	121	0,095
KODK	50	0,108	0	0,464	0	0,502	292	0,103
NOK	0	0,142	0	0,211	0	0,534	78	0,062
PLTR	58	0,154	56	0,229	0	0,366	179	0,250

**Tabelle 10: Auszug Kreuzkorrelation Aktienkurs gegenüber Kommentarkurs**

Unternehmen (Aktienkürzel)	Aktienschlusskurs zu Kommentaranzahlkurs		Gewinn/Verluste Aktienkurs zu Differenz Unternehmenserwähnungen zum Vortag	
	Peak	Peak-Wert	Peak	Peak-Wert
BB	1	0,594	1	0,543
GME	1	0,424	1	0,507
NOK	176*	0,476	0	0,812

*\*Peak-Werte mit derart hohen Zeitversätzen können nicht als eventuelle kausale Zusammenhänge zur jeweilig gegenübergestellten Größe gewertet werden, sondern sind als Anomalien, entstanden aus der*

*zugrundeliegenden Datenbasis, zu werten. Aufgrund der Effizienz der Finanzmärkte können derartige „Lags“ systematisch als nicht relevant ausgeschlossen werden.*

Im Kapitel 7 der gegenständlichen Arbeit wurde die Anwendung des zuvor vorgestellten Berechnungsmodells zum Korrelationsnachweis beschrieben. Initial erfolgte eine Prüfung der vorliegenden Daten auf Normalverteilung der Informationen. Hierbei kamen Histogramme, Schiefe- und Kurtosis-Berechnungen, sowie der Shapiro-Wilk-Test zum Einsatz. Aufgrund der daraus gewonnenen Erkenntnisse wurde anstelle des Pearson-Korrelationskoeffizienten der Spearman-Rangkorrelationskoeffizient eingesetzt, um die Größen trotz fehlender Normalverteilung auf statistische Zusammenhänge zu untersuchen. Im Rahmen der linearen Korrelationsprüfung konnten sodann je nach Unternehmen, geringe bis starke Merkmalszusammenhänge festgestellt werden, welche anhand zweiseitiger T-Tests erfolgreich auf das Vorliegen von Signifikanz geprüft wurden. Anschließend erfolgte eine Autokorrelationsuntersuchung des jeweiligen einzelnen Kommentaranzahl- und Aktienschlusskurses, bei der durch Anwendung des Durbin-Watson-Tests hohe Autokorrelationswerte festgestellt werden konnten. Jene Autokorrelationen wurden mittels Python weiter untersucht, und eine stark ausgeprägte Saisonalität der beiden genannten Größen mit den jeweiligen Vortageskursen sichtbar gemacht. Als letzter Berechnungsschritt erfolgte zu den einzelnen Kommentarkurstypen, sowie den Aktienkursinformationen eine normalisierte Kreuzkorrelationsberechnung je Unternehmen mit Untersuchung etwaiger Time-Shifts. Dabei konnte erhoben werden, dass auch bei Betrachtung der gegenständlichen Untersuchungsgrößen in einer Zeitreihe, eine geringe statistische Zusammenhangsausprägung nachweisbar ist. Diese tritt für die einzelnen Betrachtungsunternehmen jedoch mit unterschiedlichen Lags auf. Für die Hälfte der Unternehmen wurde auf diese Weise ersichtlich, dass der Kommentarkurs in einem geringen Ausmaß einen statistischen Zusammenhang mit den auf den direkten Folgetagen (<7 Kalendertage) auftretenden Aktienschlusskursen aufweist.

## 8. Schlussfolgerungen aus den Ergebnissen

Im Rahmen dieses Abschnitts werden die Einzelergebnisse aus dem angewandten Berechnungsmodell aggregiert, interpretiert und Schlussfolgerungen daraus abgeleitet.

Beginnend mit dem Kommentarexport konnte bereits festgestellt werden, dass einige vorausgewählte Unternehmen mangels adäquater Stichprobengröße an Kommentarinformationen im Untersuchungszeitraum für die weitere Betrachtung nicht in Frage kamen. Somit reduzierte sich die Anzahl der zu prüfenden Firmen auf zwölf Stück.

Anhand der Anfertigung und Prüfung der Histogramme inklusive der zugehörigen Funktionen Schiefe und Kurtosis konnte für die verbliebenen Unternehmen nachgewiesen werden, dass die Annahme, dass es sich bei den Aktienschlusskursen, sowie den erhobenen Kommentarkursen um normalverteilte Daten handelt, nicht korrekt ist und verworfen werden musste. Der abschließend durchgeführte Shapiro-Wilk-Test untermauert diese Erkenntnisse.

Aufbauend darauf wurde zur Untersuchung der Daten auf etwaige lineare Zusammenhänge nicht wie geplant auf den Pearson-Korrelationskoeffizienten, sondern auf den Spearman-Rangkorrelationskoeffizienten zurückgegriffen.

### Lineare Zusammenhänge

Die Ergebnisse der linearen statistischen Zusammenhangsermittlung anhand des Spearman-Korrelationskoeffizienten zeigen, dass zwischen den Erwähnungen einzelner Unternehmen in Kommentaren des Subreddits „Wallstreetbets“ auf der Plattform [www.reddit.com](http://www.reddit.com) und deren Aktienschlusskurs im Untersuchungszeitraum ein gering bis stärker ausgeprägter, signifikanter linearer Zusammenhang existiert. Die Ausprägung der Korrelation ist je Unternehmen unterschiedlich und könnte mit der höheren Erwähnungshäufigkeit der Unternehmen in Kommentaren auf der Plattform in Verbindung stehen. Dieser Verdacht äußert sich dadurch, dass die Ausprägung der Korrelation bei häufiger erwähnten Unternehmen höher ausfällt.

Daraus lässt sich schließen, dass je öfter ein Unternehmen erwähnt wird, desto höher dessen Chancen stehen, dass ein signifikanter linearer statistischen Zusammenhang zwischen den täglichen Erwähnungen und Aktienschlusskursen vorliegt.

Ein signifikanter statistischer Zusammenhang für den Kommentaranzahlanstieg bzw. -abfall in Korrelation mit den Tagesgewinnen und -verlusten der Aktie konnte jedoch nicht festgestellt werden.

Ebendiese Ergebnisse wurden anhand der mittels Sentimentanalyse gewonnen Polaritäts- und Kommentardetailinformationen weiter untersucht. Mit Ausnahme der Unternehmen Kodak (KODK), Nokia (NOK) Palantir (PLTR) und Virgin Galactic

(SPCE), konnte aufgrund fehlender signifikanter Zusammenhänge, die Erkenntnis aus den vorangegangenen Spearman-Korrelationsberechnungen bestätigt werden. Im Untersuchungszeitraum besteht zwischen den Kommentaren (unabhängig von deren Polaritätsausprägungen) nur in Ausnahmefällen ein signifikanter linearer statistischer Zusammenhang zu den Aktienkursgewinnen bzw. -verlusten.

Die lineare Zusammenhangsprüfung führt somit zusammengefasst zu dem Schluss, dass anhand der eingesetzten Berechnungsmethodik im Untersuchungszeitraum ein geringer signifikanter statistischer Zusammenhang zwischen den Kommentaren auf der Plattform [www.reddit.com/r/wallstreetbets](http://www.reddit.com/r/wallstreetbets) und den zugehörigen Unternehmensaktienkursen besteht. Jener existiert allerdings nur bei Betrachtung der allgemeinen Kommentaranzahlen mit Unternehmenserwähnungen gegenüber der täglichen Aktienschlusskurse. Dies bedeutet fortschreitend, dass die seitens Medien propagierte Robin Hood Geschichte zu GameStop und weiteren betroffenen Unternehmen nicht gänzlich an den Haaren herbeigezogen zu sein scheint, die tatsächliche Ausprägung des statistischen Zusammenhangs in der medialen Berichterstattung jedoch vollkommen überzogen wurde.

Weiterführend zeigt der generelle lineare Zusammenhang zwischen den Größen auch den erschreckenden Einfluss moderner Kommunikationstools und -apps im Finanzmarktsektor und welche finanzpolitische Macht jenen Unternehmen zuteilwird, die diese Services bereitstellen und die Informationsverteilung darauf gegebenenfalls in ihrem Sinne manipulieren können. Dies beginnt bei Social Media Plattformen und endet bei der Gamification des Aktienmarkts anhand diverser Trading Apps (Chaudhry und Kulkarni 2021). Gerade aktuell (April 2022) sei hier auf den Kauf von Twitter durch Elon Musk verwiesen (Susarla 2021). Gezielte Kontrolle der Informationsverteilung, sowie die Festlegung spezieller Design Patterns können die gegenständlichen Erkenntnisse zukünftig weiter intensivieren (Jameson et. al. 2014).

### Autokorrelation

Weiters haben sich die Annahmen bestätigt, dass es sich bei den Größen Kommentarkurs und Aktienschlusskurs, um stark autokorrelierte Daten handelt. Diese Erkenntnisse wurden anhand des Durbin-Watson-Tests mit dem Großteil der Ergebnisse im Bereich 0.001 bis 0.2 eindrucksvoll untermauert.

Eine Vertiefung der Untersuchungen mit der Python-statsmodels-Library hat zum Vorschein gebracht, dass die beiden Kurse mit einer Saionalität von einem Tag sehr stark autokorreliert sind. Diese Korrelationsausprägung reduziert sich mit jedem weiteren verstrichenen Vortag.

Es lässt sich daraus schlussfolgern, dass der Aktienschlusskurs, als auch der Kommentaranzahlkurs maßgeblich von seinem eigenen Vortageskurs abhängig ist und rapide Kursanstiege, wie es etwa im Jänner 2021 für GamesStop oder AMC Entertainment Inc. der Fall war, die Ausnahme darstellen (Ariva 2021). Der jeweilige

Vortageskurs ist wiederum mit seinem Vorgänger und dessen Vorgängern, in immer stärker abnehmender Ausprägung korreliert, was die zuvor getätigte Schlussfolgerung weiter stützt.

### Kreuzkorrelation

Anhand der Kreuzkorrelationsuntersuchung wurde festgestellt, dass bei allen verbliebenen Unternehmen im Betrachtungszeitraum ein statistischer Zusammenhang zwischen den Zeitreihen Kommentarkurs und Aktienkurs zueinander vorliegt. Der Zeitversatz (Time-Shift), in welchem diese Korrelation am höchsten ist, variiert jedoch stark zwischen den einzelnen Firmen und lässt keine pauschalierten Schlüsse zu.

Hervorzuheben ist, dass die Peak-Werte der Korrelation bei fünf der Untersuchungsunternehmen mit einer mäßigen bis stärkeren Korrelationsausprägung bei einem Time-Shift von bis zu sieben Kalendertagen auftritt. Daraus ist zu schließen, dass im unmittelbaren zeitlichen Naheverhältnis eine Korrelation des Kommentarkurses mit dem Aktienschlusskurs besteht. Dieser stellt somit in gewisser Weise einen Vorlaufindikator des Aktienschlusskurses dar. Insbesondere bei den Firmen BB, BBBY, KODK und MAC ist dies innerhalb eines Time-Shifts von zwei Tagen zu beobachten. Vergleichbare Ergebnisse wurden auch für die Untersuchung der Zeitreihen der Differenz der Kommentarerwähnungen zum Vortag des jeweiligen Unternehmens und dessen Tagesgewinnen und -verlusten festgestellt.

Die Berechnungsergebnisse des Polaritätsszenarios 3. zeigen zudem bei der Hälfte der untersuchten Unternehmen eine geringe bis mittlere Korrelation zwischen den beiden Zeitreihen ohne Lag - noch am selben Kalendertag. Dies bekräftigt die bereits im Rahmen der konventionellen Kreuzkorrelationsberechnung ermittelte und aufgestellte These, dass der Kommentarkurs und dabei insbesondere auch die positive Ausprägung der Kommentare, einen möglichen Vorlaufindikator des Aktienschlusskurses bzw. für potentielle Unternehmenskursgewinne darstellen könnte.

Im Zuge der gegenständlichen Ausarbeitung sollte die Hypothese untersucht werden, ob zwischen den Erwähnungen ausgewählter zum Short Squeeze aufgerufener Unternehmen auf der Internetplattform [www.reddit.com/r/wallstreetbets](http://www.reddit.com/r/wallstreetbets) und dem Aktienschlusskurs ebendieser Unternehmen ein statistischer Zusammenhang existiert.

Diese Hypothese konnte anhand der dargelegten Ergebnisse, mit Ausnahme der aus der Berechnung aufgrund zu geringer Kommentarzahlen exkludierter Firmen, bestätigt werden. Eine Korrelation der beiden Hauptuntersuchungsgrößen liegt nicht nur im linearen, sondern auch im Kontext der Zeitreihen vor.

Zusammenfassend kann daher festgehalten werden, dass bei häufig diskutierten Unternehmensaktien auf der Plattform [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets) ein signifikanter linearer Zusammenhang, sowie ein zeitversetzter Wirkungszusammenhang zwischen dem Kommentarkurs und dem Aktienkurs von

ebendiesen Unternehmen auftreten kann. Die Ausprägung dieser Korrelation ist jedoch von der Häufigkeit der Erwähnungen und somit auch der Popularität des jeweiligen Unternehmens in der „Wallstreetbets“-Community abhängig.

## 9. Conclusio

Im Rahmen der gegenständlichen Untersuchung wurde geprüft, ob im Betrachtungszeitraum 01.06.2020 bis 01.06.2021 ein statistischer Zusammenhang zwischen den Erwähnungen von ausgewählten Unternehmen im Subreddit „Wallstreetbets“ auf der Internetplattform [www.reddit.com](http://www.reddit.com) und deren Aktienkurs besteht. Zusätzlich wurden Korrelationsprüfungen auf Basis von durch Text Mining und Sentimentanalyse gewonnenen Daten angestellt. Dabei wurde unter anderem untersucht, ob zwischen Kommentaren positiver Polarität und den Unternehmensgewinnen (Anstieg des Aktienschlusskurses gegenüber des Vortags), sowie Kommentaren negativer Polarität und Unternehmensverlusten (Reduktion des Aktienschlusskurses gegenüber dem Vortag) statistische Zusammenhänge vorliegen.

Auslöser für die Durchführung der Untersuchung waren die Ereignisse rund um den exorbitanten Kursanstieg der GameStop Aktie (GME) zu Jahresbeginn 2021, sowie die zunehmende Gamification des Finanzmarkts und die damit einhergehende Entwicklung der immer jünger werdenden Anlegerschaft.

Zur Ermittlung der statistischen Zusammenhänge wurden die Finanzmarktdaten zu den Einzelaktien im Betrachtungszeitraum exportiert und für die Weiterverarbeitung aufbereitet. Dabei wurden die täglichen Aktienschlusskurse, sowie deren Gewinne und Verluste herangezogen.

Diese Daten wurden, den durch Text Mining und Sentimentanalyseverfahren gewonnen Informationen, von der Plattform [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets), gegenübergestellt. Anhand der durchgeführten Textanalyseverfahren konnten so zwei Kommentarkursdateien erstellt werden: Eine Kommentardatei, welche die Informationen, sowie den Inhalt der einzelnen Kommentare je Unternehmen enthält und eine Kommentarkursdatei, welche die Anzahl der absoluten, positiven, negativen und neutralen Kommentare pro Tag, sowie die Differenz der absoluten Kommentarzahl mit Unternehmenserwähnungen pro Tag zur Vortageskommentaranzahl beinhaltet. Die so generierten Kommentarkurse wurden den Aktienkursen je Unternehmen entgegengesetzt und untersucht.

Dabei wurde nach dem unter Kapitel 5 angeführten Berechnungsmodell initial das Vorliegen von Normalverteilungen bei den genannten Größen geprüft. Anschließend wurden lineare Merkmalszusammenhänge und abschließend statistische Korrelationen in einem zeitlichen Kontext mittels normalisierter Auto- und Kreuzkorrelation mit Zeitversatz untersucht.

Die unter Kapitel 7 und 8 beschriebenen Ergebnisse zeigen, dass zwischen den Reddit Aktivitäten (Kommentaranzahlkurs) und dem Aktienkurs im gegenständlichen Betrachtungszeitraum signifikante statistische Zusammenhänge nachweisbar sind und somit die aufgestellte Hypothese bestätigt werden kann.

Weiters wurde festgestellt, dass sowohl der Aktienkurs als auch der Kommentarkurs mit einem Versatz von einem Kalendertag stark autokorrelieren.

Es sei an dieser Stelle jedoch erwähnt, dass dies nicht für alle in Punkt 2 angeführten Untersuchungsunternehmen der Fall ist, da elf Unternehmen, aufgrund mangelnder

kommentarischer Erwähnungen auf [www.reddit.com/r/Wallstreetbets](http://www.reddit.com/r/Wallstreetbets), von der Untersuchung ausgenommen wurden (siehe auch Kapitel 7).

Zusätzlich zeigt sich, dass die lineare Korrelation zwischen Kommentaranzahlkurs und Aktienschlusskurs umso stärker ausgeprägt ist, je mehr eine Unternehmensaktie auf [www. reddit.com](http://www.reddit.com) diskutiert wird. Anhand der angestellten Kreuzkorrelationsberechnungen konnte zudem festgestellt werden, dass bei bestimmten Unternehmen der Kommentarkurs mit einem Zeitversatz von unter drei Tagen mit dem Aktienschlusskurs in gering bis mäßiger Ausprägung korreliert. Dies deutet darauf hin, dass die im Januar 2021 angestellten Behauptungen von diversen Medienagenturen und Zeitschriften – die Reddit-User hätten die großen Hedgefonds beim Handel mit der GME-Aktie in die Knie gezwungen – nicht gänzlich einer medienwirksamen Utopie entsprechen.

Anhand der ausgewerteten Kommentarpolaritäten und der generierten Polaritätskurse war es zudem möglich nachzuweisen, dass zwischen positiven, sowie negativen Kommentaren und dem Verlauf der Unternehmenstagesgewinne und deren -verlusten eine Korrelation in der Zeitreihe mit einem Zeitversatz von null bis sieben Kalendertagen im Untersuchungszeitraum festgestellt werden konnte. Die Ausprägung dieser Korrelation fällt jedoch je Unternehmen unterschiedlich stark aus. Abschließend konnte bewiesen werden, dass zwischen Tagen mit Unternehmenserwähnungen im positiven Kontext und Tagen mit Kursgewinnen Korrelationen bestehen.

Die Ergebnisse der gegenständliche Studie unterstreichen zudem die von Dowling und Lucey (Dowling und Lucey 2003) und Burnette (Burnette 2021) festgestellten Erkenntnisse, dass Investitionsentscheidungen vom Herdenverhalten und emotionalen Komponenten geprägt sind (Jameson et. al. 2014). Es bestätigt sich, dass jene gerade im Social Media Bereich möglicherweise bis zur Prognostizierung von eventuellen Aktienkursverläufen führen könnten (Burnette 2021). Dabei ist das blinde Vertrauen der jüngeren Anlegerchaft in sozialen Medien, sowie die Beeinflussbarkeit durch die Wahl entsprechender Design Patterns, etwa bei Trading Apps, erschreckend (Chaudhry und Kulkarni 2021).

Weiterführend, aber nicht mehr im Rahmen dieser Arbeit behandelt, könnte aufbauend auf die ermittelten statistischen Zusammenhänge ein Granger-Kausalitätstest durchgeführt und so der Grundstein für ein Reddit-Aktienkurs-Prognosemodell geschaffen werden. Theoretisch wäre denkbar, dass anhand der Unternehmenserwähnungen an einem Kalendertag, die Aktienkursentwicklung dieses Unternehmens in der Zukunft, anhand eines eigenen Reddit-Vorlaufindikators, prognostiziert werden könnte. Zusätzlich wäre es möglich noch in verstärkter Form auf die Sentimentanalysedaten einzugehen, um etwa Prognosen aus den Polaritäten zu ermitteln. Die Untersuchung der Intraday-Aktienkurse, sowie Kommentarkurse unter Einbeziehung des Handelsvolumens ist eine weitere Möglichkeit zur Vertiefung der vorliegenden Berechnungen.

Aus volkswirtschaftlicher Sicht ist es ratsam Überlegungen anzustellen, wie mit der Verjüngung der Anlegerschaft und einer der Ursachen davon - der steigenden Gamification am Finanzmarkt – zukünftig umgegangen werden soll. Es gilt geeignete Mechanismen zu implementieren, um das Auftreten von Short Squeezes und damit einhergehenden Phantasiepreisaktienkursen, wie es etwa im Falle von Gamestop Anfang des Jahres 2021 aufgetreten ist, zu verhindern und die gegenständlich gezeigten statistischen Zusammenhänge zu reduzieren.

## Literaturverzeichnis

- Anand, Kshitij. 2020. "As lockdown blues hit, most brokerages see mobile trading doubling in 2020". In *moneycontrol*. Abgerufen am 7. August.  
<https://www.moneycontrol.com/news/business/markets/as-lockdown-blues-hit-most-brokerages-see-mobile-tradingdoubling-in-2020-5758791.html>
- Anand, Abhinav and Pathak, Jalaj. 2021. "WallStreetBets Against Wall Street: The Role of Reddit in the GameStop Short Squeeze" *IIM Bangalore Research Paper No. 644*. <http://dx.doi.org/10.2139/ssrn.3873099>
- Angel, James. 2021. "GameStonk: What happened and what to do about it". Georgetown McDonough School of Business Research Paper No. 3782195. University of Goergetown. <http://dx.doi.org/10.2139/ssrn.3782195>
- Allen, Franklin, Lubomir Litov und Jianping Mei. 2006. „Large Investors, Price Manipulation, and Limits to Arbitrage: An Anatomy of Market Corners". *Review of Finance*, 10(4):645-693. <https://doi.org/10.1007/s10679-006-9008-5>
- Allen, Franklin, Marlene Haas, Eric Nowak und Angel Tengulov. 2017. „Market efficiency and limits to arbitrage: Evidence from the biggest short squeeze in history". Whitepaper. University of Lugano
- Affde. 2021. „Reddit-Nutzungs- und Wachstumsstatistik: Wie viele Menschen nutzen Reddit im Jahr 2021?". In *Affde*. Abgerufen am 03. November.  
<https://www.affde.com/de/reddit-users.html>
- Ariva. 2021. „Aktiensuche". in *Ariva Aktien*. Abgerufen am 14. November.  
<https://www.ariva.de/aktien/suche>
- Arnold, Marc. 2021. "Trading Apps verleiten zu risikoreichem Anlageverhalten" In *Universität St. Gallen*. Abgerufen am 20. September.  
<https://www.unisg.ch/de/wissen/newsroom/aktuell/rssnews/forschung-lehre/2021/februar/trading-apps-verleiten-zu-risikoreichem-anlageverhalten-8februar2021>
- Barber, Brad M. und Terrance Odean. 1998. „Boys will be Boys: Gender, Overconfidence, and Common Stock Investment". Whitepaper. University of California
- Bary, Emily. 2017. "Videogame Sales Are Fading and It's Chrushing GameStop" *Barron's. Dow Jones & Company, Inc.* 13. Januar.  
<https://www.barrons.com/articles/u-s-energy-prices-breaking-historic-pattern-51635879750>

- Bendel, Oliver. 2013. "Gamification. Definition: Was ist "Gamification"?". In *Gabler Wirtschaftslexikon* Springer Gabler. Abgerufen am 4. August.  
<https://wirtschaftslexikon.gabler.de/definition/gamification-53874>
- Blanz, Mathias. 2021. „Forschungsmethoden und Statistik für die Soziale Arbeit Grundlagen und Anwendung“. Deutschland. W. Kohlhammer. S. 113 - 194
- Bollen, Johan, Huina Mao und Xiao-Jun Zeng. 2010. "Twitter mood predicts the stock market". *Journal of Computational Science*, 2(1): 1-8.  
 doi:10.1016/j.jocs.2010.12.007.
- Bundesverband Geothermie. 2020. Kreuzkorrelation. In *Bundesverband Geothermie Lexikon der Geothermie*. Abgerufen am 04. Dezember.  
<https://www.geothermie.de/bibliothek/lexikon-der-geothermie/k/kreuzkorrelation.html>
- Burnette, Riley. 2021. „What Were the Factors that led to the GameStop Short Squeeze?“. Information Systems Undergraduate Honors Theses. University of Arkansas
- Cambridge Dictionary. 2021. "Short-Sale". Abgerufen am 03. November.  
<https://dictionary.cambridge.org/de/worterbuch/englisch/short-sale>
- Chaudhry, Sayan und Chinmay Kulkarni. 2021. "Design Patterns of Investing Apps and Ther Effects on Investing Behaviors". Paper präsentiert bei der Designing Interactice Systems Conference 21, Virtuelles Event, USA, 28. Juni bis 2. Juli.  
<https://doi.org/10.1145/3461778.3462008>
- Dong, Hang und Javier Gil-Bazo. 2020. „Sentiment stocks“. *International Review of Financial Analysis* vol.72. Niederlande. Elsevier  
 doi: 10.1016/j.irfa.2020.101573
- Dounis, Filippos. 2020. „Elon's Golden Gift: Predicting The Stock Price Of Tesla With Twitter And Machine Learning“. In *Citadel*. Abgerufen am 26. Oktober.  
<https://thecitadelpub.com/elons-lethal-mistake-predicting-the-stock-price-of-tesla-with-twitter-and-machine-learning-5e89282ce75f>
- Dowling, Michael M. und Brian M. Lucey. 2003. "The Role of Feelings in Investor Decision-Making". Whitepaper. Dublin City University
- Ebermann, Erwin. 2010a. " Optischer Nachweis einer Normalverteilung: das Histogramm". In *Grundlagen statistischer Auswertungsverfahren*. Universität Wien. Abgerufen am 14. Dezember.  
<https://www.univie.ac.at/ksa/elearning/cp/quantitative/quantitative-60.html>
- Ebermann, Erwin. 2010b. "Signifikanz der Korrelation". In *Grundlagen statistischer Auswertungsverfahren*. Universität Wien. Abgerufen am 23. November.  
<https://www.univie.ac.at/ksa/elearning/cp/quantitative/quantitative-108.html>

- Evangelos, Vasileiou. 2021. „Does the Short Squeeze lead to market abnormality and anti-leverage effect? Evidence from the Gamestop case.“ Whitepaper. University of the Aegean
- Evangelos, Vasileiou, Bartzou Eleftheria und Tzanakis Polydoros. 2021. „Explaining Gamestop Short Squeeze using Intraday Data and Google Searches“ Whitepaper. University of the Aegean
- Fabio, Momo. 2015. „Die Auswirkungen von Herdenverhalten auf Finanzmärkte. Wissenschaftliche Zuordnung und theoretische Grundlagen“. Deutschland: GRIN Verlag
- Gäbler, Jonas. 2021. “The Influence of Neo-Broker on the Stock Market and Investment Behavior”. Thesis. Novia University of Applied Sciences, S. 8
- Jameson, Anthony, Bettina Berendt, Silvia Gabrielli, Federica Cena, Cristina Gena, Fabiana Venero, and Katharina Reinecke. 2014. “Choice architecture for humancomputer interaction”. *Foundations and Trends in Human-Computer Interaction* 7, S. 1–235.
- Jianakoplos, Nancy Ammon und Alexandra Bernasek. 1998. “Are women more risk averse?” Whitepaper. California State University.
- Jiao, Peiran, Andre Veiga und Ansgar Walther. 2018. „Social Media, News Media and the Stock Market“. Whitepaper. Maastricht University. <http://dx.doi.org/10.2139/ssrn.2755933>
- Kolb, Peter. 2011. “Was ist statistische maschinelle Übersetzung?” Artikel. Universität Potsdam.
- Lin, Ying. 2021. “10 Twitter Statistics Every Marketer Should Know in 2021”. In *Oberlo*. Abgerufen am 03. November. <https://www.oberlo.com/blog/twitter-statistics>
- Lockton, Dan. 2012. „Cognitive biases, heuristics and decision-making in design for behaviour change“. Whitepaper. Royal College of Art.
- Long, Cheng, Brian M. Lucey und Larisa Yarovaya. 2021. “I Just Like the Stock” versus “Fear and Loathing on Main Street”: The Role of Reddit Sentiment in the GameStop Short Squeeze. Whitepaper. Trinity College Dublin <http://dx.doi.org/10.2139/ssrn.3822315>
- Mackintosh, Phil. 2020. “Who counts as a Retail Investor?”. Nasdaq. Abgerufen am 27. August. <https://www.nasdaq.com/articles/who-counts-as-a-retail-investor-2020-12-17>
- Malhotra, Suzanee. 2020. „Study of Features of Mobile Trading Apps: A Silver Lining of Pandemic“. *Journal of Global Information & Business Strategy (JGIBS)*, 12(1), 75-80. DOI: 10.5958/2582-6115.2020.00009.0

- Matplotlib. 2021. "matplotlib Users Guide". In *matplotlib*. Abgerufen am 24. November.  
<https://matplotlib.org/stable/users/index>
- McCrank, John. 2021. „Factbox: The U.S. retail trading frenzy in numbers". Reuters.  
 Abgerufen am 14. August. <https://www.reuters.com/article/us-retail-trading-numbers-idUSKBN29Y2PW>
- NumPy. 2021. "NumPy Quickstart". In *NumPy*. Abgerufen am 24. November.  
<https://numpy.org/doc/stable/user/quickstart.html>
- PRAW. 2021. " PRAW: Configuring PRAW". In *Python Reddit API Wrapper*. Abgerufen am 24. September.  
[https://praw.readthedocs.io/en/latest/getting\\_started/configuration.html](https://praw.readthedocs.io/en/latest/getting_started/configuration.html)
- Pushshift 2021. „Pushshift Reddit API Documentation". In *GitHub*. Abgerufen am 14. April. <https://github.com/pushshift/api>
- Razali, Mohd Nornadiah und Yap Bee Wah. 2011. "Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests." *J. Stat. Model. Analytics*. 2.
- Reddit.2021.“ I was about to check on my stocks I bought in margin with my life savings and see this shit in the mail". In */r/Wallstreetbets*.  
 Abgerufen am 11. Dezember  
[https://www.reddit.com/r/wallstreetbets/comments/rdsulq/i\\_was\\_about\\_to\\_check\\_on\\_my\\_stocks\\_i\\_bought\\_in/](https://www.reddit.com/r/wallstreetbets/comments/rdsulq/i_was_about_to_check_on_my_stocks_i_bought_in/)
- Reider, Rob. 2020. „Python Tutorial: Correlation of Two Time Series". In *DataCamp*.  
 Abgerufen am 07. November.  
<https://www.youtube.com/watch?v=XVV6IVPUIKU>
- Reitz, Kenneth und Tanya Schuster. 2016a. „The Hitchhiker’s Guide to Python. Best Practices for Development". USA: O’Reilly Media, S.4
- Reitz, Kenneth und Tanya Schuster. 2016b. „The Hitchhiker’s Guide to Python. Best Practices for Development". USA: O’Reilly Media, S.271-283
- Riedl, Dominique. 2021. "Neobroker im Vergleich: Wo liegen die Stärken, wo die Schwächen". In *JustETF*. Abgerufen am 30. Oktober.  
<https://www.justetf.com/de/academy/neo-broker-im-vergleich-wo-liegen-die-staerken-wo-die-schwaechen.html>
- Shapiro, Sam S. und Martin Bradbury Wilk. 1965. "An analysis of variance test for normality (for complete samples)". *Biometrika*, 52(3/4). S. 591–611
- Shead, Sam. 2021. "Elon Musk’s tweets are moving markets — and some investors are worried". In *CNBC*. Abgerufen am 26. Oktober.  
<https://www.cnbc.com/2021/01/29/elon-musks-tweets-are-moving-markets.html>

- Statsmodels. 2021. "statsmodels Users Guide". In *statsmodels v0.13.1*. Abgerufen am 24. November. <https://www.statsmodels.org/stable/user-guide.html>
- Staudinger, Martin. 2021. „MT122: Statistische Datenauswertung und -analyse“. Österreich. Ferdinand Porsche FernFH, S.72 – 112
- Stocker, Herbert. 2021. „Kap. 12 Autokorrelation“ in *Methoden der empirischen Wirtschaftsforschung*. Österreich. Universität Innsbruck
- Susarla, Anjana.2022.“ Twitter has a misinformation problem. Here’s how Elon Musk’s plans for the platform could make it worse”. In *PBS*.  
Abgerufen am 27. April  
<https://www.pbs.org/newshour/nation/twitter-has-a-misinformation-problem-heres-how-elon-musks-plans-for-the-platform-could-make-it-worse>
- Tanaka, Takumi und Hideaki Kawabata. 2020. „Interface predictability changes betting behavior in computerized gambling“. *Computers in Human Behavior 110 (2020)*, 106387. <https://doi.org/10.1016/j.chb.2020.106387>
- Thorbecke, Catherine.2021 “GameStop timeline: A closer look at the saga that upended Wall Street” In *abcNews*. Abgerufen am 31.12.2021  
<https://abcnews.go.com/Business/gamestop-timeline-closer-saga-upended-wall-street/story?id=75617315>
- Textblob. 2021. “ Textblob: Quickstart”. In *Textblob*. Abgerufen am 26. September.  
<https://textblob.readthedocs.io/en/dev/quickstart.html>
- United States Securities and Exchange Commission. 2016. Formular 10-K GameStop Corp. United States Securities and Exchange Commission.
- Walther, Björn. 2020. “Pearson-Korrelation – Signifikanz in Excel berechnen – Daten analysieren in Excel (58)”. In *Statistik am PC*. Abgerufen am 04. November.  
<https://www.youtube.com/watch?v=BhrFe4O0jGo>
- Whitebox. 2020. “Was sind Leerverkäufe”. Abgerufen am 03.November.  
<https://www.whitebox.eu/whitepedia/investieren/was-sind-leerverkaeufe>
- Wikipedia.2021. “ GameStop”. In *Wikipedia, The Free Encyclopedia*.  
Bearbeitungsstand: 29. September . Abgerufen am 28. Oktober.  
<https://de.wikipedia.org/wiki/GameStop>.
- Wursthorn Michael und Euirim Choi. 2020. „Does Robinhood Make It Too Easy to Trade? From Free Stoks to Confetti:” In *The Wallstreet Journal*. Abgerufen am 27. Oktober. <https://www.wsj.com/articles/confetti-free-stocks-does-robinhoods-design-make-trading-too-easy-11597915801>

Yarovaya, Larisa, Janusz Brzeszczyński, John W. Goodell, Brian Lucey und Chi Keung Marco Lau. 2020 "Rethinking Financial Contagion: Information Transmission Mechanism during the COVID-19 pandemic." Whitepaper.  
<http://dx.doi.org/10.2139/ssrn.3602973>

## 10. Anhang

Im gegenständlichen Anhang zur Untersuchung werden die einzelnen Ergebnisse zu den jeweiligen Unternehmen aus dem Berechnungsmodell dargestellt. Dies soll als Nachschlagewerk im Falle von Detailfragen und für weitere Analysen dienen. Die Gliederung des Abschnitts erfolgt nach den geprüften Einzelunternehmen.

### 10.1 Schiefe & Kurtosis

Tabelle 11: Schiefe & Kurtosis

Schiefe und Kurtosisberechnung				
Unternehmen (Aktienkürzel)	Aktienchlusskurs		Kommentaranzahlkurs	
	Schiefe	Kurtosis	Schiefe	Kurtosis
AAL	0,6	-1,1	7,1	58,5
AMC	2,1	6,6	6,9	50,9
BB	1,4	3,5	5,9	38,2
BBBY	0,1	-0,4	6,4	44,0
EXPR	1,3	2,0	6,6	50,7
GME	1,2	0,4	6,3	44,4
KODK	2,0	13,1	9,5	111,1
MAC	0,5	-0,2	5,2	34,3
NOK	0,9	1,8	8,7	84,4
PLTR	-0,1	-1,5	4,9	32,7
SPCE	1,6	2,4	5,0	37,4
TR	1,7	4,7	4,7	27,1

### 10.2 Shapiro-Wilk-Test

Tabelle 12: Shapiro-Wilk-Test

Shapiro-Wilk-Test ( $\alpha=5\%$ )				
Unternehmen (Aktienkürzel)	Aktienchlusskurs		Kommentaranzahlkurs	
	Teststatistik	p-Wert	Teststatistik	p-Wert

AAL	0.8941	3.1797e-15	0.3345	2.8036e-34
AMC	0.8061	1.3617e-20	0.2248	3.1577e-36
BB	0.8352	4.5983e-19	0.2596	1.2357e-35
BBBY	0.9376	3.0381e-11	0.2500	8.4472e-36
EXPR	0.8474	2.3214e-18	0.3636	1.0192e-33
GME	0.7313	7.3730e-24	0.3103	9.9596e-35
KODK	0.7922	2.8857e-21	0.2705	1.9177e-35
MAC	0.9235	1.0612e-12	0.5095	1.4996e-30
NOK	0.9109	7.4580e-14	0.1638	3.2449e-37
PLTR	0.8527	4.8269e-18	0.50338	1.0676e-30
SPCE	0.8273	1.6878e-19	0.5388	7.9587e-30
TR	0.8816	3.7803e-16	0.4869	4.3501e-31

### 10.3 Spearman Rangkorrelationskoeffizient

Tabelle 13: Spearman-Rangkorrelationskoeffizient Text Mining

Spearman-Rangkorrelationskoeffizient ( $\alpha=5\%$ )				
Unternehmen (Aktienkürzel)	Aktienchlusskurs Kommentaranzahlkurs		& Gewinn/Verluste Aktienkurs & Differenz Unternehmenserwähnungen zum Vortag	
	Korrelation	p-Wert	Korrelation	p-Wert
AAL	0.1601	0.0022	0.07287	0.1646
AMC	0.4576	2.7035e-20	0.0449	0.3921
BB	0.6710	4.3167e-49	0.0524	0.3179
BBBY	0.2229	1.7215e-05	0.0631	0.2290
EXPR	0.2688	1.8542e-07	0.0343	0.5142
GME	0.7596	8.446e-70	0.1546	0.0031
KODK	0.4014	1.4558e-15	0.0021	0.9685

MAC	0.4238	2.4168e-17	0.0746	0.1550
NOK	0.1738	0.0009	0.1053	0.0449
PLTR	0.7995	2.2931e-82	-0.0704	0.178
SPCE	0.3394	2.7134e-11	0.0821	0.1176
TR	-0.2606	4.4072e-07	-0.0347	0.5092

### 10.3.1 Spearman Rangkorrelationskoeffizient Sentimentdaten

Tabelle 14: Spearman-Rangkorrelationskoeffizient für Sentimentinformationen

Spearman-Rangkorrelationskoeffizient für Sentimentinformationen ( $\alpha=5\%$ )	Unternehmen		AAL	AMC	BB*	BBBY*	EXPR	GME*	KODK*	MAC*	NOK*	PLTR*	SPCE*	TR
	Szenario 1	Korrelation	-0,015	-0,059	-0,051	-0,001	-0,047	0,009	<b>-0,135</b>	0,058	0,029	-0,100	<b>0,110</b>	0,000
		p-Wert	0,772	0,263	0,330	0,988	0,367	0,858	<b>0,0100*</b>	0,272	0,587	0,057	<b>0,0364*</b>	0,999
	Szenario 2	Korrelation	0,026	-0,044	-0,072	-0,054	-0,019	-0,004	<b>-0,108</b>	0,072	<b>-0,106</b>	<b>-0,140</b>	0,041	0,042
		p-Wert	0,623	0,405	0,170	0,305	0,717	0,942	<b>0,0398*</b>	0,170	<b>0,0433*</b>	<b>0,0075*</b>	0,430	0,419
	Szenario 3	Korrelation	0,034	0,068	0,068	0,014	0,015	<b>0,134</b>	0,029	<b>0,153</b>	0,098	<b>0,304</b>	<b>0,193</b>	0,088
		p-Wert	0,514	0,197	0,194	0,784	0,774	<b>0,0106*</b>	0,586	<b>0,0033*</b>	0,061	<b>2,9965e-09*</b>	<b>0,0002*</b>	0,093
	Szenario 4	Korrelation	-0,023	-0,093	<b>-0,176</b>	<b>-0,118</b>	-0,058	<b>-0,124</b>	<b>-0,174</b>	-0,032	<b>-0,138</b>	<b>-0,408</b>	-0,045	-0,046
		p-Wert	0,658	0,075	<b>0,0007*</b>	<b>0,0244*</b>	0,266	<b>0,0181*</b>	<b>0,0008*</b>	0,538	<b>0,0084*</b>	<b>4,5339e-16*</b>	0,396	0,378

## 10.4 Autokorrelation

Tabelle 15: Autokorrelation

Unternehmen (Aktienkürzel)	Durbin Watson Test		Autokorrelationswert Vortag	
	Aktienkurs	Kommentaranzahlkurs	Aktienkurs	Kommentaranzahlkurs
AAL	0,0013	0,9492	0,9830	0,4440
AMC	0,0263	0,1986	0,9214	0,8953
BB	0,0092	0,1875	0,9650	0,9000
BBBY	0,0056	0,4155	0,9719	0,7795
EXPR	0,0422	0,3517	0,9278	0,7846
GME	0,0380	0,4563	0,9633	0,9095
KODK	0,0341	0,6460	0,8962	0,6477
MAC	0,0023	0,2067	0,9748	0,8227
NOK	0,0016	0,2970	0,9112	0,8464
PLTR	0,0033	0,2295	0,9930	0,8576
SPCE	0,0034	0,2991	0,9852	0,7811
TR	0,0005	0,2690	0,9269	0,8328

## 10.5 Kreuzkorrelation

**Tabelle 16: Zusammenhang Kommentaranzahlkurse zu Aktienschlusskursen**

Unternehmen (Aktienkürzel)	Kommentaranzahlkurs zu Aktienschlusskurs		Differenz Unternehmenserwähnungen zum Vortag zu Gewinn/Verluste Aktienkurs	
	Peak	Peak-Wert	Peak	Peak-Wert
AAL	359*	2,157	356*	0,516
AMC	119*	0,822	119*	0,503
BB	0	0,588	0	0,379
BBBY	2	0,340	2	0,488
EXPR	50*	0,493	0	0,405
GME	42*	0,518	40*	0,126
KODK	0	0,652	0	0,746
MAC	0	0,419	167*	0,187
NOK	113*	0,438	0	0,812
PLTR	61*	0,485	61*	0,036
SPCE	7	0,376	339*	0,173
TR	63*	0,579	60*	0,287

*\*Peak-Werte mit derart hohen Zeitversätzen können nicht als eventuelle kausale Zusammenhänge zur jeweilig gegenübergestellten Größe gewertet werden, sondern sind als Anomalien, entstanden aus der zugrundeliegenden Datenbasis, zu werten. Aufgrund der Effizienz der Finanzmärkte können derartige „Lags“ systematisch als nicht relevant ausgeschlossen werden.*

**Tabelle 17: Sentimentdaten – Analyse der Polaritäten zu Gewinnen und Verlusten**

Unternehmen (Aktienkürzel)	Positive Kommentare zu Gewinn bzw. Verlust Aktienschlusskurs		Negative Kommentare zu Gewinn bzw. Verlust Aktienschlusskurs		Tage mit pos. Kommentarkurs vs. Gewinntage Aktienschlusskurs		Tage mit neg. Kommentarkurs vs. Verlusttage Aktienschlusskurs	
	Peak	Peak- Wert	Peak	Peak- Wert	Peak	Peak- Wert	Peak	Peak-Wert
AAL	356*	1,117	335*	0,392	356*	1,107	352*	0,309
AMC	116*	0,461	116*	0,451	117*	0,841	361*	0,550
BB	118*	0,075	118*	0,072	0	0,303	109*	0,067
BBBY	12*	0,239	12*	0,130	12*	0,413	121*	0,095

EXPR	358*	0,306	348*	0,347	358*	0,316	361*	0,373
GME	361*	0,253	361*	0,218	39*	0,280	361*	0,225
KODK	50*	0,108	0	0,464	0	0,502	292*	0,103
MAC	116*	0,117	216*	0,163	0	0,191	241*	0,164
NOK	0	0,142	0	0,211	0	0,534	78*	0,062
PLTR	58*	0,154	56*	0,229	0	0,366	179*	0,250
SPCE	0	0,176	352*	0,187	0	0,263	243*	0,158
TR	105*	0,122	60*	0,144	60*	0,318	136*	0,123

*\*Peak-Werte mit derart hohen Zeitversätzen können nicht als eventuelle kausale Zusammenhänge zur jeweilig gegenübergestellten Größe gewertet werden, sondern sind als Anomalien, entstanden aus der zugrundeliegenden Datenbasis, zu werten. Aufgrund der Effizienz der Finanzmärkte können derartige „Lags“ systematisch als nicht relevant ausgeschlossen werden.*

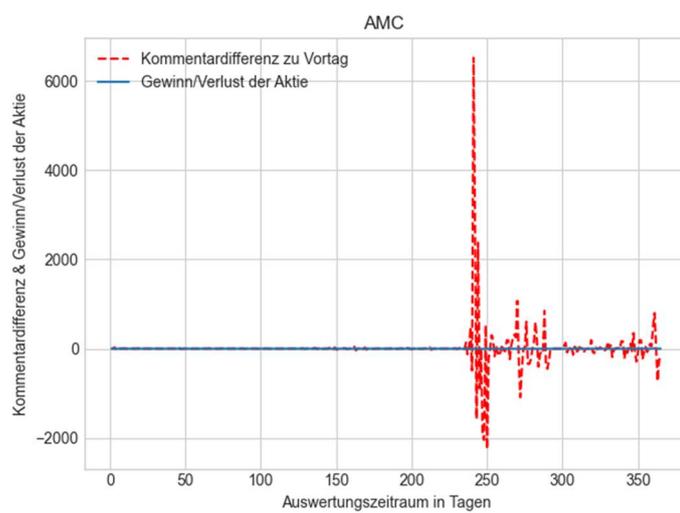
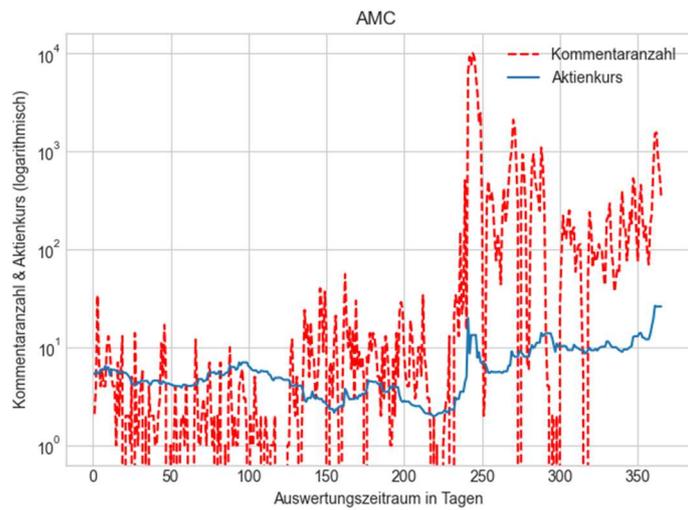
**Tabelle 18: Zusammenhang Aktienschlusskurse mit Kommentaranzahlkursen**

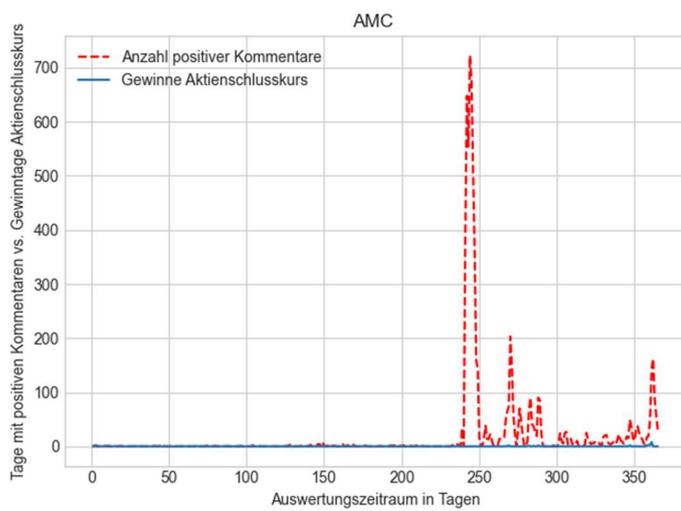
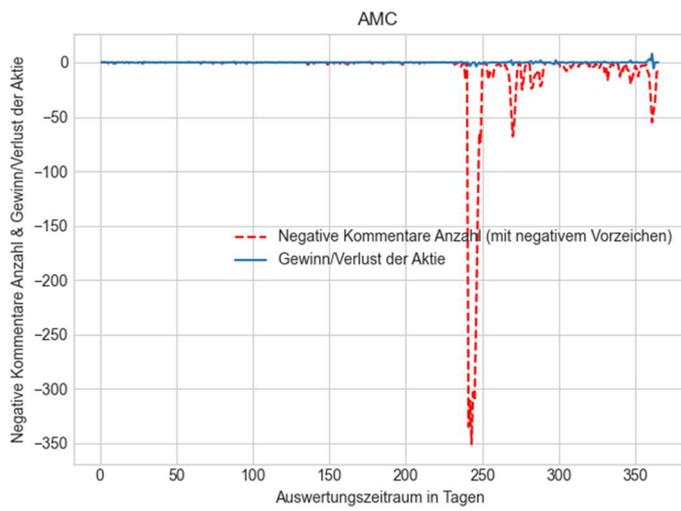
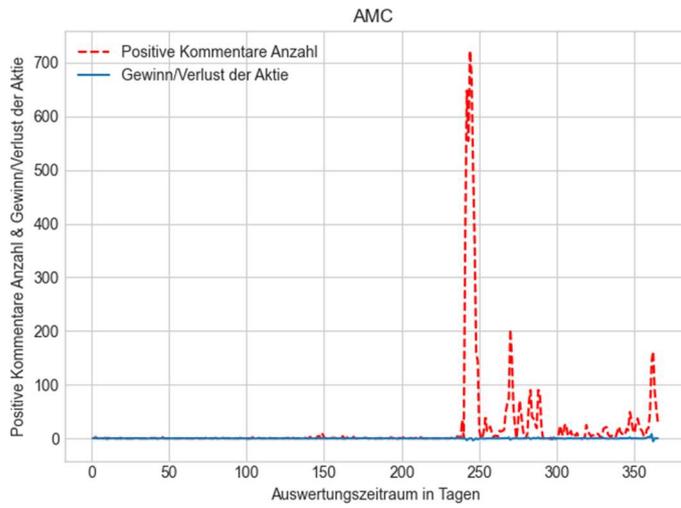
Unternehmen (Aktienkürzel)	Aktienschlusskurs zu Kommentaranzahlkurs		Gewinn/Verluste Aktienkurs zu Differenz Unternehmenserwähnungen zum Vortag	
	Peak	Peak-Wert	Peak	Peak-Wert
AAL	363*	0,428	260*	1,204
AMC	0	0,333	353*	-0,162
BB	1	0,594	1	0,543
BBBY	1	0,344	1	0,382
EXPR	0	0,462	0	0,405
GME	1	0,424	1	0,507
KODK	0	0,652	0	0,746
MAC	364*	0,564	240*	0,240
NOK	176*	0,476	0	0,812
PLTR	364*	0,533	0	0,166
SPCE	105*	0,316	358*	2,030
TR	234*	0,623	2	0,310

*\*Peak-Werte mit derart hohen Zeitversätzen können nicht als eventuelle kausale Zusammenhänge zur jeweilig gegenübergestellten Größe gewertet werden, sondern sind als Anomalien, entstanden aus der zugrundeliegenden Datenbasis, zu werten. Aufgrund der Effizienz der Finanzmärkte können derartige „Lags“ systematisch als nicht relevant ausgeschlossen werden.*

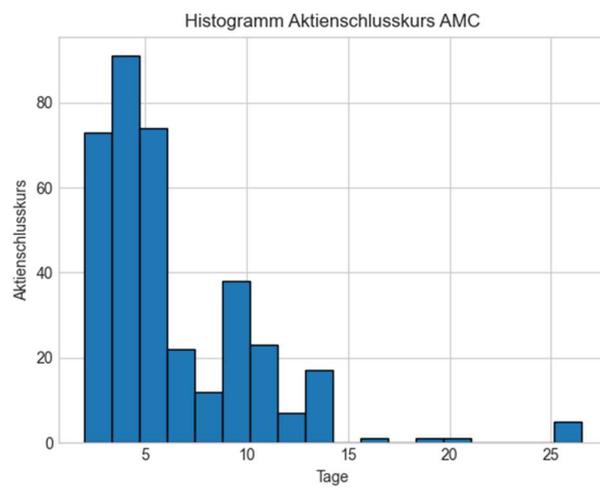
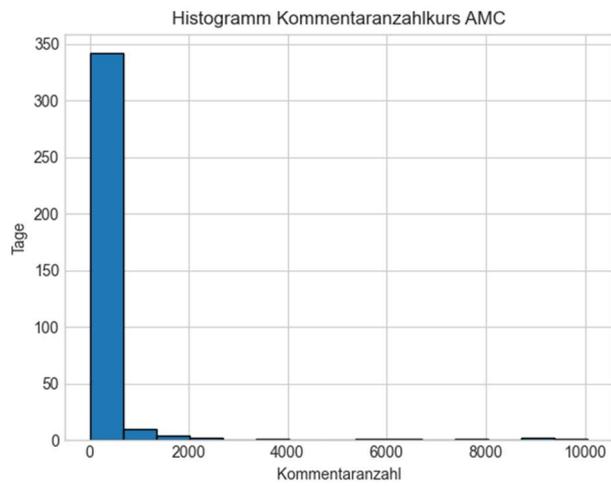
## 10.6 AMC Entertainment Holdings, Inc. (AMC)

### Kursvisualisierungen:



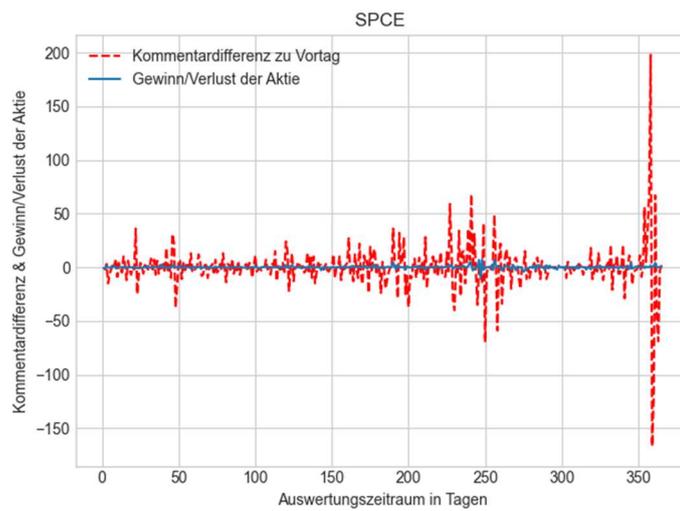
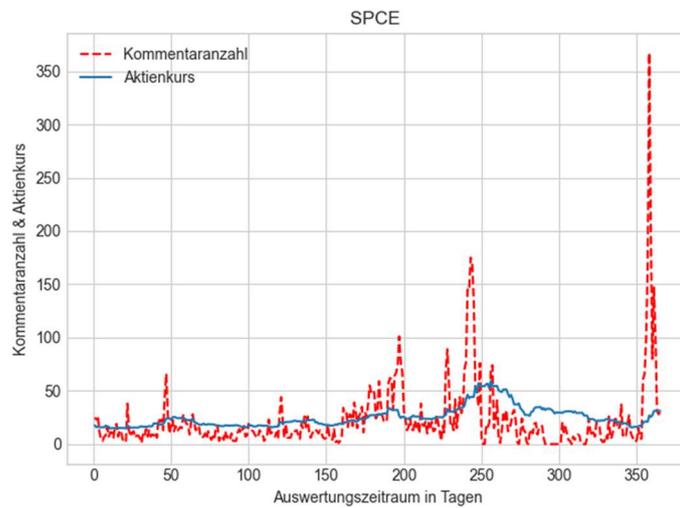


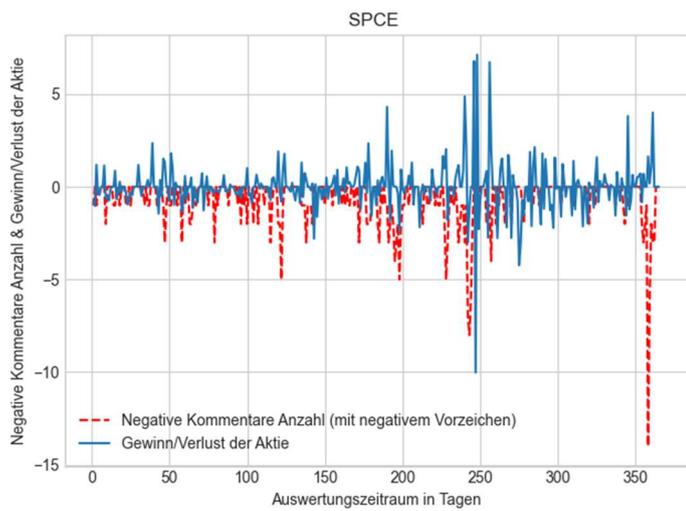
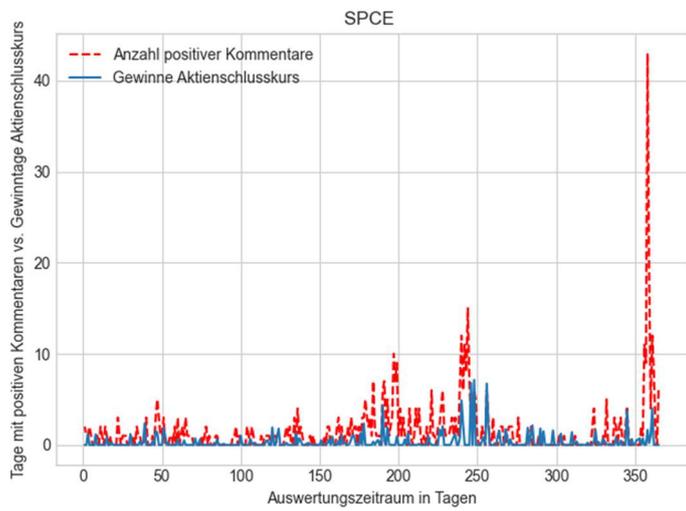
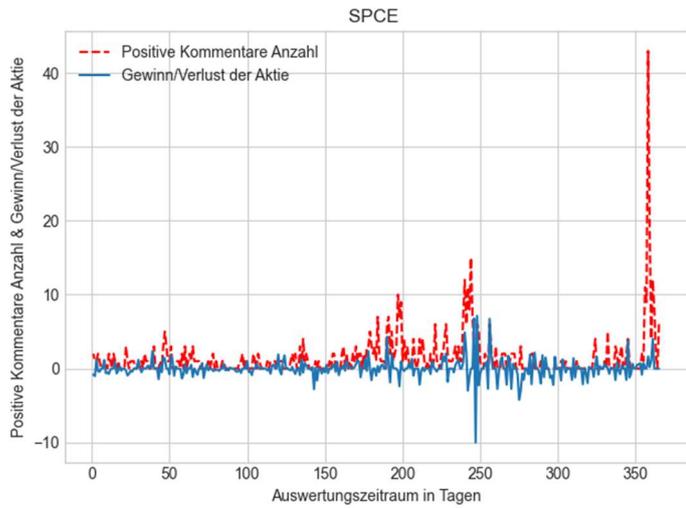
## Histogramme:



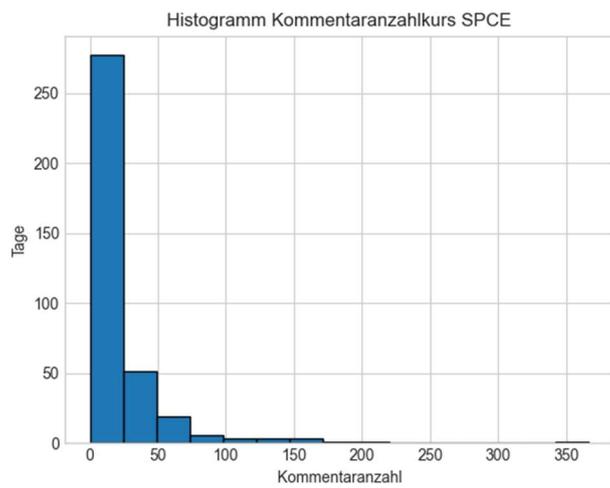
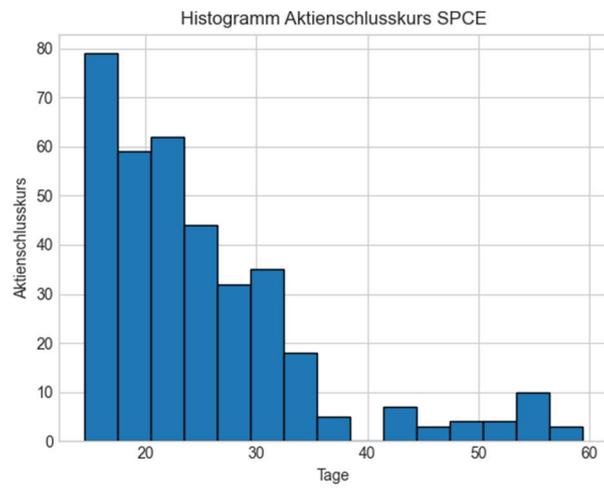
## 10.7 Virgin Galactic Holdings, Inc. (SPCE)

### Kursvisualisierungen:



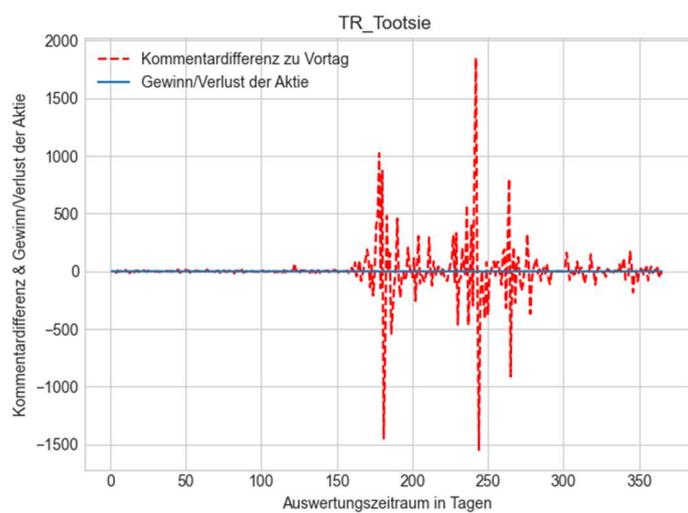
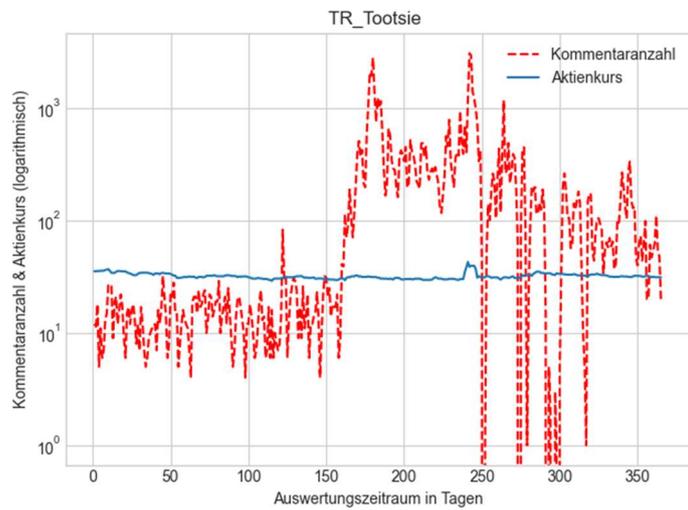


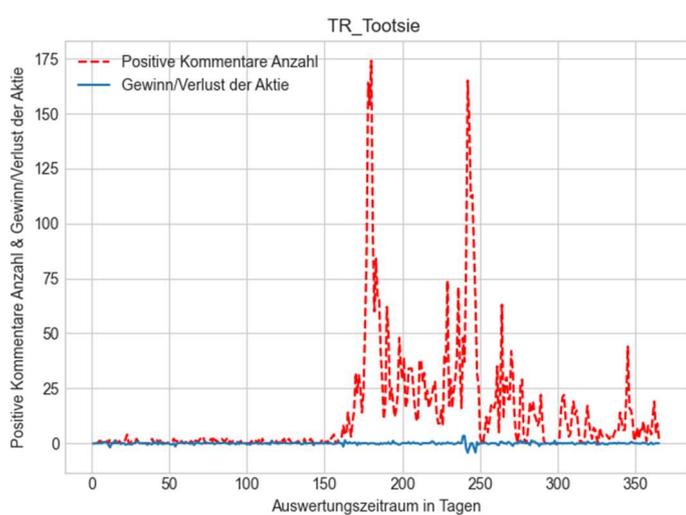
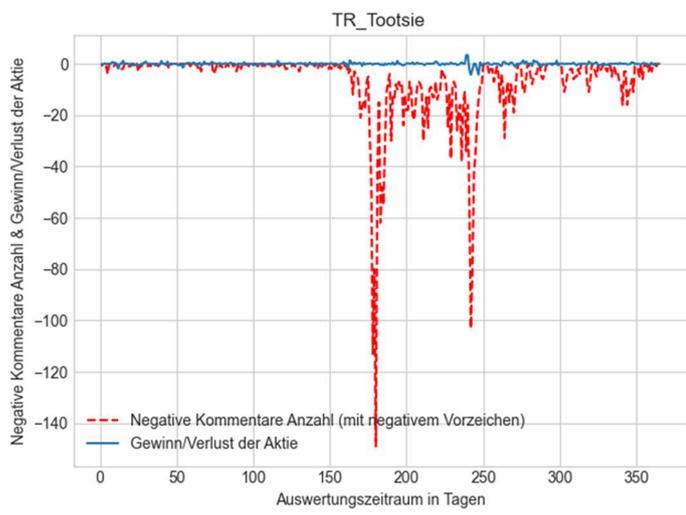
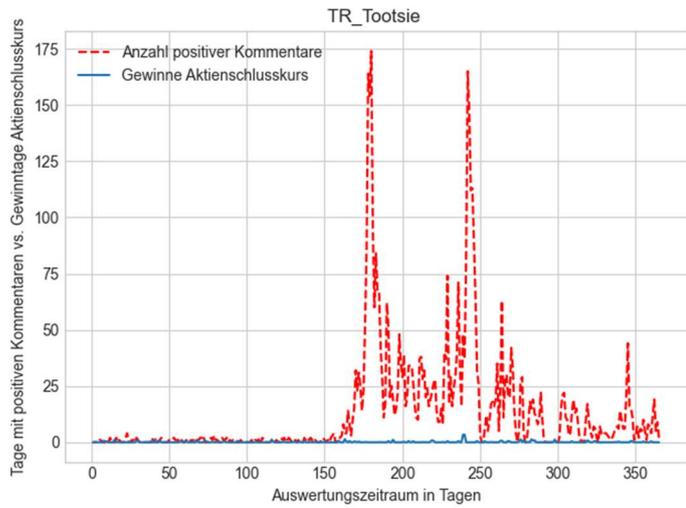
## Histogramme:



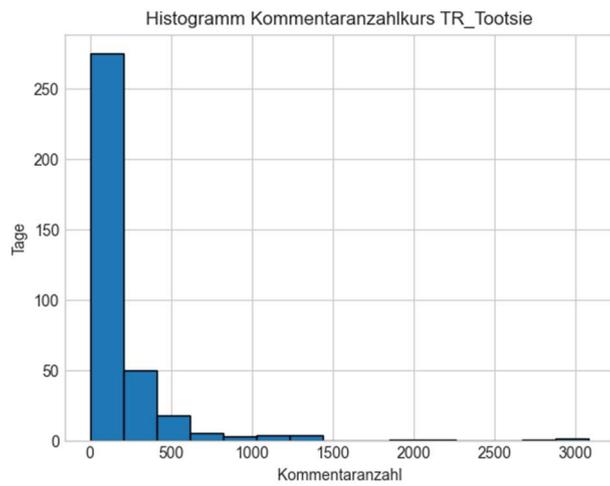
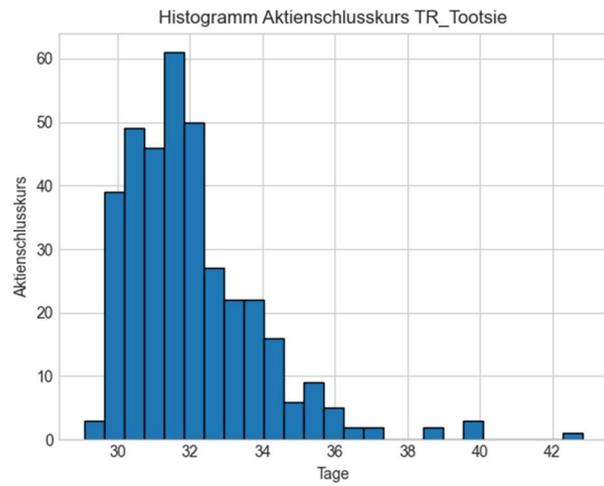
## 10.8 Tootsie Roll Industries, Inc. (TR)

### Kursvisualisierungen:



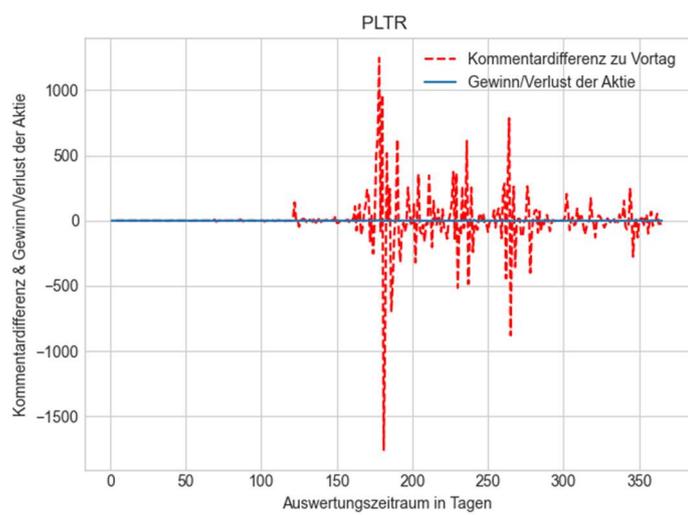
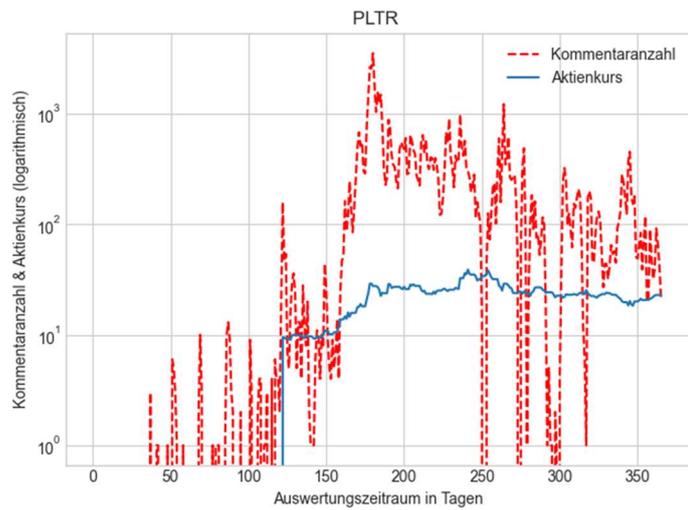


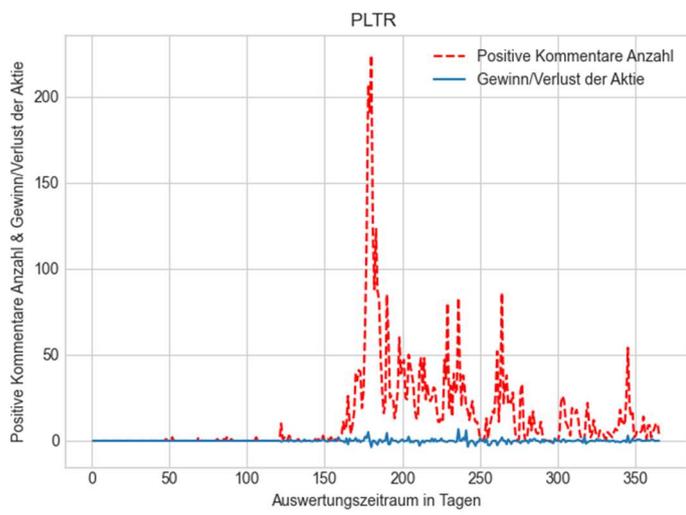
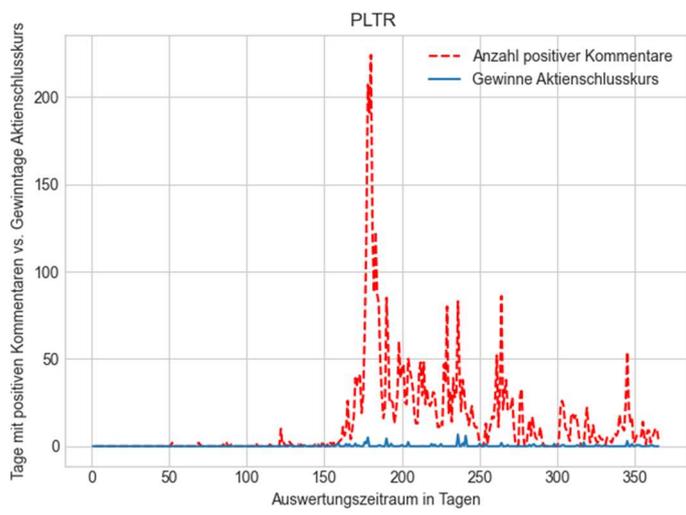
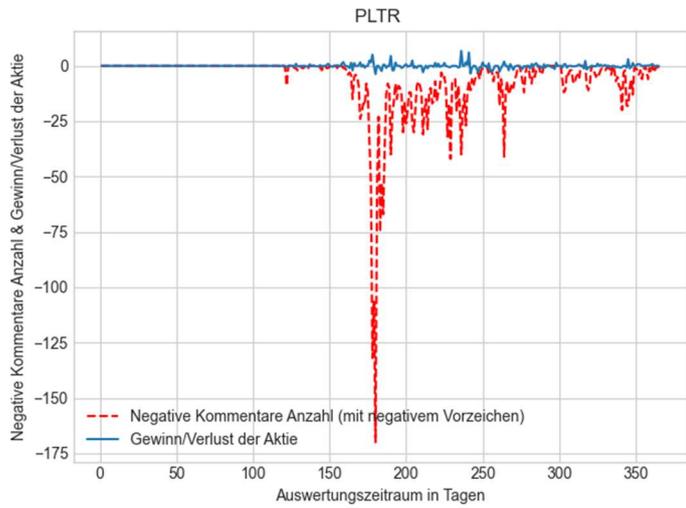
## Histogramme:



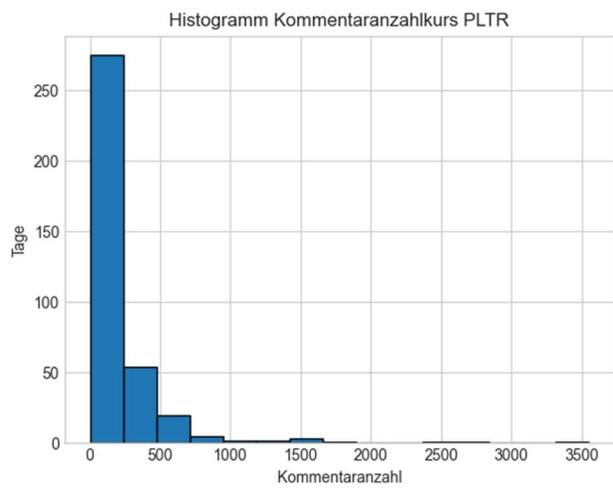
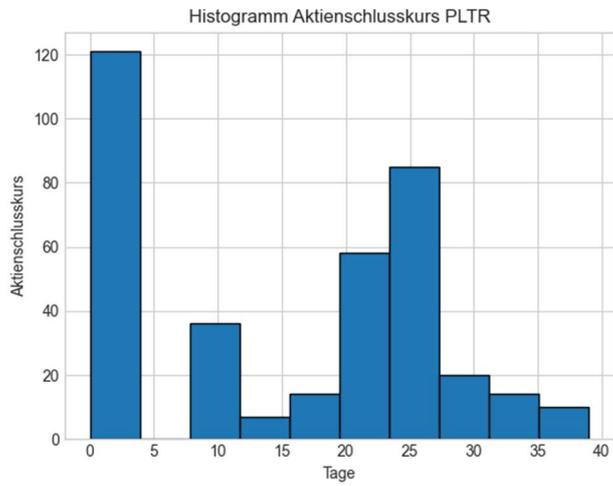
## 10.9 Palantir Technologies Inc. (PLTR)

### Kursvisualisierungen:



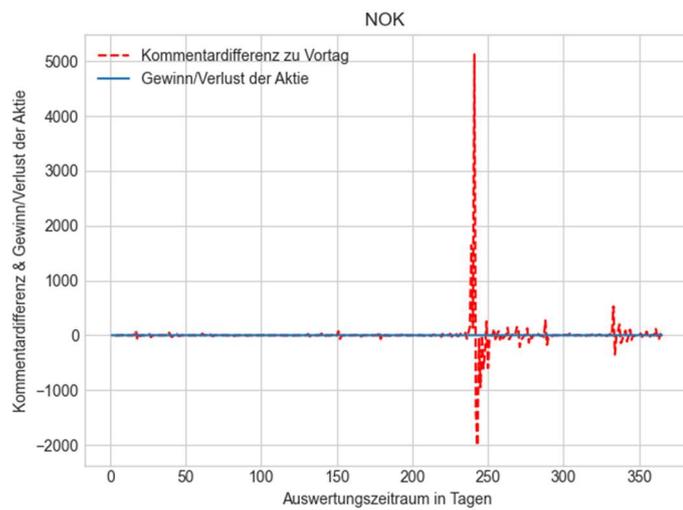
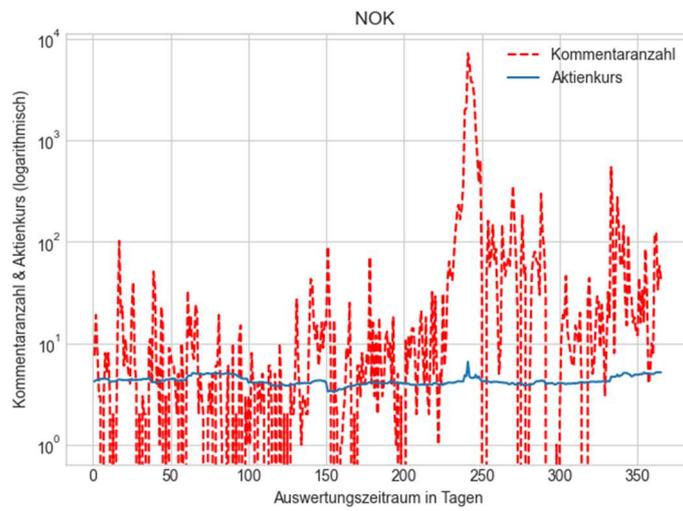


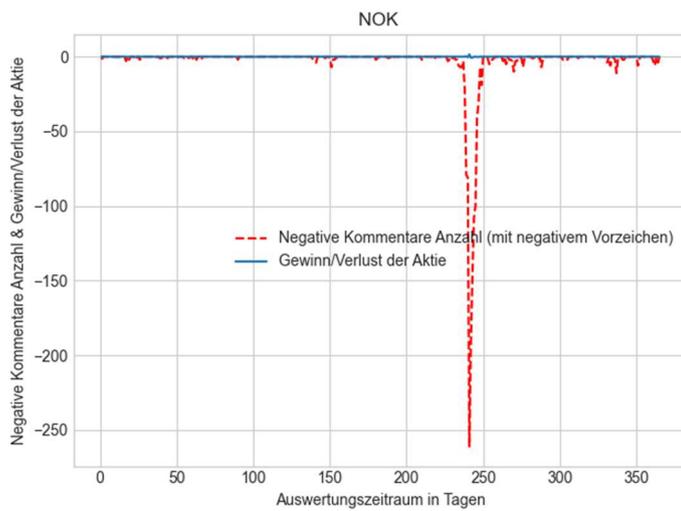
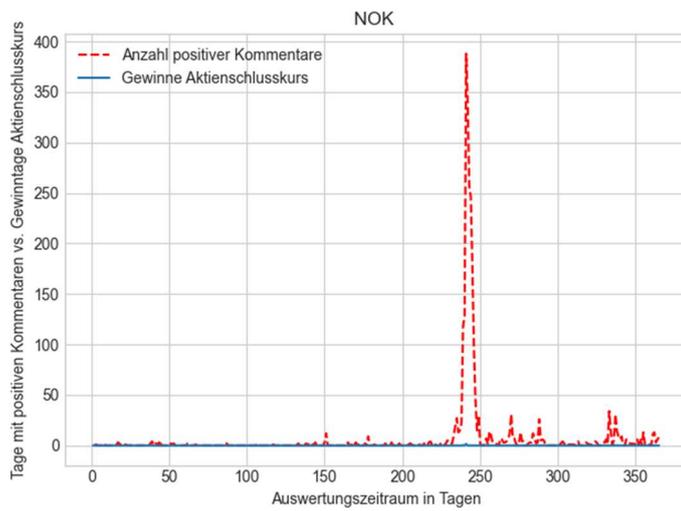
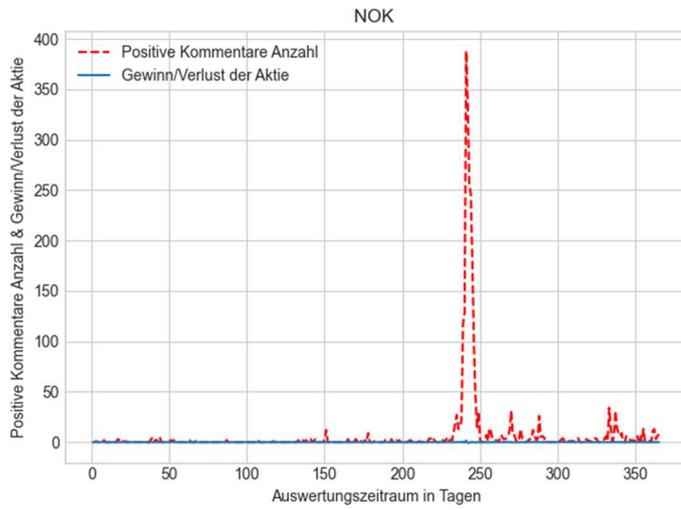
## Histogramme:



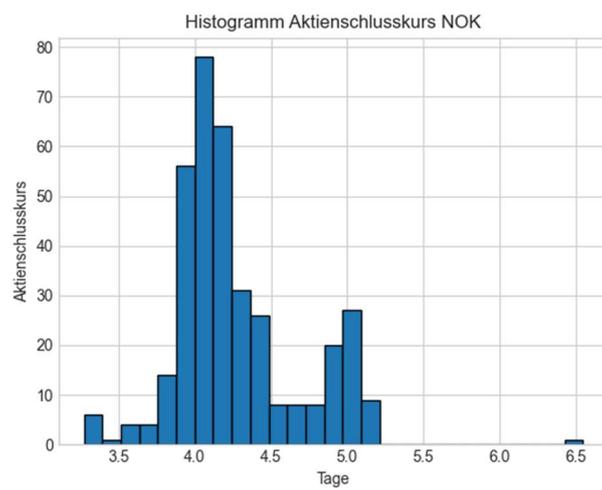
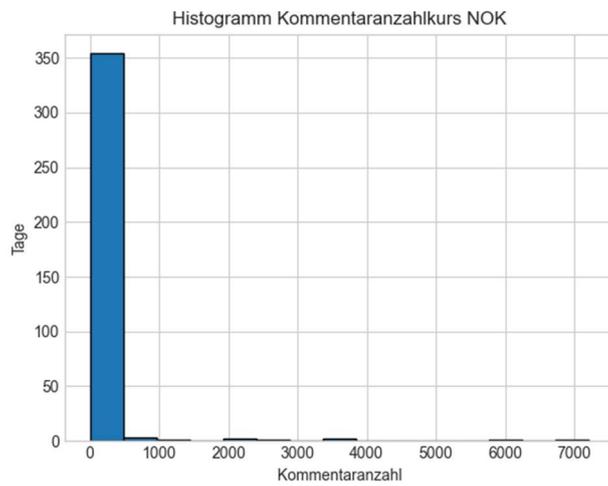
## 10.10 Nokia Oyj (NOK)

Kursvisualisierungen:



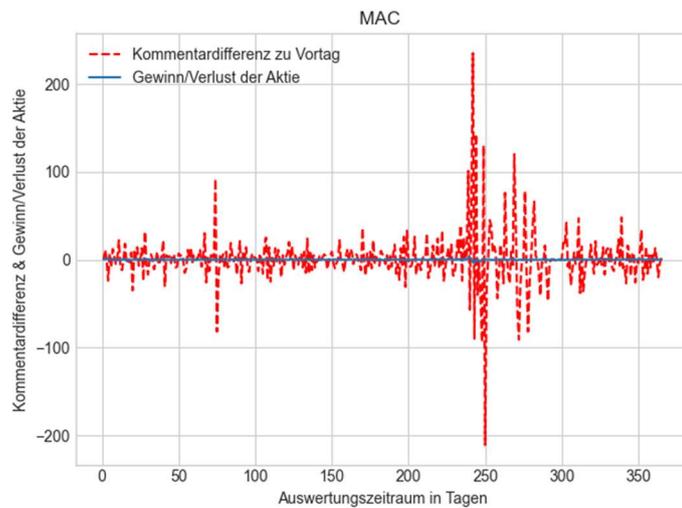
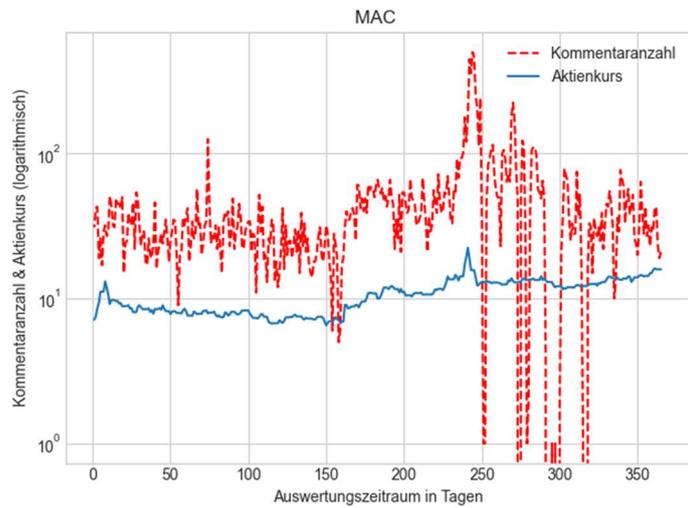


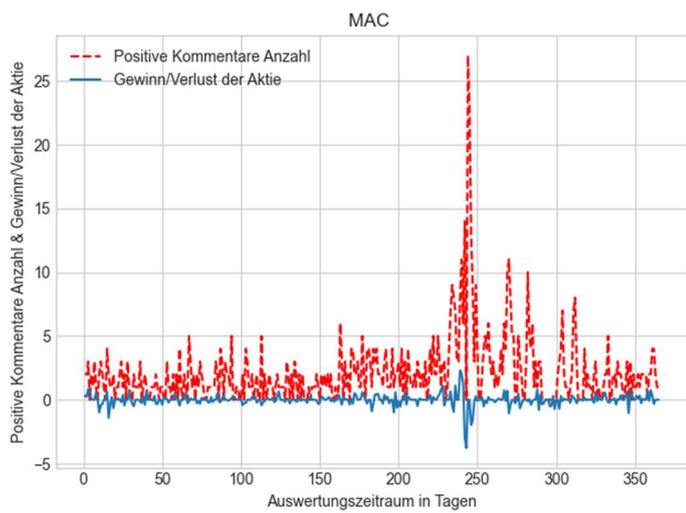
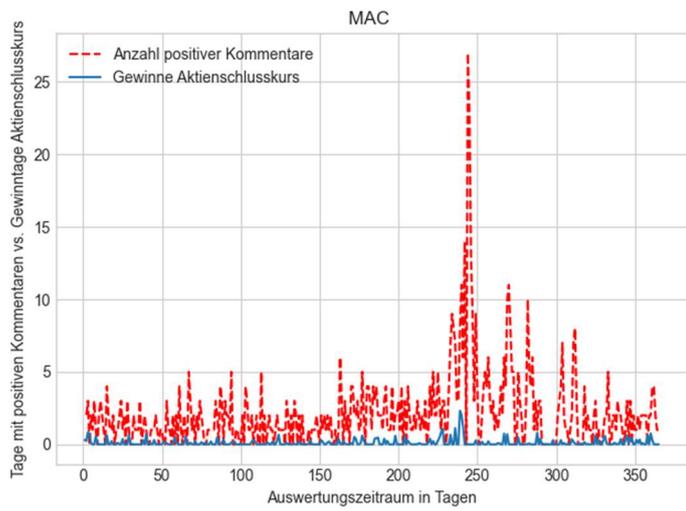
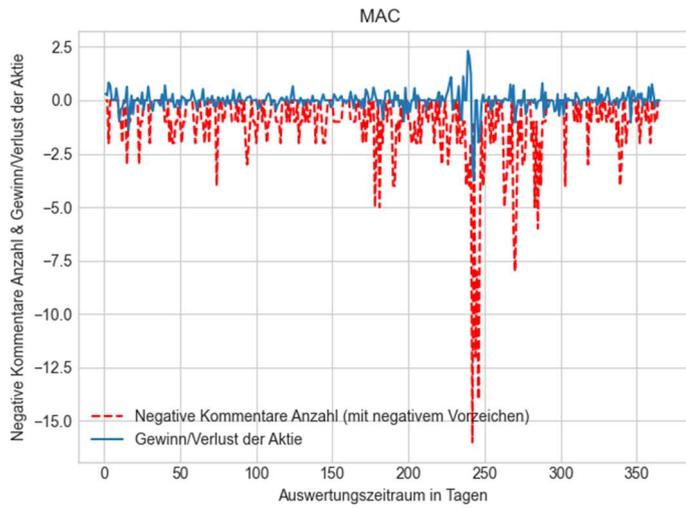
## Histogramm:



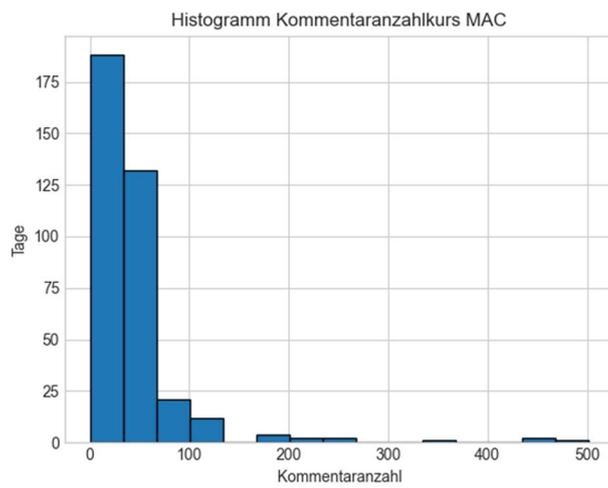
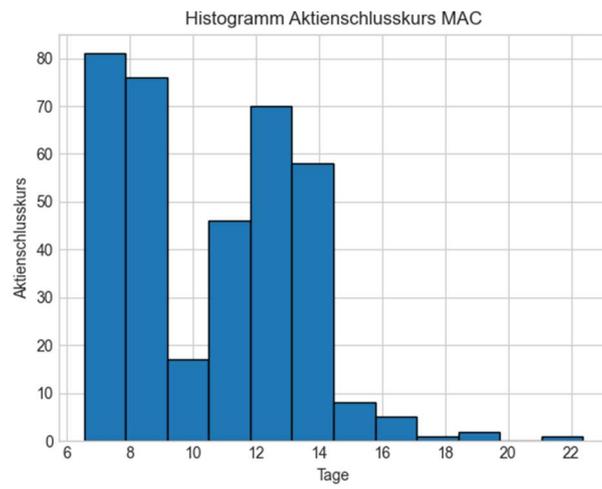
## 10.11 The Macerich Company (MAC)

### Kursvisualisierungen:



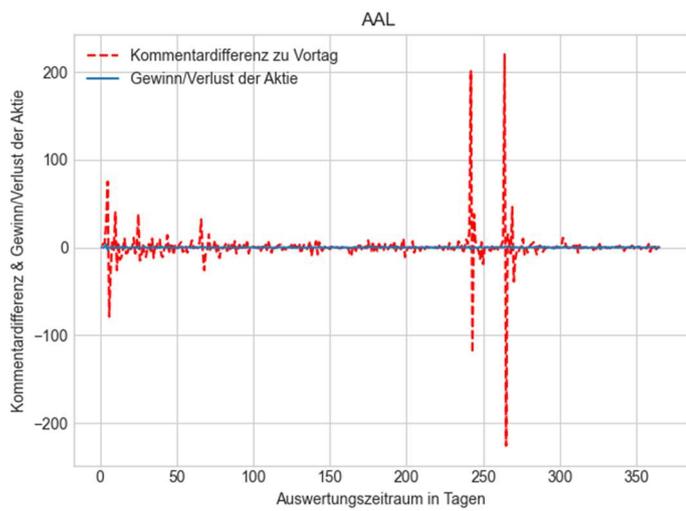
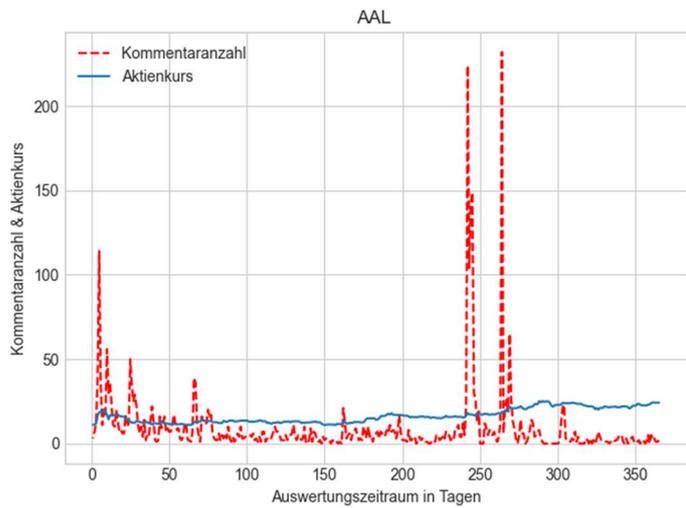


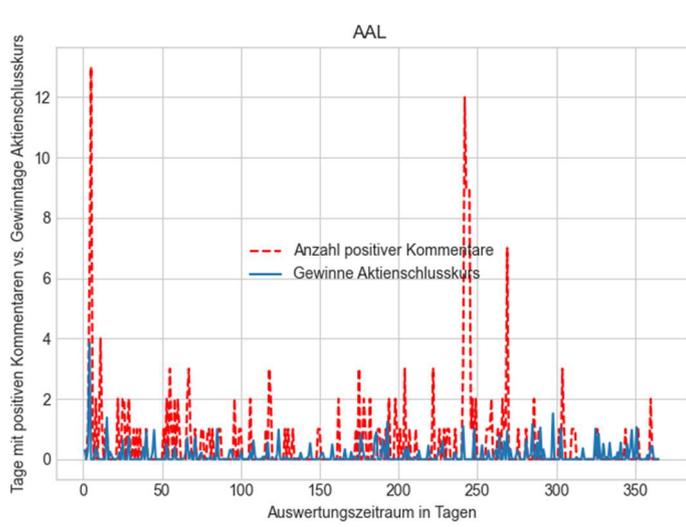
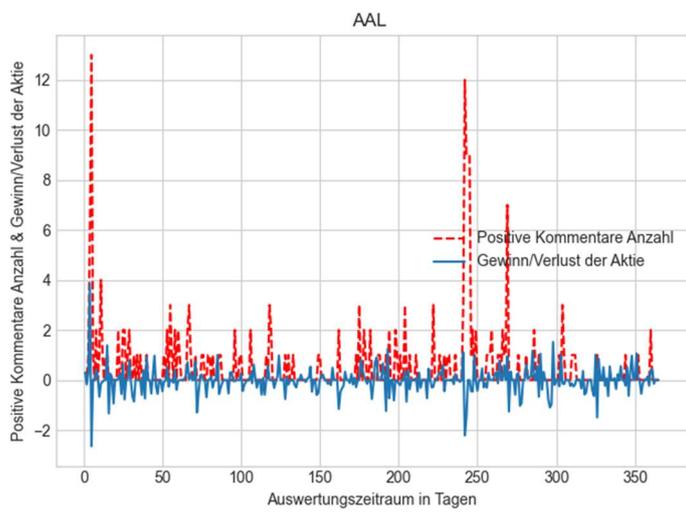
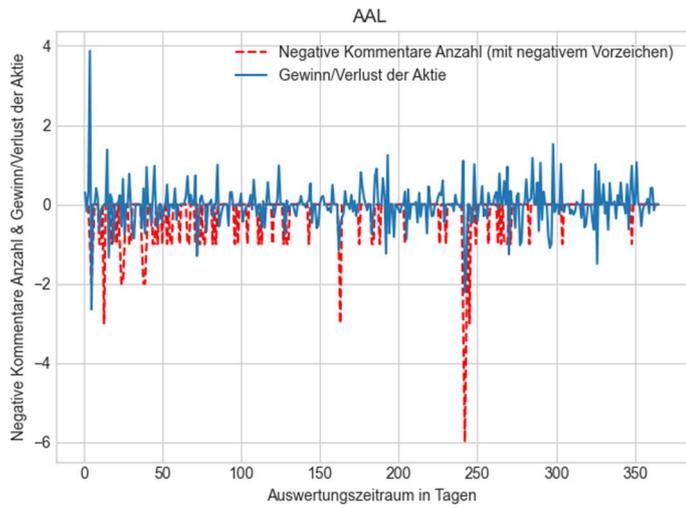
## Histogramme:



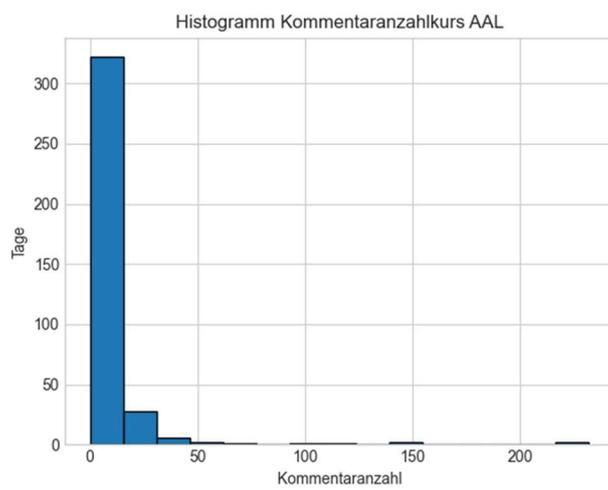
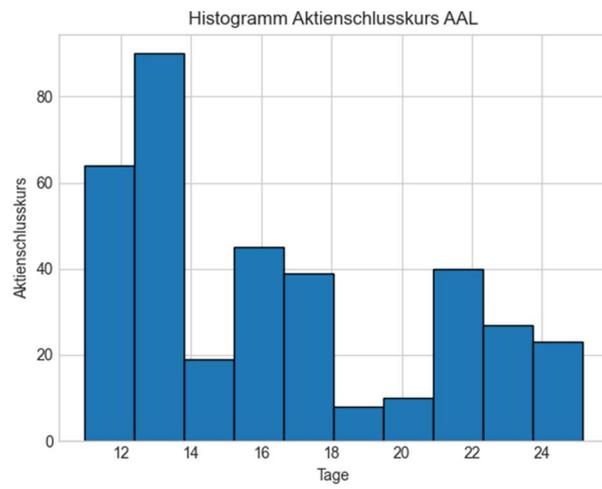
## 10.12 American Airlines Group Inc. (AAL)

### Kursvisualisierungen:



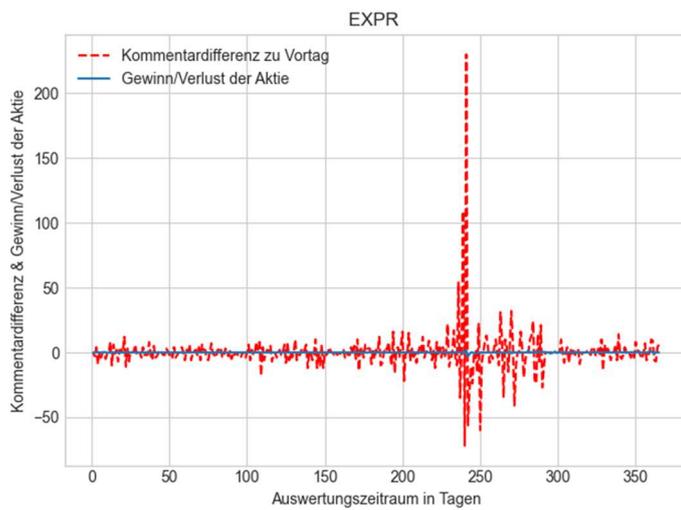
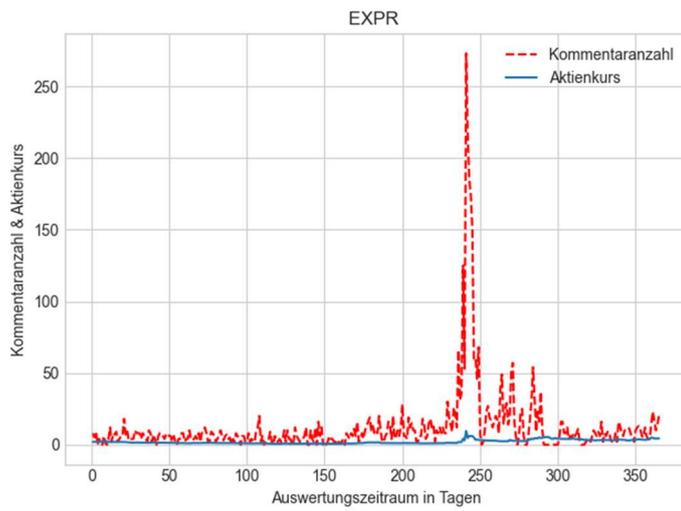


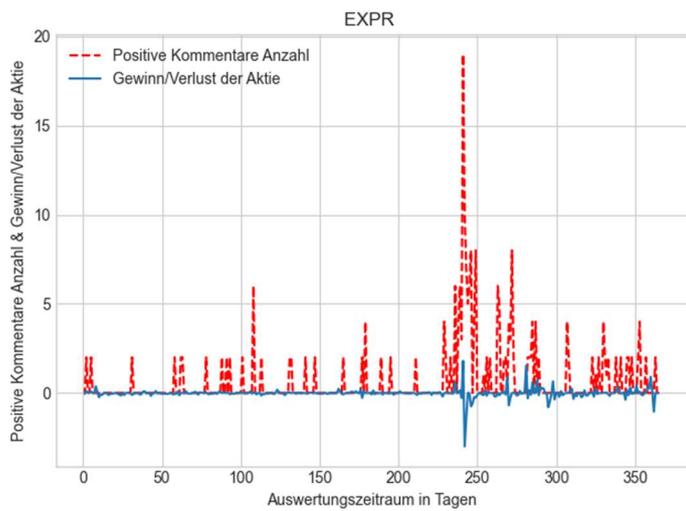
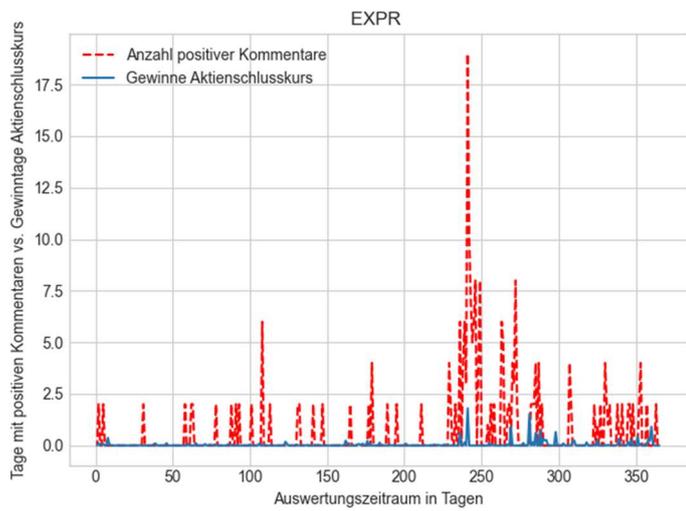
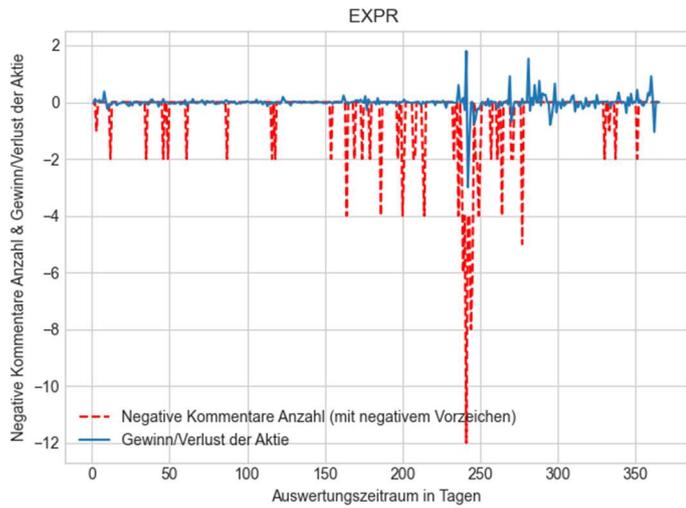
## Histogramm:



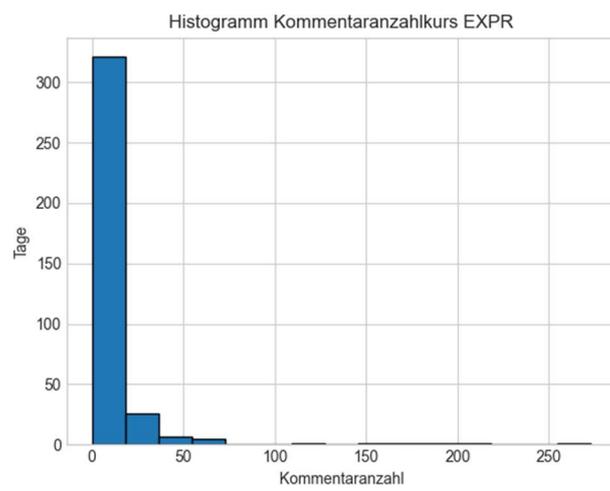
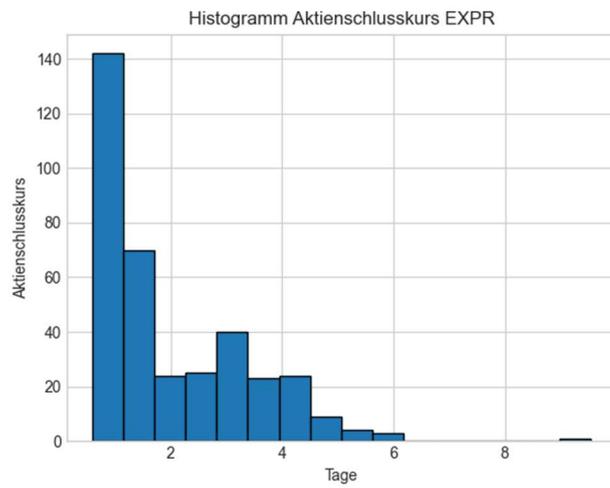
## 10.13 Express, Inc. (EXPR)

### Kursvisualisierungen:



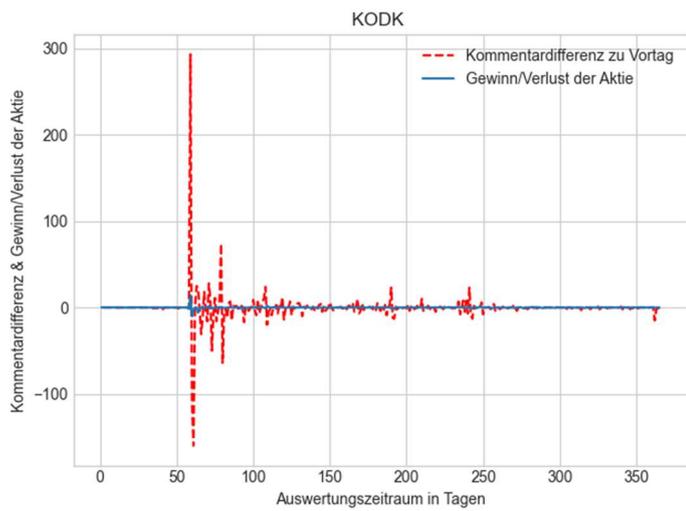
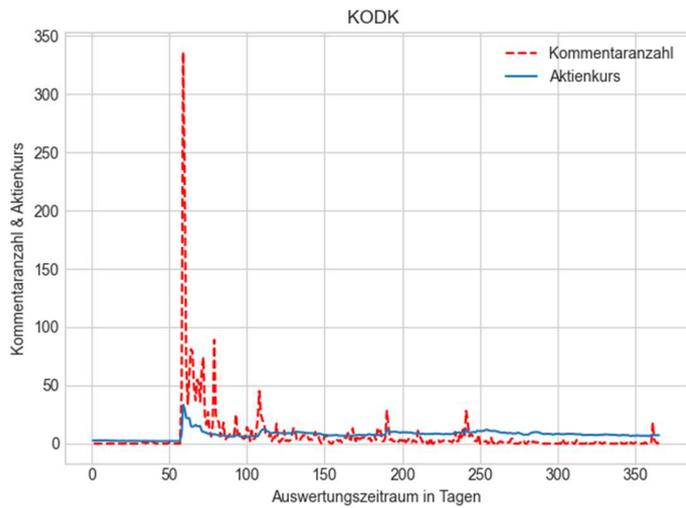


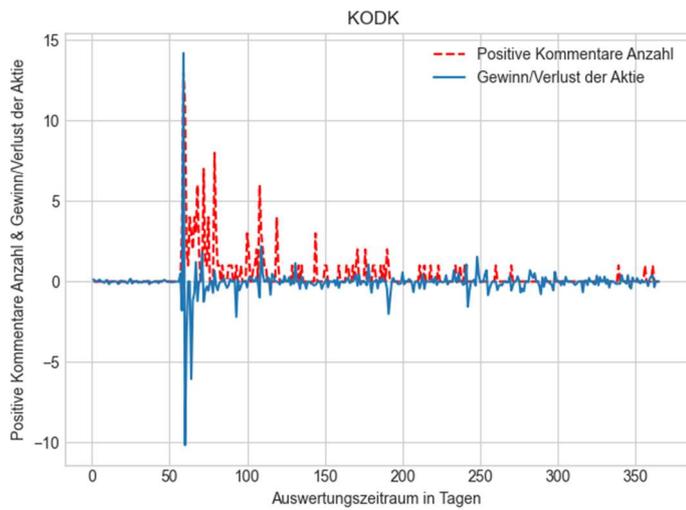
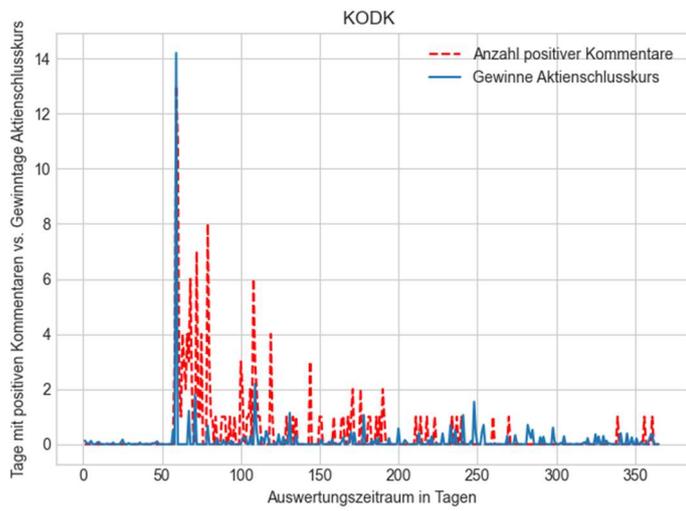
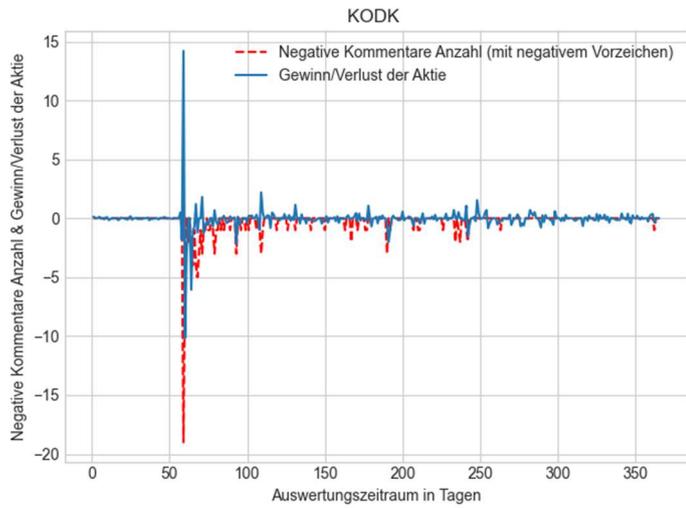
## Histogramme:



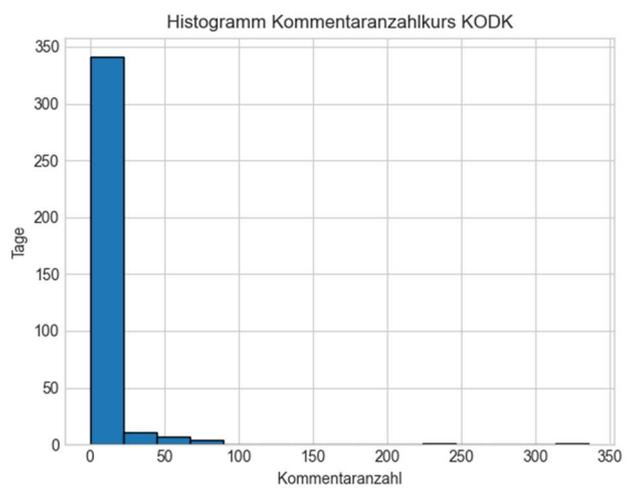
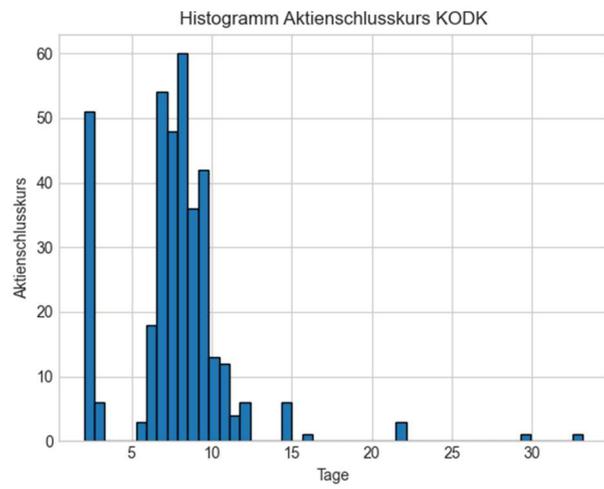
## 10.14 Eastman Kodak Company (KODK)

### Kursvisualisierungen:



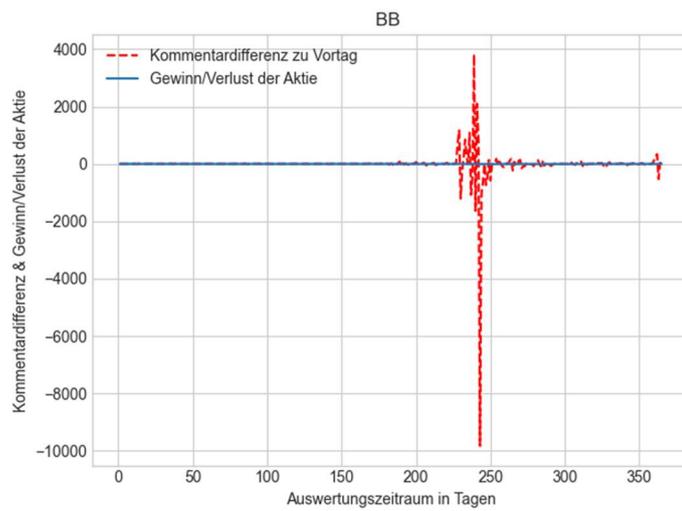
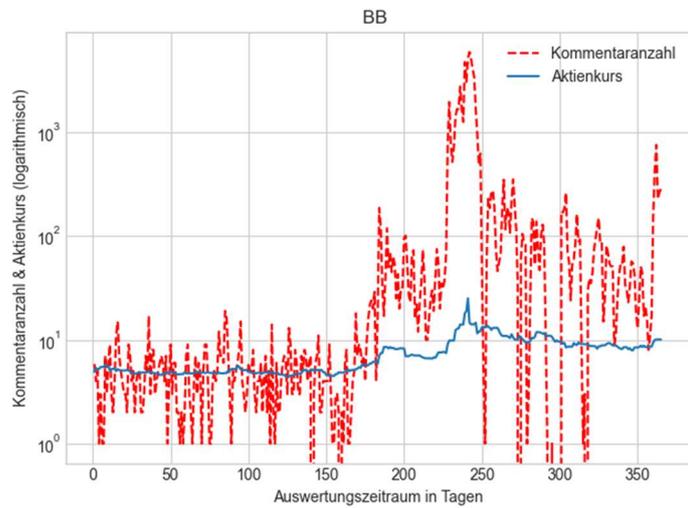


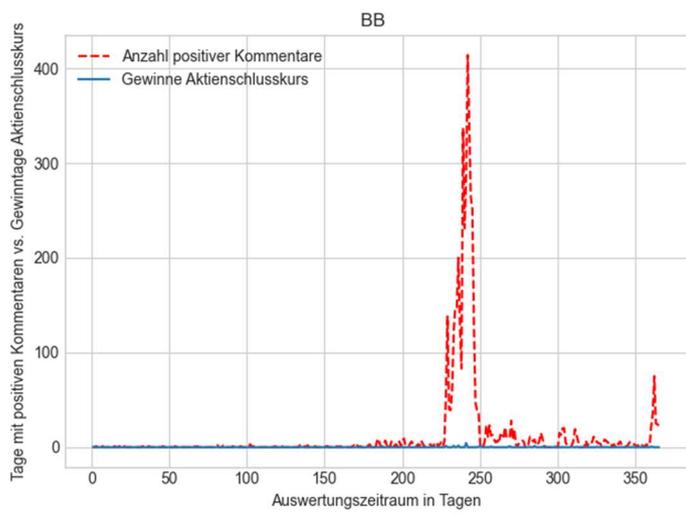
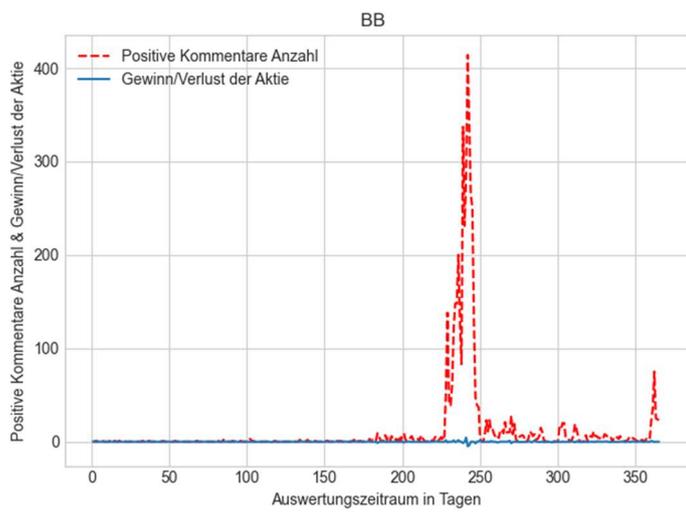
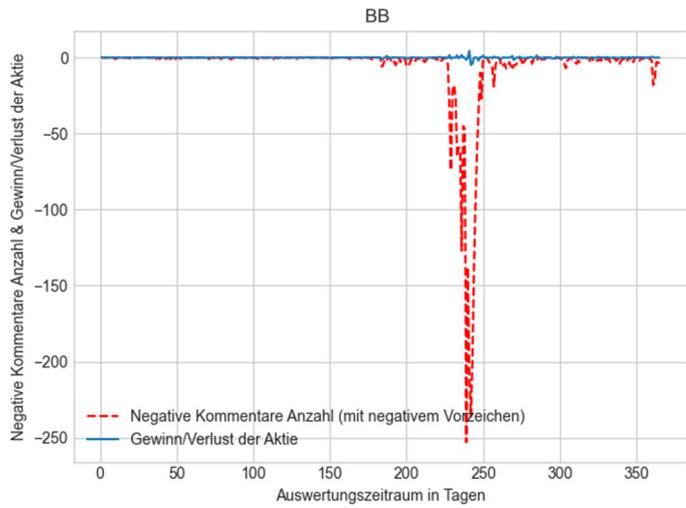
## Histogramme:



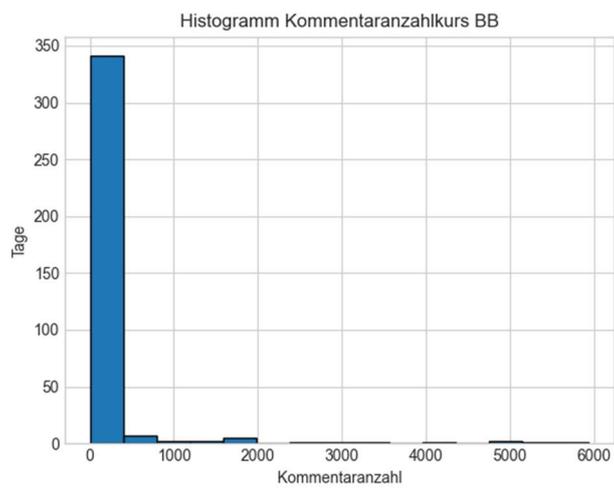
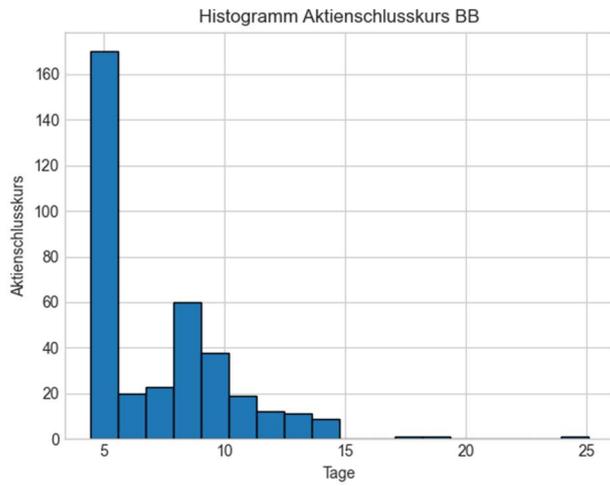
## 10.15 BlackBerry Limited (BB)

### Kursvisualisierungen:



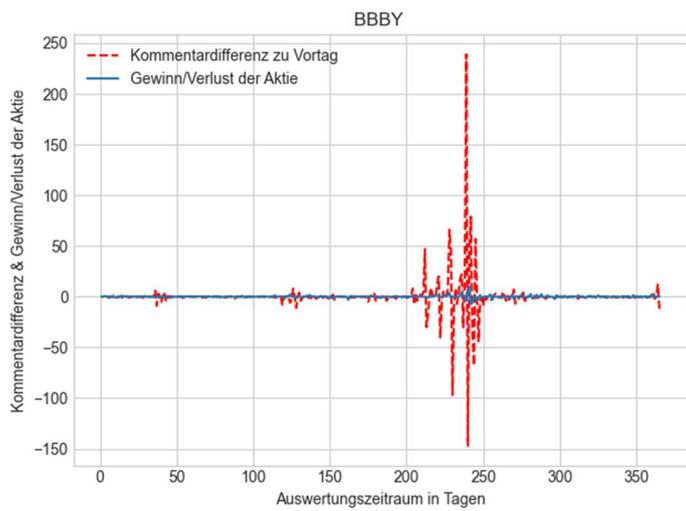
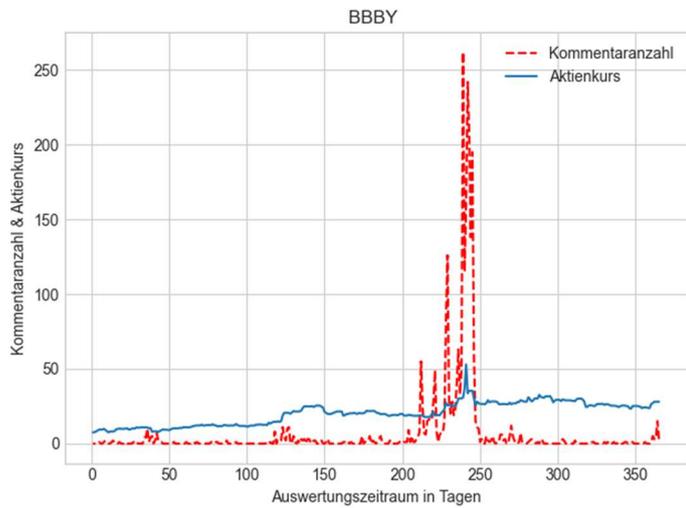


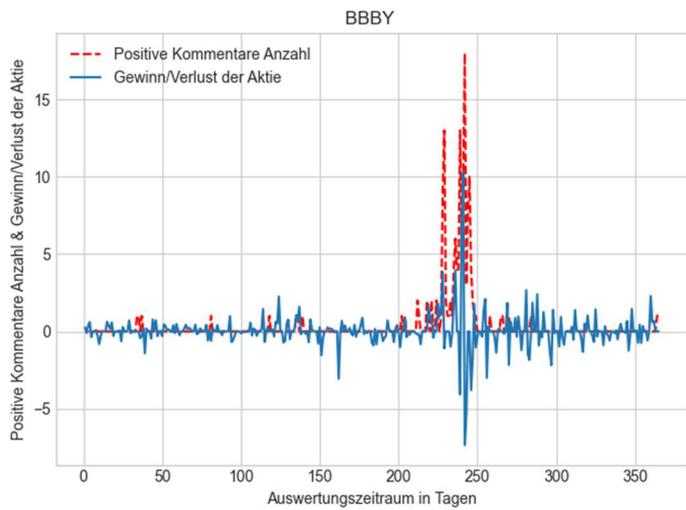
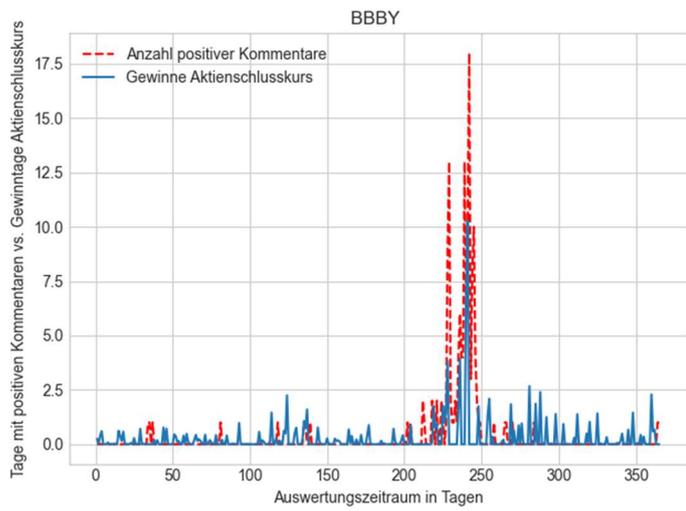
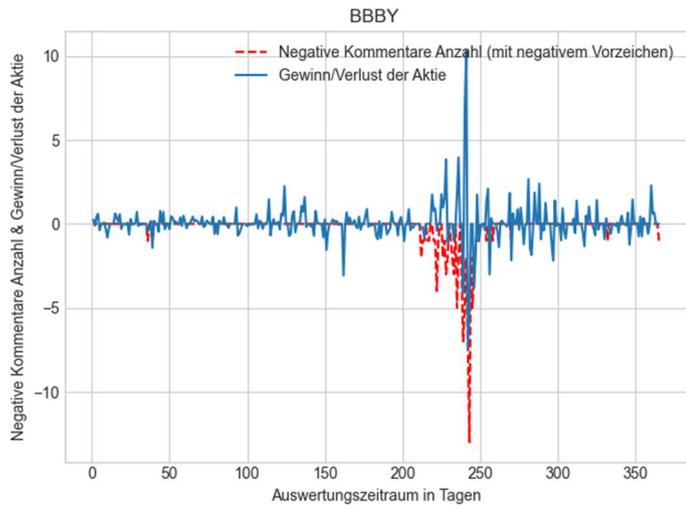
## Histogramme:



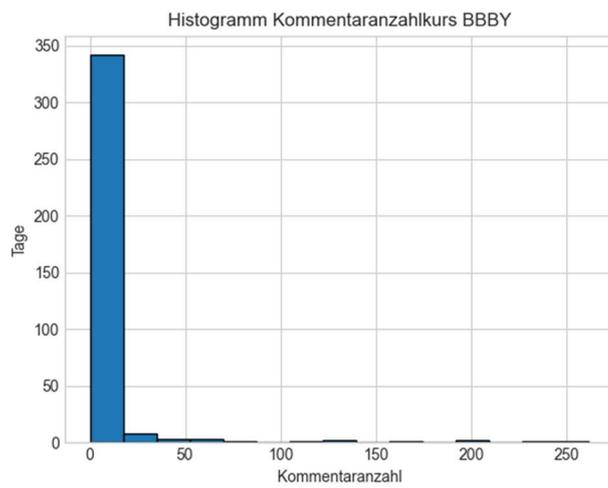
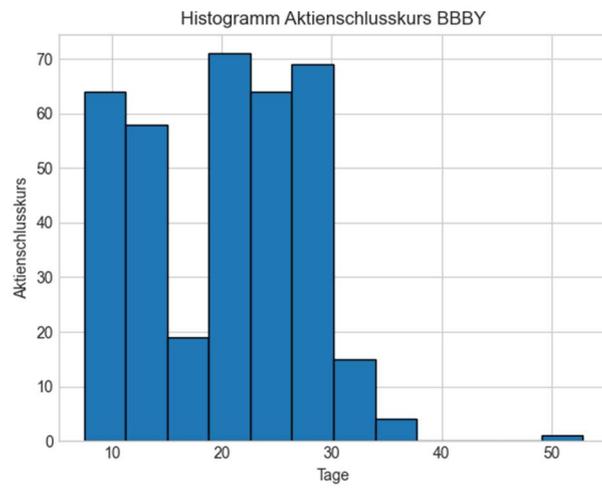
## 10.16 Bed Bath & Beyond Inc. (BBBY)

### Kursvisualisierungen:



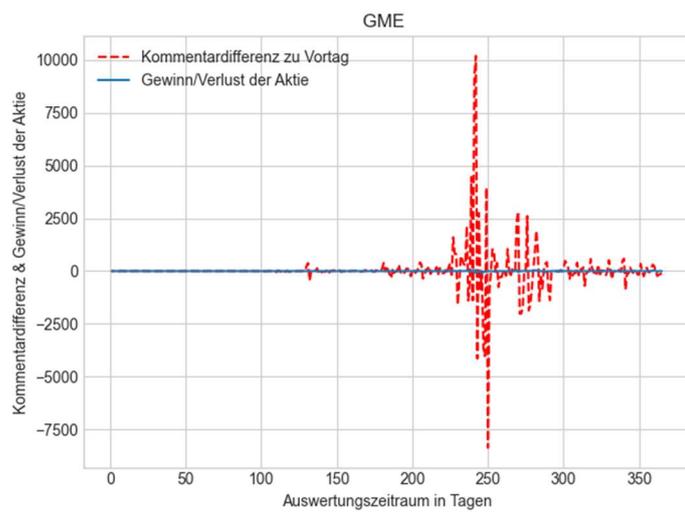
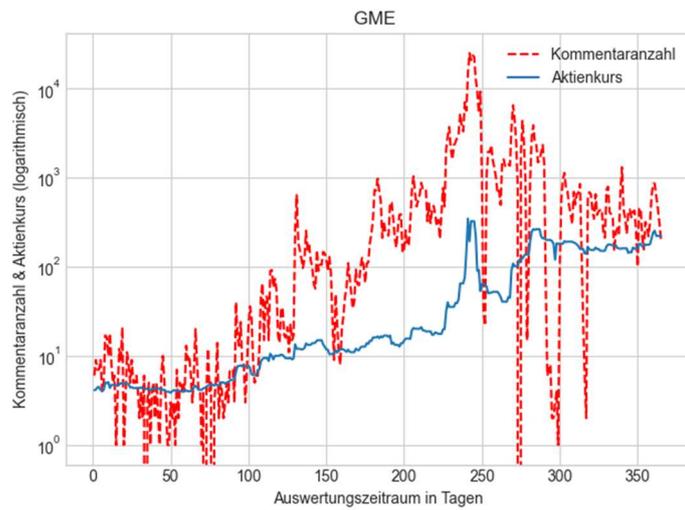


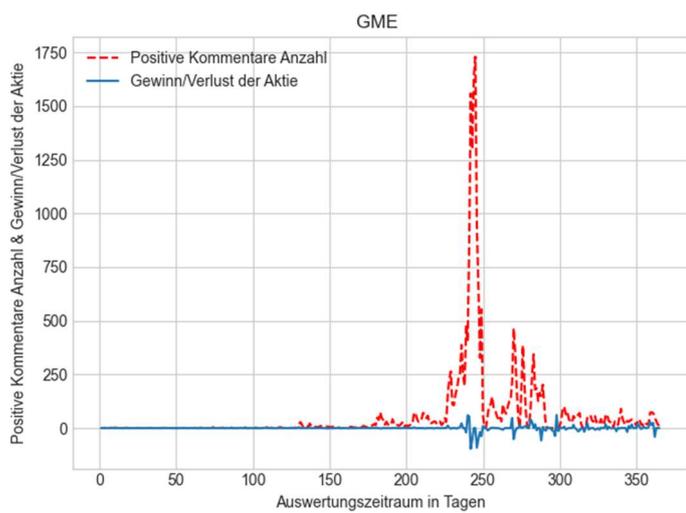
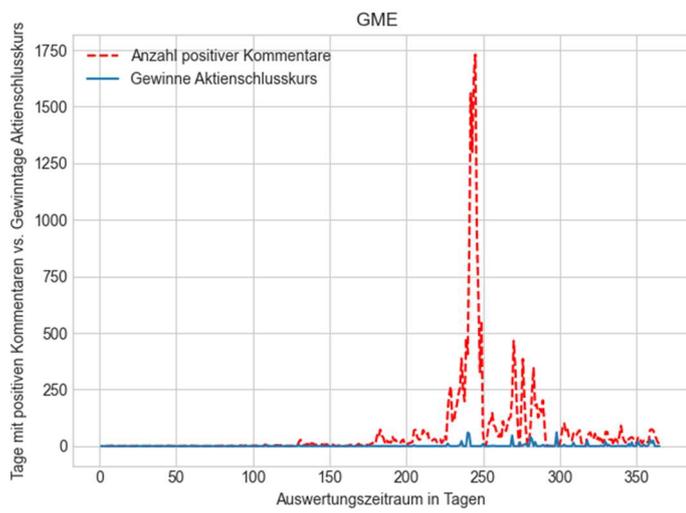
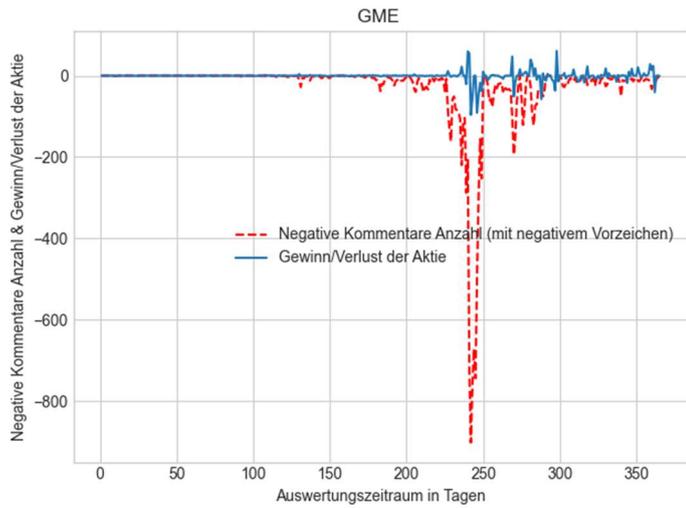
## Histogramme:



## 10.17 GameStop Inc. (GME)

### Kursvisualisierungen:





## Histogramme:

