

**Big Data in der Chefetage**  
**Der Einfluss von Datenanalyse auf den  
strategischen Entscheidungsprozess und den  
Unternehmenserfolg**

Masterarbeit

im Studiengang:

M.A. Unternehmensführung – Executive Management

an der FH Wien der WKW

Vorgelegt von:

Matthias Wagenstaller

Adresse:  
Am Sonnwend 18  
94577 Winzer  
Deutschland

# Inhalt

Abstract: .....	4
1. Die Entstehung von Big Data .....	6
2. Big Data: Kriterien, Ausprägungen und Folgen .....	9
2.1. Begriffsdefinition und Kriterien für Big Data .....	9
2.1.1. Velocity, Variety und Volume als Grundlage für die Wertschöpfung: .....	9
2.2. Radikale Innovation oder inkrementelle Veränderung? .....	13
2.2.1. Von hypothesengetriebenen zu datengetriebenen Ansätzen .....	14
2.2.2 Von der deskriptiven zur präskriptiven Datennutzung .....	15
3. Theoretische Verortung im strategischen Management .....	17
3.1. Market-Based-View .....	19
3.2. Ressource-Based-View .....	22
3.3. Grundlagen der Entscheidungstheorie .....	25
4. Die Entwicklung des Modells und Operationalisierung .....	28
4.1. Unabhängige Variablen .....	28
4.1.1. Qualität der Datenanalyse .....	28
4.1.2. Quantität der Datenanalyse .....	30
4.2. Weitere Variablen .....	31
4.2.1. Unternehmensgröße .....	31
4.2.2. Branche .....	34
4.2.3. Jurisdiktion und Datenschutzbestimmungen .....	35
4.2.4. Eigentümerstruktur .....	36
4.3. Abhängige Variable - Unternehmenserfolg .....	39
4.4. Das Modell .....	41
5. Methoden der Datenerhebung und Vorgehensweise .....	43
5.1. Kontaktaufnahme und Erhebungsmethodik .....	43
6. Erkenntnisse und Interpretation .....	52
6.1 Generelle Informationen über die teilnehmenden Unternehmen .....	53
6.2. Informationen zur Datenanalyse in den Unternehmen .....	56
6.3. Finanzinformationen der Unternehmen .....	72
7. Hypothesentests und Induktive Verfahren .....	76

7.1 H <sub>1</sub> .....	77
7.2 H <sub>2</sub> .....	78
7.3 H <sub>3</sub> .....	80
7.4 H <sub>4</sub> .....	81
7.5 Gesamtmodell.....	82
8. Zusammenfassende Analyse und Ausblick.....	83
Literaturverzeichnis:.....	86
Anhang 1: Fragebogen .....	89
Anhang 2: Variablenübersicht.....	100
Anhang 3: Begleitschreiben Fragebogenaussendung.....	106
Anhang 4 .....	108
Anhang 5 .....	109
Anhang 6 .....	110
Anhang 7 .....	111
Anhang 8 .....	117
Anhang 9 .....	119
Anhang 10 .....	120
Anhang 11 .....	123
Anhang 12 .....	124
Anhang 13 .....	125
Anhang 14 .....	126
Anhang 15 .....	127
Anhang 16 .....	128
Anhang 17 .....	130

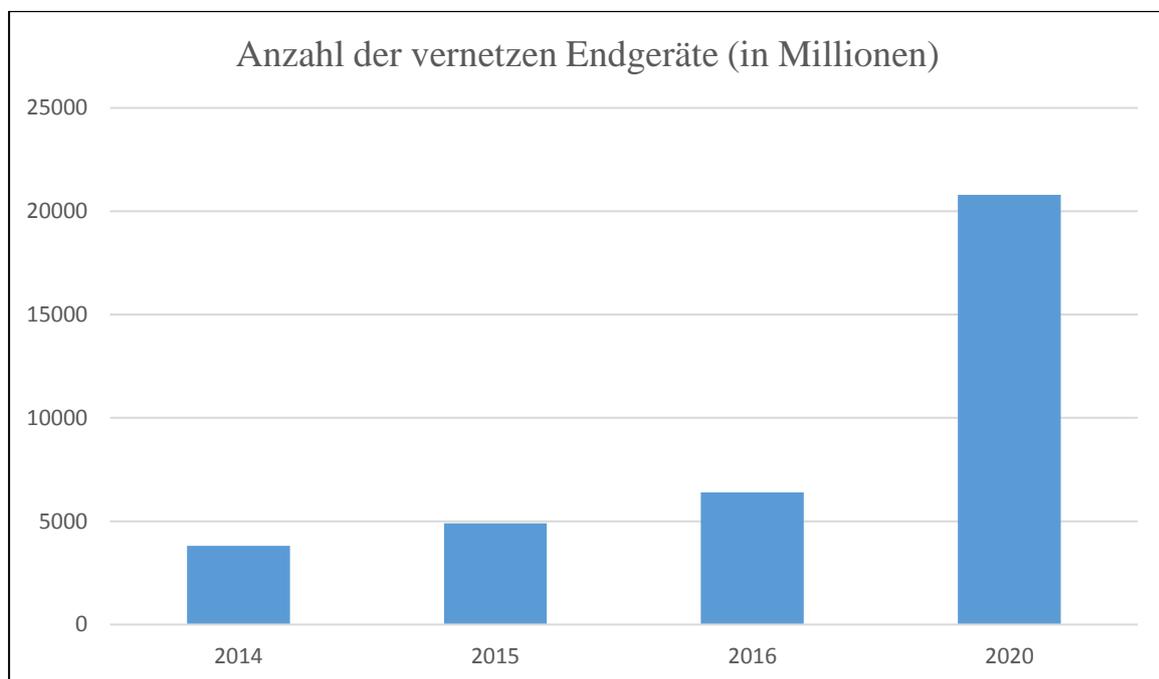
## Abstract:

Durch die technologische Entwicklung und der damit einhergehenden Vernetzung all unserer Lebensbereiche entstehen mittlerweile enorme Mengen an Daten. Diese können mit statistischen Methoden ausgewertet werden, was erstaunlich zuverlässige Vorhersagen über die Zukunft erlaubt. Immer häufiger werden solche Methoden von Unternehmen genutzt, um ihre Prozesse zu optimieren. Diese Arbeit soll den Einfluss von Big Data Methoden auf den strategischen Entscheidungsprozess in Unternehmen untersuchen, wobei primär überprüft werden soll, wie sich diese auf das Finanzergebnis der Unternehmen auswirken. Hierfür wurde, aufbauend auf Theorien des strategischen Managements, ein Fragebogen entworfen, der an ca. 13000 Unternehmen in der DACH-Region verschickt wurde. Bei der Auswertung der gewonnen Daten zeigte sich, dass sowohl die Branche, als auch die Eigentümerstruktur entgegen den Erwartungen keinen Einfluss auf die Nutzung von Big Data Technologien zu haben scheint. Eine Ausnahme hiervon stellt die Erkenntnis dar, dass öffentliche Unternehmen die Datenanalyse deutlich seltener nutzen. Die Hypothese, dass die Verbreitung solcher Technologien positiv von der Größe des Unternehmens beeinflusst wird, bestätigte sich, ebenso wie der positive Effekt verschiedener Werkzeuge auf den Unternehmenserfolg. Leider zeigte sich, dass das zugrundeliegende Datenmaterial nicht ausreicht um eine genaue Quantifizierung der Effekte vorzunehmen, was sich unter anderem in sehr großen Konfidenzintervallen widerspiegelt und für sämtliche berechneten Modelle gilt. Eine Erkenntnis die sich aus der deskriptiven Analyse ergab, war, dass Unternehmen bisher noch hauptsächlich auf altbekannte Werkzeuge und Datenquellen zurückgreifen. Mit diesen wird allerdings immer stärker versucht in den Bereich Big Data vorzustößen, um die hier ruhenden Potentiale zu heben. Dieses Vorgehen dürfte allerdings irgendwann an seine Grenze stoßen und es dürften in Zukunft größere Investitionen in die IT-Infrastruktur und qualifiziertes Personal notwendig sein, um die existierenden Erwartungen an die Effekte der Datenanalyse zu erfüllen.

Due to the technological development and the increasing interconnectedness of all areas of our live, huge amounts of data are created. These can be analysed statistically, creating amazingly accurate predictions about the future. Which are increasingly used by companies to optimize their processes. This paper wants to analyse the influence of such Big Data methods on strategic decision making in companies and will focus on the on the effects of these methods on the financial effects. Based on theories of strategic management, a questionnaire was created, which was sent to around 13000 companies in the DACH-region. The analysis of the gathered data showed, that neither branch nor equity structure have a significant effect on the use of Big Data technologies. Only public companies use methods of data analysis significantly less frequently. The hypothesis that company size has a positive effect on the existence of such technologies was proven, just like the positive effect of several Big Data Tools on company success. Unfortunately the database of this study was not sufficient enough to achieve a precise quantification of these effects, which is reflected in quite huge confidence intervals and concerns all used models. The finding of the descriptive analysis was, that companies are still using well-known tool and data sources for their analysis and try to use them to cope with the challenges of Big Data and to raise its potentials. But these attempts will be limited, so huge investment into IT-infrastructure and qualified personal will be needed in the near future to keep up with the expectations of Big Data.

## 1. Die Entstehung von Big Data

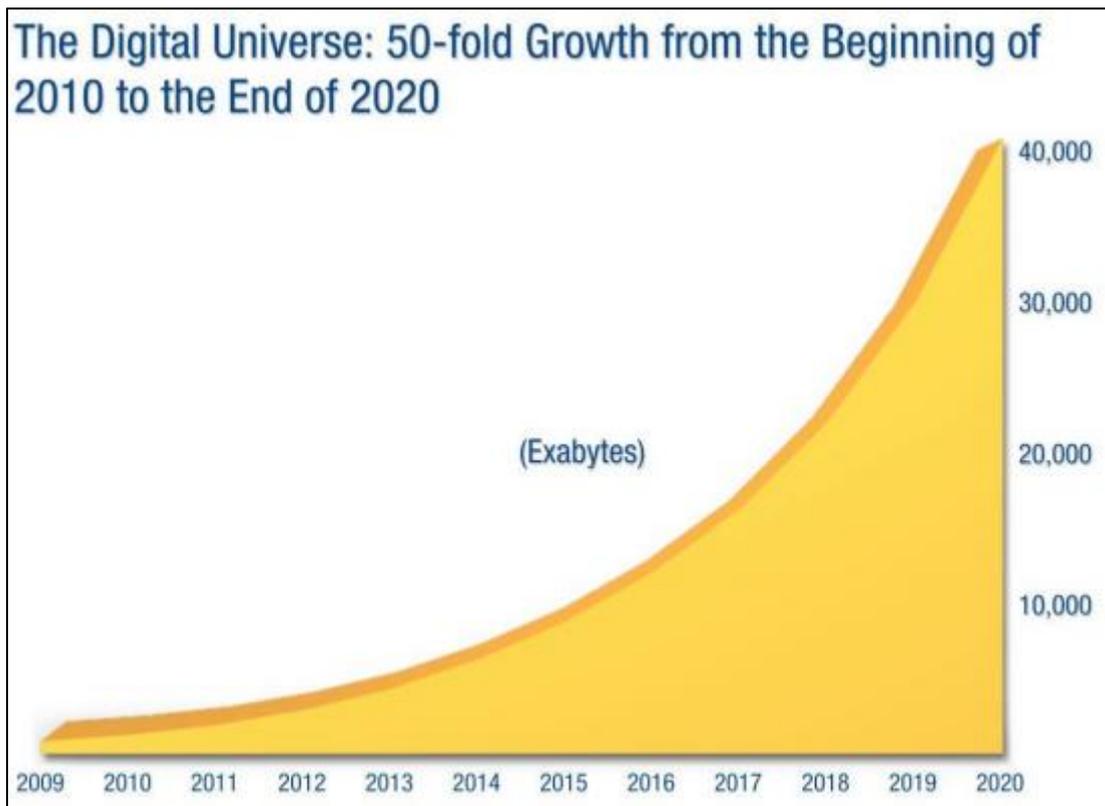
Wir befinden uns in einer Zeit, in der Digitalisierung und Vernetzung die wichtigsten Treiber des Wirtschaftswachstums darstellen. Die Verlagerung des Einzelhandels auf Onlineplattformen wie Amazon oder Ebay, die Ablösung des analogen Telefonanschlusses durch VoIP-Telefonie oder die massive Zunahme von mobilen Endgeräten sind nur einige der Indikatoren für diese Entwicklung. So stieg die Anzahl der vernetzten Geräte bisher exponentiell an und es wird geschätzt, dass diese Zahl bis 2020 die Marke von 20 Mrd. Geräten überschreitet.



Vgl. Gartner, 2016.

Dieser rasante Anstieg ist sowohl Ursache, als auch Folge einer immer stärkeren Vernetzung all unserer Lebensbereiche. Vor allem die Entwicklung sogenannter smarter Geräte und deren Anbindung an das Internet (Internet of Things) hat zur Folge, dass wir eine gigantische Menge an Daten produzieren, die unser ganzes Leben bis ins Detail aufzeichnen bzw. dessen Rekonstruktion möglich macht. Aufgrund dieser Vielzahl an Informationen ist es außerdem erstmals möglich, durch Verfahren der induktiven Statistik Vorhersagen über unser zukünftiges Verhalten zu treffen. Dies wird beispielsweise von Unternehmen wie Google oder Facebook intensiv genutzt, um die angezeigte Werbung möglichst genau an den Nutzer anzupassen. Dabei handelt es sich jedoch nur um einen ersten Schritt, denn aggregiert man diese Daten, so ist es möglich, die Entwicklung und Attraktivität gesamter Märkte zu analysieren, was nun von Unternehmen genutzt und in den strategischen

Entscheidungsprozess eingebaut werden kann. Der geschätzte Effekt scheint enorm zu sein. So glauben Ebner et al., dass das jährliche Potential im Gesundheitswesen bei \$ 300 Mrd. und im öffentlichen Bereich bei \$ 250 Mrd. liegt, sowie privatwirtschaftliche Unternehmen ihre Margen um 60% steigern könnten.<sup>1</sup>



Ganz, Reinsel, 2012, S.3.

Allgemein kann man sagen, dass es sich bei der Thematik Big Data um ein sehr junges Forschungsfeld handelt, da die gesamte Problematik, wie eben beschrieben, ein neues Symptom der technologischen Entwicklung darstellt. Die Hauptgründe hierfür sind, dass zum einen erst in den letzten Jahren die Datenbasis so breit wurde, dass entsprechende Verfahren angewandt werden können bzw. diese Verfahren notwendig wurden, um die Menge an Daten bewältigen zu können. Zum anderen verfügen wir erst durch die neuesten Verbesserungen im Bereich der Hardware über die notwendigen Rechen- und Speicherkapazitäten, um diese Menge an Daten analysieren zu können.<sup>2</sup> Trotz der erst kurzen Zeit, die sich Wissenschaft und Wirtschaft mit diesem Feld befassen, ist bereits eine große Menge an Publikationen verfügbar, welche sich grob in zwei Bereiche – technische und betriebswirtschaftliche Publikationen – unterteilen lässt. Hierbei ist anzumerken, dass

<sup>1</sup> Vgl. Ebner et al., 2014, S.2.

<sup>2</sup> Vgl. Mayer-Schönberger, Cukier, 2013, S.19.

sich die betriebswirtschaftliche Literatur fast ausschließlich auf qualitative Analysen beschränkt. Der Hauptanteil der Datengrundlage über das Forschungsfeld Big Data kommt daher nicht aus dem Bereich der rein wissenschaftlichen Publikationen, sondern wird sehr häufig von Unternehmensberatungen erhoben. Diese liefern meist deskriptive Statistiken darüber, wie Entscheidungsträger die aktuelle Situation beurteilen, was diese zu einer wichtigen Quelle für diese Arbeit macht.

Zum Abschluss dieser Einführung sollen hier noch zwei Beispiele konkreter Anwendungen die Vorhersagekraft von Big Data Methoden verdeutlichen, um Lesern einen ersten Eindruck zu verschaffen, wie entscheidend diese Entwicklung sämtliche Bereiche unseres Lebens schon heute und noch um ein Vielfaches stärker in der Zukunft beeinflussen wird:

About a year after Pole created his pregnancy prediction model, a man walked into a Minnesota Target and demanded to see the manager. He was clutching an advertisement. He was very angry. "My daughter got this in the mail!" he said. "She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?" The manager didn't have any idea what the man was talking about. He looked at the mailer. Sure enough, it was addressed to the man's daughter and contained advertisements for maternity clothing, nursery furniture, and pictures of smiling infants gazing into their mothers' eyes. The manager apologized profusely, and then called, a few days later, to apologize again. The father was somewhat abashed. "I had a talk with my daughter," he said. "It turns out there's been some activities in my house I haven't been completely aware of." He took a deep breath. "She's due in August. I owe you an apology."<sup>3</sup>

Das zweite Beispiel kommt aus dem medizinischen Bereich. Hier arbeitet Dr. Carolyn McGregor zusammen mit einem Team des Institute of Technology der University of Ohio und IBM an einer Software, die medizinische Daten, die im Krankenhaus erfasst werden, nutzt, um Frühgeborene besser behandeln zu können. Hierdurch wird es möglich, kleinste Veränderungen der Körperfunktionen wahrzunehmen, die mit bisherigen Methoden nicht erkannt werden konnten. Dies erlaubt es den Ärzten beispielsweise früher auf eine Infektion zu reagieren, was weniger Eingriffe notwendig und mögliche ineffektive Behandlungen schneller erkennbar macht. Außerdem ist es durch die so gewonnenen Daten auch möglich, neue Erkenntnisse darüber zu erhalten, welche Vorzeichen potentiell schädliche Folgen haben können, selbst wenn diese sogar für ausgebildete Ärzte kontraintuitiv erscheinen. Ein Beispiel hierfür ist, dass sehr stabile Vitalfunktionen, die in der Regel als gutes Zeichen gewertet wurden, oft ein Anzeichen für eine schwere Infektion sein können. Diese Erkenntnis liefert zwar keine Erklärung, wieso dieser Zusammenhang besteht – darüber lässt

---

<sup>3</sup> Duhigg, 2012, S. 198.

sich lediglich spekulieren – dennoch bewahrt sie Ärzte vor Fehlentscheidungen, die vor allem bei Frühgeborenen verheerende Folgen haben können.<sup>4</sup>

## 2. Big Data: Kriterien, Ausprägungen und Folgen

### 2.1. Begriffsdefinition und Kriterien für Big Data

Der Begriff Big Data wurde erstmals von Doug Laney in seinem Aufsatz „3D Data Management: Controlling Data Volume, Velocity, Variety“ genannt und beschreibt bereits die drei Kernmerkmale von Big Data – eine hohe Varianz der Quellen, großes Volumen und eine hohe Zugriffsgeschwindigkeit (Variety, Volume, Velocity). Diese werden seither oft auch als 3Vs beschrieben.<sup>5</sup> Neben diesen allgemein anerkannten Merkmalen gibt es auch einige weitere, die diskutiert werden: Value, Variability und Veracity – wobei die drei erstgenannten die unumstrittenen Kernkriterien darstellen<sup>6</sup> und deshalb im Rahmen dieser Arbeit als Definitionskriterien genutzt werden sollen.

#### 2.1.1. Velocity, Variety und Volume als Grundlage für die Wertschöpfung:

##### **Velocity**

Die Geschwindigkeit, mit der fundierte Entscheidungen in einem Unternehmen getroffen werden können, spielt eine bedeutende Rolle für die Konkurrenzfähigkeit dieses Unternehmens. Je schneller Entscheidungsträger in der Lage sind, auf Entwicklungen und Ereignisse zu reagieren, desto flexibler kann das Gesamtunternehmen reagieren, wodurch die Verluste geringer bzw. die Gewinne höher ausfallen. Insbesondere wenn man in der Lage ist, schneller als die Konkurrenz zu reagieren.

Generell wird angenommen, dass das Verhältnis zwischen verstrichener Zeit zwischen Ereignis, erfolgter Reaktion und dem Wert für das Unternehmen einer Verfallsfunktion entspricht. Somit muss ein Ziel für Unternehmen darin bestehen, die Reaktionszeit immer weiter gegen Null zu verringern. Einen entscheidenden Beitrag hierzu können real-time

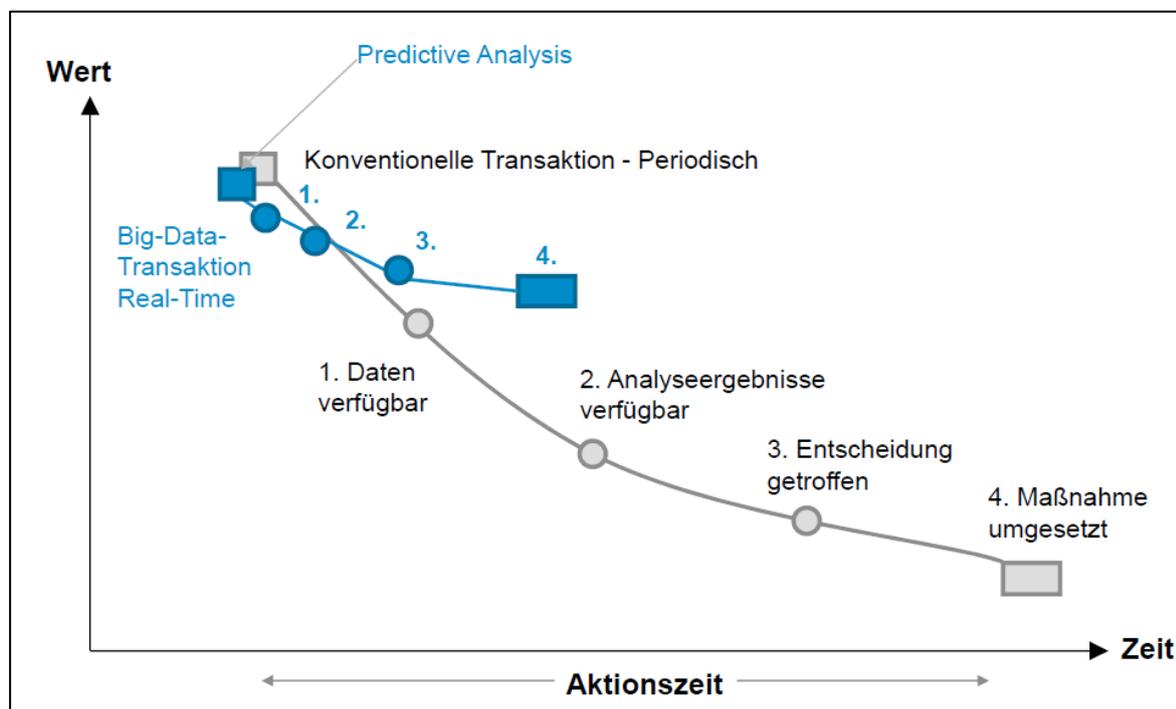
---

<sup>4</sup> Vgl. Mayer-Schönberger, Cukier, S. 60f

<sup>5</sup> Vgl. Laney, 2001.

<sup>6</sup> Vgl. Janssen et al., 2017, S. 339; Mayerl, 2015.

Datawarehousing sowie eine automatisierte Aufbereitung der Daten liefern.<sup>7</sup> Betrachtet man die folgende Grafik, bedeutet dies, dass erstens die Datenverzögerung, zweitens die Analyseverzögerung auf nahezu Null reduziert wird.



Horváth & Partners, 2016 (b), S.12.

Insbesondere die beiden genannten Faktoren spielen hier eine bedeutende Rolle. Im klassischen Berichtswesen werden periodisch Berichte verfasst, beispielsweise am Ende jedes Monats, die dafür auf Daten zurückgreifen, die erst aufbereitet werden müssen. Also im schlechtesten Fall nicht einmal aus dem aktuellen, sondern aus dem Vormonat stammen, was somit zu einer Verzögerung von zwei Monaten führen kann, bevor Entscheidungsträgern überhaupt bewusst wird, dass ein Problem vorliegt. Durch eine entsprechende Echtzeiterfassung der Geschäftsdaten, beispielsweise im Verkauf und eine standardisierte Aufbereitungsform, kann diese Zeit deutlich reduziert werden. So sind die aktuellen Zahlen möglicherweise in Echtzeit in einem Management Cockpit einsehbar, was die Zeit zur Informationsübertragung vom Ereignis bis zur Information des zuständigen Managers nahezu gegen Null bringt. Dies mag zwar eine entsprechende Analyse und damit einhergehende Kommentierung in den periodischen Berichten nicht vollständig ersetzen, da an diesem Punkt noch eine klare Ursachenanalyse fehlt, jedoch ermöglicht es, bei zu starken

<sup>7</sup> Vgl. Hackathorn, 2002, S. 23f.

Abweichungen von der Norm frühzeitig zu reagieren und außerhalb des üblichen Zyklus gezielte Ursachenforschung zu betreiben.

Im Bereich der Entscheidungsverzögerung dagegen muss man nach Entscheidungsebenen differenzieren. Auf der operativen Ebene ist eine Automatisierung durchaus denkbar und findet in manchen Bereichen auch schon statt. Ein Beispiel hierfür stellt der automatisierte Handel von Wertpapieren dar.<sup>8</sup> Auf der strategischen Ebene dagegen muss zunächst davon ausgegangen werden, dass eine solche Automatisierung aufgrund der Einzigartigkeit solcher Entscheidungen nicht, oder nur sehr eingeschränkt möglich ist. Ergänzendes hierzu soll im Kapitel 2.2.2. behandelt werden.

## **Variety**

Variety beschreibt den Umstand, dass die Datenbasis, die für eine Analyse genutzt wird, sich aus einer Vielzahl an Quellen speist und nicht, wie es tendenziell bei klassischen Berichten der Fall ist, aus einer einzigen, wie beispielweise den Verkaufszahlen eines Unternehmens. Bei solchen Quellen kann es sich um Social Media, Wetterdaten, Staumeldungen, Maschinenprotokolle, Verkaufszahlen oder auch Verhaltensmuster handeln, was dazu führt, dass die Daten einerseits in unterschiedlichen Formaten vorliegen können, auch als Text, Bild oder Tonaufnahmen, sowie in strukturierter, teilstrukturierter oder unstrukturierter Form. Hierdurch entstehen drei Dimensionen der Vielfalt: Quelle, Format und Strukturiertheitsgrad.<sup>9</sup> Dies führt zu einem sehr vielschichtigen Bild, bzgl. dessen man davon ausgehen kann, dass es klar zu einer besseren Erfassung der Realität durch die genutzten Modelle beiträgt. Jedoch ergeben sich auch eine Reihe von Schwierigkeiten. Insbesondere die unterschiedlichen Quellen sorgen dafür, dass eine Kooperation zwischen verschiedenen Abteilungen oder sogar Organisationen notwendig wird. Hierbei bestehen zurzeit jedoch noch massive Hürden, die dazu führen, dass Daten nicht einheitlich und zentral in einem Unternehmen vorliegen, sondern auf eine Vielzahl an abgeschotteten Silos verteilt sind.<sup>10</sup> Eine Möglichkeit um hier eine Verbesserung zu erreichen, sind sogenannte Datawarehouses. Diese sind nach Bill Immon üblicherweise folgendermaßen definiert: „A data warehouse is a subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management’s decisions.“<sup>11</sup> Für den Aspekt der Vielfältigkeit spielt hierbei insbesondere die

---

<sup>8</sup> Vgl. Ruta, 2014, S.824.

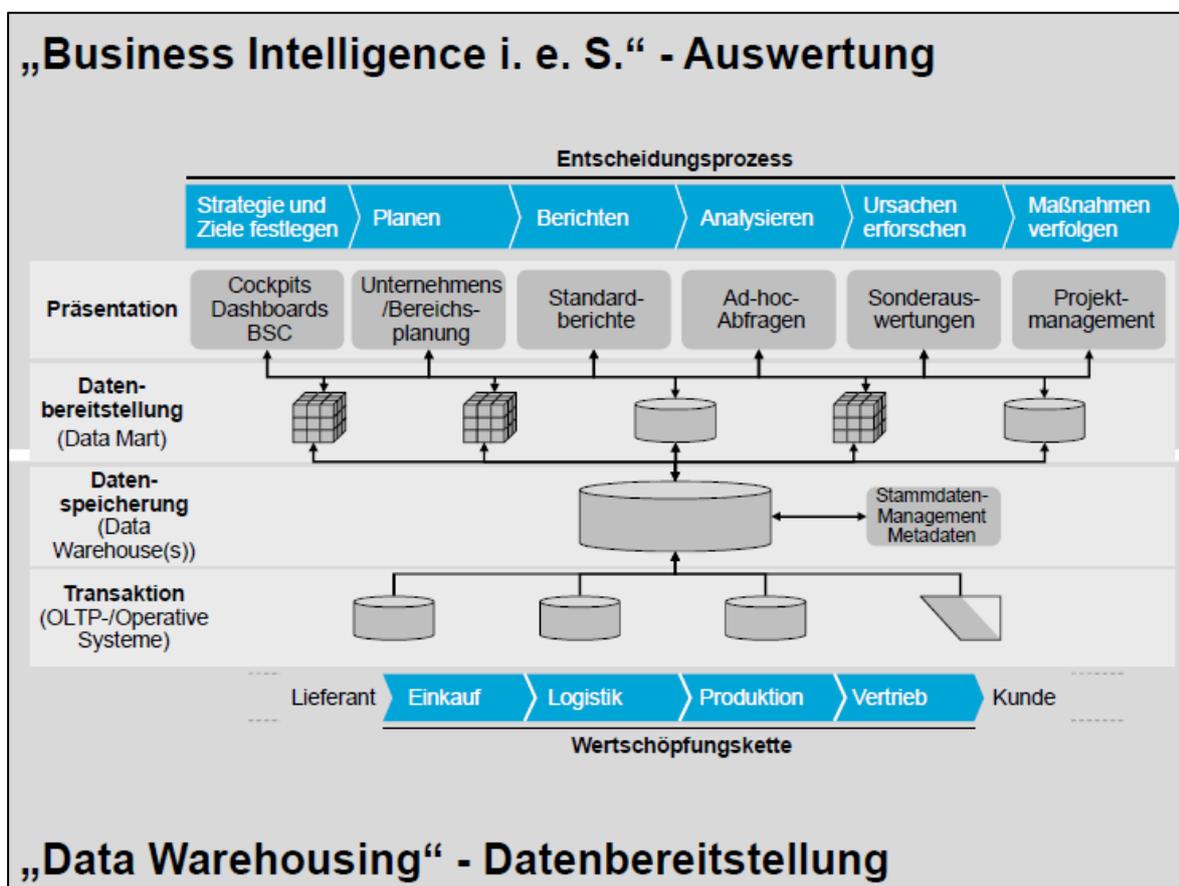
<sup>9</sup> Vgl. Qin et al., 2015, S. 355.

<sup>10</sup> Vgl. CapGemini, 2012, S.5.

<sup>11</sup> Immon, 2002, S. 31.

Integration eine bedeutende Rolle. Dieser Aspekt umfasst die Extraktion der Daten aus den ursprünglichen Quellen und die Transformation in ein einheitliches Meta-Schema, was eine anschließende Analyse oft erst möglich macht.<sup>12</sup> Insgesamt ergeben sich hieraus mehrere Vorteile. Durch die Vielfältigkeit der Datengrundlage entsteht ein entsprechend vielschichtiges Bild, das die Realität besser abbildet als bisherige Daten, wodurch man davon ausgehen kann, dass sich auch die Entscheidungen, die auf dieser Basis getroffen werden, verbessern. Die Einführung eines Datawarehouses hat außerdem den Vorteil, dass sämtliche Daten eines Unternehmens gebündelt und damit schnell einsehbar zur Verfügung stehen, sowie dass sämtliche Entscheidungsträger eines Unternehmens ihre Entscheidungen auf derselben Datenbasis treffen, wodurch eine gewisse Konsistenz der Entscheidungen sichergestellt werden kann.

### Beispiel für den Aufbau eines Datawarehouses



Horváth & Partners, 2016 (a), S.11.

<sup>12</sup> Vgl. Qin et al. 2015, S. 355f.

## **Volume:**

Das große Datenvolumen ist das wohl markanteste und bekannteste Merkmal von Big Data. Der wichtigste Punkt hierbei ist, dass die Anzahl der einzelnen Beobachtungen in einem Datensatz die Grenze, die noch mit herkömmlichen Methoden in akzeptablen Zeiträumen bearbeitet werden kann, bei weitem überschreitet. Amazon, Google oder Facebook verarbeiten täglich Daten im Petabyte-Bereich, aber nicht nur Unternehmen, auch in der Forschung oder im US-amerikanischen Wahlkampf werden solche Mengen an Daten generiert und ausgewertet.<sup>13</sup> Grundsätzlich wurden solche Daten zwar auch in der Vergangenheit schon genutzt, beispielsweise mit Hilfe von Datawarehouses, jedoch mussten sie hierfür aggregiert werden, wodurch die Granularität verringert wurde und ein deutlicher Teil des Informationsgehalts verloren ging. Außerdem war eine nachfolgende Analyse lediglich im Bereich der deduktiven Statistik möglich, also beispielsweise Auflistungen, die Kunden nach Profitabilität und Region ordneten. Durch die Nutzung der gesamten, nicht aggregierten Datengrundlage ist es jedoch möglich, auch Methoden der induktiven Statistik zu nutzen, durch die es nun möglich wird, bisher unbekannte Zusammenhänge zu erkennen oder lediglich vermutete zu bestätigen.<sup>14</sup> Jedoch nicht nur die Anzahl der Beobachtungen, auch die Anzahl der erhobenen Variablen erreicht eine gänzlich neue Dimension. Auf die Auswirkungen, die dies mit sich bringt, soll in Kapitel 2.2.2 näher eingegangen werden. Diese Steigerung des verarbeiteten Datenvolumens hat nicht nur den Vorteil, dass es ähnlich wie die Datenvielfalt unsere Modelle von der Realität erweitert und damit ein genaueres Abbild generiert, sie stellt neben der notwendigen Rechenleistung zu Verarbeitung die fundamentale Grundlage für sämtliche neuen Anwendungsmöglichkeiten des Feldes Big Data dar.

## **2.2. Radikale Innovation oder inkrementelle Veränderung?**

Betrachtet man diese vier Kriterien, kann man durchaus argumentieren, dass sie im Grundsatz auch bei bisherigen Business Analytics Methoden von Bedeutung waren und es sich bei der großen Aufmerksamkeit, die Big Data zur Zeit erfährt, lediglich um eine Modeerscheinung handelt. Diese Meinung vertritt beispielsweise Lothar Burow, Leiter der Abteilung Business Intelligence der Bayer AG: „Ich persönlich halte das Label „Big Data“ für einen Hype, der derzeit im Markt besteht. „Business Analytics“ stellt jedoch einen

---

<sup>13</sup> Vgl. Rossouw, 2013, S. 23.

<sup>14</sup> Vgl. Ortega, 2013.

langfristigen Paradigmenwechsel dar.<sup>15</sup> Allerdings muss man auch sehen, dass es sich bei Big Data, wenn auch nicht um einen Paradigmenwechsel, so doch um eine deutlich qualitative Weiterentwicklung im Bereich der Business Intelligence handelt. Insbesondere die Möglichkeit auch zukünftige Entwicklungen zu prognostizieren erhält eine gänzlich neue, empirisch fundierte Komponente, die im Wesentlichen auf den folgenden zwei methodischen Veränderungen aufbaut.

### 2.2.1. Von hypothesengetriebenen zu datengetriebenen Ansätzen

Eine fundamentale Veränderung von Big Data Analysen besteht darin, dass sie mit einem der Grundsätze empirischer Forschung – der hypothesengetriebenen Forschung – brechen. Diese besagt, dass Zahlen und Daten alleine wenig aussagekräftig sind, wenn sie nicht auf einem theoretischen Fundament aufbauen, was sich unter anderem darin zeigt, dass Korrelationen keine kausale Richtung vorgeben. Welche der beiden miteinander korrelierenden Variablen die Reaktion der anderen auslöst, lässt sich somit nicht aus dem Datenmaterial ableiten, sondern lediglich anhand der theoretischen Vorarbeit. Außerdem besteht jederzeit die Gefahr eines  $\alpha$ -Fehlers, also dass ein Zusammenhang gefunden wird, obwohl es keinen gibt. Jochen Mayerl fasst diesen Punkt sehr schön zusammen:

„In einem „data-driven“-Ansatz ersetzt die große Datenmasse und -vielfalt sowie deren Durchforstung mittels hochkomplexer Algorithmen demnach die klassische deduktiv-nomologische Logik, der zufolge wissenschaftliche Erkenntnis immer zunächst mit einem Problem und einer allgemeingültig formulierten Hypothese beginnt, und die empirische Beobachtung dann nachrangig erfolgt, um die theoretische Aussage vorläufig zu akzeptieren oder zu falsifizieren.“<sup>16</sup>

Ein Beispiel aus dem Sales Bereich führen Mayer-Schönberger und Cukier auf: Sie beschreiben, dass früher Verkäufern aller Branchen beigebracht wurde, dass sie verstehen müssen, was die Kunden zum Kauf bewegt und wieso sie dies tun. Dementsprechend wurde viel Wert auf Erfahrung und Ausbildung gelegt. Dies verglichen sie mit dem Empfehlungssystem von Amazon, welches basierend auf den Käufen anderer Kunden Produkte miteinander in Beziehung setzt und somit zusätzliche Kaufempfehlungen zum aktuellen Produkt gibt. Dabei versteht Amazon selbst nicht, wie diese beiden Produkte zusammenhängen, die Empfehlung basiert lediglich auf einer statistischen Korrelation bei vergangenen Einkäufen. Dennoch zeigte sich in der Vergangenheit des Unternehmens, dass

---

<sup>15</sup> Burow et al., 2014, S.20.

<sup>16</sup> Mayerl, 2015.

in Bezug auf die Verkaufszahlen dieses rein datengetriebene System menschlicher Erfahrungen und Empfehlungen weit überlegen ist. So wird heute davon ausgegangen, dass etwa ein Drittel des Umsatzes von Amazon durch dieses Empfehlungssystem generiert wird.<sup>17</sup> Ein solcher Daten getriebener Ansatz weist außerdem die Vorteile auf, dass der persönliche Bias kaum mehr eine Rolle spielt, da die notwendigen Variablen nicht mehr manuell ausgesucht werden, sondern einfach eine riesige Menge an Modellen berechnet und anschließend die besten ausgewählt werden. Außerdem verkürzt sich der „Forschungsprozess“ und damit die Zeit bis darauf aufbauende Entscheidungen getroffen werden können deutlich.<sup>18</sup>

### 2.2.2 Von der deskriptiven zur präskriptiven Datennutzung

Der zweite große Unterschied besteht darin, dass der Fokus nicht mehr auf den üblichen deskriptiven Messgrößen liegt, wie die absolute Anzahl der Beobachtungen oder die Anteile unterschiedlicher Kategorien, sondern primär auf der Korrelation von Variablen untereinander. Bereits unter den Punkten Velocity und Volume wurde angedeutet, dass sich die Nutzung von Daten durch moderne Datenverarbeitungssysteme und besonders durch das Aufkommen von Big Data klar verändert hat. Hierbei gilt es zu beachten, dass es eine klare Trennlinie gibt, die nicht nur zwischen alten und neuen Methoden, sondern auch quer durch das Feld Big Data verläuft. Diese trennt nämlich, mit welchen Methoden die Daten verarbeitet werden. Zunächst sind hier Methoden der deskriptiven Statistik zu nennen, welche schon immer eine bedeutende Rolle in der Entscheidungsfindung von Unternehmen spielen. Hierbei werden vergangenheitsbezogene Daten analysiert und aufbereitet. Beispiele dafür sind klassische Berichte, die die Verkaufszahlen vergangener Perioden enthalten, mit welchen Kunden und Produkten Gewinne oder Verluste erwirtschaftet wurden oder wie sich Ressourcenkosten veränderten. Also der Bereich in der folgenden Grafik, der mit Descriptive Analytics umschrieben ist. Dies wird teilweise durch eine mehr oder weniger ausführliche Kommentierung in den Berichten näher analysiert, um die Ursachen für diese Zahlen herauszufinden, was Diagnostic Analytic genannt wird. In diesen beiden Bereichen besteht der große Mehrwert von Big Data – wie bereits erwähnt – darin, dass die Zeit bis die Analysen den Entscheidungsträgern zur Verfügung stehen, massiv verkürzt wird und die

---

<sup>17</sup> Vgl. Mayer-Schönberger, Cukier, S. 51f

<sup>18</sup> Vgl. Ebd., S. 55

Genauigkeit und Granularität deutlich höher ausfällt als bei bisherigen Methoden, da nicht nur Aggregatdaten, sondern auch Individualdaten deutlich effizienter verarbeitet werden können<sup>19</sup> und außerdem oft nicht mehr auf eine Stichprobe zurückgegriffen werden muss, sondern eine Vollerhebung durchgeführt werden, bzw. die Größe der Stichprobe so enorm erweitert werden kann, dass sie nahe an die Datenqualität einer Vollerhebung kommt.<sup>20</sup> Hierdurch wird jedoch auch das Tor zu induktiven Statistik aufgestoßen, die einen gänzlich neuen Aspekt der Datenanalyse erlaubt. Diese ermöglicht es nämlich auf Basis von in der Vergangenheit erhobenen Daten, statistische Modelle zu entwickeln und zu testen, um damit Vorhersagen über die Zukunft zu treffen und stellt damit eine deutliche Erweiterung der bisherigen Möglichkeiten dar.<sup>21</sup> Dies wird durch den Begriff Predictive Analytics abgedeckt. Solche statistischen Modelle sind zwar grundsätzlich keine gänzlich neue Erscheinung und wurden schon in der Vergangenheit vor allem von Forschern und Versicherungsmathematikern genutzt. Die Qualität der Modelle, die auf der Basis von Big Data entwickelt werden können, übertrifft jedoch sämtliche bisherigeren Möglichkeiten. Der Grund hierfür liegt vor allem in der enormen Anzahl von Variablen, die in die Berechnungen miteinbezogen werden können. Während sehr ausgefeilte klassische Modelle in den beiden bereits genannten Bereichen in der Regel eine niedrige zweistellige Anzahl an Variablen oder sogar noch weniger betrachten, ist es möglich, mit Big Data Methoden Modelle mit mehreren hundert Variablen zu entwickeln.<sup>22</sup> Dies führt zu einer wesentlich realistischeren Modellierung der Realität, was wiederum die Qualität der Ergebnisse deutlich verbessert. Ergänzt man dieses Vorgehen anschließend noch mit Analysen, welche die Ausgangsvariablen betrachten und versuchen zu verstehen, welche dieser Variablen ein Unternehmen am sinnvollsten beeinflussen sollte, um das angestrebte Ergebnis zu erhalten, ähnlich wie bei der Kommentierung von Berichten, bewegt man sich im Bereich, der Prescriptive Analytics genannt wird. Die Wertschöpfung für ein Unternehmen steigt hierbei bei jeder der genannten Stufen an. Zum einen, da das Verständnis für die Wirkungsweise der eigenen Handlungen auf die Umwelt sowie diejenige der Umwelt auf das Unternehmen besser verstanden wird, wodurch bessere Entscheidungen getroffen werden können, da deren Konsequenzen besser abgeschätzt werden können.<sup>23</sup> Zum anderen aber auch parallel zum Argument das bereits beim Punkt Velocity genannt wurde, weil bei jeder Stufe die Reaktionszeit des Unternehmens verkürzt wird.

---

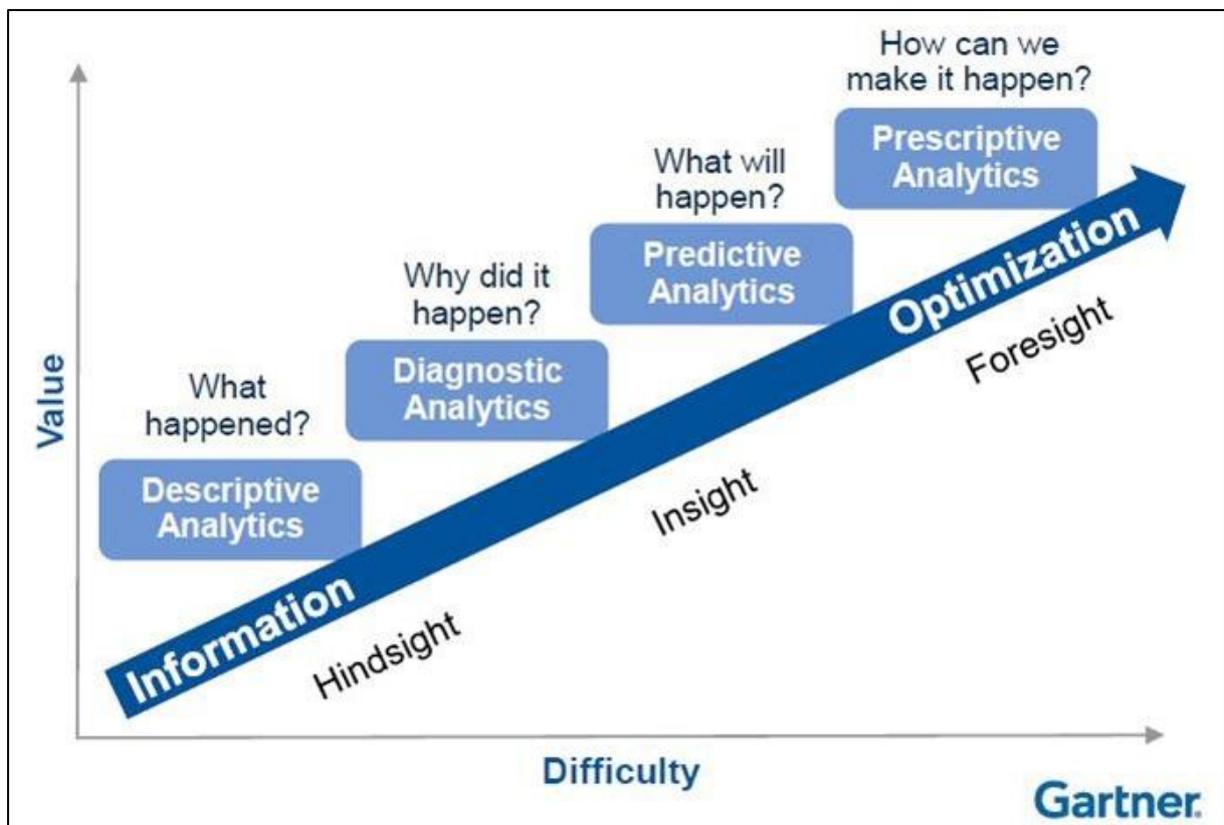
<sup>19</sup> Vgl. Meyerl, 2015.

<sup>20</sup> Vgl. Mayer-Schönberger, Cukier, S. 26

<sup>21</sup> Vgl. Iffert, 2016, S.17.

<sup>22</sup> Vgl. Blue Yonder, 2016, S.5.

<sup>23</sup> Vgl. Meyer, 2000, S.99.



Mauerer, 2015.

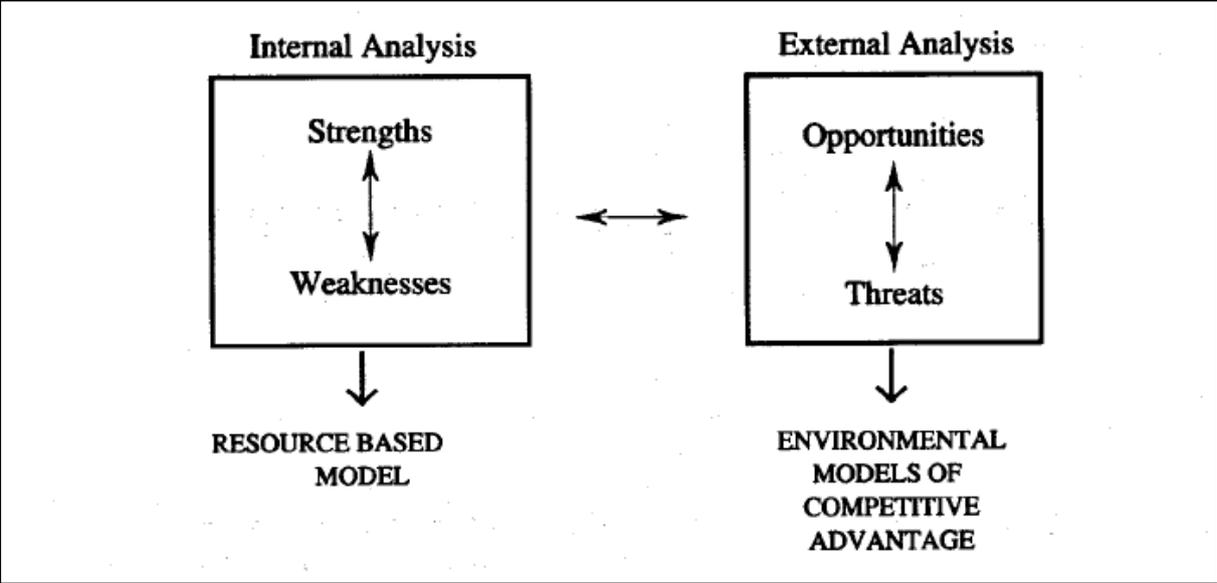
Zusammenfassend lässt sich damit sagen, dass man davon ausgehen kann, dass die Etablierung und Integration von Big Data in die Entscheidungsprozesse von Unternehmen einen signifikanten positiven Effekt auf den Unternehmenserfolg hat, woraus sich die zentrale Hypothese dieser Arbeit ergibt:

H<sub>1</sub>: Je höher die Quantität und Qualität der Datenanalyse in einem Unternehmen ist, desto erfolgreicher ist dieses Unternehmen.

### 3. Theoretische Verortung im strategischen Management

Um die folgende Analyse auf einem soliden theoretischen Fundament aufzubauen, ist es zunächst notwendig, einen Blick auf die großen Theorien des strategischen Managements zu werfen. Hier tun sich zwei große Schulen hervor, welche sich aus den möglichen Blickwinkeln auf ein Unternehmen ergeben. Zum einen der Market-Based-View, der die externe Umwelt aus der Sicht eines Unternehmens betrachtet und zum anderen der

Ressource-Based-View, der die Fähigkeiten und Kapazitäten eines Unternehmens betrachtet und damit den Blick auf die internen Gegebenheiten darstellt.



Barney 1991, S.100.

### 3.1. Market-Based-View

Bei der Betrachtung des Market-Based-Views soll vor allem auf die Werke ihres wohl prominentesten Vertreters, Michael Porter, zurückgegriffen werden. Dieser liefert vor allem mit seinen Five-Forces ein Werkzeug zur Analyse der externen Umwelt eines Unternehmens und des Marktes, in dem sich dieses bewegt.



Mind Tools.

Anders als beim Ressource-Based-View, in dem die Fähigkeiten eines Unternehmens Wettbewerbsvorteile durch die geschickte Kombination von Ressourcen zu erlangen im Zentrum der Betrachtung steht, legt der Market-Based-View den Fokus auf die externe Umwelt eines Unternehmens und seine Fähigkeit sich in dieser in eine möglichst vorteilhafte Position zu bringen. Dies bedeutet, sich im existierenden Markt so zu positionieren, dass der Druck externer Kräfte minimal ausfällt, um selbst einen großen Teil der Wertschöpfung abschöpfen zu können und diese nicht auf Zulieferer oder Kunden transferieren zu müssen.

Diese Sichtweise beruht auf zwei Grundannahmen: Zum einen, dass Unternehmen innerhalb eines Marktes in Bezug auf die ihnen zur Verfügung stehenden Ressourcen und Fähigkeiten identisch sind und zum anderen, dass für den Fall, dass Unterschiede in diesen Bereichen entstehen, diese nur von kurzer Dauer sein können, da Ressourcen und Fähigkeiten eine hohe Mobilität aufweisen und somit von anderen Unternehmen eingekauft werden können.<sup>24</sup>

Laut Porter resultiert dies in drei generischen Strategien für ein Unternehmen:

### **Kostenführerschaft**

Diese Strategie beruht darauf, die eigenen Kosten zu minimieren, um diese Einsparungen zum Teil an die Kunden weitergeben zu können und somit einen Preisvorteil auf dem Markt zu erlangen. Dies erfordert eine konsequente Steigerung der Effizienz. Erreicht werden kann dies vor allem durch Skaleneffekte, wie reduzierte Produktions- und Beschaffungskosten aufgrund des hohen Volumens, eine stärkere Streuung von Overheadkosten auf eine höhere Stückzahl, sowie Lerneffekten. Deshalb ist ein hoher Marktanteil eine entscheidende Voraussetzung für die Nachhaltigkeit dieser Strategie.<sup>25</sup> Hier können durchaus Parallelen zum VRIO Konzept des Resource-Based-Views gesehen werden, da die geschickte Kombination an Ressourcen, die zu solch niedrigen Produktionskosten führt, durchaus auch als komparativer Vorteil angesehen werden kann. Jedoch unterscheidet sich der Blick auf die Konsequenzen aus diesem Vorteil bei Porter deutlich. So wird hier argumentiert, dass die zu erwartenden Gewinne eines Unternehmens darin begründet sind, dass zum einen die Kundenmacht dadurch eingeschränkt ist, dass diese maximal das billigste Produkt auf dem Markt kaufen können, welches man somit selbst bereit stellt. Den Zulieferern gegenüber erhöht sich die Macht durch das hohe Abnahmenvolumen und die höhere Toleranz gegenüber Preiserhöhungen, die an den eigenen höheren Margen liegt. Konkurrenten und Anbieter von Substituten sowie potentielle Markteintritte haben zum einen mit den niedrigen Preisen zu kämpfen, die man den Kunden anbieten kann und zum anderen mit den hohen Anfangsinvestitionen, die notwendig sind, um ähnliche Skaleneffekte zu erzielen und damit konkurrenzfähig zu sein.<sup>26</sup> Big Data kann hierbei eine Reihe von Vorteilen liefern. Nicht nur erhält man Vorteile in Bezug auf die Effizienzsteigerung der eigenen Produktion, da die generierten Maschinendaten besser ausgewertet werden können und damit Ineffizienz besser erkannt werden kann. Auch das Absatzvolumen kann besser prognostiziert werden, was positive Auswirkungen auf die Kosten der Lagerhaltung hat. Außerdem ist eine bessere

---

<sup>24</sup> Vgl. Barney 1991 S. 100.

<sup>25</sup> Vgl. Porter, 1980, S.35f.

<sup>26</sup> Vgl. Ebd., S.36f.

Analyse des Marktes denkbar, was bei der Preisgestaltung und damit der Optimierung der Margen hilfreich sein kann.

### **Differenzierung**

Differenzierung zielt darauf ab den Kunden ein Produkt zu bieten, das in Bezug auf mindestens eine, optimalerweise aber mehrere Dimensionen einzigartig ist. Diese könnte beispielweise eine besonders effiziente Vertriebs- oder Supportstruktur sein, die nahe an den Kunden und schnell erreichbar ist, aber genauso ein herausragendes Design oder spezielle Leistungsmerkmale. Entscheidend ist, dass die Kunden den Wert zu schätzen wissen und ihn nicht von der Konkurrenz erhalten können. Dies führt zu einer höheren Kundenbindung und einer geringeren Preissensibilität der Kunden, woraus sich wiederum höhere Margen generieren lassen. In Bezug auf die Marktkräfte heißt dies, dass die hohe Kundenbindung Barrieren für Neueintritte, Konkurrenten und Substitute schafft, während die Zulieferer durch die höheren Margen, wie bereits bei der Kostenführerschaft, befriedigt werden können.<sup>27</sup> Big Data kann hierbei erneut bei der Analyse des Marktes und der Preisgestaltung helfen. Besonders interessant sind an diesem Punkt jedoch die Möglichkeiten zur Entwicklung neuer Geschäftsmodelle,<sup>28</sup> die der Treiber einer Differenzierungsstrategie sein können.

### **Fokus**

Eine Fokusstrategie zielt darauf ab, nur einen gewissen Bereich eines Marktes zu bedienen. Dies kann eine Beschränkung auf eine Region, eine spezielle Kundengruppe oder auch eine spezielle Produktgruppe sein. Hierdurch versetzt sich ein Unternehmen in die Lage, besser auf die speziellen Bedürfnisse seiner Kunden eingehen zu können und diese effektiver und effizienter zu bedienen als die Konkurrenz, die im gesamten Markt aktiv sind. Innerhalb einer solchen Fokusstrategie kann ein Unternehmen erneut eine Position der Kostenführerschaft oder Differenzierung anstreben, aufgrund des Fokus ist es jedoch auch möglich beide Ziele zu erreichen. Ein Unternehmen das eine solche Strategie wählt muss sich jedoch der Tatsache bewusst sein, dass es durch den entsprechenden Fokus auch seinen potentiellen Anteil am Gesamtmarkt deutlich einschränkt.<sup>29</sup> Die Big Data Anwendungsmöglichkeiten für eine solche Strategie umfassen zum einen die bereits genannten Möglichkeiten. Zum anderen ist es denkbar, dass durch die Analyse von Daten gewisse Kundengruppen erst identifiziert werden können, auf welche sich das Unternehmen

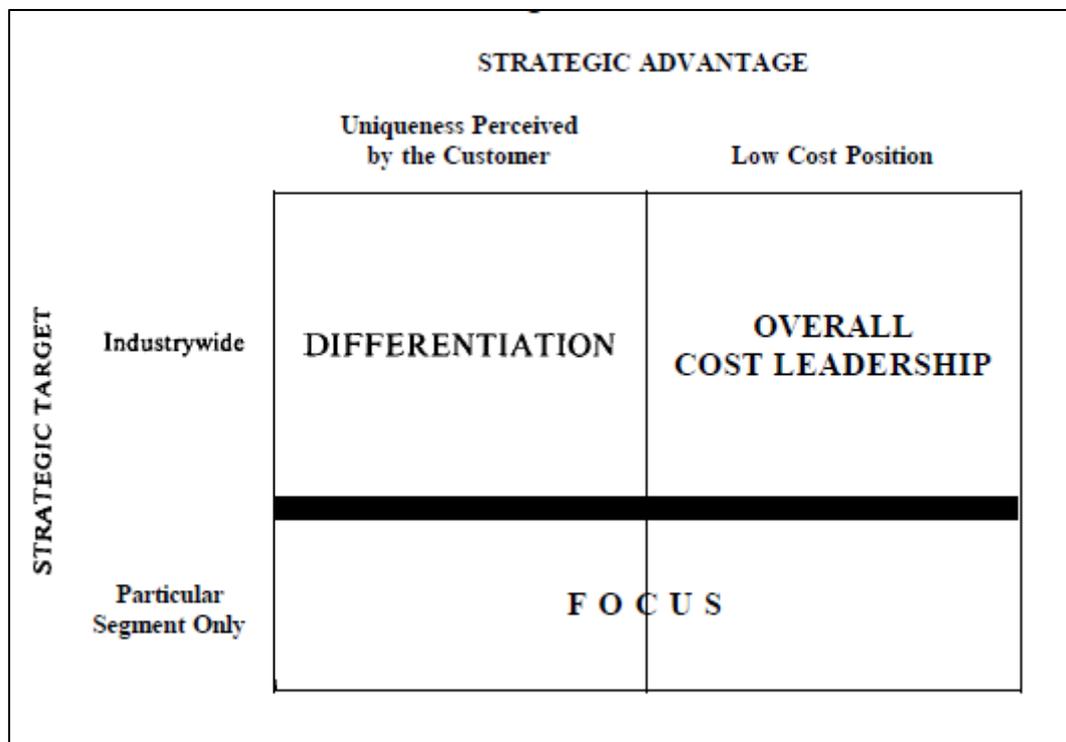
---

<sup>27</sup> Vgl. Ebd., S.37f.

<sup>28</sup> Vgl. Brownlow, 2015, S.1.

<sup>29</sup> Vgl. Porter, 1980, S.38f.

in Zukunft fokussieren will, da sie entweder besonders profitabel sind, oder viele Gemeinsamkeiten aufweisen, so dass ihre Wünsche mit reduzierten Kosten bedient werden können.



Porter, 1980, S.39.

Insgesamt kann man somit festhalten, dass der Hauptvorteil den Big Data aus Sicht des Market-Based-Views liefert, auf den bereits beschriebenen drei Merkmalen von Big Data beruht, was eine schnellere und gezieltere Reaktion auf Marktereignisse erlaubt und somit eine Verschiebung der eigenen Position in einen weniger umkämpften Bereich ermöglicht. Allerdings sollte auch das Phänomen der Blue-Oceans erwähnt werden, das bereits bei der Fokusstrategie angedeutet wurde. In einigen Fällen kann man davon ausgehen, dass Unternehmen durch die Nutzung von Big Data ein neues Geschäftsmodell entwickeln können und damit einen neuen Markt erschließen, in dem noch kein Konkurrent tätig ist. Dies erlaubt es ihnen zumindest für eine gewisse Zeit, bis Markteintritte von Konkurrenz den Druck erhöhen, hohe Gewinne abzuschöpfen.

## 3.2. Ressource-Based-View

Der Ressource-Based-View beschäftigt sich genau entgegen dem Market-Based-View mit der internen Perspektive von Unternehmen. Es wird versucht zu analysieren, aufgrund welcher Ressourcen ein Unternehmen in der Lage ist, sich Wettbewerbsvorteile gegenüber Konkurrenten zu erarbeiten und dadurch höhere Gewinne zu erzielen. Damit dies überhaupt möglich ist, müssen die beiden Grundannahmen des Market-Based-Views jedoch abgelehnt werden, da sonst keinerlei dauerhafte Wettbewerbsvorteile aufgrund von Ressourcenverfügbarkeit möglich wären. Der Ressource-Based-View geht somit davon aus, dass der Zugang von Unternehmen innerhalb eines Marktes zu wichtigen Ressourcen heterogen verteilt ist und dass diese in einigen Fällen nur eingeschränkt mobil sind und somit nicht oder nur unter hohen Kosten, von anderen Unternehmen erworben werden können.<sup>30</sup>

Die entscheidenden Kriterien um solche Ressourcen zu identifizieren, die einem Unternehmen dauerhafte Wettbewerbsvorteile sichern, finden sich im VRIO-Konzept. Sie müssen also wertvoll, selten und nicht imitierbar sein, sowie organisationale Unterstützung erfahren. Außerdem muss es sich dabei nicht zwingend um einzelne Ressourcen handeln, sondern auch eine Kombination aus Ressourcen, die diese Kriterien erfüllt, kann einen entsprechenden Vorteil liefern.<sup>31</sup>

In Bezug auf Big Data zeichnet Barney ein zwiespältiges Bild. Zum einen beschreibt er Wissenschaftler als nicht imitierbare Ressource, was an der Pfadabhängigkeit der Ursachen liegt, die zu den konkreten Durchbrüchen geführt haben: „A firm with scientists who are uniquely positioned to create or exploit a significant scientific breakthrough may obtain an imperfectly imitable resource from the history-dependent nature of these scientist’s individual human capital.“<sup>32</sup> Dies lässt sich durchaus auch auf sogenannte Data-Scientists übertragen, also auf Statistiker sowie Informatiker und die von ihnen entwickelten Algorithmen, Auswertungsmethoden und Managementinformationssysteme, aber auch auf die Datenbasis, die zu diesen Durchbrüchen führte.

Außerdem stellt das Programmieren von Algorithmen einen hoch kreativen Prozess dar, der sich in derselben Art wohl kaum in zwei Firmen auf dieselbe Art vollziehen wird (Anforderungen, Spezifikationen, Vorstellungen, Machbarkeit, Management Support, Ressourcenverfügbarkeit) und damit zu einmaligen Ergebnissen führt. Dies entspricht

---

<sup>30</sup> Vgl. Barney, 1991, S.101.

<sup>31</sup> Vgl. Ebd., S.110.

<sup>32</sup> Ebd., S. 108.

insgesamt dem Aspekt der sozialen Komplexität, womit der Faktor der Inimitability durchaus als erfüllt angesehen werden kann. Offen bleibt hierbei jedoch, wie es um die Sustainability steht. Mit Sicherheit ist es möglich, hierbei auf unterschiedlichem Weg gleichwertige Ergebnisse zu erzielen. Aufgrund der enormen Menge an Einflussfaktoren ist es jedoch sehr schwierig zu beurteilen ob dies auch real vorkommt.

Andererseits schreibt Barney aber auch klar, dass Informationsverarbeitungssysteme keine nachhaltigen Wettbewerbsvorteile generieren können, da sie imitierbar sind. Programmierer können abgeworben und Daten oftmals eingekauft werden. Er erkennt zwar durchaus an, dass sich aus der Einbettung von Informationssystemen in ein Unternehmen und dessen Entscheidungsprozesse komplexe soziale Systeme entwickeln können, die wiederum nicht imitierbar sind. Allerdings sieht er die Möglichkeit, dass dieselbe Leistung auch von einem erfahrenen Management Team erbracht werden kann, da der Hauptvorteil von Informationssystemen in der effizienten Weitergabe von Informationen und der Fähigkeit große Mengen an Informationen in kurzer Zeit in den Entscheidungsprozess einzubeziehen, bestehe.<sup>33</sup> Diese Sichtweise muss in Bezug auf moderne Managementinformationssysteme im Allgemeinen und auf Big Data im Speziellen jedoch aus mehreren Gründen kritisch hinterfragt werden. Zum einen besteht der Nutzen von Big Data nicht nur in der bloßen Weitergabe und Aufbereitung von Informationen, sondern es werden neue Informationen aus dem bestehenden Datenmaterial generiert und dies in einer Art, zu der selbst das fähigste Managementteam nicht in der Lage wäre, da die drei Kernkriterien von Big Data sowie die statistischen Vorgänge zur Auswertung nur durch leistungsstarke Soft- und Hardwarelösungen bewältigt werden können.<sup>34</sup> Des Weiteren muss man berücksichtigen, dass moderne Informationssysteme um ein vielfaches leistungsstärker sind als diejenigen, auf die sich Barney bei seiner Publikation bezog. Man kann durchaus davon ausgehen, dass vor allem Big Data Lösungen aufgrund der Komplexität, die sie durch das Zusammenspiel von Datengrundlage, Auswertungsmethoden und Integration in den Entscheidungsprozess aufweisen, nicht oder nur sehr schwer zu imitieren sind. Außerdem könnte man durchaus auch mit einer gewissen Pfadabhängigkeit in Bezug auf den Aufbau solcher Systeme in einem Unternehmen argumentieren.

Der dritte Punkt, der für die Generierung von nachhaltigen Wettbewerbsvorteilen durch Big Data Anwendungen spricht, besteht darin, dass es nicht direkt notwendig ist, einen

---

<sup>33</sup> Vgl. Ebd., S.114.

<sup>34</sup> Dies ergibt sich schlicht aus der riesigen Anzahl von Variablen (Vgl. Blue Yonder, 2016, S.5.) und der hohen Geschwindigkeit, in manchen Fällen sogar Echtzeit, in die Daten verarbeitet werden (Vgl. Mohamed, 2014, S.305).

nachhaltigen Vorteil im herkömmlichen Sinne zu erzielen. Da der Vorteil solcher Anwendungen unter anderem in der Geschwindigkeit der Datenverarbeitung und der stetigen Weiterentwicklung der dafür genutzten Algorithmen besteht, kann man argumentieren, dass der „nachhaltige“ Vorteil aus einer Abfolge einzelner nicht nachhaltiger Vorteile besteht. Dies bestätigt sich unter anderem durch die hohe Bedeutung, die stetige Weiterentwicklung und Flexibilität haben, um in diesem Bereich erfolgreich zu sein.<sup>35</sup>

Barney beschreibt außerdem sehr ausführlich die Wirkung kausaler Ambiguität. Diese beschreibt das Problem, dass die kausale Verknüpfung zwischen Ressourcen und komparativen Vorteilen oft nicht eindeutig sind und so Situationen entstehen können, in denen zwar ein Vorteil besteht, jedoch nicht verstanden wird, auf welchen Ressourcen dieser beruht. Barney fokussiert sich hierbei jedoch vor allem darauf, wie fremde Unternehmen diese Barriere überwinden können, um die notwendigen Ressourcen imitieren zu können, wodurch erneut ein Gleichgewicht entstehen würde.<sup>36</sup> Dabei geht er davon aus, dass Unternehmen, die diesen Vorteil besitzen, ein besseres Verständnis für die kausale Verbindung haben als solche die ihn nicht besitzen. Dies mag zwar durchaus der Fall sein, jedoch kann man nicht davon ausgehen, dass Unternehmen immer ein perfektes Verständnis dafür haben, wie sich ihre Ressourcen auf ihre Vorteile auswirken. Somit kann die genaue Analyse einer Vielzahl an Daten hier auch dazu beitragen, dass Unternehmen sich selbst besser verstehen und damit zielgerichteter Verbesserungen und Entwicklungen vornehmen können.

### 3.3. Grundlagen der Entscheidungstheorie

Als Abschluss der theoretischen Überlegungen soll noch etwas näher auf entscheidungstheoretische Überlegungen eingegangen werden um besser zu verstehen, an welchen Stellen in einem Entscheidungsprozess Big Data einen Mehrwert generieren kann und wie dies geschieht. Hierbei müssen zunächst drei fundamental unterschiedliche Blickweisen auf den Entscheidungsprozess unterschieden werden: Der normative Ansatz betrachtet einen theoretisch optimalen Entscheidungsprozess, der logisch nachvollziehbar

---

<sup>35</sup> Die Bedeutung dieser beiden Punkte wird unter anderem dadurch deutlich, dass Big Data Lösungen in vielen Fällen nicht auf festen Modellen beruhen, wie sie beispielweise in den meisten wissenschaftlichen Erhebungen genutzt werden, sondern über Methoden des Machine-Learnings die Algorithmen sich selbst auf Basis des Datenmaterials weiterentwickeln und optimieren. (Vgl. Witten et al., 2013, S.8f)

<sup>36</sup> Vgl. Barney, 1991, S.109.

ist und genau ein optimales Ergebnis liefert.<sup>37</sup> Die präskriptive Sichtweise geht nicht davon aus, dass der Mensch zu einer solchen kognitiven Leistung fähig ist und bezieht diese beschränkte Leistungsfähigkeit in ihre Überlegungen mit ein. Dies führt in der Folge zu deutlich vereinfachten Entscheidungsprozessen, welche nicht mehr zwingend optimale Ergebnisse liefern. Die letzte Option stellt die deskriptive Sichtweise dar, welche versucht, den real existierenden Entscheidungsprozess zu erfassen, welcher in vielen Fällen weder ein optimales, noch ein eindeutiges Ergebnis liefert.<sup>38</sup>

Ein Unternehmen hat jedoch ein Interesse daran über den Entscheidungsprozess zu einem optimalen Ergebnis zu gelangen, was folglich bedeutet, dass das Ziel eines Unternehmens bei der Gestaltung von Entscheidungsprozessen eine möglichst starke Annäherung an dieses normative Modell sein muss. Die größte Herausforderung bei der realen Verwirklichung dieses Ziels besteht dabei darin, dass die Mehrheit der Entscheidungen die Führungskräfte zu treffen haben, sich auf komplexe, nicht triviale Probleme beziehen. Die menschlichen kognitiven Fähigkeiten stellen sich dabei in der Regel als unzureichend heraus, um diese in ihrer Gesamtheit zu erfassen und zu einer optimalen Entscheidung zu kommen.<sup>39</sup> Aus diesem Grund neigen Menschen in realen Situationen zu wesentlich einfacheren Entscheidungsmodellen. Ein solches Modell stammt von Soelberg und ist unter dem Namen Implicit-Favorite-Model bekannt. Hierbei wird davon ausgegangen, dass sich Entscheidungsträger schon vor einer genaueren Analyse implizit für eine Option entscheiden, die im Weiteren als Favorit dient. Dieser wird mit sämtlichen anderen möglichen Optionen verglichen. Jedoch wird hier kein objektiver Maßstab angelegt, sondern die Entscheidungskriterien werden ex post so gewählt, dass Vorurteile gerechtfertigt werden und der Favorit bestätigt wird. Dabei ist den Entscheidungsträgern die Existenz eines solchen Favoriten und ihr an diesem orientierten Handeln in der Regel nicht bewusst.<sup>40</sup> Ein Datengetriebenes Entscheidungsmodell kann hierbei einen klaren Beitrag zur Verschiebung solcher willkürlichen, im Nachhinein gerechtfertigten Entscheidungen hin zu einem normativen, oder diesem zumindest nahestehenden Modell liefern. Zum Einen, da die einfließenden Variablen schon im Vorfeld festgelegt und in Algorithmen festgehalten werden müssen. Was eine nachträgliche Beeinflussung zumindest offensichtlich macht. Zum Anderen können durch Korrelationen und Regressionsanalysen die Zusammenhänge

---

<sup>37</sup> Vgl. Kirchler, 2003, S.46f.

<sup>38</sup> Vgl. Ebd., S.57.

<sup>39</sup> Vgl. Simon, 1955, S.101.

<sup>40</sup> Vgl. Soelberg, 1967.

zwischen abhängigen und unabhängigen Variablen klar berechnet werden, sodass die Subjektivität der Bewertung der Datengrundlage eliminiert wird.

Weitere Einblicke in das Verhalten von Entscheidungsträgern unter Unsicherheit geben die Arbeiten von Lipshitz und Strauss. Diese untersuchten Unsicherheiten und Risiken bei der Entscheidungsfindung. Aufgrund der Komplexität der Umwelt in der Unternehmen operieren, treffen diese Bedingungen nahezu bei jeder Entscheidung zu und die Folgen können nie gänzlich vorhergesehen werden. Für den Umgang mit solchen Bedingungen identifizierten sie drei große Strategien: Die Reduktion von Unsicherheit, die Akzeptanz von Unsicherheit sowie das Ignorieren von Unsicherheit.<sup>41</sup> Für die meisten Produktionsunternehmen stellt dabei die Reduktion von Unsicherheiten sicher die wünschenswerteste Möglichkeit dar. Hierzu kann eine gezielte Datenanalyse beitragen, da sie die vorhandenen Daten in einer Form aufbereitet, in der sie zum einen für Entscheidungsträger überhaupt erst verwertbar werden und zum anderen, da dabei gänzlich neue Zusammenhänge und Daten gefunden werden können. Außerdem kann dabei nicht nur qualitativ die Existenz solcher Zusammenhänge erkannt, sondern diese auch quantifiziert werden, sodass sie eine zuverlässige Grundlage für beispielweise Wirtschaftlichkeitsberechnungen liefern. Die zweite Option, das Akzeptieren von Risiko, ist dagegen vor allem für die Finanzbranche interessant, da das Geschäftsmodell von Banken und Versicherungen darauf aufbaut ein kalkuliertes Risiko einzugehen. Für solche Unternehmen ist insbesondere die eben erwähnte Quantifizierung von Zusammenhängen von Bedeutung, da sich daraus Zinssätze, Versicherungsprämien und Ausfallrisiken berechnen lassen. Außerdem besitzt hier der Bereich Big Data, im Gegensatz zur herkömmlichen Datenanalyse, den Vorteil, dass durch den datengetriebenen Ansatz Zusammenhänge erkannt werden können, die sich nicht logisch erschließen lassen oder zumindest nicht offensichtlich sind und in der Folge in die Berechnungen eingebaut werden können. Als Beispiel könnte hierfür die bereits erwähnte Supermarktkette genannt werden, die die Wahrscheinlichkeiten für eine Schwangerschaft ihrer Kundinnen berechnet:

Jenny Ward, a twenty-three-year-old in Atlanta who bought cocoa butter lotion, a purse large enough to double as a diaper bag, zinc, magnesium, and a bright blue rug? There's an 87 percent chance that she's pregnant and that her delivery date is sometime in late August.<sup>7.11</sup> Liz Alter in Brooklyn, a thirtyfive-year-old who purchased five packs of washcloths, a bottle of "sensitive skin" laundry detergent, baggy jeans, vitamins containing DHA, and a slew of moisturizers? She's got a 96 percent chance of pregnancy, and she'll probably give birth in early May. Caitlin Pike, a thirty-nine-year-old in San Francisco who purchased a \$250 stroller, but nothing else?

---

<sup>41</sup> Vgl. Lipshitz, Strauss, 1997, S.156.

She's probably buying for a friend's baby shower. Besides, her demographic data shows she got divorced two years ago.<sup>42</sup>

Hier zeigt sich auch, dass der Datenanalyse je nach Branche eine unterschiedliche Bedeutung zukommt und die Erhebung dieser als Kontrollvariable daher unerlässlich ist. Die dritte Strategie dagegen stellt für Unternehmen immer eine Gefahr dar und ist auf jeden Fall zu vermeiden. Hier besteht der Mehrwert einer systematischen Datenanalyse für Unternehmen darin, dass die Menge der Unknown-unknowns reduziert werden kann und diese Faktoren, wenn auch nicht immer eliminiert, so zumindest in die Entscheidung miteinbezogen werden können, anstatt sie unbewusst zu ignorieren.

## 4. Die Entwicklung des Modells und Operationalisierung

### 4.1. Unabhängige Variablen

#### 4.1.1. Qualität der Datenanalyse

Das Konstrukt „Qualität der Datenanalyse“ soll die Art und Weise abbilden, wie Datenanalyse im Unternehmen genutzt wird, um Entscheidungsprozesse zu fördern. Hierbei steht im Zentrum der Betrachtung, wie die Daten verarbeitet werden und wie schnell sie den Entscheidungsträgern zu Verfügung stehen.

Wie sich aus den vorhergehenden Kapiteln ergibt, stellt die Geschwindigkeit der Datenerfassung und Auswertung einen wichtigen Punkt für die Wertschöpfung dar bzw. die Zeit zwischen der Realisierung eines Ereignisses und dem Erhalt der relevanten Informationen durch das Management, um darauf reagieren zu können. Deshalb soll dieser Punkt auch in die Qualität der Datenanalyse einfließen. Hierbei sind zwei Möglichkeiten relevant. Einerseits, das durchschnittliche Alter der in einem Bericht verwendeten Daten und andererseits der Automatisierungsgrad der Dateneinspeisung und Aufbereitung durch die Managementinformationssysteme, falls solche vorhanden sind. Ersteres spielt eine Rolle, da Berichte nicht nur eine reine Ansammlung von Daten darstellen, sondern durch ihre Kommentierung auch eine Ursachenanalyse betreiben können und somit die reinen Zahlen

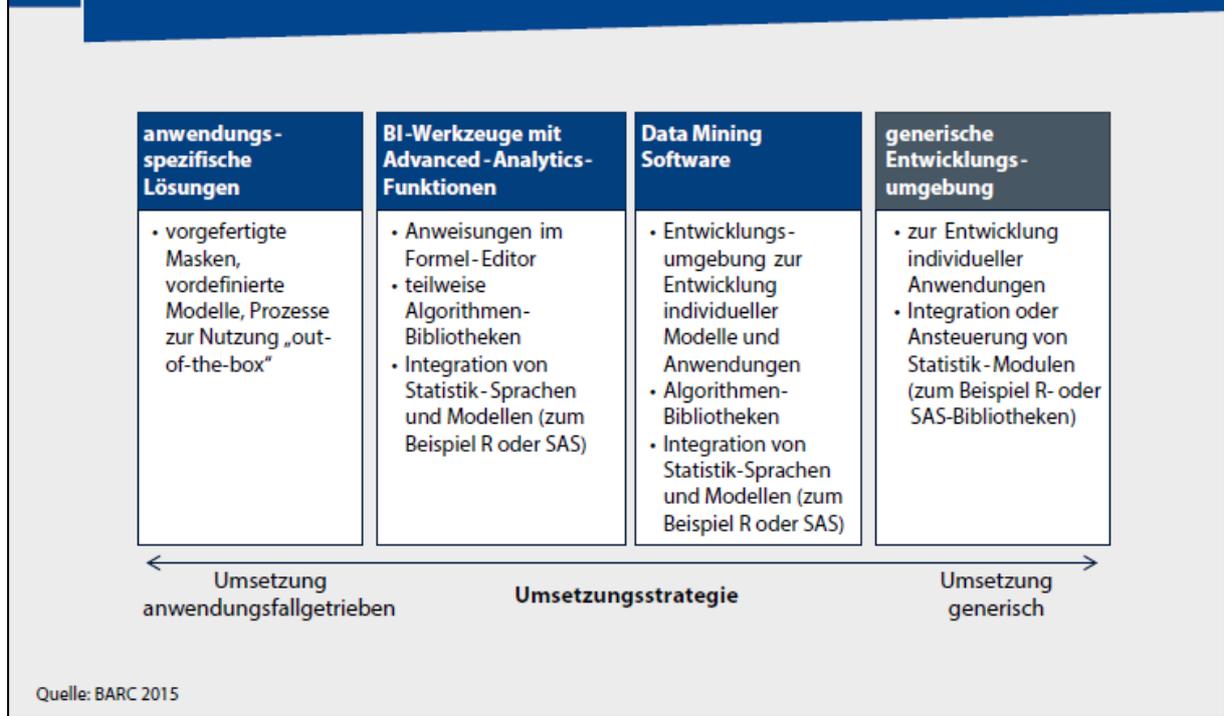
---

<sup>42</sup> Duhigg, 2012, S. 197.

um eine Kausalanalyse bereichern, parallel der daten- und hypothesengetriebenen Forschung. Letzteres spielt eine Rolle, da hierdurch Manager im Optimalfall in Echtzeit über Geschehnisse informiert werden und bei zu starken Abweichungen von der Norm bereits erste Gegenmaßnahmen oder auch nähere Analysen initiieren können.

Ein weiterer Punkt ist die Art der Software-Lösung, die verwendet wird, um die Daten zu analysieren. Wie die folgende Grafik zeigt, reicht das mögliche Spektrum hierbei von allgemeinen Standardlösungen, die fertig gekauft werden, bis hin zu Eigenentwicklungen. Generell lässt sich sagen, dass diejenigen Lösungen, die in der Grafik weiter links aufgeführt werden, die Vorteile bieten, dass sie nutzerfreundlicher sind, weniger Einarbeitungszeit benötigen und als fertige Lösung auf dem Markt erworben werden können. Alle diese Punkte verringern die Kosten bei der Einführung einer solchen Lösung. Jedoch sind die Anwendungsmöglichkeiten in solchen Fällen sehr beschränkt, auf den Zweck, wofür sie erworben wurden. Dies führt dazu, dass ein bedeutender Vorteil von Big Data verloren geht, nämlich die konstante Weiterentwicklung der zugrundeliegenden Algorithmen, um laufend die Prognosequalität zu erhöhen. Diese Möglichkeit steht in ihrer gesamten Breite nur offen, wenn man die notwendigen Qualifikationen im Unternehmen aufbaut, um diese Algorithmen selbst zu entwickeln. Solche Lösungen werden in der Grafik als generische Entwicklungsumgebung bezeichnet und stehen nur größeren Unternehmen zu Verfügung, da sie mit enormen Kosten, insbesondere Personalkosten für die notwendigen Experten, verbunden sind. Der dabei entstehende Vorteil besteht jedoch nicht nur in der konstanten Weiterentwicklung und damit Effizienzsteigerung der Modelle, sondern auch darin, dass diese genau auf die Bedürfnisse des Unternehmens angepasst sind.

**Abb. 2 Tools und Methoden zur Umsetzung von Predictive Analytics**



Iffert, 2016, S.19.

#### 4.1.2. Quantität der Datenanalyse

Die Quantität der Datenanalyse stellt insofern einen bedeutenden Einflussfaktor für die Analyse dar, da selbst die beste Datengrundlage keinen Einfluss auf die Entscheidungsqualität nehmen kann, wenn sie nicht genutzt wird. Hierbei sollen zwei Ebenen unterschieden werden: Die Horizontale beschreibt, wie häufig aufbereitete Daten als Grundlage für Entscheidungen genutzt werden, also ob sie das übliche Prozedere in einem Unternehmen darstellen oder ob andere Faktoren wie persönliche Ziele, Unternehmenskultur oder Emotionen einen starken Einfluss auf die getroffenen Entscheidungen haben. Lothar Burow unterscheidet hierbei beispielsweise zwei Arten von Entscheidungen in einem Unternehmen. Die einen, die auf der Grundlage von Fakten getroffen werden und die anderen, die nach Bauchgefühl getroffen werden. Er räumt beiden ein Existenzrecht ein und geht davon aus, dass ersteren in Zukunft ein deutlich breiteres und genaueres Spektrum an Informationen zur Verfügung steht, zweite allerdings kaum von der Entwicklung beeinflusst werden.<sup>43</sup>

<sup>43</sup> Vgl. Burow, 2014, S. 18.

Auf der vertikalen Ebene dagegen soll betrachtet werden, wer Zugang zu den Daten hat und somit in die Lage versetzt wird, die eigenen Entscheidungen auf dieser Basis zu treffen. Dies ist ein entscheidender Punkt in diesem Bereich, da traditionellerweise Mitarbeiter unterhalb der obersten Führungsebene nur eingeschränkten Zugang zu Informationen über das Unternehmen haben. Dies mag aus vielerlei Gründen durchaus sinnvoll sein, da somit vertrauliche Informationen besser geschützt sind und Neid und Missgunst zwischen Mitarbeitern und Abteilungen vermieden wird, beispielsweise bzgl. Budgetverteilungen oder Bonuszahlungen. Jedoch stellt diese Situation eine große Herausforderung für die Etablierung datenbasierter Entscheidungen in einem Unternehmen dar, da hierfür ein breiter Zugang zu Datenmaterial erforderlich ist. Insbesondere im Zuge aktueller Strömungen im Bereich der Personalentwicklung, beispielsweise dem Job Enrichment, könnte es nötig werden, selbst einfachen Mitarbeitern weitreichenden Zugang zu unternehmensinternen Daten zu ermöglichen, um ihnen die bestmögliche Grundlage für ihre Entscheidungen zur Verfügung zu stellen, was diese dazu befähigt, die bestmöglichen Entscheidungen im Sinne des Unternehmens treffen zu können. Dieser Prozess wird üblicherweise unter dem Begriff der Demokratisierung des Zugangs zu Datenmaterial diskutiert.<sup>44</sup>

Für die konkrete Erhebung bedeutet dies, dass der Anteil derjenigen Entscheidungen, die auf der Basis von Big Data gewonnenen Informationen getroffen werden, erhoben werden muss. Da hierbei kaum damit zu rechnen ist, dass es konkrete Aufzeichnungen oder Aggregatdaten hierzu gibt, muss auf die Einschätzung der befragten Personen zurückgegriffen werden.

## 4.2. Weitere Variablen

### 4.2.1. Unternehmensgröße

Die Unternehmensgröße nimmt in dieser Betrachtung eine besondere Stellung ein, da nicht davon ausgegangen wird, dass die Größe des Unternehmens einen direkten Zusammenhang mit dem Unternehmenserfolg aufweist oder die Wirkung der unabhängigen Variablen beeinflusst. Jedoch muss davon ausgegangen werden, dass die Unternehmensgröße einen deutlichen Einfluss auf die Fähigkeiten eines Unternehmens, ausgefeilte Big Data Methoden zu implementieren, ausübt. So heben die Autoren des Big Data Reviews schon auf der ersten Seite hervor: „2015 was the year that Big Data went from being something that a few bigger

---

<sup>44</sup> Vgl. Computing Research, 2015, S. 11.

organisations were doing to being something that a majority of organisations were either doing or at the very least actively considering“<sup>45</sup> Schon diese Aussage zeigt, dass Big Data eine Domäne von Großunternehmen ist, denn selbst wenn ein großer Teil der Unternehmen mittlerweile überlegt, solche Methoden einzuführen, so stellt sich dennoch die Frage, wann diese Einführungsprojekte erfolgreich abgeschlossen sind und ihre Wirkung entfalten können. Gründe für die Dominanz großer Unternehmen in diesem Bereich lassen sich viele nennen: So kann man davon ausgehen, dass Großunternehmen auch ein größeres Datenvolumen im eigenen Unternehmen erheben und somit zunächst einmal einen besseren Zugang zur Kernressource von Big Data haben. Zwar lassen sich Daten auch extern beschaffen, jedoch zeigt sich, dass solche Daten zumindest aktuell für Entscheidungsträger weniger attraktiv sind, als interne Daten zur Prozessoptimierung.



Computing Research, 2015, S.9.

Ein denkbarer Grund mag auch sein, dass große Unternehmen leichter die Kosten und das Risiko für die Einführung solcher Methoden stemmen können. Wie bereits bei der Qualität der Datennutzung beschrieben, kommen zwar auch immer mehr anwendungsbezogene Softwarelösungen auf den Markt, jedoch ist hier der Gebrauch auf vordefinierte Einsatzgebiete beschränkt, die selbst auch in der Regel nicht allzu komplex sind, also eine eher geringe Anzahl an Variablen aufweisen. Außerdem ist es nur in sehr eingeschränktem

<sup>45</sup> Ebd., S.3.

Rahmen möglich, solche Lösungen und die zugrunde liegenden Modelle weiter zu entwickeln, wodurch ein großer Teil des Nutzens, den Big Data erbringen kann, verloren geht.<sup>46</sup>

Als drittes Argument lässt sich anführen, dass die notwendigen Kenntnisse für den Aufbau einer Big Data Struktur, also primär Kenntnisse in den Bereichen Programmierung und Statistik, bisher vor allem in der IT- und Finanzbranche zu finden waren. Während in den letzten Jahrzehnten sämtliche Branchen IT-Abteilungen aufgebaut und vergrößert haben, gilt dies für Statistiker nicht. Somit müssen beispielweise Industrie- oder Handelsunternehmen diese Kompetenzen neu auf dem Arbeitsmarkt suchen und entsprechende Abteilungen erst aufbauen. Da die Gehaltsstatistiken zeigen, dass Großunternehmen in der Regel höhere Gehälter und generell attraktivere Vergütungspakete anbieten<sup>47</sup> sowie oft auch auf eine wesentlich größere Bekanntheit auf dem Arbeitsmarkt zählen können, kann davon ausgegangen werden, dass es ihnen leichter fällt, entsprechende Bewerber für sich zu gewinnen und damit die notwendigen Kompetenzen in ihrem Unternehmen schneller aufzubauen. Dies stellt vor allem insofern einen wichtigen Aspekt dar, da die Anzahl an Arbeitskräften mit dem notwendigen Fachwissen sehr beschränkt ist und die Nachfrage das Angebot deutlich übersteigt.<sup>48</sup>

Allerdings sollte nicht ignoriert werden, dass schon heute verschiedene Faktoren zu einer starken Ausweitung der Nutzung von Big Data geführt haben. So hat sich die Wahrnehmung insofern verändert, dass es nicht mehr als etwas wahrgenommen wird, dass nur große Unternehmen können. Als Ursache hierfür lässt sich primär die Entstehung von Cloudlösungen nennen. Außerdem haben Publikationen und Fallstudien dazu beigetragen, den Nutzen von Big Data besser zu verstehen und zu kommunizieren. Zusätzlich haben die Unternehmen, die solche IT-Lösungen vertreiben, diese weiterentwickelt und dadurch die Nutzerfreundlichkeit erhöht, sowie ein Angebot an Weiterbildungen und Schulungen in ihre Verkaufsstrategie mitaufgenommen.<sup>49</sup>

Trotz dieser Entwicklungen soll zunächst die Hypothese beibehalten werden, dass Big Data in Großunternehmen signifikant stärker vertreten ist, was aber anhand des erhobenen Datenmaterials überprüft werden soll.

---

<sup>46</sup> Vgl. Iffert, 2016, S.17f.

<sup>47</sup> Vgl. Gehalt.de, 2015.

<sup>48</sup> Vgl. CapGemini, 2012, S.20.

<sup>49</sup> Vgl. Computing Research, 2015, S.7.

H<sub>2</sub>: Die Implementierung von Big Data und damit die Qualität und Quantität der Datenanalyse ist in großen Unternehmen stärker vorangeschritten, als in kleinen Unternehmen.

#### 4.2.2. Branche

Die Branche, in der das befragte Unternehmen tätig ist, soll als Drittvariable herangezogen werden. Da sich aus den Überlegungen des Market-Based-Views klar ergibt, dass die externen Kräfte, die auf ein Unternehmen einwirken und damit entscheidend zu seiner Profitabilität beitragen, von Markt zu Markt deutlich unterschiedlich sind, wäre es ein großer Fehler diesen Faktor zu ignorieren. Dies würde dazu führen, dass die Ursachen für eine potentiell generell profitable Branche verschleiert werden würden, was zu einer systematischen Überschätzung des Effektes der Datenanalyse auf den Unternehmenserfolg in einer solchen Branche führen würde. Außerdem muss beachtet werden, dass gewisse Einflüsse unterschiedliche Effekte auf unterschiedliche Branchen aufweisen. Ein Beispiel hierfür wäre das Zinsniveau, das auf unterschiedliche Branchen klar entgegengesetzte Auswirkungen haben kann. Während ein niedriges Zinsniveau die Finanzbranche generell vor Schwierigkeiten stellt, kommt es der produzierenden Industrie und ganz besonders dem Baugewerbe zugute, da es ihnen erlaubt, ihre Vorhaben kostengünstig zu finanzieren. Würde diese Variable vernachlässigt, würde man außerdem unterstellen, dass die Anwendungsmöglichkeiten und die Effekte des Einsatzes von Datenanalysen in allen Branchen dieselben wären. Zudem würde man indirekt unterstellen, dass in sämtlichen Branchen die Einführung solcher Methoden mit derselben Geschwindigkeit voranschreitet. Schon alleine aufgrund der sich stark unterscheidenden Geschäftsmodelle, sowie einer grundsätzlich konservativeren Kultur in manchen Branchen, muss jedoch davon ausgegangen werden, dass dies nicht der Fall ist. Als Beispiel kann hier die durchschnittliche Amortisierungsdauer von Investitionen gesehen werden. Während in der IT-Branche in der Regel kurze Entwicklungszyklen vorliegen, was zu verhältnismäßig vielen aber kleinen Investitionen führt, die sich schnell amortisieren müssen, muss in der Öl-Branche von langfristigen Investments ausgegangen werden, die sich erst über einen Zeitraum von 10–20 Jahren auszahlen. Im Falle solcher langfristigen Investments kann man wohl davon ausgehen, dass die breite Integration neuer Technologien in ein Unternehmen entsprechend länger dauert, als bei Unternehmen, die dazu gezwungen sind die neuesten Technologien direkt in ihre aktuellen Entwicklungen einfließen zu lassen. Zuletzt gilt auch der bereits

erwähnte Aspekt, dass gewisse Branchen einen „Startvorteil“ besitzen, da sie die benötigten Ressourcen oder Qualifikationen bereits in ihrem Kerngeschäft einsetzen und ihnen die erfolgreiche Umsetzung von Big Data Projekten leichter fallen dürfte, da sie mit dem Feld besser vertraut sind.

Für die konkrete Formulierung des Fragebogens sollte ursprünglich für diese Variable auf die Branchenkategorisierung der Wirtschaftskammer Österreich zurückgegriffen werden, welche folgende Branchen umfasst: Bank und Versicherung, Gewerbe und Handwerk, Handel, Industrie, Information und Consulting, Tourismus und Freizeitwirtschaft sowie Transport und Verkehr.<sup>50</sup> Nach den Pretests des Fragebogens stellte sich jedoch heraus, dass diese einigen Testpersonen zu grob erschien und sie nicht in der Lage waren ihr Unternehmen zufriedenstellend einzuordnen. Aus diesem Grund wurde im Folgenden die Branchengliederung der BARC Big Data Use Cases<sup>51</sup> übernommen, die eine deutlich feinere Untergliederung aufweist.

Schließlich ergeben sich aus diesen Überlegungen zwei konkrete Hypothesen, welche anhand des erhobenen Datenmaterials überprüft werden sollen:

H<sub>3</sub>: Die Einführung von Big Data Methoden weist in den verschiedenen Branchen signifikante Unterschiede in ihrem Fortschritt auf.

H<sub>3a</sub>: Branchen, die klassischerweise mit den Kernbereichen von Big Data vertraut sind (Datenverarbeitung und Statistik, also Finanz- und IT-Branche), weisen einen höheren Fortschritt in der Einführung von Big Data Methoden auf.

#### 4.2.3. Jurisdiktion und Datenschutzbestimmungen

Eine Kontrollvariable deren Einfluss schwer abzuschätzen ist, die jedoch eine bedeutende Rolle spielen könnte, ist die lokale Jurisdiktion. Insbesondere die örtlichen Datenschutz-Richtlinien können die Möglichkeiten eines Unternehmens, ihre Umwelt gezielt zu analysieren, signifikant beeinflussen. Hieraus ergibt sich allerdings auch, dass sich der Einfluss auf diejenigen Faktoren beschränken müsste, die im Kapitel zum Market-Based-View behandelt wurden. In gewissem Umfang könnte jedoch auch die Analyse interner Faktoren beeinflusst werden, da vor allem personenbezogene Daten der Mitarbeiter sowie

---

<sup>50</sup> Vgl. WKO, 2016.

<sup>51</sup> Vgl. Big Data Use Cases 2015, S.43.

deren Verhaltensweisen während ihrer Arbeitszeit durchaus auch Gegenstand von Datenschutzbestimmungen sein können.

Es ist zu erwarten, dass der Einfluss dieser Variable aufgrund der hier untersuchten Länder keinen allzu großen Effekt aufweisen dürfte. Der Grund hierfür ist der stark angegliche Datenschutz innerhalb der EU, was auch für die Schweiz gilt, die in vielen Bereichen der Gesetzgebung stark mit der EU zusammenarbeitet. Dennoch soll die Variable im Modell aufgenommen werden, da sie bei der Anwendung des Modells auf andere Regionen durchaus einen bedeutenden Einfluss haben könnte.

Ganz konkret soll in dieser Variable schlicht das Land vermerkt werden, in dem der Befragte seine Hauptanstellung hat, da davon ausgegangen wird, dass er dort auch den größten Teil seiner Tätigkeit verrichtet und seine Aussagen damit die Situation im jeweiligen Land widerspiegeln. Dies ermöglicht es außerdem, die Variable als Kontrollvariable für die Institutionenlandschaft des jeweiligen Landes zu nutzen. Die Bedeutung dieser für das Unternehmensergebnis ist theoretisch auf den Institution-Based-View zurückzuführen, der bisher nicht näher erläutert wurde, da er für diese Arbeit, außerhalb dieser Variable keine große Bedeutung aufweist. Grundlegend kann man sagen, dass dieser Bereich des strategischen Managements die Wirkungsweise von politischen und gesellschaftlichen Institutionen auf Unternehmen untersucht. Hierzu zählen Bereiche wie die Gesetzgebung und Kultur.<sup>52</sup> Insgesamt ist diese Variable somit als Kontrollvariable zu verstehen, die zur Kontrolle einer Vielzahl an landesspezifischen Einflüssen dient, die teilweise kaum isoliert werden können.

#### 4.2.4. Eigentümerstruktur

Die Eigentümerstruktur ist im Kontext dieser Arbeit von Bedeutung, da die Überlegungen zur Entscheidungstheorie zum einen darauf hindeuten, dass es im Zusammenhang mit der Operationalisierung des Unternehmenserfolgs gewisse Probleme bei der korrekten Auswertung der Ergebnisse geben könnte. Um einen systematischen Fehler zu vermeiden muss deshalb diese Variable erhoben werden und auf eventuelle Abweichungen zwischen den Ausprägungen getestet werden.

Bezüglich weniger technischer und mehr allgemeiner Gründe dieser Variable lässt sich festhalten, dass die Auswirkungen der Eigentümerstruktur auf die Steuerung von

---

<sup>52</sup> Vgl. Peng et al., 2009, S.64.

Mitarbeitern und damit deren Entscheidungen kaum erforscht sind.<sup>53</sup> Überlegungen hierzu lassen sich nur indirekt von Modellen und der zu Familienunternehmen existierenden Literatur ableiten. Von besonderer Bedeutung dabei ist in beiden Fällen das Principal-Agent-Modell, welches für das Verständnis von Entscheidungen innerhalb hierarchischer Systeme eine große Bedeutung hat. Hierbei lässt sich aus zweierlei Perspektive argumentieren: Zum einen, dass Familien- sowie andere inhabergeführte Unternehmen, durch die Personalunion von Eigentümern und Unternehmensleitung kaum Agency-Kosten für die Unternehmensüberwachung entstehen, da die Eigentümerinteressen direkt in Entscheidungen umgesetzt werden können.<sup>54</sup> Dies führt auf der anderen Seite jedoch auch zu einigen Problemen, welche primär in der mangelnden Kontrolle der Geschäftsführung liegen. Zum einen wird ein solches Unternehmen in wesentlich geringerem Ausmaß vom Kapitalmarkt diszipliniert und kontrolliert, zum anderen müssen sich die Führungspersonen nicht auf dem Arbeitsmarkt für Manager behaupten und unterliegen deshalb nicht dessen disziplinierender Wirkung.<sup>55</sup> So können Geschäftsführer selbst nach schwerwiegenden Fehlentscheidungen oft ihre Position erhalten, da diese durch die familiäre Bindung abgesichert ist.<sup>56</sup> Als Folge werden Entscheidungen in solchen Unternehmen daher öfter intuitiv und weniger fakten- und kennzahlenbasiert getroffen. Außerdem werden folglich auch entsprechende entscheidungsunterstützende Systeme wie die Datenanalyse weniger genutzt.<sup>57</sup>

Betrachtet man die zweite Perspektive, so gilt es zunächst festzuhalten, dass insbesondere Großkonzerne meist nicht inhabergeführt sind, sondern zu diesem Zweck einen Vorstand berufen. Dieser kann abgesehen von der Berufung und der jährlichen Entlastung in der Regel unabhängig von den Eigentumsverhältnissen agieren. Hieraus ergibt sich aber auch eine Rechenschaftspflicht des Vorstandes gegenüber den Eigentümern. Da willkürliche Entscheidungen ohne fundierte Entscheidungsbasis kaum im Interesse der Eigentümer sein dürften, insbesondere falls sich die Situation des Unternehmens verschlechtert, ist es für den Vorstand von großer Bedeutung fundiertes Datenmaterial zur Verfügung zu haben, mit dessen Hilfe die getroffenen Entscheidungen gerechtfertigt werden können. Somit ist davon auszugehen, dass die Einführung von Big Data und damit der Zugang zu einer besseren Datengrundlage für solche Unternehmen nicht nur wirtschaftliche Bedeutung hat, sondern

---

<sup>53</sup> Vgl. Posch, Speckbacher, 2012 S.9.

<sup>54</sup> Vgl. Fama, Jensen, 1983, S.306.

<sup>55</sup> Vgl. Posch, Speckbacher, 2012, S.6.

<sup>56</sup> Vgl. Schulze et al., 2001, S.103.

<sup>57</sup> Vgl. Posch, Speckbacher, 2012, S. 4.

auch für die Rechtfertigung von Entscheidungen genutzt werden kann und somit ein Eigeninteresse des Vorstandes an der Implementierung solcher Methoden besteht.

Beide Perspektiven kommen somit zu dem Schluss, dass die Datenanalyse in managementgeführten Unternehmen stärker ausgeprägt ist als in inhabergeführten. Theoretisch müsste sich aus diesen Überlegungen, unter Beachtung des Hauptaspekts Disziplinierung durch einen freien Markt, folgende Abstufung in der Anwendung von Datenanalyseinstrumenten in absteigender Reihenfolge ergeben:

- Aktienunternehmen im Streubesitz
- Aktienunternehmen mit Mehrheitseigentümer
- Inhabergeführt ohne familiären Zusammenhang
- Familienunternehmen
- Inhabergeführt mit einem einzigen Inhaber
  
- Öffentliche Unternehmen (Der Vollständigkeit halber)

Es ist jedoch fragwürdig, ob diese Abstufung im Detail statistisch nachgewiesen werden kann, weshalb der signifikante Unterschied vermutlich zwischen den Clustern managementgeführt (1-2) und inhabergeführt (3-5) zu finden sein wird.

Die These hierfür lautet:

H<sub>4</sub>: Aufgrund der Pflicht das eigene Handeln zu rechtfertigen, ist die Implementierung von Big Data und damit die Qualität und Quantität der Datenanalyse in inhabergeführten Unternehmen weniger weit vorangeschritten als in managementgeführten Unternehmen.

Ein weiterer möglicher Effekt ergibt sich aus den Überlegungen zur Entscheidungstheorie. Geht man von einem normativen Modell aus, so besteht eine zentrale Voraussetzung darin, dass es ein einziges dominantes Ziel geben muss, da mehrere Ziele untereinander inkompatibel sein könnten.<sup>58</sup> An diesem Punkt besteht jedoch die Möglichkeit, dass inhabergeführte Unternehmen im allgemeinen, insbesondere jedoch Familienunternehmen, eine von der im Folgenden ausgeführten Operationalisierung des Unternehmenserfolgs abweichen, welche stark auf einen rendite- und damit investorenorientierten Erfolgsbegriff

---

<sup>58</sup> Vgl. Kirchler, 2003, S.46.

ausgelegt ist. Inhabergeführte Unternehmen dagegen könnten durchaus bereit sein, eine nachhaltige Bewirtschaftung des Unternehmens über aktuelle Renditen zu stellen, da sie wesentlich stärker an das Unternehmen gebunden sind damit die theoretischen Kosten für die Verschiebung ihrer Investments bedeutend höher wären, als dies bei externen Investoren, beispielweise Aktienbesitzern der Fall ist. Dies würde sich durch eine systematisch niedrigere Ausprägung der abhängigen Variable bei Inhabergeführten Unternehmen zeigen. Deshalb ist die Erhebung der Eigentümerstruktur nicht nur aufgrund der potentiellen Wirkung auf die unabhängige Variable von Bedeutung, sondern auch aufgrund der potentiellen direkten Wirkung auf die abhängige Variable.

#### 4.3. Abhängige Variable - Unternehmenserfolg

Die Operationalisierung des Unternehmenserfolgs als zentrale abhängige Variable des Modells stellt eine gewisse Herausforderung dar, da dieser aus sehr unterschiedlichen Blickwinkeln betrachtet werden kann und somit sehr viele Ansichten darüber existieren, wann ein Unternehmen erfolgreich ist. Die vermutlich wichtigsten Faktoren hierfür dürften die eigene Position im Kreise der Stakeholder eines Unternehmens sowie der persönliche Hintergrund sein. Ersteres spielt eine Rolle, da man davon ausgehen kann, dass Shareholder den Erfolg eines Unternehmens vermutlich über die Rendite ihrer Anteile definieren werden, wohingegen Gewerkschaften dies über Arbeitsbedingungen und die Schaffung von Arbeitsplätze tun werden und Umweltorganisationen über die Nachhaltigkeit und Umweltfreundlichkeit eines Unternehmens. Der zweite, individuelle Punkt dürfte eine Rolle spielen, da man davon ausgehen kann, dass jemand mit liberalem Weltbild den Erfolg anders definieren wird als jemand mit sozialistischer Prägung, um nur ein Beispiel zu nennen. In dieser Arbeit soll ein klar wirtschaftlicher Ansatz vertreten werden, der den Shareholder Value in den Vordergrund stellt. Grund hierfür ist zum einen der Aufbau der Arbeit, die auf einem Fundament klassischer betriebswirtschaftlicher Theorien fußt, zum anderen aber auch die im Vorfeld festgelegten Ziele der Arbeit, die das Finanzergebnis ins Zentrum der Betrachtung stellen.

Hierbei gilt es insbesondere das ökonomische Prinzip zu beachten, dass zwei zentrale Grundsätze für die Entscheidungen und das Handeln von Menschen definiert. Dies ist zunächst die Rationalität von Entscheidungen. Suboptimale Entscheidungen werden demnach nicht bewusst gewählt, sondern sind lediglich als Ergebnis imperfekter Entscheidungsprozesse, mangelnder Informationen und als Folge von Risiken zulässig. Der

zweite Punkt betrifft das Prinzip der Nutzenmaximierung. Dies bedeutet zum einen, dass ein klar nutzenorientiertes und daher beispielsweise nicht altruistisches Handeln angestrebt wird und zum anderen, dass Sättigungseffekte ausgeschlossen werden.<sup>59</sup> Hierbei ist allerdings explizit darauf hinzuweisen, dass dies noch nicht definiert worin der erzielte Nutzen liegt. So ist es durchaus denkbar, dass in der gängigen Ansicht als altruistisch geltende Handlungen wie beispielsweise Spenden oder Unterstützung von Schwachen einen Nutzen im Sinne emotionaler Befriedigung oder Prestige generieren. Demzufolge ist auch die mögliche unterschiedliche Zielsetzung von Familienunternehmen (z.B. Erhalt des Unternehmens für spätere Generationen) und Shareholder Value orientierten Unternehmen (z.B. Maximierung der Rendite) kein Widerspruch zu diesem Prinzip. Steht jedoch ein Familienunternehmen vor zwei Entscheidungsalternativen, welche beide denselben Beitrag zum Überleben des Unternehmens liefern, eine davon jedoch eine höhere Rendite verspricht, so verlangt das ökonomische Prinzip die zweite Alternative zu wählen. Ein drittes, allerdings durchaus umstrittenes Prinzip besteht darin, dass Entscheidungen nach dem ökonomischen Prinzip bei voller Information, also unter vollem Bewusstsein sämtlicher möglicher Alternativen und der damit verbundenen Konsequenzen, getroffen werden. Diese Annahme ist für makroökonomische Modelle, bei denen die unsichtbare Hand des Marktes, wirkend durch das Konzept des Grenznutzens und die Beteiligung einer großen Anzahl an Marktteilnehmern, durchaus sinnvoll. Für Entscheidungen auf betrieblicher Ebene muss dies jedoch strikt abgelehnt werden, da kein Entscheidungsträger, selbst nach intensivster Analyse, behaupten kann sämtliche Konsequenzen seines Handelns zu kennen. Jedoch ist hier erneut anzumerken, dass die Datenanalyse einen bedeutenden Beitrag zur Informationsgewinnung liefert und zur Folge hat, dass sich reale Entscheidungen der theoretischen Optimalsituation stärker annähern.

Trotz dieser ersten Einschränkung ist die Auswahl von Indikatoren des latenten Konstrukts Erfolg alles andere als trivial. So sind Operationalisierungen über das Geschäftsvolumen, die Rendite oder auch die Effizienz von Unternehmen denkbar. Im Sinne des Shareholder Value Ansatzes, der insbesondere die Interessen von Investoren berücksichtigt, soll hierbei zunächst die Kapitalrendite (Return on Investment) als zentrale Messgröße des Erfolgs herangezogen werden. Die zentrale Bedeutung dieser Kennzahl lässt sich außerdem aus dem Du-Pont-Schema ableiten, welches bis heute das wohl bekannteste und verbreitetste Kennzahlensystem darstellt und insbesondere im Controlling von großer Bedeutung ist. Die Bedeutung dieser Kennzahl zur Messung von Unternehmenserfolg kann man auch daran

---

<sup>59</sup> Vgl. Meyer, 2000, S.10.

erkennen, dass sie von der wohl größten Studie zur Erfolgsfaktoren Forschung, Profit Impact of Market Strategies (PIMS), als Operationalisierung für den Unternehmenserfolg verwendet wird.<sup>60</sup> Allerdings besteht eine berechtigte Überlegung darin, ob der ROI wirklich den Erfolg des Unternehmens abbildet, da hierin auch die Zinsaufwendungen für Fremdkapital dem positiven Ergebnis zugerechnet werden. Im Sinne eines holistischen Ansatzes, der den Fokus auf die Wertschöpfung eines Unternehmens legt, ist dies durchaus sinnvoll. Im Sinne eines reinen Shareholder-Value Ansatzes dagegen muss dies abgelehnt werden, da die Fremdkapitalkosten den Gewinn schmälern. Hierfür wäre der Return on Equity (ROE) die treffendere Kennzahl.

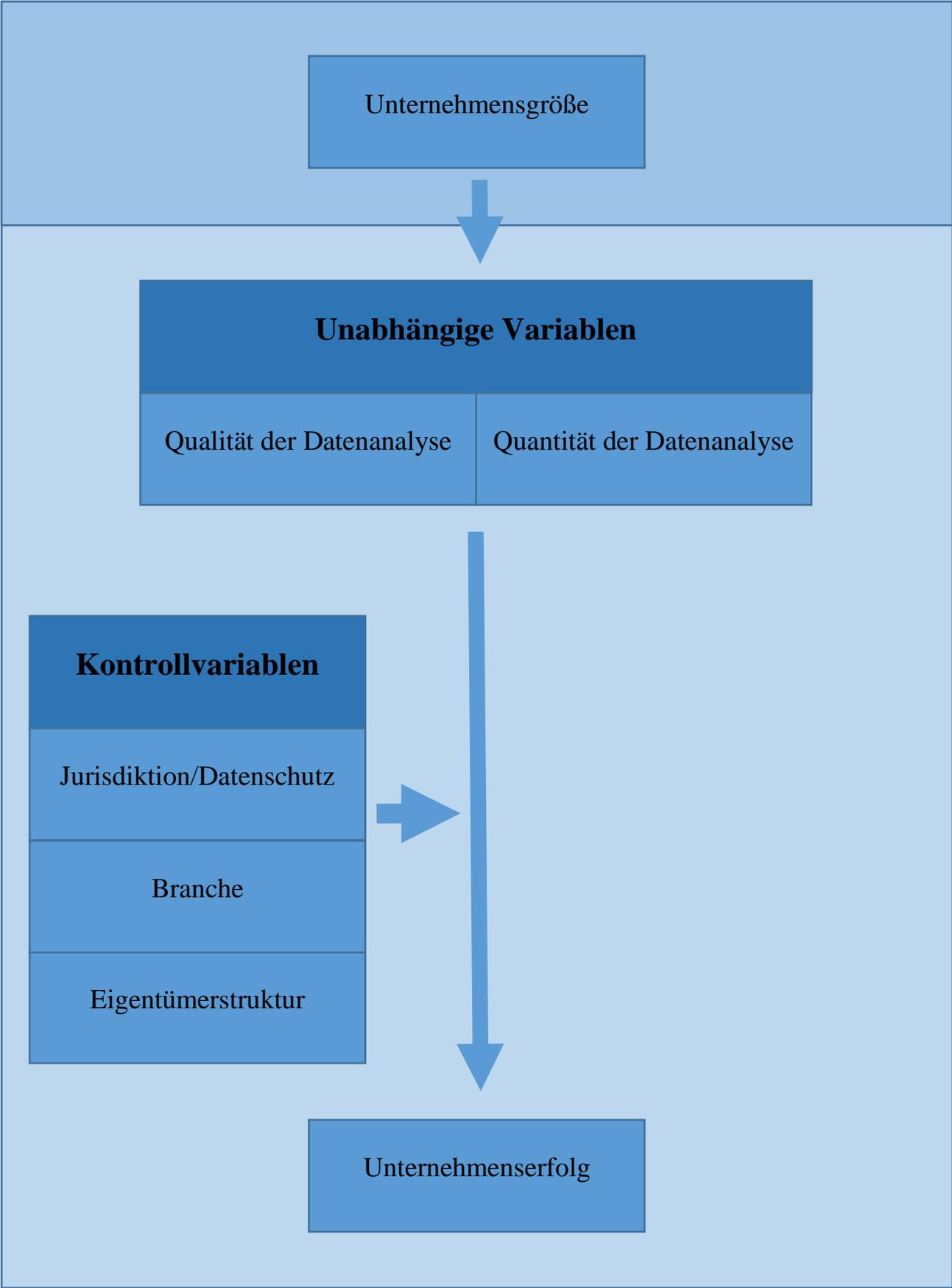
Jedoch muss durchaus angezweifelt werden, ob diese Kennzahl für langfristige, an das Unternehmen gebundene Eigentümer, wie beispielsweise bei Familienunternehmen die optimale Definition von Unternehmenserfolg darstellt. Deshalb soll noch eine weitere Kennzahl zur Bewertung des Unternehmenserfolgs herangezogen werden, nämlich der absolute Gewinn, der insbesondere die Interessen von Familienunternehmen und anderen inhabergeführten Unternehmen abdecken soll.

#### 4.4. Das Modell

Aus diesen Variablen ergibt sich ein Gesamtmodell, das die Qualität und Quantität der Datenanalyse in einem Unternehmen als unabhängige Variablen nutzt, und daraus den Einfluss auf den Unternehmenserfolg berechnet. Dabei sollen außerdem die Effekte der Variablen Jurisdiktion oder auch, das Land, in dem das Unternehmen tätig ist, Eigentümerstruktur sowie Branche kontrolliert werden. Dieses Kernmodell soll außerdem um den Faktor der Unternehmensgröße erweitert werden, der zwar nicht direkt auf die kausale Beziehung zwischen unabhängigen und abhängiger Variable wirkt, jedoch davon auszugehen ist, dass ein gewisser Einfluss auf die unabhängigen Variablen vorliegt, der nicht außer Acht gelassen werden soll. Somit ergibt sich folgende grafische Darstellung des Modells:

---

<sup>60</sup> Vgl. Buzzel Gale, 1987, S.22.



Überblick über die Hypothesen:

- H<sub>1</sub>: Je höher die Quantität und Qualität der Datenanalyse in einem Unternehmen ist, desto erfolgreicher ist dieses Unternehmen.
- H<sub>2</sub>: Die Implementierung von Big Data und damit die Qualität und Quantität der Datenanalyse, ist in großen Unternehmen stärker vorangeschritten, als in kleinen Unternehmen.
- H<sub>3</sub>: Die Einführung von Big Data Methoden weist in den verschiedenen Branchen signifikante Unterschiede in ihrem Fortschritt auf.
- H<sub>3a</sub>: Branchen die klassischerweise mit den Kernbereichen von Big Data vertraut sind (Datenverarbeitung und Statistik, also Finanz- und IT-Branche), weisen einen höheren Fortschritt in der Einführung von Big Data Methode auf.
- H<sub>4</sub>: Aufgrund der Pflicht das eigene Handeln zu rechtfertigen, ist die Implementierung von Big Data und damit die Qualität und Quantität der Datenanalyse in inhabergeführten Unternehmen weniger weit vorangeschritten als in managementgeführten Unternehmen.

## 5. Methoden der Datenerhebung und Vorgehensweise

### 5.1. Kontaktaufnahme und Erhebungsmethodik

Um Zugang zu geeigneten Interviewpartnern zu bekommen, wurde zunächst versucht den Kontakt zu den Arbeitgeberverbänden in den drei relevanten Ländern Deutschland, Österreich und Schweiz herzustellen. Dies wären konkret der Deutsche Arbeitgeberverband, die Wirtschaftskammer Österreich und der Schweizerische Arbeitgeberverband. Leider war von keinem dieser Verbände trotz vielfacher Nachfrage eine Antwort zu erhalten. Lediglich die WKO erteilte der Anfrage eine konkrete Absage, jedoch auch erst nach intensivsten Bemühungen. Dabei wurde als Begründung Datenschutz und Unparteilichkeit angeführt.

Aufgrund dieses Fehlschlags wurde in der Folge versucht, Zugriff auf eine Marketing Datenbank zu erhalten, was in Form der Amadeus-Datenbank, die vom Creditreform e.V. betrieben wird, gelang. Der Zugriff auf diese Datenbank erfolgte über die Lizenz der Wirtschaftsuniversität Wien. Hier ist der Vollständigkeit halber anzumerken, dass die

Amadeus-Datenbank an der Wirtschaftsuniversität Wien nicht direkt erreicht werden kann, sondern in der Datenbank Orbis integriert ist, die zusätzlich M&A Daten enthält. Die Extraktion erfolgte zu zwei Zeitpunkten: Am 20.04.2017 wurde die erste Extraktion vorgenommen, hierbei wurden die Daten von 16557 Unternehmen entnommen, womit die genutzten Kontaktdaten dem Stand dieses Datums entsprechen. Um eine möglichst breite und aktuelle Datengrundlage zu erhalten wurde dieser Vorgang nach Abschluss der Datenerhebung am 06.05.2017 wiederholt um eine höhere Anzahl an Finanzdaten von 2016 zu erhalten. Durch Veränderungen in der Datengrundlage wurden hierbei trotz gleicher Suchkriterien lediglich die Daten von 16431 Unternehmen erhoben. Dies führte auf der einen Seite dazu, dass 5 Fragebögen nicht verwendet werden konnten, da die dazugehörigen Finanzdaten im aktuellen Datensatz nicht enthalten waren. Auf der anderen Seite wurden je nach Variable dadurch 20-40 neue, auswertbare Beobachtungen für das Jahr 2015 gewonnen. Das eigentliche Ziel allerdings, auswertbare Daten für 2016 zu gewinnen, konnte nicht realisiert werden, da hier noch immer niedrige zweistellige Beobachtungszahlen für die Finanzdaten vorhanden waren, was keine sinnvolle Analyse zuließ. Die Entscheidung die Finanzdaten aus der Orbis-Datenbank zu entnehmen und nicht auf die selbst im Fragebogen erhobenen Finanzdaten zurückzugreifen, wurde aufgrund von zwei Problemen getroffen. Zum einen zeigte der stichprobenartige Vergleich einiger Datensätze mit den öffentlichen Geschäftsberichten, dass die Genauigkeit der Orbis-Daten deutlich höher ist und die selbst erhobenen in einigen Fällen Falschangaben enthielten. Zum anderen war die Datenbasis hierdurch deutlich größer, da eine hohe Anzahl von Unternehmen nicht bereit war, ihre Finanzdaten anzugeben. Insgesamt konnte somit die Fülle des auswertbaren Datenmaterials deutlich erhöht werden. Genauere Angaben hierzu finden sich in Kapitel 5.3..

Die Suchkriterien für die Datenbank waren diejenigen, die entsprechend europäischem Recht eine große Kapitalgesellschaft ausmachen und im nationalen Recht in Deutschland in § 267 HGB und in Österreich in § 221 UGB verankert sind:

- 20 Millionen € Bilanzsumme
- 40 Millionen € Umsatz
- 250 Mitarbeiter

Hierbei mussten jedoch gewisse Zugeständnisse an das vorliegende Datenmaterial gemacht werden. So stellte sich heraus, dass die Angaben zur Mitarbeiterzahl in der Orbisdatenbank bei vielen Unternehmen kaum zuverlässig erscheinen, so sind beispielweise für keines der Unternehmen des Allianzkonzerns Mitarbeiterdaten vorhanden. Bedenkt man außerdem den

Fokus dieser Arbeit auf den finanziellen Aspekt von Unternehmen, erschien es sinnvoll Bilanzsumme und Umsatz als Selektionskriterien zu nutzen. Des Weiteren wurde nicht im strengen Sinne des Gesetzes darauf geachtet, dass die Kriterien in zwei aufeinanderfolgenden Jahren überschritten wurden, sondern es wurden die jeweils aktuellsten vorhandenen Daten genutzt. Außerdem ist anzumerken, dass es sich bei Unternehmen im Sinne dieser Arbeit um Legaleinheiten handelt und keine Konzernstrukturen beachtet wurden. So wird beispielsweise Audi als selbstständiges Unternehmen und nicht unter dem Dach von VW geführt. Bedingt ist dies zum einen technisch, durch die Art der Gewinnung der Kontaktdaten, da dies der Form entspricht, in der Orbis die Unternehmen führt. Andererseits macht diese Art der Erhebung auch aus theoretischer Sicht Sinn, da zum einen nicht pauschal beurteilt werden kann wie eigenständig ein Tochterunternehmen agiert und lieber eine quasi Doppelbeteiligung sehr stark ins Mutterunternehmen integrierter Töchter in Kauf genommen wird, als einen großen Teil der Unternehmen, die per se eigenständig unter einem gemeinsamen Konzerndach operieren, auszuschließen. Zum anderen erlaubt dies auch eine wesentlich klarere Zuordnung der gewonnenen Daten zu verschiedenen Branchen, da viele Konzerne, die in unterschiedlichen Branchen aktiv sind, für diese auch unterschiedliche Tochterunternehmen gründen. Somit müsste ohne eine solche Trennung mit einer starken Verzerrung des Datenmaterials gerechnet werden.

Das letzte „Selektionskriterium“ dagegen war rein technischer Natur. So waren auch viele Unternehmen in der Datenbank aufgeführt, zu denen schlicht keine Kontaktmailadresse aufgeführt war und welche somit nicht kontaktiert werden konnten. Dies führte zu einer Anzahl von 14300 Emailadressen. In einem weiteren Schritt wurden sämtliche doppelt geführten Adressen eliminiert. Nach diesem Schritt verblieben noch 13068 Emailadressen, die schlussendlich kontaktiert wurden.

Die konkrete Kontaktaufnahme erfolgte ebenso wie die Fragebogenerstellung, über die Onlineplattform [socisurvey.de](https://www.socisurvey.de). Hierbei sind insbesondere drei Funktionen hervorzuheben: Zum einen die Möglichkeit personalisierte Links zu verschicken, was erstens sicherstellte, dass die Daten einem Unternehmen eindeutig zugeordnet werden konnten, was wiederum wichtig war, um die gewonnenen Daten im späteren Verlauf mit den Finanzdaten aus der Orbis-Datenbank zu kombinieren. Zweitens wurde damit sichergestellt, dass eine Mehrfachbeteiligung eines einzelnen Unternehmens, indem mehrere Mitarbeiter den Fragebogen ausfüllen, verhindert wurde, womit jede Beobachtung ein eigenes Unternehmen widerspiegelt und damit keine Mehrfachteilnahme möglich war. Drittens ist es so den

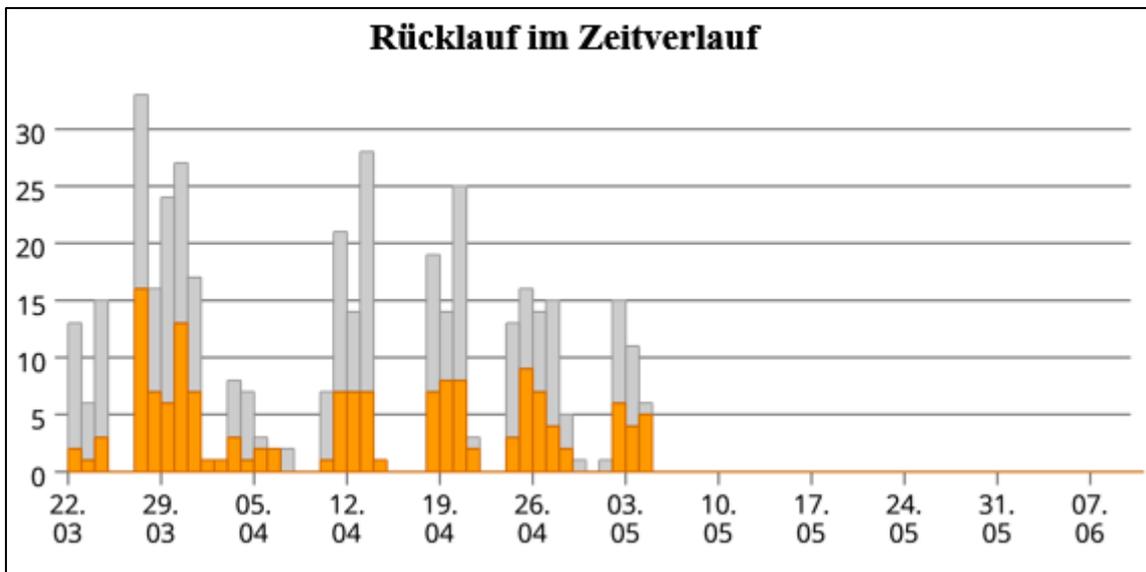
Umfrageteilnehmern möglich, auch über die unterschiedlichen Erinnerungswellen hinweg, bereits begonnene Datensätze abzuschließen.

Der Versand der Fragebögen erfolgte mit Hilfe der Serienmailfunktion von Socisurvey. Um zu vermeiden, dass die Anfragen durch einen massenhaften Versand als Spam klassifiziert werden, wurden die Emailadressen in Cluster unterteilt, wobei jeder zwischen 1689 und 2174 Adressen enthielt. Die unterschiedlichen Größen sind zum einen dadurch bedingt, dass die Eliminierung der Doppelten Emailadressen mit den vorhandenen Werkzeugen erst nach der Einteilung der Cluster möglich war, die ursprünglich alle 2000 Emailadressen enthielten. Der letzte Cluster mit 2174 Emailadressen ergab sich durch den Restwert von wenigen hundert Adressen, die nicht eigens geführt, sondern zum letzten Cluster hinzugefügt wurden. Anschließend erfolgte der Versand der Einladung an jeweils einen Cluster täglich um 10:30 Uhr vom 22.03.2017 - 30.03.2017 wobei Samstag und Sonntag keine Einladungen verschickt wurden, um zu vermeiden, dass diese in den über das Wochenende kommenden Emails untergehen.

Vom 10.-18.04.2017 wurde eine erste Welle von Erinnerungsmails verschickt. Dabei wurden sämtliche Adressaten kontaktiert, die bisher keinen vollständigen Datensatz generiert hatten. Jedoch wurden, anders als bei der ersten Aussendung, nicht alle Adressaten wie in der ursprünglichen Email gleichzeitig kontaktiert, sondern ein randomisierter Versand zwischen 10:30 Uhr und 15:00 Uhr durchgeführt. Erneut wurden Samstag und Sonntag ausgelassen.

Die Zweite und letzte Erinnerungswelle wurde zwischen 24.04.2017 und 03.05.2017 verschickt, wobei erneut vom 29.04.2017 - 01.05.2017 aufgrund des Wochenendes und Feiertages keine Anfragen verschickt wurden. Es wurde, wie bei der ersten Erinnerung, erneut ein zeitlich randomisierter Versand zwischen 10:30 Uhr und 15:00 Uhr gewählt.

Insgesamt führte dieses Vorgehen bis zum 05.06.2017 zu einer Gesamtanzahl von 978 Aufrufen, 401 angelegten Datensätzen und 153 vollständig abgeschlossenen Datensätzen. Die zeitliche Abfolge der Rückmeldungen kann der nachfolgenden Grafik entnommen werden. Angelegte, aber nicht abgeschlossene Datensätze sind dabei grau dargestellt, abgeschlossene Datensätze orange:



Sieht man von den Rückmeldungen am 27.04.2017 und 30.04.2017 ab, lässt sich kein Rückgang der Rückmeldungen (abgeschlossene Datensätze) erkennen. Dies lässt zum einen darauf schließen, dass die Anzahl der Rückmeldungen durch erneute Erinnerungen noch deutlich gesteigert werden hätte können, was allerdings aufgrund des zeitlichen Rahmens dieser Arbeit nicht möglich war. Zum anderen muss deshalb davon ausgegangen werden, dass trotz Maßnahmen wie Clusterbildung und zeitlicher Randomisierung eine Vielzahl der Anfragen in SPAM-Ordner gelandet ist, die Zielunternehmen die Anfrage also erst durch eine Erinnerungsmail bekamen.

Wie in der nachfolgenden Grafik zu sehen ist, schloss leider nur etwa ein Drittel derjenigen, die mit der Befragung begonnen hatten, diese auch ab (153 von 404). Die Gründe hierfür können vielfältig sein, zwei Punkte stechen jedoch besonders hervor. Zum einen scheint es eine größere Anzahl an Personen gegeben zu haben, die den Fragebogen nur aus einem ersten Interesse heraus aufrufen und abbrechen sobald die Fragen begannen. (Abbruch auf Seite 4) Zum anderen zeigt sich auf den darauffolgenden zwei Seiten eine stark abnehmende Anzahl an Abbrüchen, wobei es bei Seite 6 möglich sein könnte, dass der Grund für die Abbrüche, der sehr technische Inhalt der Fragen war.

<b>Ausstiegsseiten</b>				
Letzte bearbeitete Seite	Datensätze abgeschlossen / gesamt / kumulativ			
Seite 13	153	155	155	
Seite 12	0	8	163	
Seite 11	0	13	176	
Seite 10	0	5	181	
Seite 9	0	16	197	
Seite 8	0	14	211	
Seite 7	0	7	218	
Seite 6	0	22	240	
Seite 5	0	48	288	
Seite 4	0	83	371	
Seite 3	0	9	380	
Seite 2	0	15	395	
Seite 1	0	2	397	
-	0	7	404	
<b>Gesamt</b>	<b>153</b>	<b>404</b>		

Insgesamt wurden 978 Aufrufe (Klicks) für diesen Fragebogen aufgezeichnet (einschließlich versehentlicher doppelter Klicks, Aufrufe durch Suchmaschinen, ...).

## 5.2. Fragebogenerstellung

Der Fragebogen wurde Großteils selbständig auf Basis der Operationalisierung erstellt. Folgende drei Fragen wurden jedoch, in leicht veränderter Form, dem Fragebogen zur BARC Studie Big Data Use Cases von 2015 entnommen, zu dem die Autoren freundlicherweise Zugriff gewährten:

- Welche der folgenden Datentypen werden in Ihrem Unternehmen für Big Data Analysen genutzt?
- Welche der folgenden Werkzeuge werden in Ihrem Unternehmen zu Datenanalyse genutzt?
- In welcher Branche ist Ihr Unternehmen tätig?

Zur letzten Frage ist anzumerken, dass, wie bereits in Kapitel 4.2.2. beschrieben wurde, ursprünglich die Branchengliederung der WKO geplant war, sich während der Pretests des Fragebogens jedoch ergab, dass diese zu grob erscheint und damit keine klare Zuordnung vieler Unternehmen ermöglicht.

Dieser Pretest fand im Vorfeld der Aussendung statt und diente zur Sicherstellung der Qualität und Objektivität des Fragebogens. Durchgeführt wurde er von vier Personen die den Fragebogen in Bezug auf Rechtschreibung, Grammatik, Inhalt, Struktur und

Plausibilität überprüfen. Dies waren eine Germanistin, ein Psychologe, eine Mitarbeiterin in der strategischen Planung und der Betreuer dieser Arbeit, der als Abteilungsleiter bei einem Energieunternehmen tätig ist.

Während der ersten Versandwelle wurde schnell klar, dass die frühe Nennung der Finanzkennzahlen einen abschreckenden Faktor darstellt, der die Teilnehmer dazu veranlasste die Befragung abubrechen. Um an belastbare Daten zu gelangen, wurden deshalb in sämtlichen folgenden Ausendungen drei konkrete Veränderungen vorgenommen. Zum einen wurde die Abfrage der Finanzkennzahlen an das Ende des Fragebogens verschoben, um im Falle eines Abbruchs zumindest die bisherigen deskriptiv nutzbaren Daten zu erhalten. Zum anderen wurde kenntlich gemacht, dass die Angabe der Finanzkennzahlen zwar erwünscht ist, jedoch keine zwingende Voraussetzung darstellt, um den Fragebogen abzuschließen. Drittens wurden alternativ zur genauen Angabe der Kennzahlen Auswahlfragen eingeführt, die diese Kennzahlen in Clusterform abfragen, um so den Datenschutzbedürfnissen der Unternehmen entgegen zu kommen.

Dabei wurden die Cluster für Umsatz, Gewinn und Kapitalvolumen generiert, indem die Daten für 2015 aus Orbis entnommen wurden, die 10% Quantile berechnet. Diese Werte wurden als Basis für eine grobe, aber runde Näherung genommen, um die Clustergrenzen zu definieren.

### 5.3. Operationalisierung konkret

#### **Qualität der Datenanalyse:**

Für die Qualität der Datenanalyse spielt, wie in Kapitel 4.1.1. aufgearbeitet, insbesondere die Zeit die zur Auswertung notwendig ist, sowie die dafür verwendete Softwarelösung eine Rolle. Ersteres fließt in die Modelle mit zwei Variablen mit ein, die durchschnittliche Berichtszeit für singuläre und diejenige für repetitive Ereignisse. Die verwendete Softwarelösung dagegen muss aufgrund der vorliegenden Kodierung mit einzelnen dichotomen Variablen miteinbezogen werden.

### **Quantität der Datenanalyse:**

Um das Konstrukt Quantität der Datenanalyse zu operationalisieren wurden Indexwerte aus den erhobenen Daten gebildet. So wurde das arithmetische Mittel der Werte der Antworten von Seite 8 (DA01\_01-DA01\_06) errechnet, was den Indexwert für Anwendungsquantität ergab. Auf dieselbe Weise wurde ein Indexwert für die Krisenerkennung durch Datenanalyse aus den Antworten von Seite 11 (DA07\_01 & DA07\_02) gebildet. Die dritte Variable die dem Konstrukt der Quantität der Datenanalyse zuzuordnen ist, ist die Demokratisierung des Datenzugriffs, diese soll im Folgenden etwas genauer dargelegt werden:

### **Demokratisierung des Datenzugriffs**

Da diese Variable als Auswahlfrage erhoben wurde, wurde durch Socisurvey automatisch eine Binärvariable für jede Auswahlmöglichkeit generiert. Grundsätzlich wäre es zwar möglich, diese alle in ein Modell zu übernehmen, jedoch soll zunächst ein Indexwert gebildet werden, der die Gesamtvariable widerspiegelt. Grund hierfür ist, dass es so wesentlich wahrscheinlicher wird einen signifikanten Zusammenhang nachzuweisen, da das Gesamtkonstrukt abgebildet wird und nicht jede einzelne Teilvariable. Außerdem ist diese Umformung auch im Sinne des Parsimonyprinzips, welches von guten Modellen eine möglichst geringe Anzahl an Variablen verlangt. Gebildet wurde dieser Indexwert, indem eine neue Variable generiert wurde, deren Wert für eine positive Antwort in den Kategorien Mitarbeiter in der strategischen Planung, Datenanalyse und IT, sowie für alle Führungskräfte um jeweils 2 Punkte erhöht wurde, für Führungskräfte auf ihren Bereich beschränkt um 1 Punkt. Für den Fall, dass angegeben wurde, dass alle Mitarbeiter Zugriff haben, wurde pauschal der Wert 10 vergeben, um eine Überschneidung dieser Kategorie und die daraus resultierende Obsoleszenz zu vermeiden.

### **Fallauswahl**

Da wie in Kapitel 5.1. gezeigt, eine große Anzahl an Interviews nicht abgeschlossen wurde spielt die Auswahl gültiger Fälle eine wichtige Rolle für die Bewertung des vorliegenden Datenmaterials. Zwar ist die Anzahl von 153 vollständig abgeschlossenen Datensätzen mehr als ausreichend, jedoch kann vor einer näheren Analyse nicht davon ausgegangen werden, dass sämtliche abgeschlossenen Datensätze auch sämtliche Informationen enthalten, da ein bewusstes überspringen von Fragen zugelassen wurde. Des Weiteren sollten möglichst sämtliche erhaltenen Informationen in die Analyse miteinfließen, weshalb entschieden wurde, auch unvollständige Datensätze miteinzubeziehen. Das entscheidende Kriterium war

dabei, dass mindestens die erste Seite mit Fragen zur Datenanalyse beantwortet wurde, da reine Informationen zum Unternehmen noch keinen Mehrwert für diese Arbeit geliefert hätten. In Bezug auf den Fragebogen bedeutet dies, dass ein Umfrageteilnehmer mindestens Seite 6 erreichen musste, um als gültiger Fall in die Analyse miteinbezogen zu werden. Hierdurch wurde die Anzahl gültiger Interviews von 153 auf 255 erhöht.

#### 5.4. Validität der Arbeit

Einen wichtigen Punkt, der an dieser Stelle diskutiert werden soll, stellt die Validität dieser Studie dar. Hierfür muss zum einen in interne und externe Validität unterschieden werden. Zu internen Validität lässt sich zunächst sagen, dass diese durch die angewandten statistischen Methoden in hohem Umfang gewährleistet ist. Hierfür wird bei der Interpretation der Ergebnisse auf die üblichen Signifikanzniveaus geachtet und diese kommuniziert. Einschränkend muss in diesem Zusammenhang jedoch auch darauf verwiesen werden, dass, wie bei jeder Feldstudie, die Existenz unbekannter Drittvariablen natürlich nicht ausgeschlossen werden kann und diese im Falle ihrer Existenz, die interne, sowie externe Validität der Ergebnisse massiv beeinträchtigen könnten. Zur externen Validität lässt sich sagen, dass sich hier durchaus ein ambivalentes Bild zeigt. Zum einen stärkt die klare Definition der befragten Gruppe an Unternehmen (über 20 € Mio. Bilanzsumme und 40 € Mio. Umsatz sowie die Tätigkeit in der DACH-Region) die externe Validität. Außerdem ist die Anzahl der Rückmeldungen mehr als ausreichend, um induktive Methoden anzuwenden<sup>61</sup> und durchaus vergleichbar mit den Zahlen professioneller Studien wie von BARC (ca. 500) oder Computer Research.(ca. 390) Schließlich sorgt auch das Forschungsdesign selbst dafür, dass eine relativ hohe externe Validität angenommen werden kann, da die Erhebung, im Gegensatz zum Laborexperiment, real existierende Unternehmen befragt und den dortigen Einsatz von Big Data Methoden erhebt. Problematisch dagegen wirkt sich die geringe Rücklaufquote auf die externe Validität aus, da dadurch eine massive Verzerrung durch einen Self-Selection-Bias nicht ausgeschlossen werden kann.

Die Validität der Datenerhebung selbst ist dagegen durch mehrere Gütekriterien gewährleistet. Zunächst wurde versucht über eine grundlegende theoretische Aufarbeitung von Theorien des strategischen Managements und die Verknüpfung dieser mit den aktuellen

---

<sup>61</sup> Dies trifft auf die meisten berechneten Modelle zu, bei solchen wo dies nicht der Fall ist, wird explizit darauf verwiesen, dass aufgrund der geringen Fallzahl keine belastbare Aussage getroffen werden kann.

Entwicklungen im Bereich Big Data und der darauf aufbauenden theoretischen Merkmalsdefinition in Form der Operationalisierung ein Fundament für die erhobenen Variablen zu schaffen. Somit wird laut Moosbrugger dem Kriterium der Inhaltsvalidität Genüge getan.<sup>62</sup> Jedoch muss hier auch explizit erwähnt werden, dass eben nicht direkt auf existierende, generell anerkannte und vielfach getestete Theorien, die den Zusammenhang zwischen latenten Konstrukten und manifesten Variablen erklären, zurückgegriffen werden konnte, da solche nicht existieren, sondern es sich aufgrund der Neuartigkeit der Entwicklung um einen rein argumentativen und explorativen Ansatz handelt.

Der Aufbau dieser Arbeit geht jedoch über das Kriterium der Inhaltsvalidität hinaus und kann aufgrund des hypothetisch-deduktiven Vorgehens den Anspruch der Konstruktvalidität erheben. Dies ergibt sich aus der theoretischen Bildung eines nomologischen Netzes, das sich aus den Ausführungen zu Inhaltsvalidität ergibt, und der hieraus folgenden Bildung von Hypothesen, welches sich schließlich in der Darstellung des Gesamtmodells widerspiegelt. Diese Hypothesen wurden anschließend durch empirische Verfahren der induktiven Statistik überprüft, so dass durch die Kombination dieser beiden Verfahren das Kriterium der Konstruktvalidität, sogar in seiner starken Form nach Cronbach und Meehl, als gegeben angesehen werden kann.<sup>63</sup>

## 6. Erkenntnisse und Interpretation

Die Präsentation der Erkenntnisse soll im Folgenden in zwei Schritten erfolgen. Zunächst sollen rein deskriptiv die Ergebnisse der Umfrage präsentiert werden, um den Lesern einen Eindruck des Datenmaterials zu vermitteln. Auch können schon hier erste Erkenntnisse über die Verbreitung und Nutzung von Big Data Methoden gewonnen werden. In einem zweiten Schritt sollen die Daten mittels statistischer Verfahren analysiert werden, um die erarbeiteten Hypothesen mit dem Datenmaterial abzugleichen.

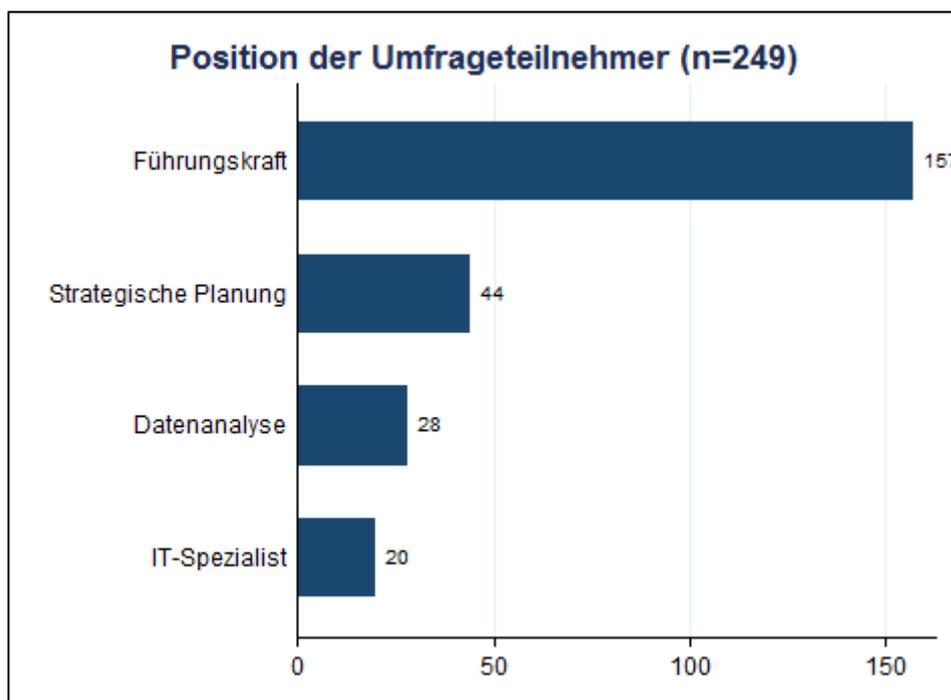
---

<sup>62</sup> Vgl. Moosbrugger, 2012, S.149.

<sup>63</sup> Vgl. Ebd., S.153-155.

## 6.1 Generelle Informationen über die teilnehmenden Unternehmen

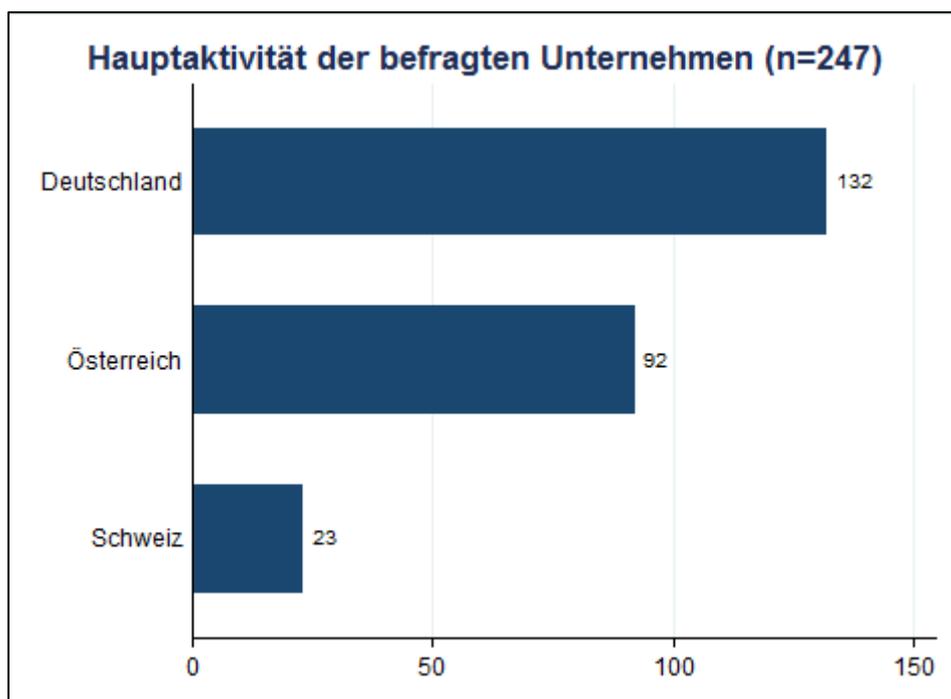
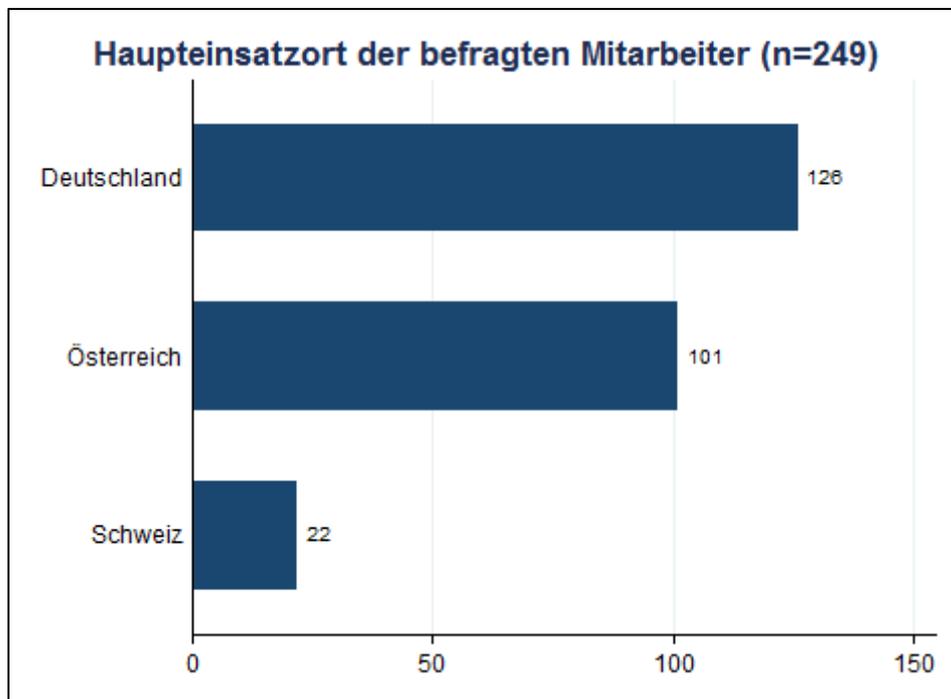
Zunächst wurden einige allgemeine Informationen zum Unternehmen und dem Mitarbeiter erhoben, der den Fragebogen ausfüllt. Hierdurch soll klar werden, aus welchem Blickwinkel die gewonnenen Informationen stammen. So zeigt beispielweise die Betrachtung der Position des Mitarbeiters, dass erstaunlich viele Führungskräfte sich die Zeit genommen haben, um die Umfrage auszufüllen:



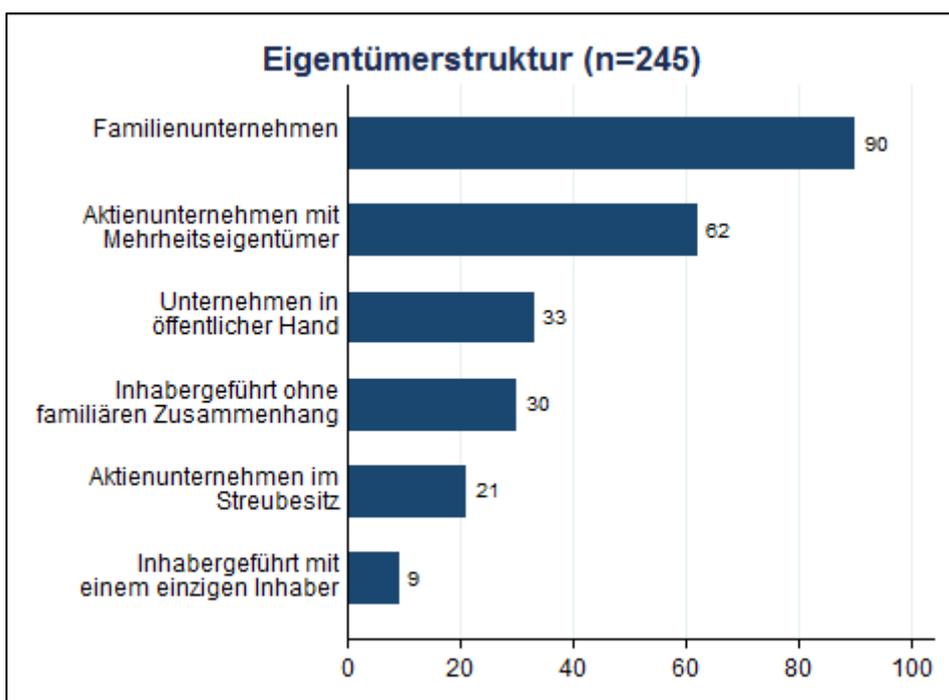
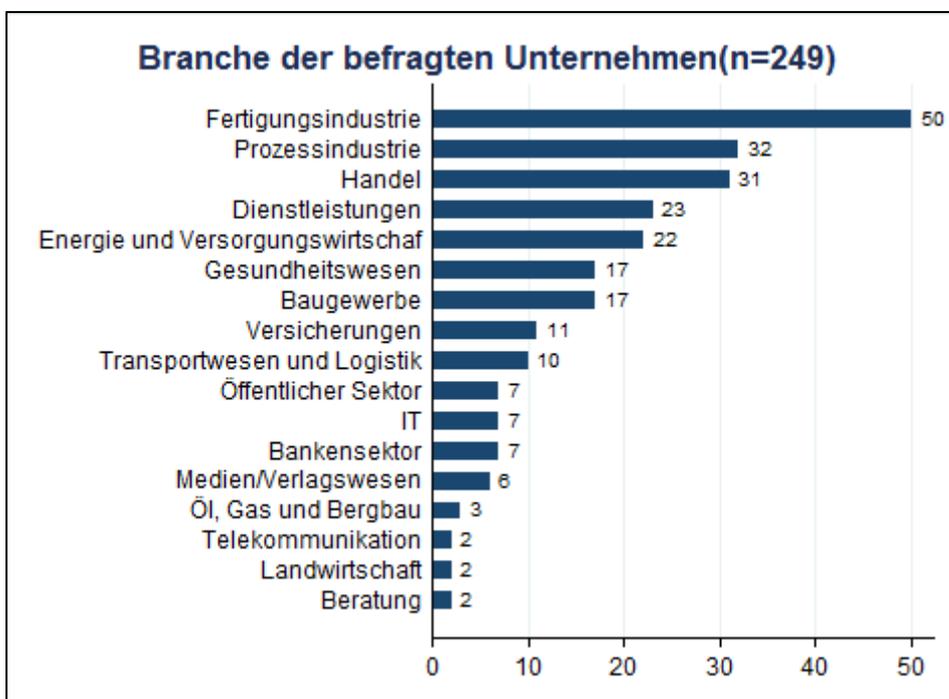
Zum einen kann man hierdurch vermuten, dass es sich bei dem Thema durchaus um ein solches handelt, dass für die Unternehmen von hohem Interesse ist, sodass sie es nicht einfach delegieren. Zum anderen zeigt es natürlich auch, aus welcher Perspektive die gewonnenen Daten entstanden sind.

Eine weitere Variable stellt das Land dar. Hierbei wurden zwei Fälle unterschieden: Das Land, in dem das Unternehmen hauptsächlich tätig ist und das Land, in dem der Mitarbeiter hauptsächlich tätig ist. Wie bereits in der Operationalisierung beschrieben, soll das Land zur Kontrolle der juristischen und institutionellen Rahmenbedingungen erhoben werden. Jedoch ist es durchaus möglich, dass hierbei die individuelle Wahrnehmung des Mitarbeiters nicht gänzlich der Realität des Gesamtunternehmens entspricht, weshalb beide Angaben erhoben wurden, um sie in der Folge in den Modellen testen zu können. Was bei der Betrachtung

der Zahlen insbesondere auffällig, ist das enorme Übergewicht österreichischer Unternehmen, welches vermutlich darauf zurückzuführen ist, dass die Studie als Masterarbeit an einer österreichischen FH verfasst wird und von einer österreichischen Stiftung gefördert wird. Diese Verbindung scheint die Bereitschaft zur Teilnahme bei österreichischen Unternehmen massiv positiv beeinflusst zu haben.



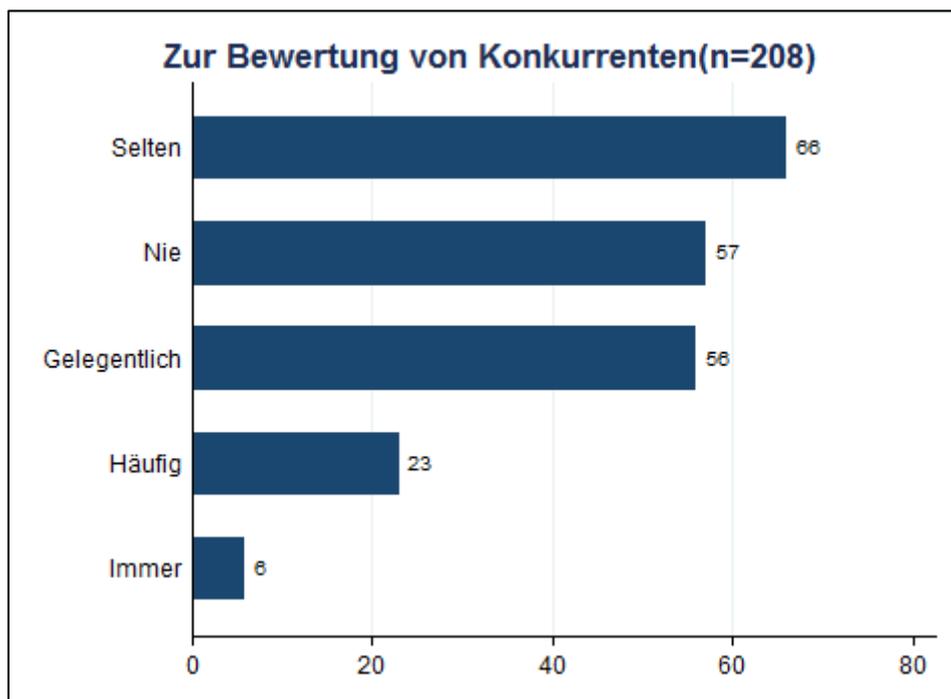
Die dritte relevante Variable stellt bei der Erhebung die Branche der Unternehmen dar. Hier ist anzunehmen, dass sich insbesondere Unternehmen aus solchen Branchen beteiligt haben, für die das Thema Big Data derzeit eine große Rolle spielt, also solche die sich im Umbruch befinden und gerade dabei sind die Möglichkeiten zur Nutzung solcher Technologien zu eruieren bzw. solche Fähigkeiten aufzubauen. Gestützt wird diese Vermutung unter anderem auch dadurch, dass eine größere Anzahl an Absagen die Information enthielt, dass die Thematik Big Data für das angeschriebene Unternehmen keine Rolle spiele und man deshalb nicht an der Umfrage teilnehmen möchte. Im Umkehrschluss lässt sich daraus die Vermutung stützen, dass alle Unternehmen, die sich beteiligten, ein gewisses Eigeninteresse an den Ergebnissen haben, insbesondere da auch explizit angeboten wurde, die gewonnenen Informationen in Form der fertigen Studie den Teilnehmern zukommen zu lassen. Dass dieses Angebot immerhin von 86 der 155 Teilnehmer, die einen vollständigen Datensatz generierten, also guten 55% genutzt wurde, scheint diese Annahme zu bestätigen. Diesen Überlegungen entsprechend ist es kaum verwunderlich, dass ein großer Teil der Teilnehmer (82) aus dem Bereich der Industrie kommt. Dieser ist vermutlich, auch im Zusammenhang mit dem Thema Industrie 4.0, zurzeit der Bereich, in dem am intensivsten nach Anwendungsmöglichkeiten für die gewonnenen Daten gesucht wird. Der Handel mit 31 Teilnehmern dagegen ist hier schon deutlich weiter, dies zeigt sich unter anderem in den verschiedenen konkreten Anwendungsbeispielen, die bereits dargelegt wurden. Verwunderlich erscheint dagegen, dass Telekommunikationsunternehmen mit gerade einmal 2 Teilnehmern kaum vertreten sind. Eine banale Erklärung hierfür könnte sein, dass dieser Markt von wenigen Großunternehmen dominiert wird und damit die absolute Anzahl an existierenden Unternehmen deutlich geringer ist als beispielweise im Bereich Industrie.

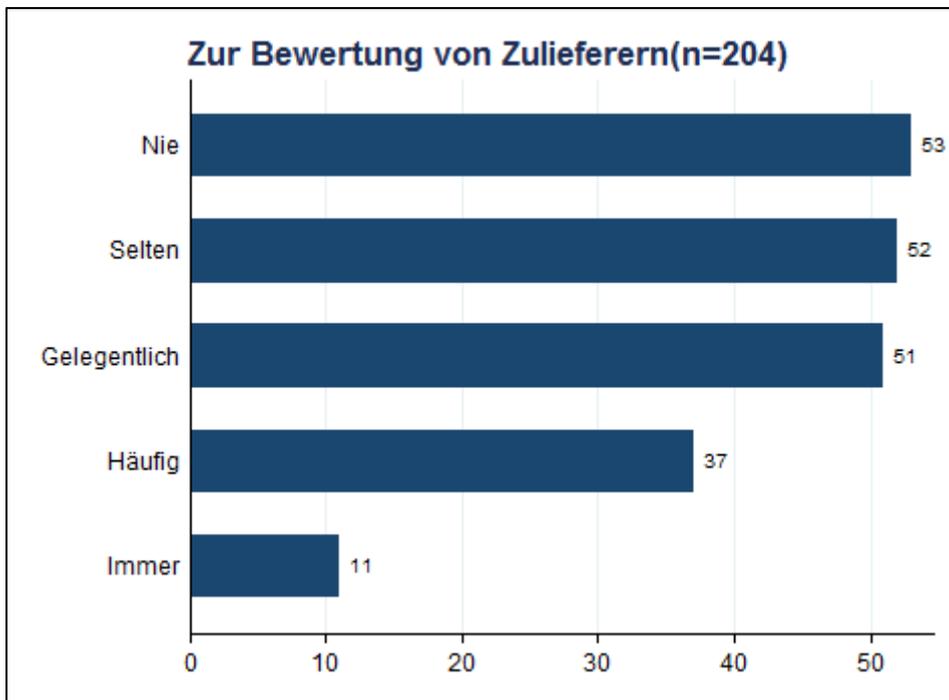


## 6.2. Informationen zur Datenanalyse in den Unternehmen

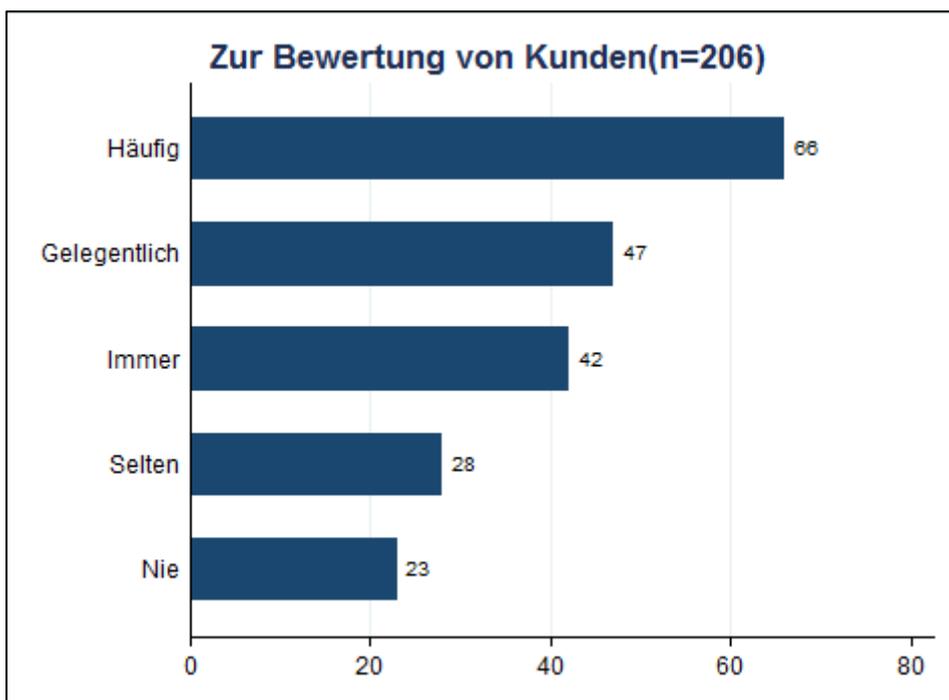
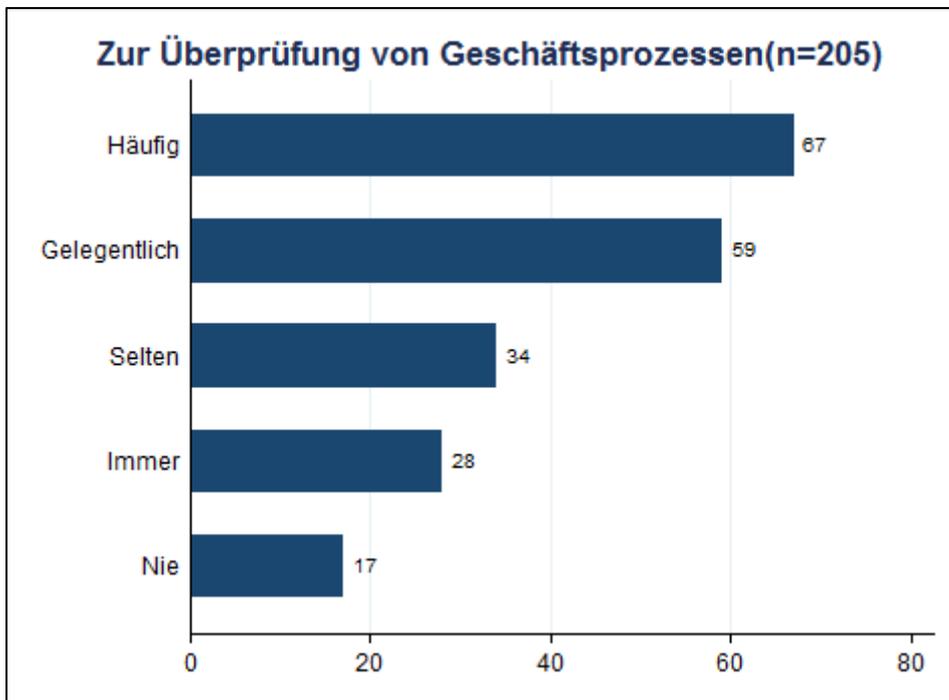
Das Kernstück dieser Arbeit stellt die Analyse der Big Data Nutzung der Unternehmen dar, weshalb im Folgenden zunächst dargestellt werden soll, wofür diese Technologien von den

Unternehmen genutzt werden und anschließend, welche Datenbasis und Werkzeuge dafür verwendet werden. Hierfür wurden zunächst 5 mögliche Zielfelder betrachtet, die analysiert werden können: Konkurrenten, Zulieferer, Kunden, Geschäftsprozesse und Projekte. Ergänzt wird dies durch die Nutzung der Datenanalyse zur Identifikation von Krisen sowohl auf strategischer, als auch auf Projektebene. Neben dem dadurch gewonnen Überblick über die aktuellen Einsatzgebiete von Big Data können hierdurch vor allem Unterschiede in der Bedeutung dieser Gebiete identifiziert werden.



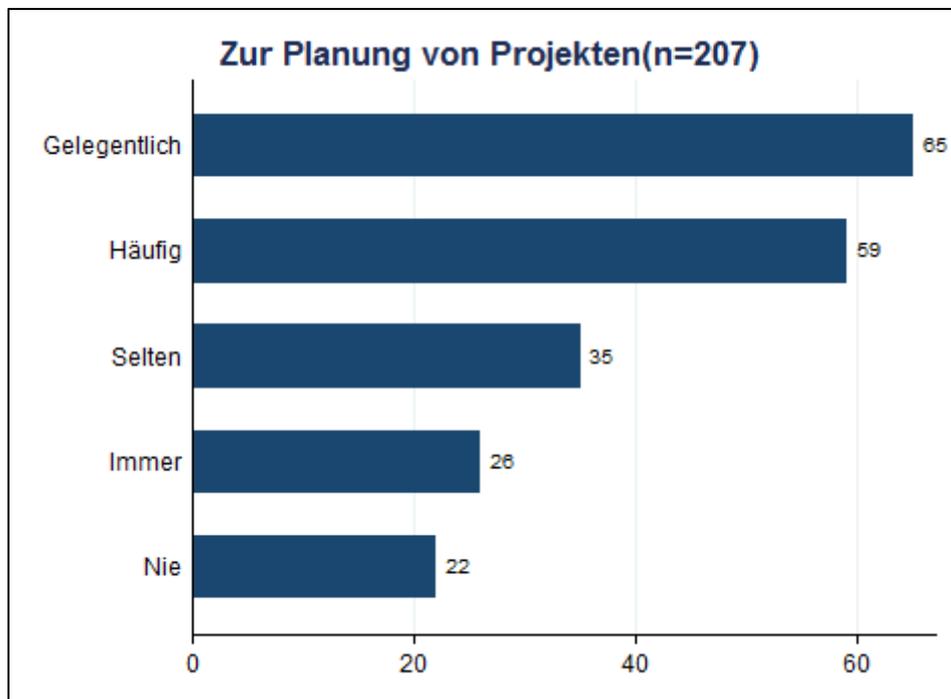


Sowohl für die Analyse von Konkurrenten mit „Selten“ als meistgenannten Punkt, als auch für die Zulieferer, wo sogar „Nie“ die häufigste Nennung darstellt, zeigt, dass in beiden Bereichen Big Data bisher eine eher geringe Rolle spielt. Allerdings ist in beiden Fällen die Nennung „Gelegentlich“ mit minimalem Abstand auf Platz 3, was immerhin zeigt, dass die Datenanalyse auch hier nicht gänzlich zu vernachlässigen ist. Nicht erkennbar ist allerdings die Ursache hierfür. Denkbar sind zum einen, dass es an konkreten Anwendungsfeldern mangelt oder dass diese Bereiche sich erst im Aufbau befinden. Am wahrscheinlichsten erscheint jedoch, dass dies schlicht diejenigen Bereiche sind, in denen es die wenigsten Daten gibt und damit eine fundierte Analyse schwierig ist und auf klassische Methoden zurückgegriffen wird. Diese Vermutung wird auch durch die Rückmeldungen zur Nutzung externer Daten, die im Folgenden noch vorgestellt werden, gestützt.

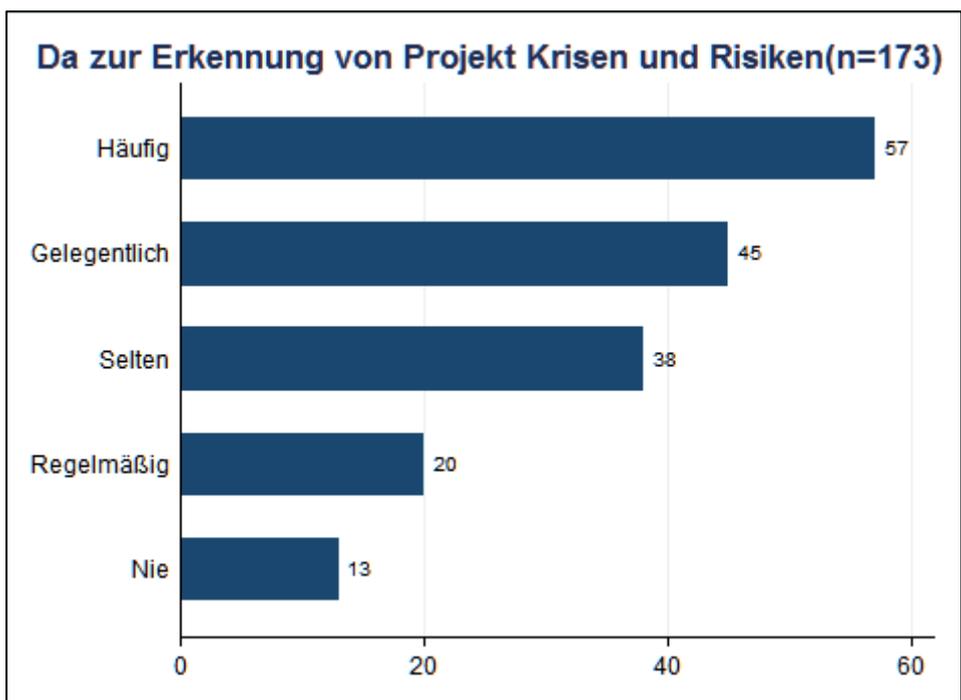


Wenig verwunderlich ist auch, dass in den beiden Bereichen Geschäftsprozesse und Kunden „Häufig“ die häufigste Nennung darstellt. Beides sind klassische Felder von Big Data Analysen, wie bereits die verschiedenen praktischen Anwendungsbeispiele, die bisher genannt wurden, zeigen. Mitverantwortlich hierfür ist aber sicher auch die existierende Datengrundlage. Bzgl. der Geschäftsprozesse liegen schlicht interne Daten vor, die „lediglich“ ausgewertet werden müssen und von Kunden wird seit einiger Zeit aktiv versucht

mehr und mehr Daten zu gewinnen, wobei dieses Phänomen vor allem den Handel betrifft, wie die genannten Beispiele zeigen. Ein anderes Beispiel hierfür stellen auch die Kundenkarten dar, die von nahezu jedem Handelsunternehmen, meist in Kombination mit Rabattaktionen angeboten werden. Dabei liegt der Hauptzweck solcher Karten in der Generierung von analysierbarem Datenmaterial.



Bei der Nutzung der Datenanalyse zur Identifikation von Krisen zeigt sich, dass diese durchaus intensiv genutzt wird, so stellen sowohl für den Bereich Strategie, als auch für Projekte die beiden meistgenannten Kategorien „Häufig“ und „Gelegentlich“ dar. Wirklich regelmäßig, also als Teil der Standardprozesse, wird sie jedoch nur in einem kleinen Teil der Unternehmen eingesetzt, nämlich bei jeweils ca. 12%. Die Mittelwertschätzer zeigen außerdem, dass die Unterschiede zwischen strategischer und Projektebene nicht signifikant sind, also die hier abgebildeten Differenzen auf zufällig auftretende Unterschiede im Sample und nicht auf real existierende in der Grundgesamtheit zurückgeführt werden müssen.

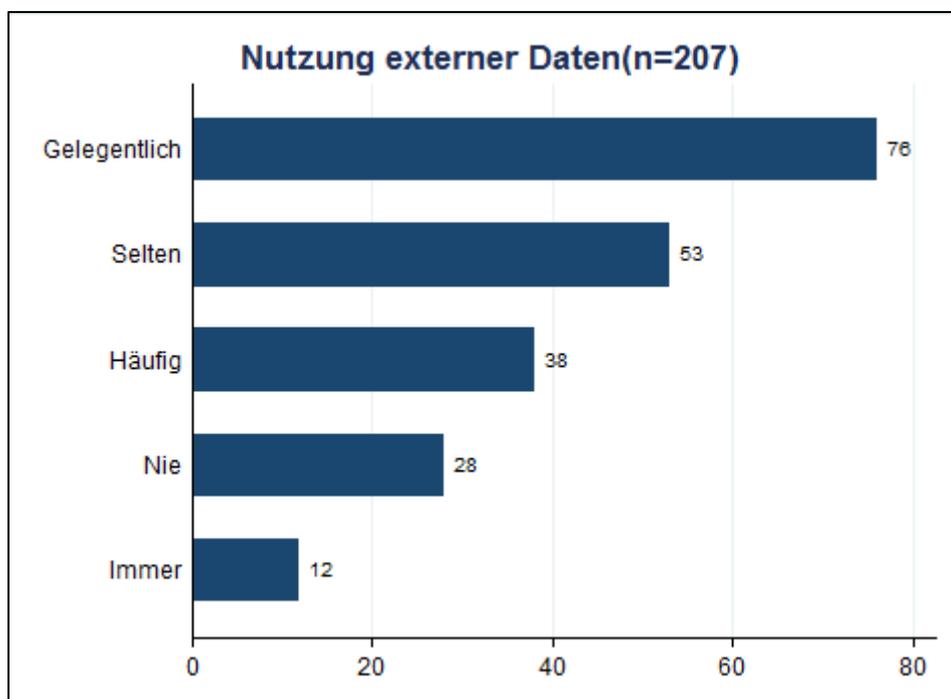


Ein Vergleich der Mittelwertschätzer dieser Daten bestätigt den klaren Unterschied in den Anwendungsfeldern der Datenanalyse. So wird diese signifikant häufiger für die Analyse von Kunden, Geschäftsprozessen und Projekten sowie für die Identifikation von Krisen, sowohl auf strategischer als auch Projektebene benutzt, als für die Analyse von Konkurrenten und Zulieferern. Wobei diese beiden Gruppen untereinander keine signifikanten Unterschiede aufweisen. Die genauen Werte können der folgenden Tabelle

entnommen werden, in der sich die Zahlenwerte aus der Kodierung für Nie=1 bis Immer=5 ergeben.

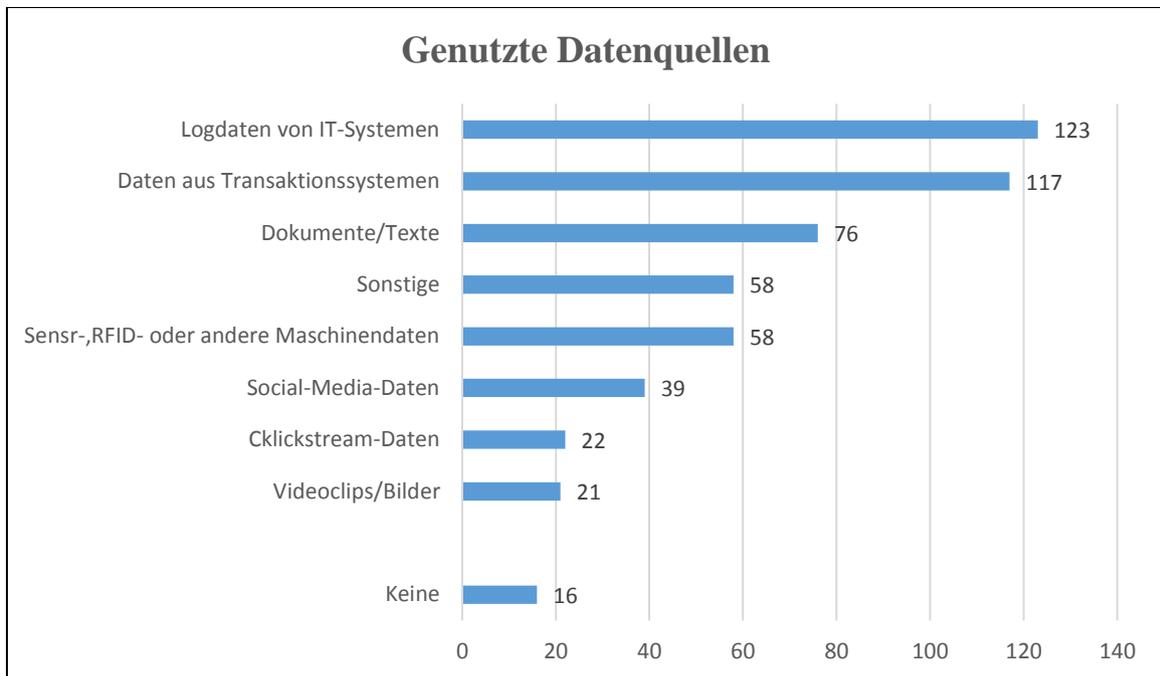
<b>Mittwertschätzung der DA Nutzung</b>				
n=171	Mean	Std. Err.	[95% Conf. Intervall]	
Konkurrenten	2,29557	0,07623	2,14526	2,44587
Zulieferer	2,52217	0,08477	2,35503	2,68931
Kunden	3,37931	0,08831	3,20518	3,55344
Geschäftsprozesse	3,26601	0,07988	3,10851	3,42351
Projekte	3,16256	0,08139	3,00208	3,32304
Krise Strategie	3,15790	0,08351	2,99304	3,32275
Krise Projekte	3,19298	0,08672	3,02179	3,36418

Ein interessanter Punkt für die praktische Nutzung und eventuell neu entstehende Geschäftsmodelle stellt die Nutzung externer Daten dar. Dies erlaubt es Unternehmen auf einen deutlich größeren Datenpool zuzugreifen und damit Analysen durchzuführen, für die die eigenen Daten schlicht unzureichend sind. Sollten deshalb die Kapazitäten und Fähigkeiten der Unternehmen im Bereich Datenanalyse wachsen, ist davon auszugehen, dass der Handel mit Daten ein immer attraktiveres Geschäftsfeld wird. Zurzeit ist allerdings eher der umgekehrte Fall Realität, dass deutlich mehr Daten verfügbar sind, als ausgewertet werden können, sodass der direkte Ankauf von zusätzlichen Daten wohl nur für wenige Unternehmen wirklich von Bedeutung ist. Außerdem spielen hier enorme Bedenken bzgl. des Datenschutzes sowie die dafür geltenden Rahmenbedingungen eine wichtige Rolle. Im Lichte dieser Überlegungen sind die Ergebnisse der Erhebung durchaus interessant. So gaben immerhin 76 Unternehmen an, gelegentlich auf externe Daten zurückzugreifen und einige wenige (12), dass sie dies immer tun würden.



In gewisser Weise spiegelt sich dies auch in der Art der genutzten Datenquellen wieder. Zwar stellt der Punkt Sonstige mit 58 Nennungen den meist genannten Wert dar, aber direkt darauf folgen Dokumente, Daten aus Transaktionssystemen und Logdaten von IT-Systemen. Dies sind üblicherweise Daten, die das Unternehmen selbst generiert und somit einen vollständigen und kostenlosen Zugang dazu hat, was zumindest als erste Anlaufstelle für den Aufbau von Datenanalysekapazitäten eine interessante Option darstellt. Damit zeigt sich, dass die Ergebnisse der Computing Research Studie von 2015 in diesem Punkt auch heute noch Gültigkeit haben und weiterhin der Hauptfokus der Big Data Nutzung auf der Analyse interner Prozesse zum Zweck der Effizienzsteigerung liegt und nicht auf Analyse externer Daten zur Eröffnung neuer Märkte.<sup>64</sup>

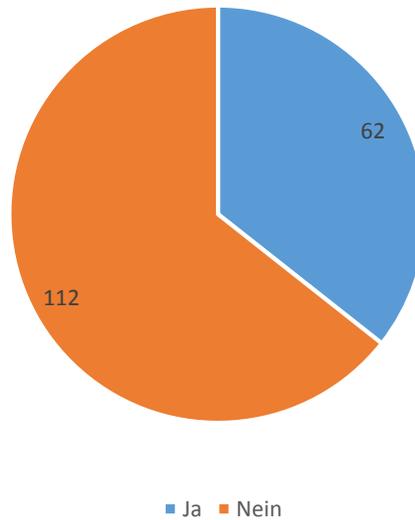
<sup>64</sup> Vgl. Computing Research, 2015, S.3.



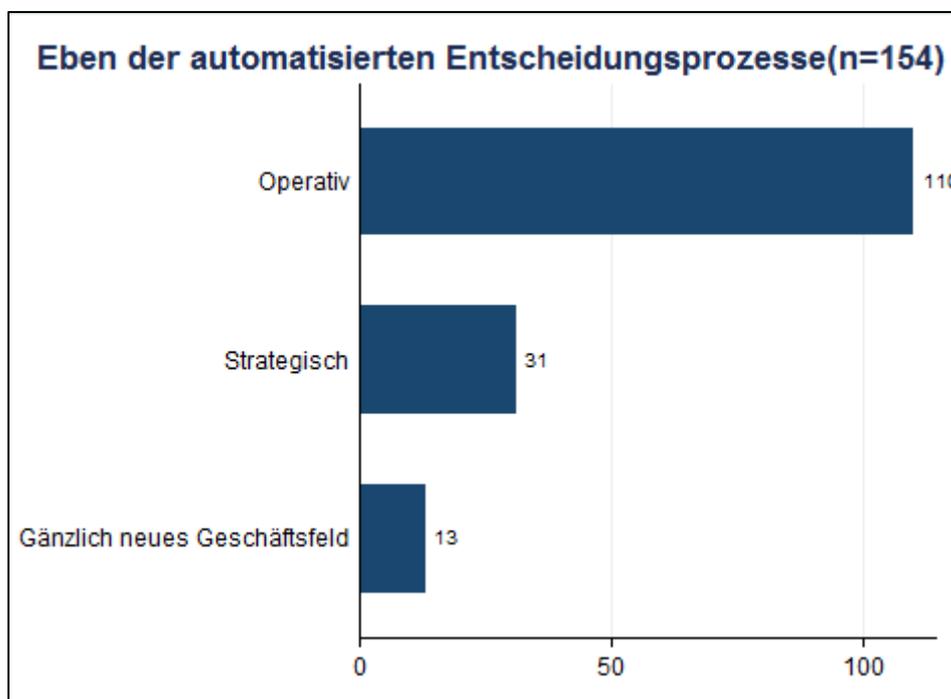
Die konkreten Datenquellen belegen erneut die Aussage, dass externe Daten für Unternehmen aktuell keine zu große Relevanz besitzen. So werden primär Logdaten und Daten von Transaktionssystemen ausgewertet. Daten aus Social-Media Plattformen oder Videos dagegen werden kaum genutzt. Ein interessanter Aspekt ist dabei auch, dass immerhin 58 Unternehmen angaben, Sensor-, RFID- und andere Maschinendaten auszuwerten. Dies zeigt, dass neben dem klassischen Big Data Anwendungsbereich Business Intelligence, auch technische Anwendungen durchaus bereits eine relevante Bedeutung haben. Ein Beispiel hierfür stellt der Bereich autonomes Fahren dar.

Interessant ist auch, dass bereits etwa ein Drittel der Unternehmen Entscheidungsprozesse automatisiert.

### Werden Entscheidungsprozesse automatisiert? (n=174)



Wie zu erwarten war, befinden sich diese Entscheidungen zum überwiegenden Teil auf der operativen Ebene, wofür die zielgerichtete Kundenansprache oder der Hochfrequenzhandel Beispiele sind. Deutlich wichtiger dagegen ist, dass ganze 31 Unternehmen angaben, bereits strategische Entscheidungen zu automatisieren. Diese Angabe muss allerdings durchaus kritisch hinterfragt werden, da vier dieser Unternehmen auch konkrete Angaben dazu machten, welche Prozesse automatisiert werden und die meisten dieser klar in den Bereich der operativen Tätigkeiten fallen, wie die Erstellung von GuV, Bilanzen, Abrechnungen und Archivierung. Prozesse die hier genannt wurden und wirklich auch auf strategischer Ebene liegen könnten, sind Financeprozesse und Vertriebsprozesse. Da jedoch die überwiegende Mehrheit der Unternehmen (27/31) keine näheren Angaben dazu machte, welche Art von Prozessen automatisiert werden, kann nicht zwingend davon ausgegangen werden, dass die vorliegenden Antworten repräsentativ sind. Ebenso interessant ist, dass immerhin 13 Unternehmen angaben, dass diese automatisierten Entscheidungsprozesse sich auf ein gänzlich neues Geschäftsfeld beziehen. Zur konkreten Art dieser neuen Geschäftsfelder liegen allerdings keine Rückmeldungen vor.

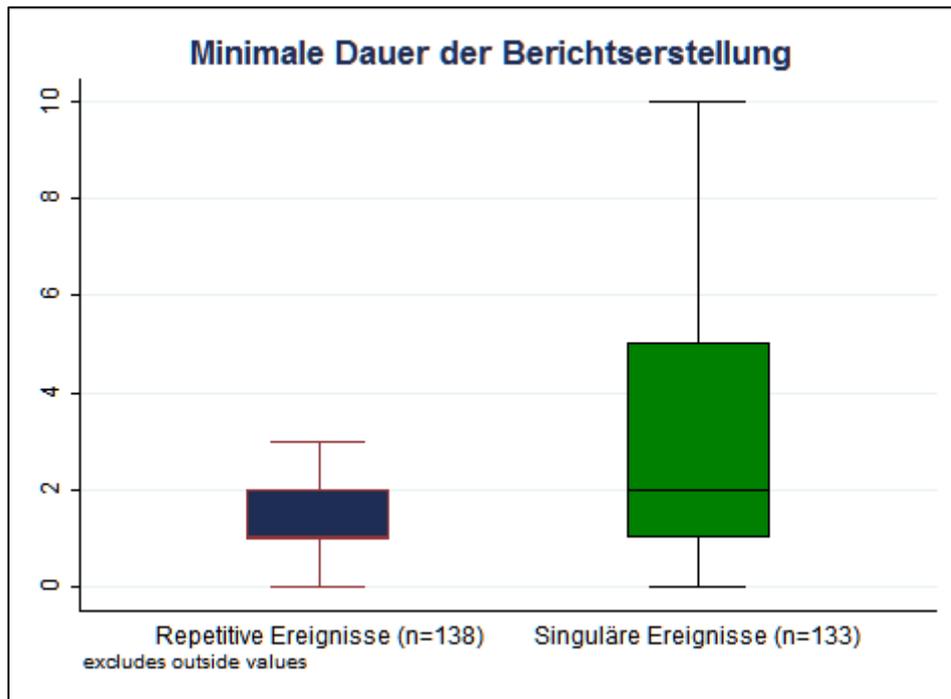


Einen wichtigen Faktor für die Geschwindigkeit mit der Unternehmen auf Ereignisse reagieren können, stellt das Berichtswesen dar, welches im Folgenden näher betrachtet werden soll. Den ersten wichtigen Punkt stellt hierbei die Zeit dar, die zwischen einem Ereignis und seiner Aufarbeitung in einem Bericht liegt. Dabei wird zwischen der minimalen, durchschnittlichen und maximalen Dauer unterschieden, um eine bessere Darstellung der Realität zu erreichen und außerdem zwischen repetitiven und singulären Ereignissen, da davon ausgegangen werden muss, dass repetitive Ereignisse deutlich leichter in moderne Informationssysteme eingebunden und automatisiert werden können. Im Sinne der Verwendbarkeit der grafischen Darstellungen wurden die Ausreißer entfernt.<sup>65</sup> Durch die Trennung zwischen den Kategorien „minimal“, „maximal“ und „durchschnittlich“ muss allerdings genau auf die Skalierung der Y-Achse geachtet werden, da diese nicht in allen drei Grafiken gleich ist. Um die Vergleichbarkeit zu fördern, werden jeweils die Werte für repetitive und singuläre Ereignisse gegenübergestellt.

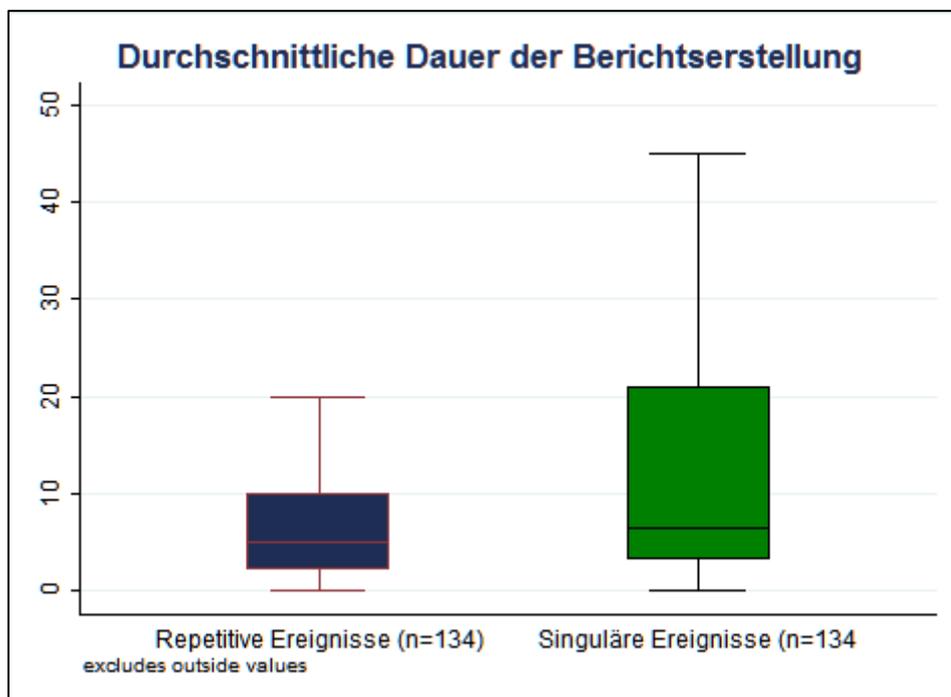
Kaum verwunderlich ist, dass wie erwartet singuläre Ereignisse, vermutlich aufgrund ihrer schwierigeren Erfassung und deren schlechte Automatisierbarkeit deutlich länger brauchen, um in Berichte einzufließen, auch die Streuung ist hierbei deutlich größer. Auffällig ist auch, dass der Abstand zwischen Median und 1. Quartil in allen Fällen sehr gering ausfällt. Was

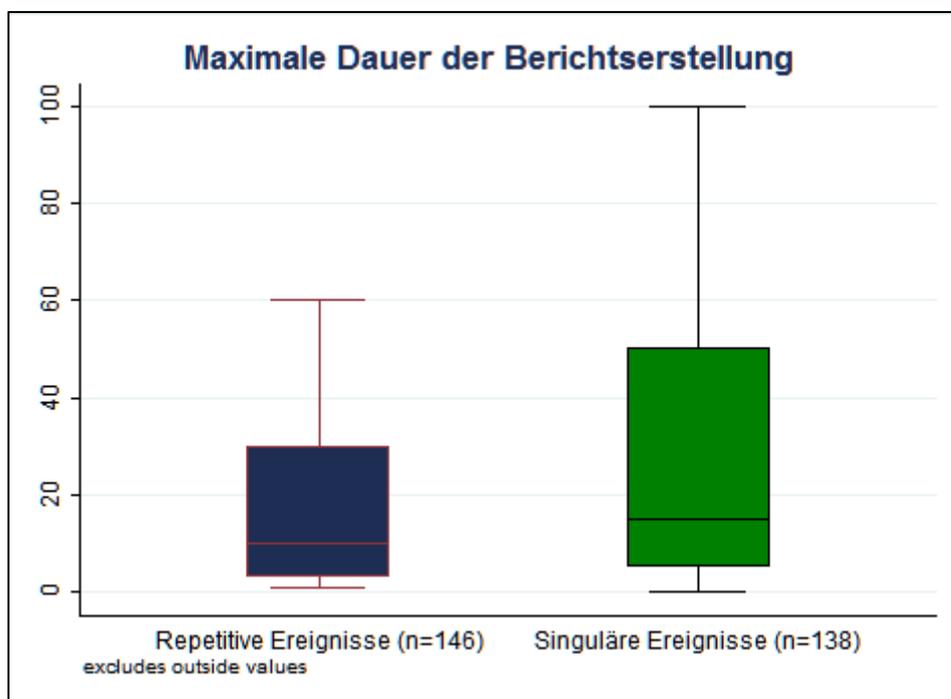
<sup>65</sup> Als Ausreißer wurden solche Werte klassifiziert, welche über dem 1,5 fachen Quartilsabstand über dem 75% Quantil liegen.

zeigt, dass die Unterschiede zwischen den Unternehmen der schnelleren Hälfte deutlich geringer sind als zwischen denjenigen der langsameren Hälfte.

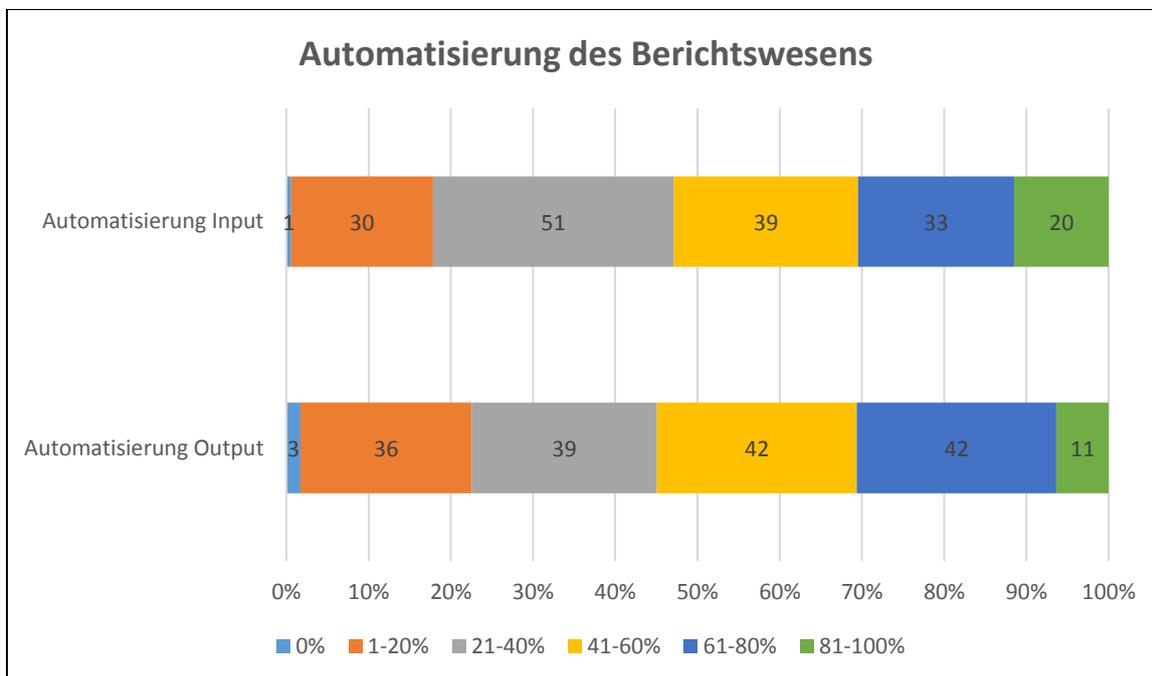


Bei dieser Abbildung ist zu beachten, dass der Median mit dem 25% Quantil bei den repetitiven Ereignissen zusammenfällt, weshalb dieser schlecht optisch erkennbar ist.





Einer der Gründe für die hier beobachtbaren Unterschiede dürfte im Grad der Automatisierung des Berichtswesens zu finden sein. Da eine Automatisierung in diesem Bereich die Informationen nicht nur schneller, sondern im Optimalfall sogar in Echtzeit verfügbar macht, wird die Verzögerung des Informationsflusses zu den Entscheidungsträgern reduziert oder gar eliminiert. Es zeigt sich jedoch, dass die Automatisierung des Berichtswesens zwar eine große Rolle spielt, so betrifft sie in nahezu allen Unternehmen Teilbereiche, eine vollständige Automatisierung ist allerdings kaum vorhanden. Außerdem zeigt sich, dass es zwar kleine Unterschiede zwischen der Automatisierung der Erfassung von Daten und der Verarbeitung dieser Daten in Berichten gibt, wie in der Folgenden Abbildung erkennbar ist.

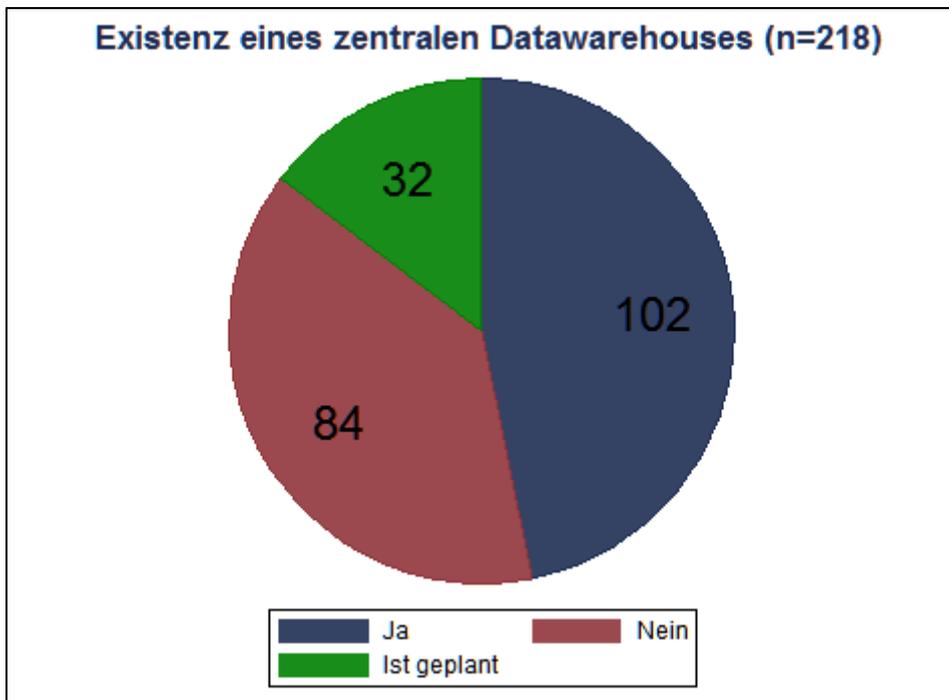


Dieser Unterschied spiegelt sich allerdings nur in der Verteilung der Kategorien wieder, im arithmetischen Mittel besteht kein signifikanter Unterschied zwischen Input und Output, dies trifft auch auf den Median zu, der in beiden Fällen bei 41-60% liegt.

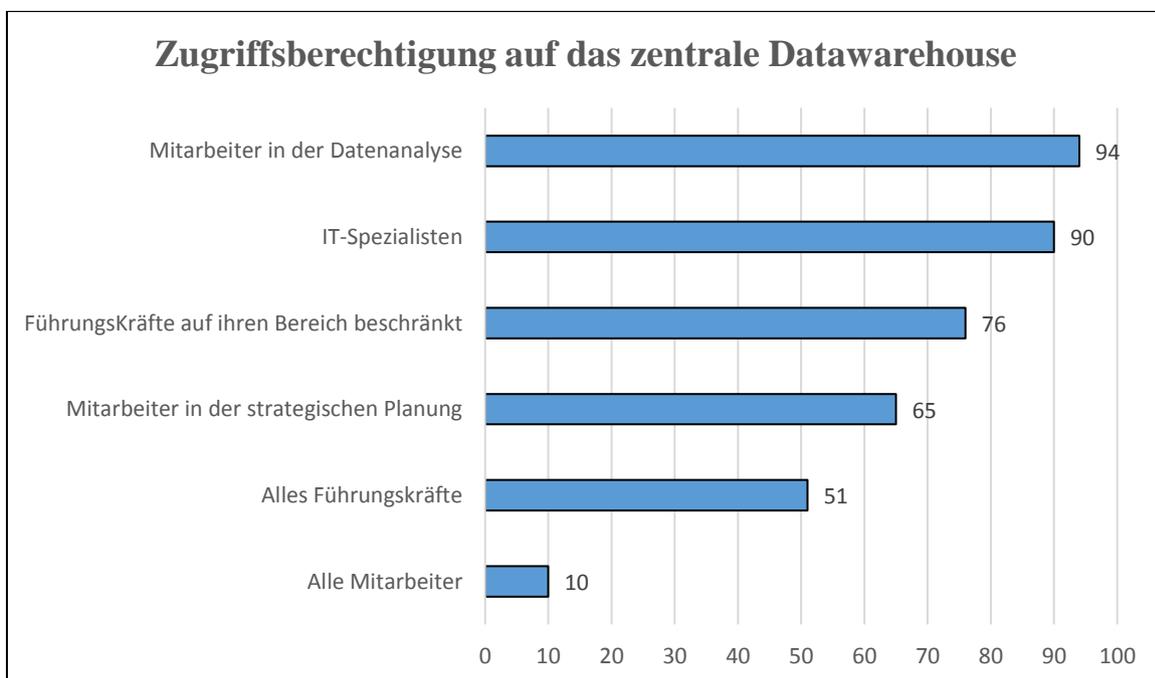
Mittelwertschätzung der Automatisierung im Berichtswesen				
n=173	Mean	Std. Err.	[95% Conf. Intervall]	
Input	3,77457	0,09704	3,58302	3,96611
Output	3,67630	0,09664	3,48556	3,86704

*Kodierung: 0%=1, 81-100%= 6*

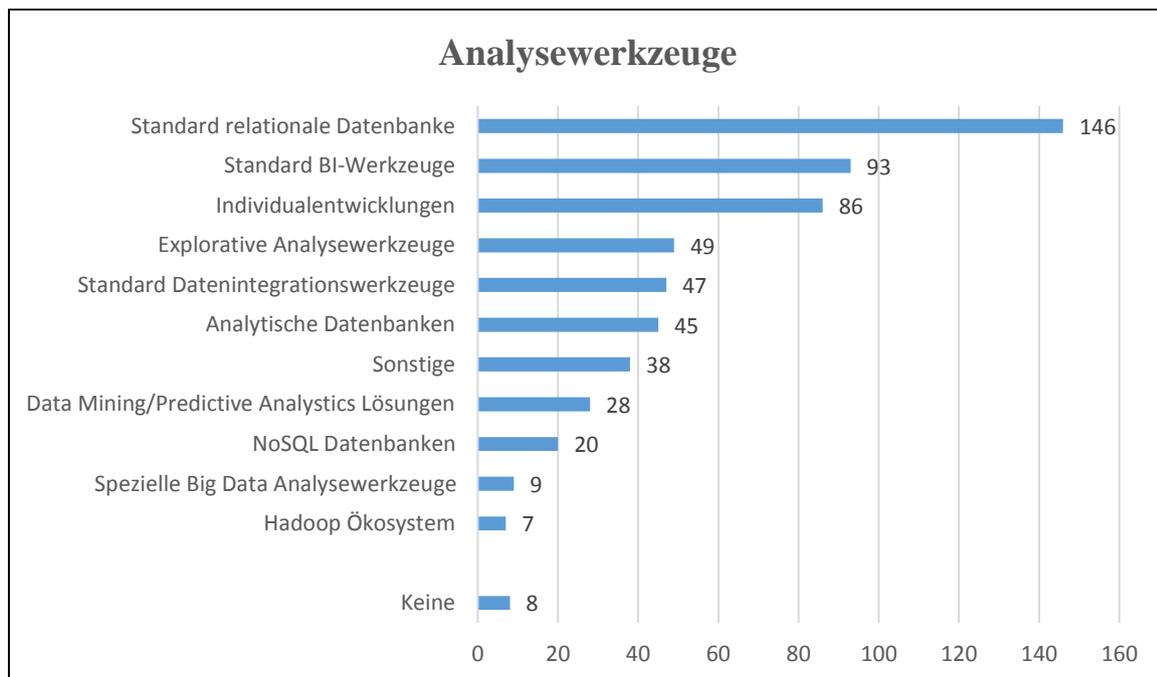
Eine wichtige Rolle für die Auswertung der Daten in einem Unternehmen spielt ein zentrales Datawarehouse. Dieses stellt sicher, dass alle Mitarbeiter auf dieselbe Datengrundlage zurückgreifen, was inkonsistente Analysen, zumindest auf Ebene der Datengrundlage, verhindert. Es stellt quasi sicher, dass allen Mitarbeitern dasgleiche „Vokabular“ zur Verfügung steht. Dass die Unternehmen die Bedeutung eines solchen zentralen Datenpools bereits erkannt haben, zeigt sich darin, dass nahezu die Hälfte der Unternehmen angeben, über ein solches zu verfügen und bei weiteren 15% ein solches bereits geplant ist.



In diesem Zusammenhang spielt jedoch nicht nur die Existenz eines Datawarehouses eine Rolle, sondern auch wer darauf Zugriff hat und damit die vorhandenen Informationen abrufen kann. Dieser Punkt wurde bereits unter dem Begriff Demokratisierung des Datenzugriffs diskutiert.



Die Verteilung der Zugriffrechte in den befragten Unternehmen zeigt, dass großteils noch eine klassische Berechtigungsstruktur vorherrscht, in denen der Zugang zu Informationen den Führungskräften bzw. denjenigen, die aufgrund ihrer Tätigkeit zwingend einen Zugriff benötigen, vorbehalten ist.



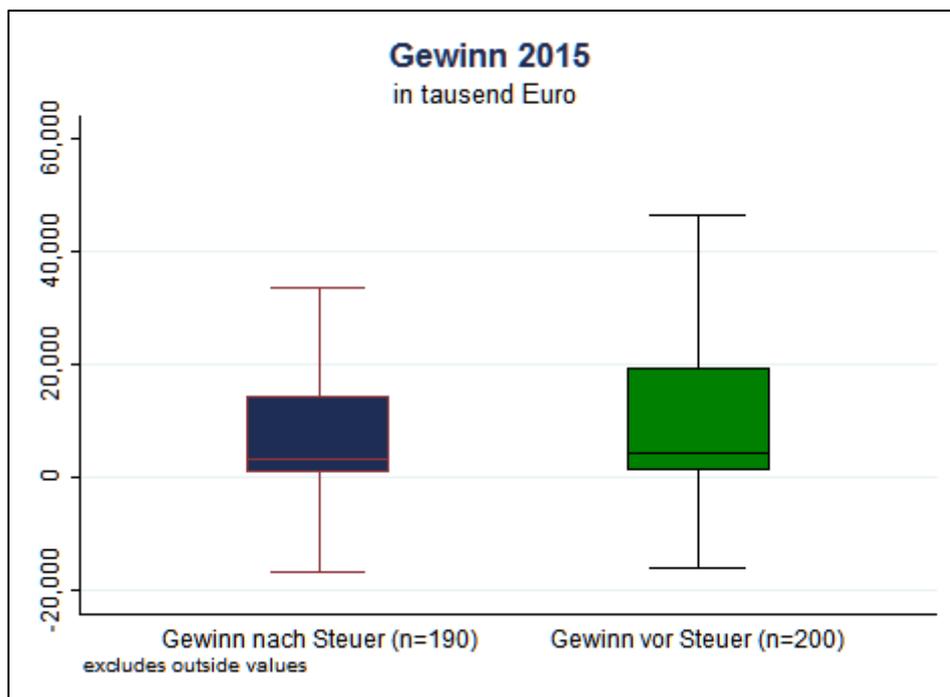
Ein Aspekt im Bereich Big Data sind die genutzten Werkzeuge, die gewisse Rückschlüsse auf den Reifegrad zulassen, da sie entscheidende Möglichkeiten eröffnen oder einschränken. Hierbei wird offensichtlich, dass die Verbreitung konkreter Big Data Technologien wie Hadoop oder Data Mining Werkzeugen noch sehr gering ist. Dagegen werden bekannte und in den Unternehmen bereits vorhandene Softwarelösungen auch für diesen Bereich genutzt. Eine spezielle Erwähnung verdient außerdem der Punkt, dass ganze 86 Unternehmen angaben, für die Datenanalyse Individualentwicklungen zu nutzen. Es lässt sich nicht feststellen, auf welchem Komplexitätsniveau diese Lösungen anzusiedeln sind, allerdings könnte man durchaus eine hohe Bereitschaft der Unternehmen in die Datenanalyse zu investieren ableiten, da solche Lösungen in der Regel deutlich teurer als Standardlösungen, im Gegenzug jedoch auf die Bedürfnisse des Unternehmens zugeschnitten sind.

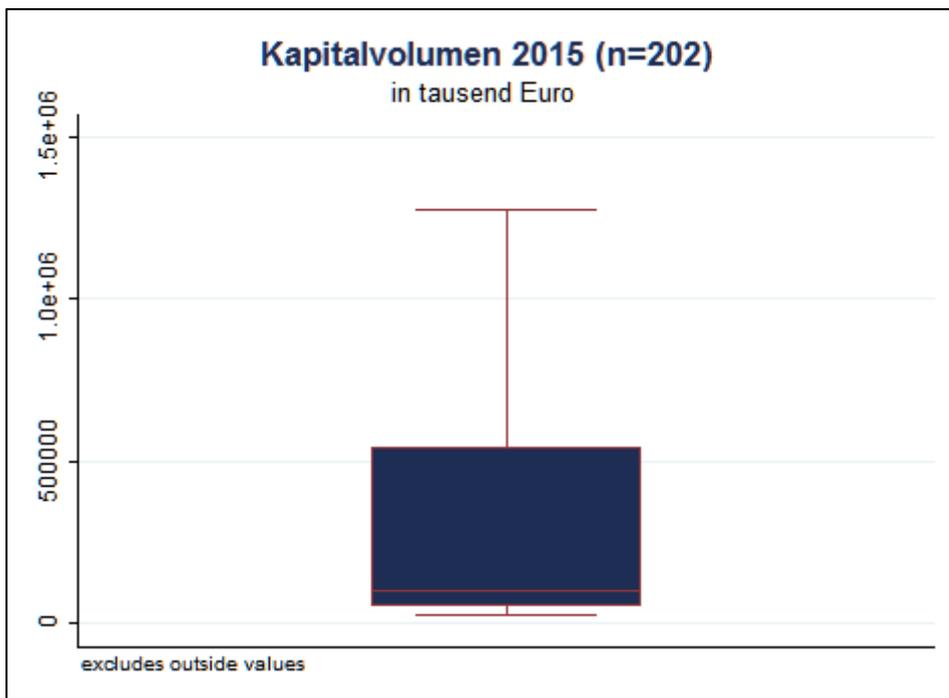
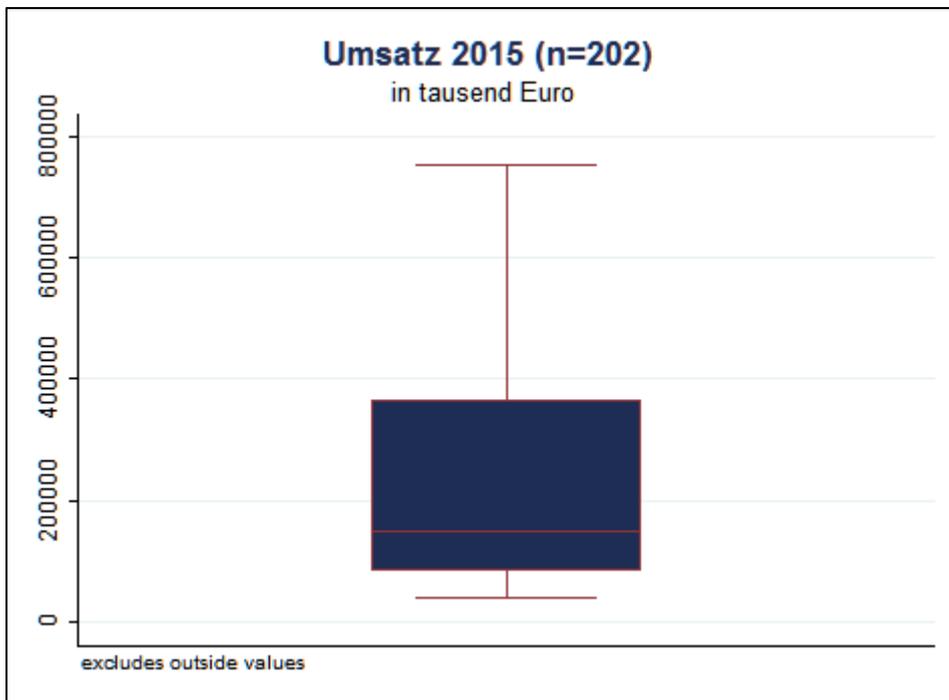
Betrachtet man diese Daten in Kombination mit denen zur Datengrundlage, so wird erkennbar, dass die meisten Unternehmen versuchen ihre bewährten Werkzeuge, in denen das vorhandene Personal geschult ist, zu nutzen und diese mit neuen Daten zu füttern. Dieses Vorgehen kann jedoch maximal eine Übergangslösung darstellen, da es entscheidend die

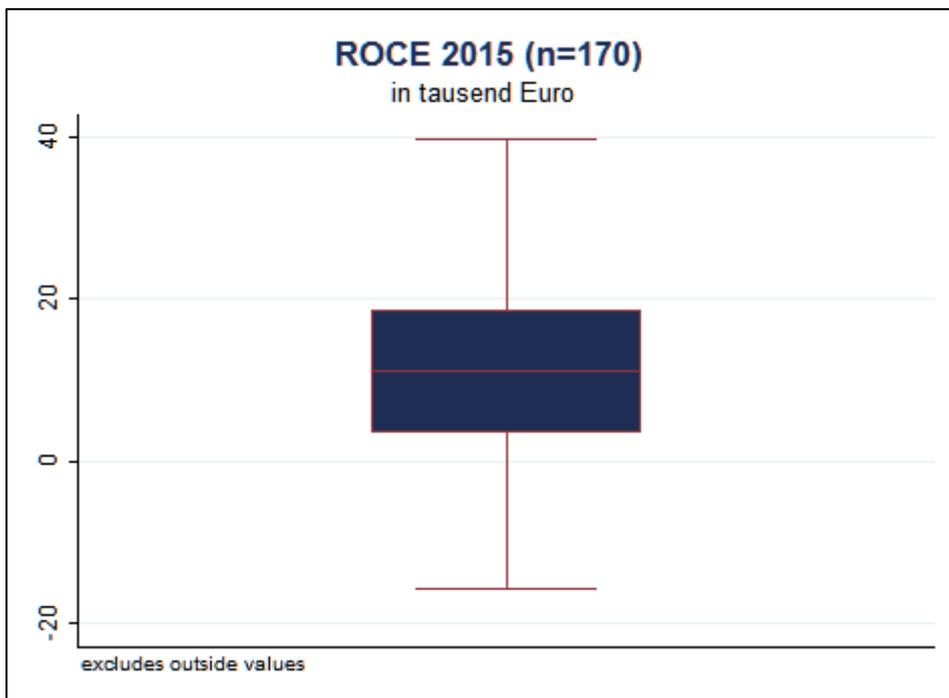
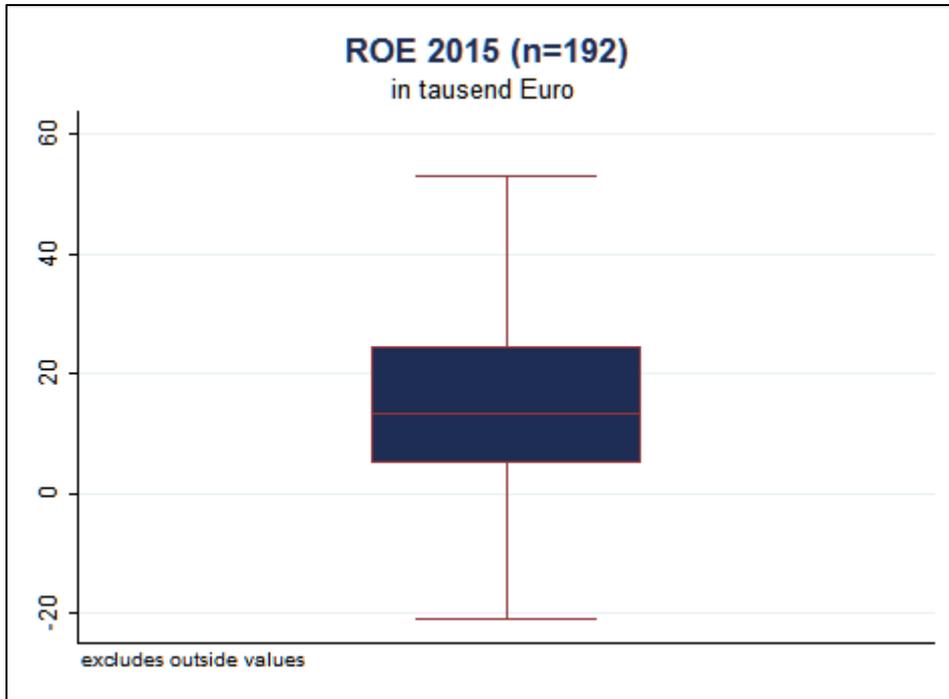
Potentiale der Datenanalyse, insbesondere im Bereich Echtzeitauswertung der Daten, einschränkt. Verständlicherweise ist es allerdings ein sehr attraktiver Zwischenschritt für viele Unternehmen, da keine großen Investitionen in die IT-Infrastruktur oder neues Personal, das ohnehin sehr schwer zu finden ist, notwendig sind.

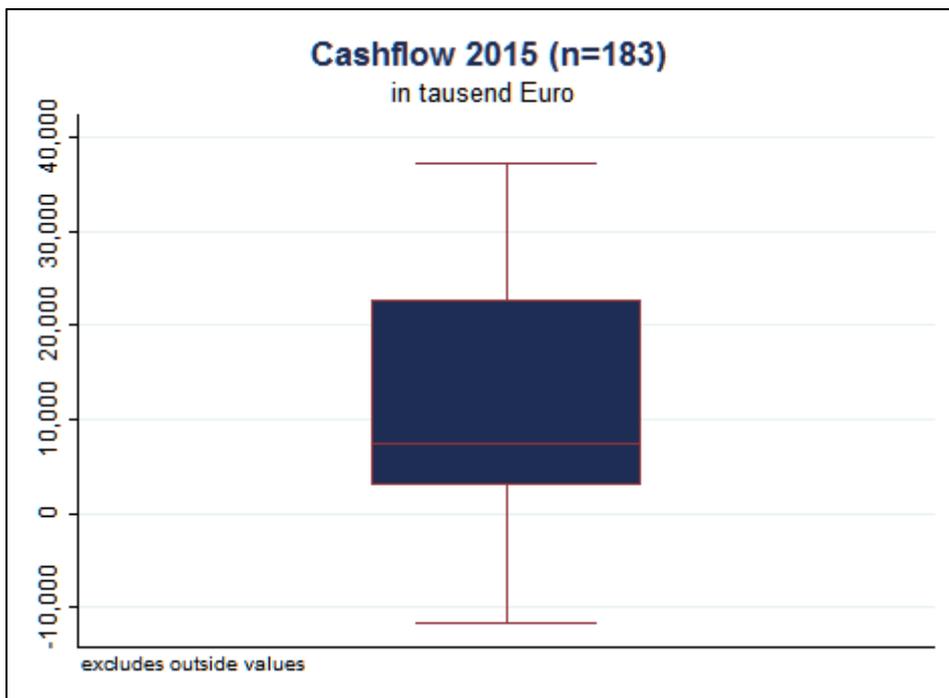
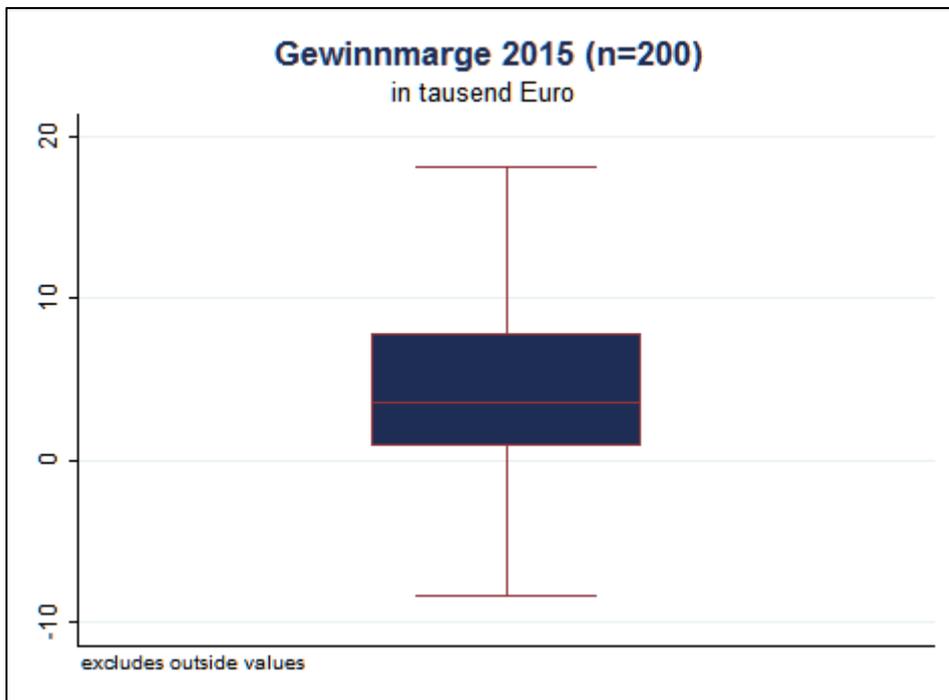
### 6.3. Finanzinformationen der Unternehmen

Um einen Eindruck zu bekommen, welche Finanzcharakteristika und Größe die befragten Unternehmen aufweisen, sollen diese im Folgenden als letzter Punkt der deskriptiven Analyse näher betrachtet werden. Diese stellen die zentralen, abhängigen Variablen der folgenden induktiven Analyse dar und sind somit ein essentieller Bestandteil dieser Arbeit. Deshalb wurde auch der Zusatzaufwand wie in Kapitel 5.1. beschrieben, in Kauf genommen, um eine möglichst hohe Güte des Datenmaterials zu gewährleisten. Wichtig war in diesem Bereich erneut, Ausreißer auszuschließen, was zwar zu einem gewissen Informationsverlust führt, jedoch wäre ansonsten keine sinnvolle Darstellung der Ergebnisse möglich. Die Unterschiede zwischen einigen sehr großen Unternehmen und der großen Masse sind schlicht zu groß.









## Lagemaße der Finanzdaten

(in tausend Euro und %)

	1. Quartil	Median	Arythm. Mittel	3. Quartil
Gewinn (v.S.)	1160	4332	22187	19270
Gewinn (n.S.)	769	3156	18939	14160
Umsatz	80391	147571	849573	366447
Kapitalvolumen	47154	99890	243291	541903
ROE (v.S.)	5,08	13,28	21,31	24,42
ROCE	3,48	11,11	12,94	18,66
Gewinnmarge	0,86	3,56	4,58	7,86
Cashflow	2364	7503	67843	22720

Wie bereits bei der Erstellung der Grafiken, wo das Problem der Darstellung mit Ausreißern bestand, zeigen sowohl die optische Analyse der Boxplots als auch die Verteilungswerte in obiger Tabelle, dass in dem vorliegenden Datensatz, ähnlich wie auch in der Realität, die meisten Unternehmen eher im unteren Größenbereich angesiedelt sind, jedoch einige um ein Vielfaches größere existieren. Dieses Verhältnis zeigt sich zum einen in der durchgängig großen Nähe des Medians zum 1. Quartil und dem deutlichen Unterschied in der Größe der Wiskers. In obiger Tabelle wird dies besonders durch die großen Unterschiede zwischen Median und arithmetischem Mittel deutlich, dass bei allen absoluten Zahlen massiv nach oben abweicht, und mit einer Ausnahme beim Kapitalvolumen sogar höhere Werte als das 3. Quartil annimmt.

## 7. Hypothesentests und Induktive Verfahren

Da nun ein genereller Überblick über das vorliegende Datenmaterial gewonnen wurde, soll in der Folge der Blick erneut auf die theoretisch generierten Hypothesen geworfen werden. Hierzu dient eine Reihe von hypothesenprüfenden, statistischen Verfahren sowie Regressionsanalysen.

## 7.1 H<sub>1</sub>

H<sub>1</sub> behauptet, dass ein Unternehmen erfolgreicher ist, wenn es sich Big Data Methoden zu eigen macht. Geprüft werden soll dies zunächst durch einige OLS Modelle, in welchen die Variablen „Dauer der Berichterstellung“ (In und Output), „generelle Nutzung der Datenanalyse“ in Form des Indexwertes, „Nutzung der Datenanalyse“ zur Krisenprävention, der Zugriffsindex sowie die genutzten Analysewerkzeuge einfließen. Die Analysen der Effekte dieser Variablen zeigten folgende Erkenntnisse: Wird der ROE als abhängige Variable genutzt, ist nahezu keine der vorliegenden Variablen signifikant. Einzig die Nutzung von Data Mining und Predictive Analytics Lösungen als Werkzeug zeugt von einem signifikanten Einfluss mit einem Koeffizienten von 116,28 und einem Standardfehler von 45,98. Diese Werte sind jedoch mit äußerster Vorsicht zu genießen. Während der positive Effekt klar erkennbar ist, erscheint die Stärke des Effekts doch deutlich zu groß, um als allgemeingültiger Wert angenommen zu werden. (Anhang 4)

Ein interessanter Effekt zeigt sich, wenn der ROCE als abhängige Variable genutzt wird, sämtliche vorherigen unabhängigen Variablen gleich bleiben und für den ROE kontrolliert wird. Zum einen bleibt die Nutzung von Data Mining und Predictive Analytics Lösungen signifikant, der Koeffizient wird jedoch negativ mit einem Wert von -19,19 und einem Standardfehler von 8,61. Eine theoretische Erklärung für diesen Effekt kann nicht gefunden werden. Wesentlich bedeutender jedoch ist, dass der Effekt des Indexwertes zur Nutzung der Datenanalyse nun signifikant ist, mit einem Koeffizienten von 8,88 und einem Standardfehler von 4,22. Eine mögliche Interpretation dieser Daten ist, dass der ROCE unter Kontrolle des ROE den Einfluss der Variablen auf die Kosten des Fremdkapitals abbildet. Es scheint so, als würde eine intensive Nutzung der Datenanalyse zu einem besseren Bild der Umgebung führen, welches zu einer höheren Bonität führt. Eventuell, da dieses bessere Bild der Umgebung in das Risikomanagement des Unternehmens miteinfließt und dies positiv von Fremdkapitalgebern gewürdigt wird. (Anhang 5)

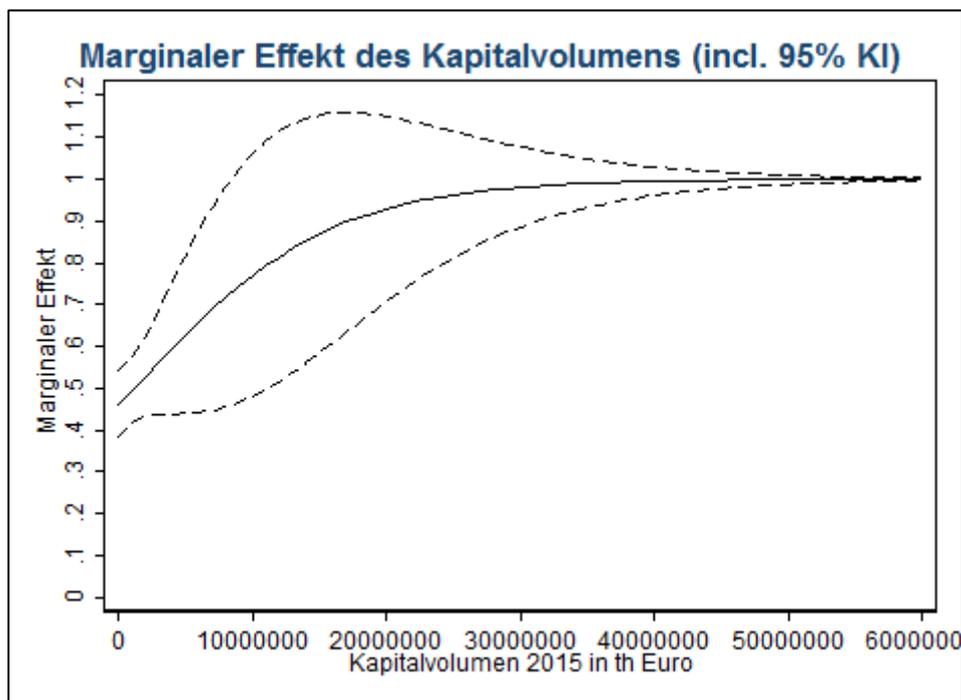
Betrachtet man dagegen den Einfluss der Variablen auf den absoluten Gewinn des Unternehmens (nach Steuer) so ergeben sich ein signifikanter Werte für die Datenanalyse zur Erkennung von Krisen (Koeffizient =-48051; Std. Fehler= 18632) und ein hochsignifikanter Wert für die Nutzung eines Hadoop Ökosystems (Koeffizient= 264838; Std. Fehler= 84647). Ersteres ist ein durchaus überraschendes Ergebnis, da zwar mit einem Effekt gerechnet wurde, jedoch einem positiven. Inwieweit sich dies erklären lässt, oder nur

auf einem  $\alpha$ -Fehler beruht, kann im Rahmen dieser Arbeit nicht geklärt werden, da hierfür eine Metastudie oder eine deutliche größere Datenbasis notwendig wäre (Anhang 6).

Insgesamt zeigt sich somit ein unklares Bild, ob die Datenanalyse einen positiven Einfluss auf den Unternehmenserfolg hat. Einzelne Variablen, insbesondere Werkzeuge, weisen zwar signifikante Effekte auf, jedoch treten diese nicht konsequent auf, sondern sind jeweils nur für eine abhängige Variable nachweisbar. Dies lässt zumindest die Wahrscheinlichkeit von  $\alpha$ -Fehlern zu, da eine große Anzahl an Variablen getestet wurde. Somit kann  $H_1$  zwar für die diskutierten Fälle angenommen werden, in denen signifikante Werte vorlagen, eine generelle Aussage zu  $H_1$  kann jedoch nicht getroffen werden.

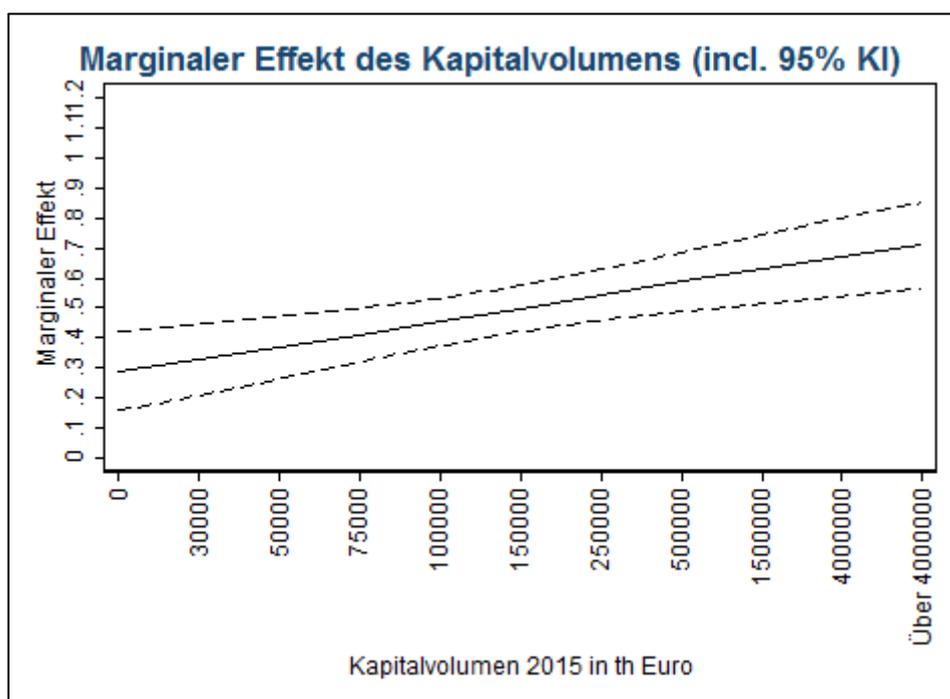
## 7.2 $H_2$

Um diese Hypothese zu untersuchen, soll zunächst die Existenz eines Datawarehouses zu Beurteilung des Fortschritts der Datenanalyse herangezogen werden. Hierfür dient ein logistisches Regressionsmodell.

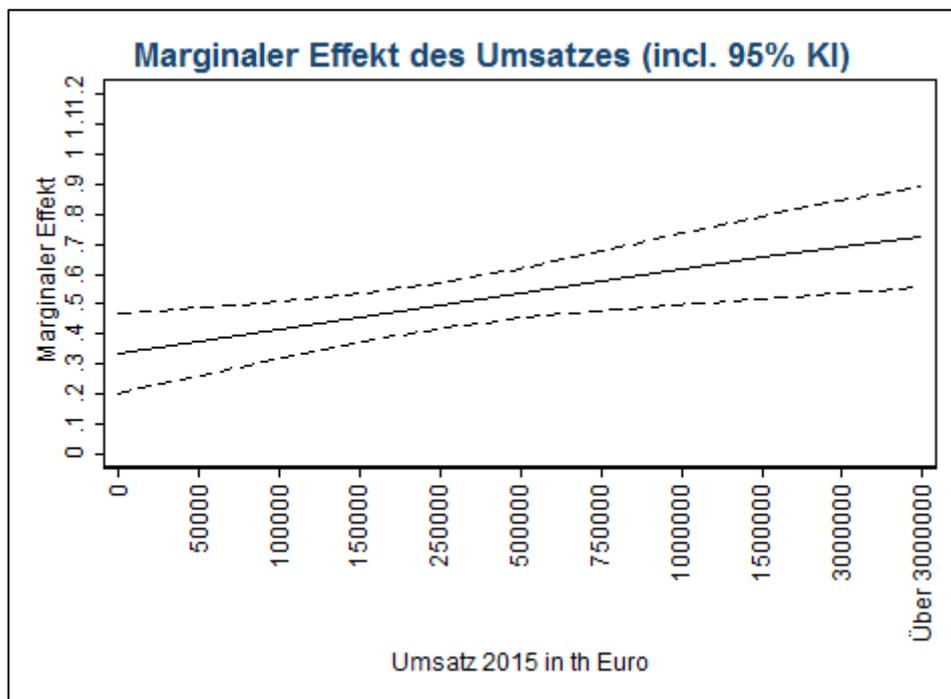


Die Analyse zeigt dabei, wie der obigen Grafik und Anhang 7 entnommen werden kann, höchstsignifikante marginale Effekte des Kapitalvolumens auf die Wahrscheinlichkeit ein zentrales Datawarehouse im Unternehmen zu haben. Die stark ansteigenden Werte im

Bereich bis zu 2 € Mrd. zeigen, dass besonders in diesem Bereich die steigende Größe eines Unternehmens einen deutlichen Einfluss darauf hat, ob es ein solches Datawarehouse gibt oder nicht. Bei Unternehmen über 2 € Mrd. steigt dieser Wert kaum weiter, da ein Sättigungseffekt eintritt. Das Problem dieser Grafik ist allerdings, dass das Grundmodell, welches dieser Grafik zugrunde liegt, mit einem P-Wert von 0,12 nicht signifikant ist. Der Grund hierfür ist vermutlich die sehr rechtsschiefe Verteilung der Werte (Schiefe=6,72), da wie bereits festgestellt deutlich mehr kleine Unternehmen dem Datensatz zu Grunde liegen als große. Dieses Problem wurde durch eine erneute Clusterbildung, anhand der bereits für den Fragebogen generierten Kategorien gelöst (Schiefe=0,97). Berechnet man das Modell mit dem geclusterten Kapitalvolumen, erhält man ein hochsignifikantes Grundmodell und erneut höchstsignifikante Effekte. In diesem Fall scheint der Zusammenhang in Bezug auf die marginalen Effekte auf den ersten Blick nahezu linear, dies täuscht jedoch, da die X-Achse durch die Cluster nicht linear gestaltet ist (Anhang 8).



Nahezu dasselbe Bild zeigt sich, wenn man den Umsatz als Grundlage der Unternehmensgröße heranzieht. Das ursprüngliche Modell ist mit einem P-Wert von 0,16 nicht signifikant, durch die Clusterbildung jedoch kann die problematische Verteilung der Datengrundlage umgangen werden und das Modell wird mit einem P-Wert von 0,013 signifikant (Anhang 9 & 10):



Zu einem gewissen Teil mag diese Ähnlichkeit der Ergebnisse durchaus mit der hohen Korrelation zwischen Umsatz und Kapitalvolumen eines Unternehmens von 0,638 zu begründen sein, jedoch nicht ausschließlich, sodass sich durchaus sagen lässt, dass die Größe eines Unternehmens sich positiv auf die Existenz eines zentralen Datawarehouses auswirkt und damit  $H_0$  verworfen werden kann.

### 7.3 $H_3$

Um  $H_3$  zu prüfen wurde zunächst ein Chi<sup>2</sup>-Test auf statistische Unabhängigkeit zwischen der Branche und der Existenz eines Datawarehouses durchgeführt. Dieser ergab jedoch kein signifikantes Ergebnis, so dass hier kein Zusammenhang postuliert werden kann (Anhang 11). Auch ein logistisches Regressionsmodell bestätigt diese Aussage. (Anhang 12). Selbst die Bildung von Clustern bzgl. IT- (IT & Telekommunikation) und Finanzbranche (Banken & Versicherungen) führt zum selben Ergebnis. Der Hauptgrund dürfte die sehr geringe Fallzahl sein, da sich im Cluster IT gerade einmal 7 Unternehmen und im Cluster Finanz 14 Unternehmen finden. Eine mögliche Erklärung hierfür wäre aber auch, dass selbst in diesen Branchen solche Technologien sich auf wenige Marktführer wie Google oder Facebook beschränken, aber nicht großflächig von allen Unternehmen genutzt werden, sodass diese

Ausreißer, zumindest in einer nicht nach Unternehmensgröße gewichteten Erhebung, untergehen.

Auf Basis der hier vorliegenden Daten muss  $H_3$  somit verworfen werden. Jedoch stellt sich die Frage, wie eben erläutert, ob diese Aussage generalisierbar oder einzig dem Datenmaterial geschuldet ist.

#### 7.4 $H_4$

Für die These, dass die Eigentümerstruktur einen Einfluss auf die Datenanalyse hat, wurde zunächst erneut ein logistisches Regressionsmodell mit der Existenz eines Datawarehouses berechnet. Mit einem P-Wert von 0,12 war dieses jedoch nicht signifikant (Anhang 13). Auch ein statistische Unabhängigkeitstest zeigt mit einem  $\chi^2=0,36$ , dass kein erkennbarer Zusammenhang zwischen der Eigentümerstruktur und der Existenz eines Datawarehouses besteht (Anhang 14). Selbst eine Umformung der Variablen im Sinne der Hypothese in inhabergeführte und managementgeführte Unternehmen verändert hier nichts, da die P-Werte sogar deutlich höher ausfallen.

Zusätzlich wurde zum Test dieser Hypothese mit einer Regressionsanalyse gearbeitet, welche die Wirkung der Eigentümerstruktur als kategoriale Variable auf die Nutzung der Datenanalyse testet. Hierbei zeigte sich, dass es zwar einen Effekt gibt, jedoch sind davon nicht wie erwartet Familien- oder inhabergeführte Unternehmen betroffen. Hier ließ sich kein signifikanter Unterschied zu Aktienunternehmen nachweisen. Öffentliche Unternehmen dagegen weisen eine signifikant niedrigere Nutzung von Big Data Methoden auf mit einem durchaus beachtlichen Koeffizienten von -0,527, was im Schnitt immerhin einer halben Stufe auf der Likertskala entspricht (Anhang 15).

Somit muss  $H_4$  klar abgelehnt werden, da kein signifikanter Unterschied wie vermutet gefunden werden konnte. Jedoch zeigte sich, dass staatliche Unternehmen deutlich weniger auf Datenanalyse setzen, was vermutlich den Verflechtungen mit dem politischen System geschuldet ist, so dass politische Zielsetzungen das ökonomische Maximierungsprinzip überlagern. Verantwortlich hierfür könnten allerdings auch die teilweise unflexiblen Strukturen und langen Genehmigungsprozesse für Investitionen sein, die in staatlichen Unternehmen vorherrschen.

## 7.5 Gesamtmodell

Um die induktive Analyse abzuschließen, soll nun noch ein Blick auf das Gesamtmodell geworfen werden. Hierzu werden die Konstrukte wie in der Operationalisierung und im Modell beschrieben, herangezogen. Außerdem sollen wie bereits in Kapitel 4.3. dargelegt, zum einen der ROE, als auch der absolute Gewinn als zwei Optionen der abhängigen Variablen herangezogen werden. Für den ROE zeigt sich als Ergebnis zunächst, dass das Modell insgesamt einen relativ hohen Erklärungswert aufweist, der sich in einem  $R^2=0,48$  widerspiegelt. Durch die hohe Anzahl der dafür notwendigen Variablen, jedoch beträgt das adjusted  $R^2$  0,17. Leider zeigt sich jedoch, dass nur wenige der verwendeten Variablen einen signifikanten Effekt aufweisen. Diese sind: Die durchschnittliche Berichtsdauer repetitiver Ereignisse ( $P=0,000$ ), die Nutzung analytischer Datenbanken ( $P=0,023$ ), die Nutzung von Data Mining und Predictive Analytics Lösungen ( $P=0,008$ ). Überraschend ist dabei, dass der Koeffizient von 0,72 bei der durchschnittlichen Berichtsdauer einen positiven Zusammenhang nachweist, also genau entgegen der Hypothese, dass eine kürzere Dauer der Berichtserstellung einen positiven Effekt auf die Rentabilität aufweist. Mit Koeffizienten von 96,20 und 148,00 entspricht die Richtung des Zusammenhangs für die beiden Werkzeuge den Erwartungen, jedoch ist fragwürdig wie zuverlässig die Höhe dieser Koeffizienten bewertet werden kann, da sie extrem hoch wirken und es unwahrscheinlich ist, dass der ROE von Unternehmen, die solche Werkzeuge nutzen 96% bzw. 148% höher ausfällt.

Der letzte signifikante Wert, nämlich derjenige von Familienunternehmen ( $P=0,007$ ), muss hiervon abweichend interpretiert werden, denn er bezieht sich auf den Unterschied zwischen Familienunternehmen und Aktienunternehmen im Streubesitz, der signifikant ist. Auch hier muss der Koeffizient sehr kritisch betrachtet werden, da dessen Aussage wäre, dass der ROE von Familienunternehmen 188% höher ist als von Aktienunternehmen im Streubesitz. Bei allen hier genannten Werten muss außerdem auf die sehr hohen Konfidenzintervalle verwiesen werden. Genaueres hierzu kann Anhang 16 entnommen werden.

Für das Gesamtmodell mit dem absoluten Gewinn als abhängige Variable zeigen sich ähnliche Gütewerte mit einem  $R^2=0,47$  und einem adj.  $R^2= 0,17$ . Jedoch weisen gänzlich andere Variablen signifikante Effekte auf, nämlich die Nutzung eines Hadoop Ökosystems ( $P=0,006$ ) sowie die Nutzung von Standard BI-Werkzeugen ( $P=0,041$ ). Dabei zeigt sich, dass der Effekt eines Hadoop Ökosystems vermutlich deutlich größer ist als derjenige der Standard BI-Werkzeuge, dies jedoch aufgrund von sich überschneidenden

Konfidenzintervallen nicht mit Gewissheit gesagt werden kann. Insbesondere, da diese auch in diesem Modell sehr groß sind.

Bzgl. der kategorialen Variablen zeigt sich, dass hier einzig ein höchstsignifikanter Unterschied zwischen dem Bankensektor und der Öl-, Gas- und Bergbaubranche nachgewiesen werden kann, und zwar im Schnitt ca. eine halbe Million geringere Gewinne pro Unternehmen. Dies ist jedoch für das Jahr 2015 kaum überraschend, betrachtet man den niedrigen Ölpreis, der dieser Branche stark zusetzt. Genauere Daten können Anhang 17 entnommen werden.

## 8. Zusammenfassende Analyse und Ausblick

Insgesamt stellte sich heraus, dass die Datenanalyse durchaus einen Einfluss auf den Finanzerfolg von Unternehmen hat. Dies zeigt sich durch verschiedene signifikante Zusammenhänge, die anhand der Regressionsanalysen nachgewiesen werden konnten. Allerdings stellte sich auch heraus, dass die errechneten Werte eine zu hohe Schwankung aufweisen, um eine verlässliche Basis für konkrete Investitionsberechnungen zu bilden. Dies ist wohl in erster Linie auf den Umfang der Datenbasis zurückzuführen. Nähere Untersuchungen zu diesem Thema wären deshalb durchaus sinnvoll, dürften sich allerdings als schwierig erweisen, da sich zeigte, dass die Thematik die internen Prozesse von Unternehmen in einer Weise berührt, die sie nur ungern nach außen kommunizieren.

Auch zeigten sich deutlich die Probleme, die mit der mangelnden einheitlichen Definition von Big Data einhergehen, so fanden die Antworten auf die offene Frage nach den automatisierten Prozessen auf Seite 10 des Fragebogens keinen Eingang in diese Arbeit, da sich zeigte, dass die Teilnehmer der Umfrage hier in vielen Fällen Prozesse aus dem Rechnungswesen nannten, die zwar automatisiert werden, aber in keinem Zusammenhang zu Big Data Lösungen stehen, wie sie hier definiert wurden.

Eine überraschende Erkenntnis ist, dass sich keine signifikanten Unterschiede zwischen den Branchen finden ließen. Während dies bei der Aufgliederung in die Einzelbranchen auf die geringe Fallzahl pro Branche zurückgeführt werden könnte, zeigen die Daten, dass es auch im Cluster „IT“ also IT- und Telekommunikationsunternehmen ebenso wie der Cluster „Finanz“ mit Banken und Versicherungen keine signifikanten Unterschiede in der Nutzung von Big Data Technologien gibt. Allerdings ist auch dieser Punkt durchaus mit Vorsicht zu

genießen, da selbst nach der Clusterbildung der Umfang der Cluster sehr gering war, was einen möglichen Grund für die nicht signifikanten Ergebnisse darstellt.

Leider konnten viele der in dieser Arbeit angedeuteten Zusammenhänge nicht endgültig bewiesen oder verworfen werden, teilweise da das zugrundeliegende Datenmaterial seine Grenzen erreichte, wie die beim Einfluss der Branche auf die Big Data Nutzung oder als Kontrollvariable war oder weil sich verschiedene Variablen als schlechte Indikatoren herausstellten. Anschließend wäre deshalb eine konfirmatorische Faktoranalyse der Daten von großem Wert, um vermutete Zusammenhänge fundierter zu bestätigen, oder falls dies nicht gelingen sollte, eine explorative Faktorenanalyse um die relevanten Faktoren im Bereich Big Data zu identifizieren, die einem gesteigerten Unternehmenserfolg zugrunde liegen. Der Vorteil der explorativen Faktoranalyse würde hierbei insbesondere darin liegen, dass eventuell vorhandenen Drittvariablen identifiziert werden könnten, die bisher noch nicht bedacht wurden. Beides würde jedoch den Umfang dieser Arbeit sprengen.

Was die Arbeit sehr konkret zeigte, ist der Zusammenhang zwischen der Größe eines Unternehmens und dem Stand der Big Data Nutzung oder konkreter, der Existenz eines Datawarehouses. Da ein Datawarehouse eine entscheidende Voraussetzung für eine konsistente Auswertung der vorhandenen Daten darstellt, kann dies zumindest in begrenztem Umfang mit der Nutzung von Big Data Methoden gleichgesetzt werden. Dies ist insbesondere deshalb von Bedeutung, da die meisten anderen Daten, u.a. vermutlich bedingt durch die bereits erwähnte Fallzahl und die Aufteilung in eine Vielzahl an Kategorien, keine signifikanten Ergebnisse lieferten. Somit muss festgehalten werden, dass der schlussendliche Hauptnutzen dieser Arbeit in einer deskriptiven Beschreibung des status quo liegt und der Versuch den Nutzen der Datenanalyse zu quantifizieren nur in Ansätzen gelungen ist. Diese Beschreibung liefert allerdings wichtige Erkenntnisse über den Stand der Entwicklung, sowie die Stoßrichtung der Unternehmen, die sich daraus ableiten lässt.

Hier ist grundsätzlich anzunehmen, dass in Zukunft massive Investitionen getätigt werden, welche in den Ausbau der IT-Infrastruktur ebenso wie in qualifiziertes Personal fließen werden. Denn die bisher verwendeten Analysetools schränken die Möglichkeiten der Datenanalyse deutlich ein. Insbesondere was den Bereich der Echtzeitanalyse betrifft, aber auch bzgl. der Komplexität der zugrundeliegenden Modelle bestehen hier Potentiale.

Ebenso schränkt zurzeit die Datenbasis die Nutzung der Technologien ein, da der Fokus stark auf internen Daten liegt. Eine Erweiterung dieser Datenbasis um externe Daten würde sowohl die Qualität, als auch die Anwendungsbereiche der Methoden deutlich erweitern. Jedoch bestehen hier durchaus auch Hürden bzw. Risiken. Zum einen würde eine

Plausibilisierung und Homogenisierung der externen und internen Daten notwendig werden, um zu belastbaren Ergebnissen zu kommen, was zwar technologisch möglich ist, jedoch einen entsprechenden Aufwand mit sich bringt. Außerdem bestehen in diesem Bereich deutliche Konflikte mit dem existierenden Datenschutzrecht, das die Nutzung externer Daten bzw. den Handel mit Daten nur sehr eingeschränkt zulässt. Eine Änderung dieser Normen im Sinne eines Datenschutzrechts, das den Handel anonymisierter Daten erlaubt, um als Grundlage für Analysen und damit dem Erkennen allgemeiner Muster zu dienen, müsste zum einen der Privatsphäre der Bürger, als auch dem Informationsbedürfnis der Unternehmen gerecht zu werden. Dies stellt jedoch ein hochpolitisches Thema dar und es ist aktuell kaum abzusehen in welche Richtung sich dies entwickelt.

## Literaturverzeichnis:

- Bange, Carsten, Grosser, Timm, Janoschek, Nikolai, Big Data Use Cases – Getting real on data monetization, 2015.
- Barney, Jay, Firm Resources and Sustained Competitive Advantage, in: Journal of Management, Vol. 17, Nr. 1, 1991, S. 99-120.
- Blue Yonder, Replenishment und Price Optimization bei OTTO – Wie der Multichannel-Händler seine Kunden begeistert und gleichzeitig den Umsatz steigert, 2016.
- Brownlow, Josh, Zaki, Mohamed, Neely, Andy, Urmetzer, Florian, Data and Analytics – Data Driven Business Models: A Blueprint for Innovation – The Competitive Advantage of the New Big Data World, 2015.
- Burow, Lothar, Leyk, Jörg, Briem, Christian, Experteninterview zum Thema “Controlling und Big Data”, in: Gleich, Roland, Grönke, Kai, Kirchmann, Markus, Leyk, Jörg (Hrsg.), Controlling und Big Data – Anforderungen, Auswirkungen, Lösungen, 2014, S.13-22.
- Buzzell, Robert D., Gale, Bradley T., Das PIMS-Programm, 1987.
- CapGemini, The Deciding Factor: Big Data & Decision Making, 2012.
- Computing Research, Big Data Review 2015 – A detailed investigation into the maturing of Big Data analytics, 2015.
- Duhigg, Charles, The Power of Habit – Why we do what we do in Life and Business, 2012.
- Ebner, Katharina, Bühnen, Thilo, Nils, Urbach, Think Big with Big Data: Identifying Suitable Big Data Strategies in Corporate Environments, in: 47<sup>th</sup> Hawaii International Conference on System Science, S. 3748-3757, 2016.
- Fama, Eugen F., Jensen, Michael C., Separation of Ownership and Control, in: Journal of Law and Economics, Vol. 26, Nr. 2, 1983, S.301-325.
- Gantz, John, Reinsel, David, The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East, 2012.
- Gehalt.de, Studie Gehaltsunterschied: 56% verdienst Du mehr in großen Unternehmen, 26.03.2015, URL: <https://www.gehalt.de/news/studie-gehaltsunterschied-56-verdienst-du-mehr-in-grossen-unternehmen>, entnommen am 07.06.2017.
- Gartner, Inc., Gartner Says 6.4 Billion Connected „Things“ Will Be in Use in 2016, Up 30 Percent From 2015, 10.11.2015, URL: <http://www.gartner.com/newsroom/id/3165317>, entnommen am: 16.09.2016.
- Hackathorn, Richard, Current practices in active data warehousing, 2002.
- Horváth & Partners, Vorlesungsunterlagen zu: Scheutz, Management-Informationssysteme zur Unternehmenssteuerung Kapitel 4: IT Systeme, 2016 (a).
- Horváth & Partners, Vorlesungsunterlagen zu: Scheutz, Management-Informationssysteme zur Unternehmenssteuerung Kapitel 5: Big Data, 2016 (b).

- Iffert, Lars, Predictive Analytics richtig einsetzen, in: Schäfer, Utz, Weber, Jürgern (Hrsg.), Controlling & Management Review Sonderheft, Vol. 1, 2016, S. 16-23.
- Immon, W.H., Building the Data Warehouse, 2002.
- Janssen, Marijn, Voort, Haiko van der, Wahyudi, Agung, Factors influencing big data decision-making quality, in: Journal of Business Research, Vol.70, 2017, S. 338-348.
- Kirchler, Erich M., Wirtschaftspsychologie – Grundlagen und Anwendungsfelder der Ökonomischen Psychologie, Göttingen, 2003.
- Laney, Doug, 3D Data Management: Controlling Data Volume, Velocity, and Variety, 2001.
- Lipshitz, Raanan, Strauss, Orna, Coping with Uncertainty: A Naturalistic Decision-Making Analysis, in: Organizational Behavior and Human Decision Processes, Vol. 69, Nr. 2, 1997, S.149-163.
- Mauerer, Jürgen, Was ist was bei Predictive Analytics, 26.08.2015, URL: <https://www.computerwoche.de/a/was-ist-was-bei-predictive-analytics,3098583,4>, entnommen am: 01.02.2017.
- Mayer-Schönberger, Viktor, Cukier, Kenneth, Big Data: A Revolution That Will Transform How We Live, Work, and Think, 2013.
- Mayerl, Jochen, Bedeutet ‘Big Data‘ das Ende der sozialwissenschaftlichen Methodenforschung? 29.12.2015, URL: <http://www.soziopolis.de/beobachten/wissenschaft/artikel/bedeutet-big-data-das-ende-der-sozialwissenschaftlichen-methodenforschung/>, entnommen am: 25.04.2017.
- Meyer, Roswitha, Entscheidungstheorie – Ein Lehr- und Arbeitsbuch, Wiesbaden, 2000.
- Mind Tools, Porter’s Five Forces, URL: [https://www.mindtools.com/pages/article/newTMC\\_08.htm](https://www.mindtools.com/pages/article/newTMC_08.htm), entnommen am: 12.09.2016.
- Mohamed, Nader, Al-Jaroodi, Jameela, Real-Time Big Data Analytics: Applications and Challenges, in: International Conference on high Performance Computing & Simulation, 2014, S.305-310.
- Moosbrugger, Helfried, Kelava, Augustin, Testtheorie und Fragebogenkonstruktion, 2012.
- Peng, Mike W., Sun, Sunny Li, Pinkham, Brian, Chen, Hao, The Institution-Based View as a Third Leg for a Strategy Tripod, in: Academy of Management Perspectives, Vol. 23, Nr. 3, 2009, S.63-81.
- Posch, Arthur, Speckbacher, Gerhard, Führung in Familienunternehmen: Besonderheiten der Entscheidungsfindung und Verhaltenssteuerung und deren Auswirkungen auf den Unternehmenserfolg, in: Zeitschrift für Betriebswirtschaft, Vol. 82, Nr. 5, 2012, S.5-23.
- Porter, Michael E., Competitive Strategy – Techniques for Analyzing Industries and Competitors, 1980.

- Quin, Hai-fei, Qian, Zhi-ming, Zhao, Yong-chao, On the Research of Data Warehouses in Big Data, in: International Conference on Network and Information Systems for Computers, 2015, S.354-357.
- Rossouw, Louis, Big Data in der Lebensversicherung, 2013.
- Ruta, Dymitr, Automated Trading with Machine Learning on Big Data, in: IEEE International Congress on Big Data, 2014, S.824-830.
- Schulze, William S., Lubatkin, Michael H., Dino, Richard N., Buchholtz, Ann K., Agency Relationships in Family Firms: Theory and Evidence, in: Organization Science, Vol. 12, Nr. 2, 2001, S.99-116.
- Simon, Herbert A., A Behavioral Model of Rational Choice, in: The Quarterly Journal of Economics, Vol. 69, Nr. 1, 1955, S. 99-118.
- Soelberg, P. O. (1967). Unprogrammed decision making. Industrial Management Review, 20, S. 19–29.
- Ortega, Ilias, Big-Data-System oder Data Warehouse?, 09.04.2013, URL: <http://www.computerwoche.de/a/big-data-system-oder-data-warehouse,2516320>, entnommen am: 02.09.2016.
- WKO, Branchenübersicht, 2016, URL: <https://www.wko.at/branchen/Branchenauswahl.html>, entnommen am: 03.10.2016
- Witten, Ian H., Frank, Eibe, Hall, Mark A., Data Mining – Practical Machine Learning Tools and Techniques, 2013.

# Anhang 1: Fragebogen

Begrüßung

Vielen Dank, dass Sie sich bereit erklären, an der Umfrage zur **Nutzung von Big Data Technologien in Großunternehmen** teilzunehmen.

Das Ausfüllen des Fragebogens wird etwa **15 Minuten** in Anspruch nehmen. Dabei können Sie Ihren Fortschritt jederzeit am rechten unteren Rand nachverfolgen. Sollten Sie unterbrochen werden, können Sie die Befragung jederzeit fortsetzen.

Im Folgenden erhalten Sie zunächst einige Hinweise zur Umfrage, bevor Sie zu den Fragen gelangen.

Verschwiegenheitserklärung

## Datenschutz und Verschwiegenheitserklärung

Zunächst möchte ich Ihnen hiermit versichern, dass sämtliche Antworten nur zum Zwecke meiner Masterarbeit verwendet und Informationen nur in aggregierter oder anonymisierter Form publiziert werden. Ihre Angaben werden keinerlei Dritten zugänglich gemacht.

## Big Data Definition

Da es sich bei Big Data zur Zeit um ein viel genutztes Schlagwort handelt, ist oft nicht klar, welche genaue Definition jeweils zugrunde liegt. Wenn im Folgenden von **Datenanalyse** generell die Rede ist, bezieht sich dies deshalb auf folgende Definition, welche auf das Business Application Research Center zurück geht:

Big Data bezeichnet Methoden und Technologien für die hochskalierbare Erfassung, Speicherung und Analyse polystrukturierter Daten\*. Big-Data-Analysen unterstützen die datengetriebene Planung, Steuerung und Kontrolle von Geschäftsprozessen und die Umsetzung digitaler Geschäftsmodelle.

\*Als polystrukturierte Daten wird eine Datengrundlage bezeichnet, die sich aus unterschiedlichen Formaten zusammensetzt, also beispielsweise Bild-, Ton-, Text- und Sensordaten kombiniert, um daraus Erkenntnisse zu erzielen.

Welche Position haben Sie in Ihrem Unternehmen inne?

AF01

- Führungskraft/Entscheidungsträger
- IT-Spezialist
- Mitarbeiter in der Datenanalyse
- Mitarbeiter in der strategischen Planung

In welchem Land sind Sie als Arbeitnehmer hauptsächlich tätig?

AF02

- Deutschland
- Österreich
- Schweiz

In welchem Land ist Ihr Unternehmen hauptsächlich tätig?

AF03

- Deutschland
- Österreich
- Schweiz

In welcher Branche ist Ihr Unternehmen tätig?

AF06

- Bankensektor
- Baugewerbe
- Beratung
- Bildung
- Prozessindustrie (z.B. Chemie, Pharma, Lebensmittelherstellung)
- Dienstleistungen (außer Beratung)
- Handel
- Fertigungsindustrie (z.B. Automotive, Maschinenbau)
- Medien/Verlagswesen
- Gesundheitswesen
- IT
- Landwirtschaft
- Öffentlicher Sektor
- Öl, Gas und Bergbau
- Telekommunikation
- Transportwesen und Logistik
- Versicherungen
- Energie und Versorgungswirtschaft

Wie ist die Eigentümerstruktur Ihres Unternehmens beschaffen?

AF05

- Familienunternehmen
- Inhabergeführt ohne familiären Zusammenhang
- Inhabergeführt mit einem einzigen Inhaber
- Aktienunternehmen mit Mehrheitseigentümer
- Aktienunternehmen im Streubesitz
- Unternehmen in öffentlicher Hand

Welche der folgenden Werkzeuge werden in Ihrem Unternehmen zur Datenanalyse genutzt?

DA10

- Standard Relationale Datenbanken
- Analytische Datenbanken (Exasol, Microsoft APS, Oracle Big Data Appliances)
- Hadoop Ökosystem
- NoSQL Datenbanken (Casandra, CouchDB, MarkLogic, MongoDB)
- Spezielle Big Data Analysewerkzeuge (Datameer, Spluk)
- Standard BI-Werkzeuge (IBM Congos, SAP BO)
- Data Mining/Predictive Analytics Lösungen (Azure ML, Blue Yonder, IBM SPSS, KXEN, R)
- Explorative Analysewerkzeuge (Infonea, OlikView, SAS Visual Analytics, Tableau, Tibco, Spotfire)
- Individualentwicklungen
- Standard Datenintegrationswerkzeuge
- Sonstige
- Keine

Welche der folgenden Datentypen werden in Ihrem Unternehmen für Datenanalysen genutzt?

DA11

- Logdaten von IT-Systemen
- Sensor-, RFID- oder andere Maschinendaten
- Clickstream-Daten
- Videoclips/Bilder
- Daten aus Transaktionssystemen
- Social-Media-Daten
- Dokumente/Texte (z.B. Emails)
- Sonstige
- Keine

Existiert in Ihrem Unternehmen ein zentrales Datawarehouse?

DA08

- Ja  
 Nein  
 Ist geplant

Welche Personen haben darauf Zugriff?

DA09

- Alle Führungskräfte/Entscheidungsträger  
 Führungskräfte/Entscheidungsträger auf Ihren Zuständigkeitsbereich beschränkt  
 IT-Spezialisten  
 Mitarbeiter in der Datenanalyse  
 Mitarbeiter in der strategischen Planung  
 Alle Mitarbeiter

Wie häufig nutzen Sie...

DA01

	Nie	Selten	Gelegentlich	Häufig	Immer
...Datenanalyse zur Bewertung Ihrer Konkurrenten?	<input type="radio"/>				
...Datenanalyse zur Bewertung Ihrer Zulieferer?	<input type="radio"/>				
...Datenanalyse zur Bewertung Ihrer Kunden?	<input type="radio"/>				
...Datenanalyse zur Überprüfung Ihrer Geschäftsprozesse?	<input type="radio"/>				
...Datenanalyse zur Planung von Projekten?	<input type="radio"/>				
...zusätzlich externe Daten für Ihre Analysen?	<input type="radio"/>				

Wie viel Zeit schätzen Sie, vergeht in Ihrem Unternehmen zwischen einem Ereignis und dem fertigen L...  
der dieses Ereignis enthält?

DA02

**Repetitive Ereignisse (z.B. Bestellungen, Verkaufsabwicklungen)**

Minimum  Tage

Maximum  Tage

Durschnitt  Tage

**Singuläre Ereignisse (z.B externe Schocks, strategische Entscheidungen)**

Minimum  Tage

Maximum  Tage

Durschnitt  Tage

Wie hoch würden Sie den Anteil der Automatisierung der Befüllung (Input) der Managementinformationssysteme in Ihrem Unternehmen einschätzen?

DA03 

- 0%
- 1-20%
- 21-40%
- 41-60%
- 61-80%
- 81-100%

Wie hoch würden Sie den Anteil der Automatisierung der Berichterstellung (Output) durch Managementinformationssysteme in Ihrem Unternehmen einschätzen?

DA12 

- 0%
- 1-20%
- 21-40%
- 41-60%
- 61-80%
- 81-100%

Nutzen Sie die Ihrem Unternehmen zugänglichen Informationen, um Entscheidungsprozesse zu automatisieren?

DA04 

- Ja
- Nein

DA05 

Falls ja, welche Prozesse automatisieren Sie?


Auf welcher Ebene würden Sie diese automatisierten Entscheidungsprozesse einordnen?

DA06

- Operativ
- Strategisch
- Gänzlich neues Geschäftsfeld

Seite 11

Nutzen Sie Datenanalyse zur Identifizierung von Krisen und Risiken...

DA07

	Nie	Selten	Gelegentlich	Häufig	Regelmäßig
...auf strategischer Ebene?	<input type="radio"/>				
...auf Projekt-Ebene?	<input type="radio"/>				

Damit ein Zusammenhang zwischen den bisher abgefragten Informationen und dem Erfolg von Unternehmen hergestellt werden kann, bitte ich Sie, im Folgenden noch einige Finanzkennzahlen Ihres Unternehmens anzugeben. Aufgrund der Sensibilität dieser Daten ist es natürlich verständlich, dass viele Unternehmen nicht in der Lage sind, eine genaue Angabe hierzu zu machen. Aus diesem Grund wird die Möglichkeit angeboten, sowohl genaue als auch geclusterte Angaben zu machen. Soweit es mit Ihren betrieblichen Richtlinien vereinbar ist, würde ich Sie bitten, beide auszufüllen. Sollten Sie keinerlei Angaben machen können, bitte ich Sie dennoch den Fragebogen fertig abzuschließen.

FK02

#### Wie hoch waren folgende Kennzahlen in Ihrem Unternehmen im Jahr 2015?

AF04

Umsatz	<input type="text"/>	€
Gewinn	<input type="text"/>	€
ROI	<input type="text"/>	%
Kapitalvolumen	<input type="text"/>	€
Anzahl der Mitarbeiter	<input type="text"/>	

#### Wie hoch waren diese Kennzahlen im Jahr 2016? (Falls bereits bekannt)

AF07

Umsatz	<input type="text"/>	€
Gewinn	<input type="text"/>	€
ROI	<input type="text"/>	%
Kapitalvolumen	<input type="text"/>	€
Anzahl der Mitarbeiter	<input type="text"/>	

#### Umsatz 2015 in Millionen €

FK05

Unter 50	50-99	100-149	150-249	250-499	500-749	750-999	1.000-1.499	1.500-3.000	Über 3.000
<input type="radio"/>									

#### Umsatz 2016 in Millionen €

Unter 50	50-99	100-149	150-249	250-499	500-749	750-999	1.000-1.499	1.500-3.000	Über 3.000
<input type="radio"/>									

Gewinn nach Steuer 2015 in Millionen €

FK06

Weniger als		-0,99 bis							
-5	-5 bis -1	0,99	1 bis 2,49	2,5 bis 4,99	5 bis 9,99	10 bis 24,99	25 bis 49,99	50 bis 99,99	Mehr als 100
<input type="radio"/>									

Gewinn nach Steuer 2016 in Millionen €

Weniger als		-0,99 bis							
-5	-5 bis -1	0,99	1 bis 2,49	2,5 bis 4,99	5 bis 9,99	10 bis 24,99	25 bis 49,99	50 bis 99,99	Mehr als 100
<input type="radio"/>									

ROI 2015

FK01

Unter -1%	-1 bis 0,99%	1 bis 2,99%	3 bis 4,99%	5 bis 7,49%	7,5 bis 10%	Über 10%
<input type="radio"/>						

ROI 2016

Unter -1%	-1 bis 0,99%	1 bis 2,99%	3 bis 4,99%	5 bis 7,49%	7,5 bis 10%	Über 10%
<input type="radio"/>						

Kapitalvolumen 2015 in Millionen €

FK03

Unter 30	30-49	50-74	75-99	100-149	150-249	250-499	500-1.499	1.500-4.000	Über 4.000
<input type="radio"/>									

Kapitalvolumen 2016 in Millionen €

Unter 30	30-49	50-74	75-99	100-149	150-249	250-499	500-1.499	1.500-4.000	Über 4.000
<input type="radio"/>									

FK04

### Angestellte 2015

Unter 250	250-499	500-999	1.000-2.499	2.500-4.999	5.000-9.999	10.000-19.999	20.000-49.999	50.000-99.999	Über 100.000
<input type="radio"/>									

### Angestellte 2016

Unter 250	250-499	500-999	1.000-2.499	2.500-4.999	5.000-9.999	10.000-19.999	20.000-49.999	50.000-99.999	Über 100.000
<input type="radio"/>									

Seite 13

Vielen Dank, dass Sie sich die Zeit genommen und diesen Fragebogen ausgefüllt haben. Falls Sie Interesse an den Ergebnissen haben, bitte ich Sie noch Ihre Email-Adresse anzugeben, damit ich Ihnen die Endfassung der Studie zukommen lassen kann.

EM01

Email-Adresse:

Letzte Seite

## Vielen Dank für Ihre Teilnahme!

Wir möchten uns ganz herzlich für Ihre Mithilfe bedanken.

Ihre Antworten wurden gespeichert, Sie können das Browser-Fenster nun schließen.

Matthias Wagenstaller, FHWien der WKW; sponsored by



## Anhang 2: Variablenübersicht

27.5.2017

Variablen-Übersicht

### Variablen-Übersicht

#### Fragebogen-interne Daten

Im Datensatz finden Sie neben Ihren Fragen folgende zusätzliche Variablen, sofern Sie die entsprechende Option beim Herunterladen des Datensatzes nicht deaktivieren.

**CASE** Fortlaufende Nummer der Versuchsperson

**REF** Referenz, falls solch eine im Link zum Fragebogen übergeben wurde

**LASTPAGE** Nummer der Seite im Fragebogens, die zuletzt bearbeitet und abgeschickt wurde

**QUESTNNR** Kennung des Fragebogens, der bearbeitet wurde

**MODE** Information, ob der Fragebogen im Pretest oder durch einen Projektmitarbeiter gestartet wurde

**STARTED** Zeitpunkt, zu dem der Teilnehmer den Fragebogen aufgerufen hat

**FINISHED** Information, ob der Fragebogen bis zur letzten Seite ausgefüllt wurde

**TIME\_001...** Zeit, die ein Teilnehmer auf einer Fragebogen-Seite verbracht hat

Bitte beachten Sie, dass Sie die Fragebogen-internen Variablen nicht mit der Funktion `value()` auslesen können. Für Interview-Nummer und Referenz stehen aber die PHP-Funktionen `PHP-Funktion caseNumber()` und `PHP-Funktion reference()` zur Verfügung.

Details über die zusätzlichen Variablen stehen in der Anleitung: [Zusätzliche Variablen in der Datenausgabe](#)

#### Rubrik AF: Allgemeine Fragen zum Unternehmen

**[AF01]** Auswahl

Position

"Welche Position haben Sie in Ihrem Unternehmen inne?"

**AF01** Position

- 1 = Führungskraft/Entscheidungsträger
- 2 = IT-Spezialist
- 3 = Mitarbeiter in der Datenanalyse
- 4 = Mitarbeiter in der strategischen Planung
- 9 = nicht beantwortet

**[AF02]** Auswahl

Land

"In welchem Land sind Sie als Arbeitnehmer hauptsächlich tätig?"

**AF02** Land

- 1 = Deutschland
- 2 = Österreich
- 3 = Schweiz
- 9 = nicht beantwortet

**[AF03]** Auswahl

Land Unternehmen

"In welchem Land ist Ihr Unternehmen hauptsächlich tätig?"

**AF03** Land Unternehmen

- 1 = Deutschland
- 2 = Österreich
- 3 = Schweiz
- 9 = nicht beantwortet

**[AF04]** Offene Texteingabe

Umsatz 2015

"Wie hoch waren folgende Kennzahlen in Ihrem Unternehmen im Jahr 2015?"

**AF04\_01** Umsatz ... €

**AF04\_02** Gewinn ... €

**AF04\_03** ROI ... %

**AF04\_04** Kapitalvolumen ... €

Offene Eingabe (Dezimalzahl)

**AF04\_05** Anzahl der Mitarbeiter

Offene Eingabe (Ganze Zahl)

**[AF05]** Auswahl

Eigentümerstruktur

"Wie ist die Eigentümerstruktur Ihres Unternehmens beschaffen?"

**AF05** Eigentümerstruktur

<https://www.socisurvey.de/admin/index.php>

1/6

1 = Familienunternehmen  
 2 = Inhabergeführt ohne familiären Zusammenhang  
 3 = Inhabergeführt mit einem einzigen Inhaber  
 4 = Aktienunternehmen mit Mehrheitseigentümer  
 5 = Aktienunternehmen im Streubesitz  
 6 = Unternehmen in öffentlicher Hand  
 -9 = nicht beantwortet

**[AF06] Auswahl****Branche**

"In welcher Branche ist Ihr Unternehmen tätig?"

**AF06 Branche**

1 = Bankensektor  
 2 = Baugewerbe  
 3 = Beratung  
 4 = Bildung  
 5 = Prozessindustrie (z.B. Chemie, Pharma, Lebensmittelherstellung)  
 6 = Dienstleistungen (außer Beratung)  
 7 = Handel  
 8 = Fertigungsindustrie (z.B. Automotive, Maschinenbau)  
 9 = Medien/Verlagswesen  
 10 = Gesundheitswesen  
 11 = IT  
 12 = Landwirtschaft  
 13 = Öffentlicher Sektor  
 14 = Öl, Gas und Bergbau  
 15 = Telekommunikation  
 16 = Transportwesen und Logistik  
 17 = Versicherungen  
 18 = Energie und Versorgungswirtschaft  
 -9 = nicht beantwortet

**[AF07] Offene Texteingabe****Umsatz 2016**

"Wie hoch waren diese Kennzahlen im Jahr 2016? (Falls bereits bekannt)"

**AF07\_01** Umsatz ... €

**AF07\_02** Gewinn ... €

**AF07\_03** ROI ... %

**AF07\_04** Kapitalvolumen ... €

Offene Eingabe (Dezimalzahl)

**AF07\_05** Anzahl der Mitarbeiter

Offene Eingabe (Ganze Zahl)

**Rubrik DA: Datenanalyse****[DA01] Skala (Zwischenwerte beschriftet)**

DA |

"Wie häufig nutzen Sie..."

**DA01\_01** ...Datenanalyse zur Bewertung Ihrer Konkurrenten?

**DA01\_02** ...Datenanalyse zur Bewertung Ihrer Zulieferer?

**DA01\_03** ...Datenanalyse zur Bewertung Ihrer Kunden?

**DA01\_04** ...Datenanalyse zur Überprüfung Ihrer Geschäftsprozesse?

**DA01\_05** ...Datenanalyse zur Planung von Projekten?

**DA01\_06** ...zusätzlich externe Daten für Ihre Analysen?

1 = Nie  
 2 = Selten  
 3 = Gelegentlich  
 4 = Häufig  
 5 = Immer  
 -9 = nicht beantwortet

**[DA02] Offene Texteingabe****Auswertungsdauer**

"Wie viel Zeit schätzen Sie, vergeht in Ihrem Unternehmen zwischen einem Ereignis und dem fertigen Bericht, d..."

**DA02\_01** Minimum ... Tage

**DA02\_02** Maximum ... Tage

**DA02\_03** Durchschnitt ... Tage

**DA02\_04** Minimum ... Tage

**DA02\_05** Maximum ... Tage

**DA02\_06** Durchschnitt ... Tage

Offene Eingabe (Dezimalzahl)
<p><b>[DA03]</b> Auswahl <b>Automatisierung</b> "Wie hoch würden Sie den Anteil der Automatisierung der Befüllung (Input) der Managementinformationssysteme i..."</p> <p><b>DA03</b> Automatisierung 1 = 0% 2 = 1-20% 3 = 21-40% 4 = 41-60% 5 = 61-80% 6 = 81-100% -9 = nicht beantwortet</p>
<p><b>[DA04]</b> Auswahl <b>Auto</b> "Nutzen Sie die Ihrem Unternehmen zugänglichen Informationen, um Entscheidungsprozesse zu automatisieren?"</p> <p><b>DA04</b> Auto 1 = Ja 2 = Nein -9 = nicht beantwortet</p>
<p><b>[DA05]</b> Offene Texteingabe <b>Auto</b> "Falls ja, welche Prozesse automatisieren Sie?"</p> <p><b>DA05_01</b> [01] <b>DA05_02</b> [02] <b>DA05_03</b> [03] <b>DA05_04</b> [04] <b>DA05_05</b> [05] <b>DA05_06</b> [06] <b>DA05_07</b> [07] <b>DA05_08</b> [08] <b>DA05_09</b> [09] <b>DA05_10</b> [10] Offene Texteingabe</p>
<p><b>[DA06]</b> Auswahl <b>Automatisierung</b> "Auf welcher Ebene würden Sie diese automatisierten Entscheidungsprozesse einordnen?"</p> <p><b>DA06</b> Automatisierung 1 = Operativ 2 = Strategisch 3 = Gänzlich neues Geschäftsfeld -9 = nicht beantwortet</p>
<p><b>[DA07]</b> Skala (Zwischenwerte beschriftet) <b>Krisen</b> "Nutzen Sie Datenanalyse zur Identifizierung von Krisen und Risiken..."</p> <p><b>DA07_01</b> ...auf strategischer Ebene? <b>DA07_02</b> ...auf Projekt-Ebene? 1 = Nie 2 = Selten 3 = Gelegentlich 4 = Häufig 5 = Regelmäßig -9 = nicht beantwortet</p>
<p><b>[DA08]</b> Auswahl <b>Warehouse</b> "Existiert in Ihrem Unternehmen ein zentrales Datawarehouse?"</p> <p><b>DA08</b> Warehouse 1 = Ja 2 = Nein 3 = Ist geplant -9 = nicht beantwortet</p>
<b>[DA09]</b> Erweiterte Auswahl

<p><b>Zugriff</b> "Welche Personen haben darauf Zugriff?"</p> <p><b>DA09_CN</b> TODO Ganze Zahl</p> <p><b>DA09x01</b> Alle Führungskräfte/Entscheidungsträger <b>DA09x02</b> Führungskräfte/Entscheidungsträger auf Ihren Zuständigkeitsbereich beschränkt <b>DA09x03</b> IT-Spezialisten <b>DA09x04</b> Mitarbeiter in der Datenanalyse <b>DA09x05</b> Mitarbeiter in der strategischen Planung <b>DA09x06</b> Alle Mitarbeiter</p> <p>1 = nicht gewählt 2 = ausgewählt</p>
<p><b>[DA10]</b> Erweiterte Auswahl Tools "Welche der folgenden Werkzeuge werden in Ihrem Unternehmen zur Datenanalyse genutzt?"</p> <p><b>DA10_CN</b> TODO Ganze Zahl</p> <p><b>DA10x01</b> Standard Relationale Datenbanken <b>DA10x02</b> Analytische Datenbanken (Exasol, Microsoft APS, Oracle Big Data Appliances) <b>DA10x03</b> Hadoop Ökosystem <b>DA10x04</b> NoSQL Datenbanken (Cassandra, CouchDB, MarkLogic, MongoDB) <b>DA10x05</b> Spezielle Big Data Analysewerkzeuge (Datameer, Splunk) <b>DA10x06</b> Standard BI-Werkzeuge (IBM Congos, SAP BO) <b>DA10x07</b> Data Mining/Predictive Analytics Lösungen (Azure ML, Blue Yonder, IBM SPSS, KXEN, R) <b>DA10x08</b> Explorative Analysewerkzeuge (Infonea, OlikView, SAS Visual Analytics, Tableau, Tibco, Spotfire) <b>DA10x09</b> Individualentwicklungen <b>DA10x10</b> Standard Datenintegrationswerkzeuge <b>DA10x101</b> Sonstige <b>DA10x1001</b> Keine</p> <p>1 = nicht gewählt 2 = ausgewählt</p>
<p><b>[DA11]</b> Mehrfachauswahl Datenbasis "Welche der folgenden Datentypen werden in Ihrem Unternehmen für Datenanalysen genutzt?"</p> <p><b>DA11</b> Datenbasis: Ausweichoption (negativ) oder Anzahl ausgewählter Optionen Ganze Zahl</p> <p><b>DA11_01</b> Logdaten von IT-Systemen <b>DA11_02</b> Sensor-, RFID- oder andere Maschinendaten <b>DA11_03</b> Clickstream-Daten <b>DA11_04</b> Videoclips/Bilder <b>DA11_05</b> Daten aus Transaktionssystemen <b>DA11_06</b> Social-Media-Daten <b>DA11_07</b> Dokumente/Texte (z.B. Emails) <b>DA11_08</b> Sonstige <b>DA11_09</b> Keine</p> <p>1 = nicht gewählt 2 = ausgewählt</p>
<p><b>[DA12]</b> Auswahl Auto Bericht "Wie hoch würden Sie den Anteil der Automatisierung der Berichterstellung (Output) durch Managementinformati..."</p> <p><b>DA12</b> Auto Bericht</p> <p>1 = 0% 2 = 1-20% 3 = 21-40% 4 = 41-60% 5 = 61-80% 6 = 81-100% -9 = nicht beantwortet</p>

**Rubrik EM: Email**

<b>[EM01]</b> Offene Texteingabe Danksagung und Email
<b>EM01_01</b> Email-Adresse Offene Texteingabe

**Rubrik FK: Kennzahlen**

<b>[FK01]</b> Skala (Zwischenwerte beschriftet) ROI
<b>FK01_01</b> ROI 2015 <b>FK01_02</b> ROI 2016 1 = Unter -1% 2 = -1 bis 0,99% 3 = 1 bis 2,99% 4 = 3 bis 4,99% 5 = 5 bis 7,49% 6 = 7,5 bis 10% 7 = Über 10% -9 = nicht beantwortet

<b>[FK03]</b> Skala (Zwischenwerte beschriftet) Kapitalvolumen
<b>FK03_01</b> Kapitalvolumen 2015 in Millionen € <b>FK03_04</b> Kapitalvolumen 2016 in Millionen € 1 = Unter 30 2 = 30-49 3 = 50-74 4 = 75-99 5 = 100-149 6 = 150-249 7 = 250-499 8 = 500-1.499 9 = 1.500-4.000 10 = Über 4.000 -9 = nicht beantwortet

<b>[FK04]</b> Skala (Zwischenwerte beschriftet) Angestellte
<b>FK04_01</b> Angestellte 2015 <b>FK04_02</b> Angestellte 2016 1 = Unter 250 2 = 250-499 3 = 500-999 4 = 1.000-2.499 5 = 2.500-4.999 6 = 5.000-9.999 7 = 10.000-19.999 8 = 20.000-49.999 9 = 50.000-99.999 10 = Über 100.000 -9 = nicht beantwortet

<b>[FK05]</b> Skala (Zwischenwerte beschriftet) Umsatz
<b>FK05_01</b> Umsatz 2015 in Millionen € <b>FK05_02</b> Umsatz 2016 in Millionen € 1 = Unter 50 2 = 50-99 3 = 100-149 4 = 150-249 5 = 250-499 6 = 500-749 7 = 750-999 8 = 1.000-1.499 9 = 1.500-3.000 10 = Über 3.000 -9 = nicht beantwortet

<b>[FK06]</b> Skala (Zwischenwerte beschriftet) Gewinn
---

**FK06\_01** Gewinn nach Steuer 2015 in Millionen €

**FK06\_02** Gewinn nach Steuer 2016 in Millionen €

- 1 = Weniger als -5
- 2 = -5 bis -1
- 3 = -0,99 bis 0,99
- 4 = 1 bis 2,49
- 5 = 2,5 bis 4,99
- 6 = 5 bis 9,99
- 7 = 10 bis 24,99
- 8 = 25 bis 49,99
- 9 = 50 bis 99,99
- 10 = Mehr als 100
- 9 = nicht beantwortet

## Anhang 3: Begleitschreiben Fragebogenaussendung

### Studie zur Big Data Nutzung in Großunternehmen

Sehr geehrte Damen und Herren,

mein Name ist Matthias Wagenstaller und ich studiere zurzeit an der FhWien der WKW im Masterstudium Unternehmensführung. In meiner Masterarbeit untersuche ich den Zusammenhang zwischen **Big Data Analysen** und dem **Finanzerfolg** von Großunternehmen. Hierfür würde ich Sie bitten, den verlinkten Fragebogen von **einer** qualifizierten Person in Ihrem Unternehmen ausfüllen zu lassen. Optimal wäre eine Person, die im Bereich der **strategischen Planung** oder in der **Datenanalyse tätig ist**.

→ [Zum Fragebogen](#)

Wenn Sie wollen können Sie direkt mit der Beantwortung des Fragebogens beginnen, ansonsten finden Sie im Anschluss einen groben Überblick über Inhalt und Hintergrund der Arbeit:

#### **Herausforderung Digitalisierung:**

Die Digitalisierung und die damit einhergehende Vernetzung aller Lebensbereiche stellt den wohl größten Umbruch für unsere Gesellschaft seit der Einführung von Computern dar und bedeutet für Unternehmen sowohl eine Herausforderung als auch Chance. Dabei bestehen die Chancen nicht nur in der Automatisierung von Fertigungsprozessen, auch die enorme Menge an Daten, die durch diese Vernetzung gewonnen wird, birgt ein gewaltiges Potential. Die Auswertung dieser Daten, üblicherweise mit dem Begriff Big Data umschrieben, ermöglicht eine deutliche Verfeinerung von Prognosemodellen und kann somit zur Steigerung der Unternehmensrentabilität beitragen. Ein Beispiel hierfür wären zuverlässigere Absatzprognosen.

#### **Die Arbeit:**

##### **Ziele:**

In meiner Masterarbeit möchte ich die Auswirkungen der Analyse solcher Daten durch Unternehmen auf deren Finanzerfolg untersuchen. Die Erkenntnisse, die dadurch gewonnen werden, sollen zum einen zu einem besseren Verständnis der Zusammenhänge und Einbettung technischer Möglichkeiten der Datenanalyse im Unternehmensumfeld dienen. Zum anderen soll Entscheidungsträgern damit ermöglicht werden, die Wirtschaftlichkeit solcher Maßnahmen besser zu beurteilen.

##### **Vorgehen:**

Im Sinne eines quantitativen Forschungskonzeptes werden sämtliche Unternehmen mittels Fragebogen interviewt, die folgende Kriterien erfüllen:

- Bilanzsumme größer als 20€ Millionen
- Umsatz größer als 40€ Millionen
- Tätigkeit in der DACH-Region

##### **Ihr Nutzen:**

Sollten Sie sich bereit erklären den in dieser Email verlinkten Fragebogen auszufüllen, lasse ich Ihnen gerne sämtliche Erkenntnisse, die daraus gewonnen wurden in Form der abgeschlossenen Arbeit zukommen. Außerdem möchte ich Sie an dieser Stelle darauf hinweisen, dass sämtliche Daten die Sie mir zukommen lassen, streng vertraulich behandelt werden und lediglich in aggregierter oder anonymisiert Form Dritten zugänglich gemacht werden.

Abschließend soll auch erwähnt werden, dass die Studie durch ein Netidee-Stipendium der Internet Foundation Austria gefördert wird und im Zuge dessen von unabhängigen Experten begutachtet wurde, die die hohe Qualität bestätigten.

Ich hoffe, ich konnte Sie vom Nutzen meiner Arbeit für die Wissenschaft im Allgemeinen und Ihr Unternehmen im Speziellen überzeugen und Sie nehmen sich ca. 15min Zeit um den Fragebogen unter folgendem Link zu beantworten:

→ [Zum Fragebogen](#)

---

Diese E-Mail wird mit SoSci Survey versendet ([Impressum](#)). Zu Datenschutz und dem Inhalt dieser E-Mail steht Ihnen [Matthias Wagenstaller](#) gerne Rede und Antwort.

Falls Sie keine weiteren E-Mails über SoSci Survey erhalten möchten, klicken Sie bitte [hier](#).

## Anhang 4

Source	SS	df	MS	
Model	273290.897	17	16075.9351	Number of obs = 102
Residual	1360149.26	84	16192.253	F( 17, 84) = 0.99
Total	1633440.15	101	16172.6748	Prob > F = 0.4743
				R-squared = 0.1673
				Adj R-squared = -0.0012
				Root MSE = 127.25

ROEb_15	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
QAN_Dat	-7.419435	25.79822	-0.29	0.774	-58.72203 43.88316
QAN_Krise	-6.611468	17.66797	-0.37	0.709	-41.74617 28.52324
DA02_02	-.2886931	.5589439	-0.52	0.607	-1.400214 .8228281
DA02_05	-.0036621	.013499	-0.27	0.787	-.0305063 .0231822
ZugriffIn	-3.353133	5.140369	-0.65	0.516	-13.57532 6.869054
DA12	-6.637483	11.67305	-0.57	0.571	-29.85063 16.57566
DA08	-5.780244	14.77343	-0.39	0.697	-35.15882 23.59833
DA10x01	37.36156	32.14903	1.16	0.248	-26.57032 101.2934
DA10x02	50.44835	36.1353	1.40	0.166	-21.41065 122.3073
DA10x03	31.09897	96.6244	0.32	0.748	-161.0493 223.2472
DA10x04	-6.64211	53.39863	-0.12	0.901	-112.8311 99.54691
DA10x05	16.32255	82.18955	0.20	0.843	-147.1204 179.7655
DA10x06	-23.8417	29.95534	-0.80	0.428	-83.41117 35.72777
DA10x07	116.2821	45.97541	2.53	0.013	24.85492 207.7092
DA10x08	-29.70431	32.35634	-0.92	0.361	-94.04844 34.63983
DA10x09	-11.10886	27.12745	-0.41	0.683	-65.05476 42.83704
DA10x10	-35.72413	30.56366	-1.17	0.246	-96.50331 25.05505
_cons	46.38308	114.5047	0.41	0.686	-181.322 274.0881

## Anhang 5

Source	SS	df	MS	Number of obs =	89
Model	23794.669	18	1321.92605	F( 18, 70) =	3.40
Residual	27235.4872	70	389.078389	Prob > F =	0.0001
Total	51030.1562	88	579.888139	R-squared =	0.4663
				Adj R-squared =	0.3290
				Root MSE =	19.725

ROCEb_15	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ROEb_15	.1142398	.0171259	6.67	0.000	.0800832	.1483963
QAN_Dat	8.878164	4.222467	2.10	0.039	.4567197	17.29961
QAN_Krise	-3.467952	2.945546	-1.18	0.243	-9.342657	2.406754
DA02_02	-.0924074	.0889831	-1.04	0.303	-.2698785	.0850637
DA02_05	.0002422	.0021301	0.11	0.910	-.0040061	.0044904
ZugriffIn	.1726043	.85272	0.20	0.840	-1.528092	1.873301
DA12	-.8611544	2.042765	-0.42	0.675	-4.935321	3.213012
DA08	.4039106	2.353367	0.17	0.864	-4.289732	5.097554
DA10x01	6.553046	5.585593	1.17	0.245	-4.587068	17.69316
DA10x02	-5.474015	6.332106	-0.86	0.390	-18.103	7.154971
DA10x03	16.29308	15.23826	1.07	0.289	-14.09867	46.68483
DA10x04	3.811354	10.72951	0.36	0.723	-17.58798	25.21068
DA10x05	9.93123	13.07325	0.76	0.450	-16.14254	36.005
DA10x06	6.122966	5.084587	1.20	0.233	-4.017922	16.26386
DA10x07	-19.18518	8.607025	-2.23	0.029	-36.35135	-2.01901
DA10x08	-1.928858	5.479034	-0.35	0.726	-12.85645	8.99873
DA10x09	-5.185951	4.67675	-1.11	0.271	-14.51343	4.141532
DA10x10	5.669287	5.112574	1.11	0.271	-4.52742	15.86599
_cons	-16.40129	18.85379	-0.87	0.387	-54.004	21.20141

## Anhang 6

Source	SS	df	MS	
Model	4.3457e+11	17	2.5563e+10	Number of obs = 104
Residual	1.5873e+12	86	1.8456e+10	F( 17, 86) = 1.39
				Prob > F = 0.1642
				R-squared = 0.2149
				Adj R-squared = 0.0598
Total	2.0218e+12	103	1.9629e+10	Root MSE = 1.4e+05

Profit_15	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
QAN_Dat	20995.19	27395.15	0.77	0.446	-33464.57	75454.94
QAN_Krise	-48051.61	18632.34	-2.58	0.012	-85091.47	-11011.75
DAO2_02	-8.77653	592.8951	-0.01	0.988	-1187.413	1169.86
DAO2_05	-9.201984	14.42977	-0.64	0.525	-37.88742	19.48345
ZugriffIn	5206.576	5373.277	0.97	0.335	-5475.144	15888.3
DA12	-6984.627	12369.82	-0.56	0.574	-31575.02	17605.77
DA08	696.2474	15560.78	0.04	0.964	-30237.57	31630.06
DA10x01	9165.719	33883.05	0.27	0.787	-58191.55	76522.99
DA10x02	22187.93	39556.05	0.56	0.576	-56446.89	100822.7
DA10x03	264837.8	84647.44	3.13	0.002	96564.27	433111.4
DA10x04	11116.02	55505.15	0.20	0.842	-99224.57	121456.6
DA10x05	8521.482	77130.89	0.11	0.912	-144809.7	161852.6
DA10x06	23647.27	31252.87	0.76	0.451	-38481.38	85775.91
DA10x07	-55212.63	47621.85	-1.16	0.250	-149881.7	39456.48
DA10x08	-17788.13	32990.42	-0.54	0.591	-83370.91	47794.65
DA10x09	-27393.91	28304.7	-0.97	0.336	-83661.78	28873.97
DA10x10	56293.17	32847.23	1.71	0.090	-9004.959	121591.3
_cons	77377.33	127087	0.61	0.544	-175263.2	330017.8

## Anhang 7

```

Iteration 0:  log likelihood = -120.59611
Iteration 1:  log likelihood = -117.30647
Iteration 2:  log likelihood = -117.15793
Iteration 3:  log likelihood = -117.15791
Iteration 4:  log likelihood = -117.15791

```

```

Logistic regression              Number of obs   =       174
                                LR chi2(1)       =         6.88
                                Prob > chi2        =       0.0087
Log likelihood = -117.15791     Pseudo R2      =       0.0285

```

Datawarehouse	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Asset_15	1.36e-07	8.77e-08	1.55	0.121	-3.58e-08	3.08e-07
_cons	-.1508174	.1629868	-0.93	0.355	-.4702656	.1686309

Adjusted predictions  
Model VCE : OIM

Number of obs = 174

Expression : Pr(Datawarehouse), predict()  
1.\_at : Asset\_15 = 0  
2.\_at : Asset\_15 = 1000000  
3.\_at : Asset\_15 = 2000000  
4.\_at : Asset\_15 = 3000000  
5.\_at : Asset\_15 = 4000000  
6.\_at : Asset\_15 = 5000000  
7.\_at : Asset\_15 = 6000000  
8.\_at : Asset\_15 = 7000000  
9.\_at : Asset\_15 = 8000000  
10.\_at : Asset\_15 = 9000000  
11.\_at : Asset\_15 = 1.00e+07  
12.\_at : Asset\_15 = 1.10e+07  
13.\_at : Asset\_15 = 1.20e+07  
14.\_at : Asset\_15 = 1.30e+07  
15.\_at : Asset\_15 = 1.40e+07  
16.\_at : Asset\_15 = 1.50e+07  
17.\_at : Asset\_15 = 1.60e+07  
18.\_at : Asset\_15 = 1.70e+07  
19.\_at : Asset\_15 = 1.80e+07  
20.\_at : Asset\_15 = 1.90e+07  
21.\_at : Asset\_15 = 2.00e+07

21._at	: Asset_15	=	2.00e+07
22._at	: Asset_15	=	2.10e+07
23._at	: Asset_15	=	2.20e+07
24._at	: Asset_15	=	2.30e+07
25._at	: Asset_15	=	2.40e+07
26._at	: Asset_15	=	2.50e+07
27._at	: Asset_15	=	2.60e+07
28._at	: Asset_15	=	2.70e+07
29._at	: Asset_15	=	2.80e+07
30._at	: Asset_15	=	2.90e+07
31._at	: Asset_15	=	3.00e+07
32._at	: Asset_15	=	3.10e+07
33._at	: Asset_15	=	3.20e+07
34._at	: Asset_15	=	3.30e+07
35._at	: Asset_15	=	3.40e+07
36._at	: Asset_15	=	3.50e+07
37._at	: Asset_15	=	3.60e+07
38._at	: Asset_15	=	3.70e+07
39._at	: Asset_15	=	3.80e+07
40._at	: Asset_15	=	3.90e+07
41._at	: Asset_15	=	4.00e+07
42._at	: Asset_15	=	4.10e+07
43._at	: Asset_15	=	4.20e+07

44.\_at : Asset\_15 = 4.30e+07  
 45.\_at : Asset\_15 = 4.40e+07  
 46.\_at : Asset\_15 = 4.50e+07  
 47.\_at : Asset\_15 = 4.60e+07  
 48.\_at : Asset\_15 = 4.70e+07  
 49.\_at : Asset\_15 = 4.80e+07  
 50.\_at : Asset\_15 = 4.90e+07  
 51.\_at : Asset\_15 = 5.00e+07  
 52.\_at : Asset\_15 = 5.10e+07  
 53.\_at : Asset\_15 = 5.20e+07  
 54.\_at : Asset\_15 = 5.30e+07  
 55.\_at : Asset\_15 = 5.40e+07  
 56.\_at : Asset\_15 = 5.50e+07  
 57.\_at : Asset\_15 = 5.60e+07  
 58.\_at : Asset\_15 = 5.70e+07  
 59.\_at : Asset\_15 = 5.80e+07  
 60.\_at : Asset\_15 = 5.90e+07  
 61.\_at : Asset\_15 = 6.00e+07

	Delta-method				
	Margin	Std. Err.	z	P> z	[95% Conf. Interval]
_at					
1	.462367	.0405159	11.41	0.000	.3829573 .5417766
2	.4962875	.0394982	12.56	0.000	.4188724 .5737025
3	.5302422	.0490209	10.82	0.000	.4341629 .6263214
4	.5639192	.0640635	8.80	0.000	.4383569 .6894814
5	.5970168	.0806484	7.40	0.000	.4389489 .7550846

7	.6603791	.1115715	5.92	0.000	.4417031	.8790551
8	.6901779	.1243319	5.55	0.000	.4464918	.9338639
9	.7184767	.1347856	5.33	0.000	.4543018	.9826515
10	.7451451	.1427933	5.22	0.000	.4652754	1.025015
11	.7700957	.1483509	5.19	0.000	.4793333	1.060858
12	.7932812	.1515582	5.23	0.000	.4962327	1.09033
13	.8146912	.1525921	5.34	0.000	.5156162	1.113766
14	.8343468	.1516825	5.50	0.000	.5370544	1.131639
15	.8522955	.1490907	5.72	0.000	.5600831	1.144508
16	.8686057	.1450905	5.99	0.000	.5842336	1.152978
17	.8833613	.139954	6.31	0.000	.6090565	1.157666
18	.896657	.1339402	6.69	0.000	.6341391	1.159175
19	.908594	.127287	7.14	0.000	.6591161	1.158072
20	.9192763	.1202066	7.65	0.000	.6836756	1.154877
21	.9288079	.1128833	8.23	0.000	.7075607	1.150055
22	.9372909	.1054728	8.89	0.000	.730568	1.144014
23	.9448232	.0981035	9.63	0.000	.7525438	1.137103
24	.9514975	.0908782	10.47	0.000	.7733794	1.129616
25	.9574009	.0838769	11.41	0.000	.7930051	1.121797
26	.962614	.0771594	12.48	0.000	.8113844	1.113844
27	.967211	.070768	13.67	0.000	.8285081	1.105914
28	.9712596	.0647309	15.00	0.000	.8443894	1.09813
29	.9748213	.0590638	16.50	0.000	.8590583	1.090584
30	.9779517	.0537731	18.19	0.000	.8725583	1.083345
31	.9807005	.0488573	20.07	0.000	.8849421	1.076459
32	.9831126	.0443089	22.19	0.000	.8962687	1.069957
33	.9852278	.0401163	24.56	0.000	.9066013	1.063854
34	.9870815	.0362643	27.22	0.000	.9160047	1.058158
35	.9887052	.0327358	30.20	0.000	.9245442	1.052866
36	.9901269	.0295121	33.55	0.000	.9322842	1.04797
37	.9913713	.026574	37.31	0.000	.9392872	1.043455
38	.99246	.0239018	41.52	0.000	.9456132	1.039307
39	.9934122	.0214763	46.26	0.000	.9513194	1.035505
40	.9942449	.0192785	51.57	0.000	.9564597	1.03203
41	.9949728	.0172903	57.55	0.000	.9610844	1.028861
42	.9956091	.0154944	64.26	0.000	.9652407	1.025978
43	.9961652	.0138742	71.80	0.000	.9689722	1.023358
44	.9966511	.0124146	80.28	0.000	.972319	1.020983
45	.9970756	.0111009	89.82	0.000	.9753181	1.018833
46	.9974464	.00992	100.55	0.000	.9780036	1.016889
47	.9977703	.0088594	112.62	0.000	.9804063	1.015134
48	.9980532	.0079077	126.21	0.000	.9825545	1.013552
49	.9983003	.0070544	141.51	0.000	.9844739	1.012127
50	.998516	.0062901	158.74	0.000	.9861877	1.010844
51	.9987045	.0056058	178.15	0.000	.9877172	1.009692
52	.998869	.0049938	200.02	0.000	.9890814	1.008657

54	.999138	.0039577	252.45	0.000	.9913811	1.006895
55	.9992475	.0035212	283.78	0.000	.9923461	1.006149
56	.9993431	.0031317	319.11	0.000	.9932052	1.005481
57	.9994266	.0027842	358.96	0.000	.9939697	1.004884
58	.9994994	.0024744	403.93	0.000	.9946496	1.004349
59	.9995631	.0021984	454.68	0.000	.9952543	1.003872
60	.9996186	.0019525	511.96	0.000	.9957917	1.003446
61	.9996671	.0017336	576.63	0.000	.9962692	1.003065

---



Expression : Pr(Datawarehouse), predict()

```

1._at      : Asset_kat      =      0
2._at      : Asset_kat      =      1
3._at      : Asset_kat      =      2
4._at      : Asset_kat      =      3
5._at      : Asset_kat      =      4
6._at      : Asset_kat      =      5
7._at      : Asset_kat      =      6
8._at      : Asset_kat      =      7
9._at      : Asset_kat      =      8
10._at     : Asset_kat      =      9
11._at     : Asset_kat      =     10

```

	Delta-method				[95% Conf. Interval]	
	Margin	Std. Err.	z	P> z		
_at						
1	.2888324	.0673217	4.29	0.000	.1568842	.4207805
2	.326983	.0607394	5.38	0.000	.2079359	.44603
3	.3675683	.0531614	6.91	0.000	.2633739	.4717626
4	.4101213	.0457572	8.96	0.000	.3204389	.4998037
5	.4540638	.0404345	11.23	0.000	.3748136	.5333139
6	.4987337	.0393662	12.67	0.000	.4215773	.5758901
7	.5434238	.0432185	12.57	0.000	.4587171	.6281306
8	.5874257	.0503986	11.66	0.000	.4886463	.6862051
9	.6300736	.0587361	10.73	0.000	.5149529	.7451943
10	.6707837	.0666607	10.06	0.000	.5401312	.8014362
11	.7090823	.0732506	9.68	0.000	.5655137	.8526509





```

Iteration 0:  log likelihood = -120.59611
Iteration 1:  log likelihood = -117.30388
Iteration 2:  log likelihood = -117.29946
Iteration 3:  log likelihood = -117.29946

```

```

Logistic regression                               Number of obs =      174
                                                  LR chi2(1)      =       6.59
                                                  Prob > chi2     =     0.0102
Log likelihood = -117.29946                    Pseudo R2      =     0.0273

```

Datawarehouse	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Orev_kat	.1659568	.0665094	2.50	0.013	.0356007	.2963129
_cons	-.6817456	.3029049	-2.25	0.024	-1.275428	-.0880629

```
. margins, at (Orev_kat=(0 (1) 10))
```

```

Adjusted predictions                               Number of obs =      174
Model VCE      : OIM

```

```
Expression    : Pr(Datawarehouse), predict()
```

```

1._at      : Orev_kat      =      0
2._at      : Orev_kat      =      1
3._at      : Orev_kat      =      2
4._at      : Orev_kat      =      3
5._at      : Orev_kat      =      4
6._at      : Orev_kat      =      5
7._at      : Orev_kat      =      6
8._at      : Orev_kat      =      7

```

10.\_at : Orev\_kat = 9  
 11.\_at : Orev\_kat = 10

	Delta-method					[95% Conf. Interval]	
	Margin	Std. Err.	z	P> z			
_at							
1	.3358718	.0675666	4.97	0.000	.2034438	.4682998	
2	.3738375	.0580605	6.44	0.000	.2600409	.487634	
3	.4134232	.048584	8.51	0.000	.3182002	.5086461	
4	.4541603	.0411846	11.03	0.000	.37344	.5348806	
5	.4955205	.038655	12.82	0.000	.4197581	.571283	
6	.5369422	.0423912	12.67	0.000	.453857	.6200274	
7	.5778602	.050627	11.41	0.000	.478633	.6770874	
8	.6177366	.0606304	10.19	0.000	.4989033	.7365699	
9	.656088	.0704995	9.31	0.000	.5179115	.7942644	
10	.692507	.0791524	8.75	0.000	.537371	.8476429	
11	.726676	.0860182	8.45	0.000	.5580835	.8952685	

## Anhang 11

AF06	Datawarehouse		Total
	0	1	
Bankensektor	3	3	6
Baugewerbe	7	8	15
Beratung	1	1	2
Dienstleistungen (auß	9	10	19
Energie und Versorgun	11	9	20
Fertigungsindustrie (	32	11	43
Gesundheitswesen	6	8	14
Handel	10	20	30
IT	2	4	6
Landwirtschaft	1	1	2
Medien/Verlagswesen	1	4	5
Prozessindustrie (z.B	13	12	25
Telekommunikation	1	0	1
Transportwesen und Lo	5	3	8
Versicherungen	2	6	8
Öffentlicher Sektor	6	1	7
Öl, Gas und Bergbau	2	1	3
<b>Total</b>	<b>112</b>	<b>102</b>	<b>214</b>

Pearson chi2 (16) = 23.6900 Pr = 0.096

## Anhang 12

```

Logistic regression
Number of obs   =      213
LR chi2(15)     =      23.88
Prob > chi2     =      0.0672
Pseudo R2      =      0.0810
Log likelihood = -135.51023
    
```

Datawarehouse	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
AF06					
Baugewerbe	.1335314	.9667077	0.14	0.890	-1.761181 2.028244
Beratung	-2.16e-15	1.632993	-0.00	1.000	-3.200608 3.200608
Dienstleistungen (außer Beratung)	.1053605	.936898	0.11	0.910	-1.730926 1.941647
Energie und Versorgungswirtschaft	-.2006707	.9320337	-0.22	0.830	-2.027423 1.626082
Fertigungsindustrie (z.B. Automotive, Maschinenbau)	-1.067841	.8881586	-1.20	0.229	-2.8086 6.729183
Gesundheitswesen	.2876821	.978945	0.29	0.769	-1.631015 2.206379
Handel	.6931472	.9036961	0.77	0.443	-1.078065 2.464359
IT	.6931472	1.190238	0.58	0.560	-1.639677 3.025971
Landwirtschaft	-2.16e-15	1.632993	-0.00	1.000	-3.200608 3.200608
Medien/Verlagswesen	1.386294	1.384437	1.00	0.317	-1.327153 4.099742
Prozessindustrie (z.B. Chemie, Pharma, Lebensmittelherstellung)	-.0800427	.9093531	-0.09	0.930	-1.862342 1.702257
Telekommunikation	0	(empty)			
Transportwesen und Logistik	-.5108256	1.095445	-0.47	0.641	-2.657859 1.636207
Versicherungen	1.098612	1.154701	0.95	0.341	-1.164559 3.361784
Öffentlicher Sektor	-1.791759	1.354006	-1.32	0.186	-4.445563 .8620443
Öl, Gas und Bergbau	-.6931472	1.471196	-0.47	0.638	-3.578136 2.191842
_cons	1.83e-15	.8164966	0.00	1.000	-1.600304 1.600304



## Anhang 14

Source	SS	df	MS
Model	3.46178811	5	.692357621
Residual	102.610065	190	.540052975
Total	106.071853	195	.543958222

Number of obs = 196  
 F( 5, 190) = 1.28  
 Prob > F = 0.2733  
 R-squared = 0.0326  
 Adj R-squared = 0.0072  
 Root MSE = .73488

	QAN_Dat						
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]		
AF05							
Aktienunternehmen mit Mehrheitseigentümer	-.2763889	.1955857	-1.41	0.159	-.6621871	.1094093	
Famliienunternehmen	-.1736642	.1846852	-0.94	0.348	-.5379809	.1906125	
Inhabergeföhrt mit einem einzigen Inhaber	-.2208333	.3074236	-0.72	0.473	-.8272351	.3855684	
Inhabergeföhrt ohne familiären Zusammenhang	-.261508	.2296072	-1.14	0.256	-.7144146	.1913987	
Unternehmen in öffentlicher Hand	-.5278985	.2246847	-2.35	0.020	-.9710954	-.0847016	
_cons	3.158333	.1643248	19.22	0.000	2.834198	3.482469	

## Anhang 15

AF05	Datawarehouse		Total
	0	1	
Aktienunternehmen im	11	9	20
Aktienunternehmen mit	22	29	51
Inhabergeführt ohne f	10	14	24
Familienunternehmen	45	38	83
Inhabergeführt mit ei	6	2	8
Unternehmen in öffent	16	10	26
Total	110	102	212

Pearson chi2 (5) = 5.5084 Pr = 0.357

# Anhang 16

Source	SS	df	MS	Number of obs =	102
Model	977310.557	37	26413.7988	F ( 37, 64) =	1.57
Residual	1077983.76	64	16843.4962	Prob > F =	0.0566
				R-squared =	0.4755
				Adj R-squared =	0.1723
				Root MSE =	129.78
Total	2055294.31	101	20349.4487		

	ROEB_15	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	Zugriffin	-2.897662	5.039336	-0.58	0.567	-12.96489 7.169568
	DA02_02	.7292406	.1954382	3.73	0.000	.3388079 1.119673
	DA02_05	-.0100573	.0147025	-0.68	0.496	-.0394289 .0193143
	QAN_Dat	-18.29708	32.83619	-0.56	0.579	-83.89492 47.30075
	QAN_Krise	15.48588	21.89365	0.71	0.482	-28.25172 59.22347
	DA10x01	-36.79852	38.59387	-0.95	0.344	-113.8986 40.30161
	DA10x02	96.20207	41.38714	2.32	0.023	13.52175 178.8824
	DA10x03	-70.12758	110.8026	-0.63	0.529	-291.4812 151.226
	DA10x04	-34.80718	59.64718	-0.58	0.562	-153.9661 84.35175
	DA10x05	25.62953	89.85163	0.29	0.776	-153.8697 205.1288
	DA10x06	-44.97565	35.28185	-1.27	0.207	-115.4592 25.50795
	DA10x07	148.0067	54.08106	2.74	0.008	39.96741 256.0461
	DA10x08	-47.05874	39.42239	-1.19	0.237	-125.814 31.69655
	DA10x09	-1.220211	32.28108	-0.04	0.970	-65.70908 63.26866
	DA10x10	-48.97747	36.12802	-1.36	0.180	-121.1515 23.19655



# Anhang 17

Source	SS	df	MS
Model	9.4599e+11	37	2.5567e+10
Residual	1.0759e+12	66	1.6302e+10
Total	2.0219e+12	103	1.9630e+10

Number of obs = 104  
 F ( 37, 66) = 1.57  
 Prob > F = 0.0553  
 R-squared = 0.4679  
 Adj R-squared = 0.1695  
 Root MSE = 1.3e+05

	Profit_15	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Zugriffin		3070.277	4965.472	0.62	0.538	-6843.608 12984.16
DA02_02		-53.84284	193.4293	-0.28	0.782	-440.0369 332.3512
DA02_05		-10.4382	14.44527	-0.72	0.472	-39.27911 18.40271
QAN_Dat		20400.35	29533.54	0.69	0.492	-38565.27 79365.97
QAN_Krise		-38539.67	20080.74	-1.92	0.059	-78632.16 1552.817
DA10x01		-68115.41	38169.43	-1.78	0.079	-144323.1 8092.324
DA10x02		46693.03	40365.81	1.16	0.252	-33899.92 127286
DA10x03		255647.1	90492.02	2.83	0.006	74973.94 436320.3
DA10x04		19495.38	57824.83	0.34	0.737	-95955.63 134946.4
DA10x05		-11395.1	81893.98	-0.14	0.890	-174901.7 152111.5
DA10x06		72282.58	34717.92	2.08	0.041	2966.017 141599.1
DA10x07		-61952.7	51635.36	-1.20	0.235	-165046 41140.63
DA10x08		2417.673	37465.28	0.06	0.949	-72384.17 77219.51
DA10x09		-734.3202	31459.32	-0.02	0.981	-63544.87 62076.23
DA10x10		44857.53	36033.07	1.24	0.218	-27084.81 116799.9

