



Autor: Andreas Schau

## Leitfaden für nachvollziehbare Schritte

### 1. Kurze Darstellung des Problembereichs / Aufriss des Themas

#### 1.1 Inhaltlich

Kern der Untersuchung:

Data Science Methodiken auf Superstore Dataset anwenden und Einsichten gewinnen.

Grobe Ziele der Arbeit:

- Verluste Identifizieren: Darunter zum Beispiel Produktkategorien, einzelne Produkte, Kundensegmente oder Regionen herausstellen, die besonders schlecht abschneiden und dringender Intervention benötigen.
- Gewinne Identifizieren: Ebenfalls nach diesen Dimensionen untersuchen um Top-Performer zu finden, die vor zu experimentellen Änderungen geschützt werden sollten. Weitere nützliche Informationen herausarbeiten, wie Zukunftstrends, Bearbeitungszeiten oder regional unterschiedliches Kundenverhalten.

#### 1.2 Begründung desThemas

##### Darstellung der Relevanz des Themas?

Viele Märkte sind heute zu Tage sehr kompetitiv. Unternehmen benötigen daher jede Erkenntnis, die sie aus ihren bisherigen Abläufen und Erfolgen gewinnen können. Dies können Erkenntnisse über unprofitable Produktkategorien oder einzelne Produkte sein, über ungünstige Verkaufsgespräche die zu hohen Retouren führen oder über kontraproduktive Discount-Strategien.

Trends der Verkäufe und Profite können ebenfalls sehr vorteilhaft für die Zukunftsplanung sein. Je mehr Daten man zusätzlich zu seinen Kunden und deren Kaufentscheidungen und Bedürfnissen hat, desto sicherer kann man auch riskantere Zukunftsinvestitionen tätigen.

### 2. Nachvollziehbare Schritte

#### 2.1 Der Stand der Forschung / Auswertung der vorhandenen Literatur / Tutorials ...

Das Superstore Dataset ist unter aufstrebenden Daten-Analysten und Daten-Wissenschaftlern sehr beliebt, da es sich gut für Business-Intelligence Untersuchungen und Visualisierungen eignet.

Deswegen gibt es zu diesem Datensatz viele Untersuchungen und Projektarbeiten, die vieles schon auf die unterschiedlichsten Weisen beleuchtet haben, meist unter Verwendung von Tableau und einem visuell-explorativem Ansatz.

## 2.2 Fragestellung

Befindet sich unser Unternehmen auf Erfolgskurs? Fahren wir Verluste ein, die vermeidbar sind? Welche Kundensegmente, Produktkategorien und Regionen sind am Profitabelsten und bieten Expansionspotenzial?

## 2.3 Stand der Forschung

Typischerweise reicht es, wenn man sich auf die allgemeinen KPIs konzentriert und via Verkaufsvolumen, Profit und eventuell Verschuldung die Trends abliest, ob es einem Unternehmen finanziell gut geht und wie es sich weiterentwickeln mag.

Will man jedoch noch etwas mehr Sicherheit in zukünftige Investitionen finden, so kann eine genauere Analyse weiterer Business-Faktoren helfen. Darunter könnte die Dauer von Bearbeitungsprozessen fallen und wie kompetitiv man in diesem Bereich mit seinen Wettbewerbern aufgestellt ist oder wie feinfühlig und trainiert das Verkaufspersonal ist, sodass es genau die richtigen Waren für die Kunden herausucht und dadurch die Verkäufe maximiert ohne die Retouren in die Höhe zu treiben.

Es ist außerdem interessant, wie das Unternehmen sich auf die Änderungen im Markt anpassen kann. Kommen neue Produkte hinzu dauert es sicher eine Weile, bis man eine profitable Herangehensweise gefunden hat, mit guten Einblicken in die Daten können solche Optimierungs und Entscheidungsprozesse zielgerichteter durchgeführt werden.

## 2.4 Wissenslücke

Der Datensatz ist umfangreich, jedoch kann man mit ihm auch nicht alle Fragen, die aufkommen könnten beantworten. Es gibt folgend die Feststellung, dass durch einen 3D-Drucker einiges an Verlusten entstanden sind. Jedoch kann man nur spekulieren, ob das zum Beispiel dadurch entstanden ist, dass diese Geräte in der Early-Adopter Phase für einen sehr hohen Preis eingekauft wurden, für den man sie dann nicht an normale Konsumenten weiter verkaufen konnte, da diese noch abgewartet haben, bis günstigere Modelle verfügbar waren oder ähnliches. Es bleibt auch offen, ob die Verluste durch Tische einem ähnlichen Problem zuzuschreiben sind oder ob es durch andere Faktoren dazu gekommen ist.

## 2.5 Methode

Das Superstore Datasets ist sehr umfangreich und hat eine hohe Qualität. Deswegen müssen nur wenige Vorbereitungs-Schritte ausgeführt werden.

Darunter sind:

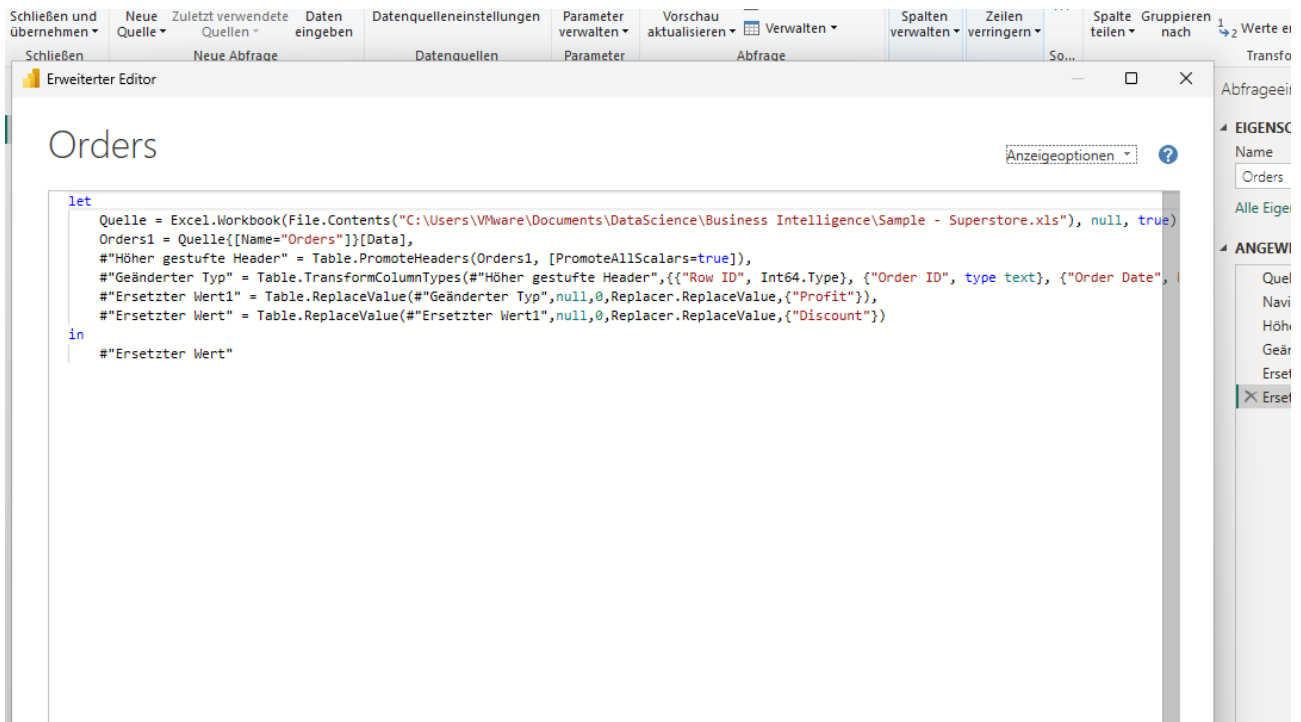
- Erste Reihen, die in Header umgewandelt werden müssen.
- Spalten-Datentypen die korrigiert werden müssen, da sie nicht automatisch richtig erkannt werden.
- In der Profit- und Discount-Spalte wurden "null" Werte jeweils durch den Zahlenwert "0" ersetzt, da

kein vorhandener Profit einem Profit von Null entspricht. Gleiches gilt für den Discount.

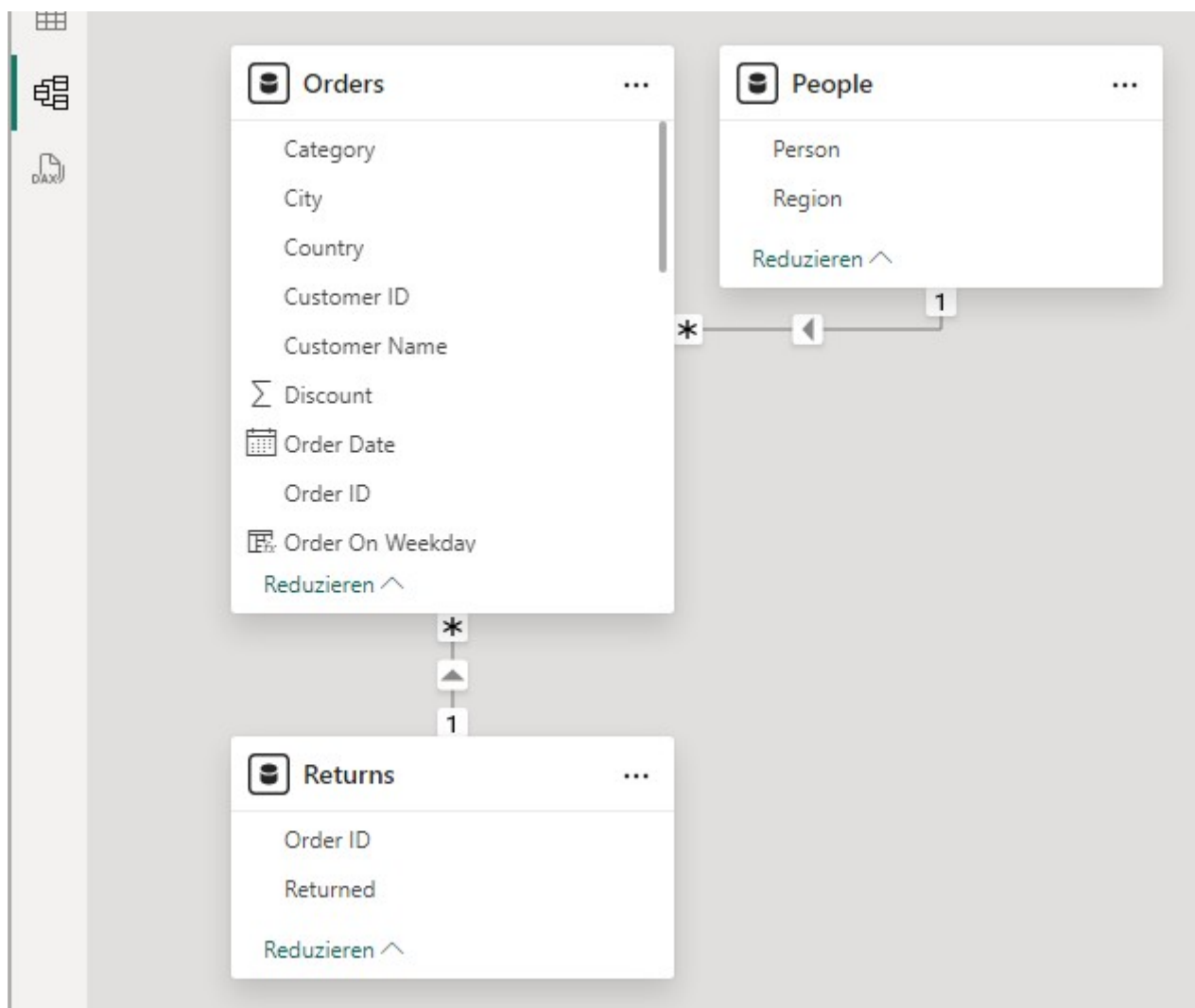
Diese Schritte wurden mit Power BI, nach dem Laden des Datensatzes aus der „Sample – Superstore.xls“ Excel-Datei, im Daten-Transformations-Fenster durchgeführt. Man kann sich alle angewendeten Schritte in der rechten Seitenleiste anschauen und über einen Klick auf „Erweiterter Editor“ bekommt man diese in der DAX Programmiersprache von Power BI zusammengefasst.

The screenshot shows the Microsoft Power BI Desktop interface. The top ribbon is set to 'Transform' (Transformieren), with options like 'Spalten verwalten' (Manage Columns), 'Zeilen verringern' (Reduce Rows), 'Spalte teilen' (Split Column), and 'Gruppieren nach' (Group By). The 'Abfrage' (Query) section shows a DAX formula: `Table,0,Replacer.ReplaceValue,{"Discount"})`. Below the formula, a table of data is displayed with columns 'Date', 'Ship Date', and 'Ship Mode'. The 'Queries' pane on the right shows the 'EIGENSCHAFTEN' (Properties) for the 'Orders' query and a list of 'ANGEWENDETE SCHRITTE' (Applied Steps), including 'Quelle', 'Navigation', 'Höher gestufte Header', 'Geänderter Typ', 'Ersetzter Wert1', and 'Ersetzter Wert'.

Date	Ship Date	Ship Mode
08.11.2016	11.11.2016	Second Class
08.11.2016	11.11.2016	Second Class
12.06.2016	16.06.2016	Second Class
11.10.2015	18.10.2015	Standard Class
11.10.2015	18.10.2015	Standard Class
09.06.2014	14.06.2014	Standard Class
09.06.2014	14.06.2014	Standard Class
09.06.2014	14.06.2014	Standard Class
09.06.2014	14.06.2014	Standard Class
09.06.2014	14.06.2014	Standard Class



Nach dem Laden der Daten musste überprüft werden, ob Power BI die Fremdschlüssel der jeweiligen Fakten und Dimensionstabellen richtig zuordnen konnte. Dafür geht man in die Modellansicht, welche wie folgt aussehen sollte und ein typisches Star-Schema mit Orders als Fakt-Tabelle und People und Returns als Dimensions-Tabellen hat:



Nun müssen noch ein paar berechnete Spalten hinzugefügt werden. Diese sind „Order on Weekday“ und „Order Processing Time“. „Order on Weekday“ transformiert via DAX Befehl das Bestelldatum von einem numerischen Datum zu einer Zahl, dem numerischen Tag der Woche und seinem Tagesnamen.

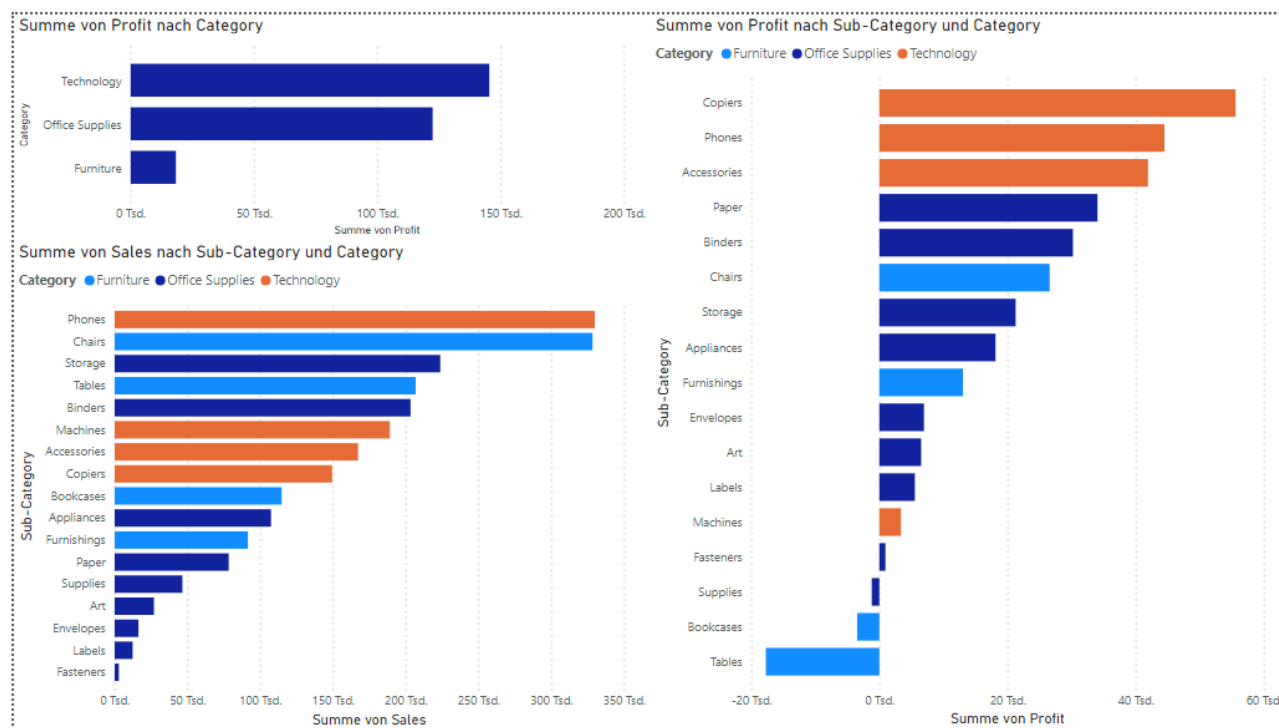
Struktur							
Formatierung							
Eigenschaften							
Sortieren							
Gruppen							
BE							
1 Order On Weekday = CONCATENATE(CONVERT(WEEKDAY(Orders[Order Date], 2), STRING), FORMAT(Orders[Order Date], " - dddd"))							
Product Name	Sales	Quantity	Discount	Profit	Order On Weekday	Order Proce	
Storage shelf, platinum	77,88	2	0	3,894	7 - Sonntag		
	6,63	3	0	1,7901	4 - Donnerstag		
	5,88	2	0	1,7052	4 - Donnerstag		
	5,46	3	0	1,5288	2 - Dienstag		
Motor-Operated Pencil Sharpener	73,2	5	0	21,228	2 - Dienstag		
Book W/Dividers/Space For Phone Numbers, 5 1/4"X8 1/2", 200/Messages	22,72	4	0	10,224	2 - Dienstag		
	45,36	7	0	21,7728	7 - Sonntag		
	11,34	3	0	5,2164	5 - Freitag		
Clear Snap-On Lid, Black Granite	80,3	5	0	20,878	5 - Freitag		
	64,74	13	0	30,4278	5 - Freitag		
ts	405,64	4	0	12,1692	5 - Freitag		

„Order Processing Time“ transformiert das Bestell- und Sendungsdatum zu einer Differenz in Tagen und repräsentiert damit die Bearbeitungszeit der Bestellung, von in Auftragsstellung, bis zum Versand.

1 Order Processing Time = DATEDIFF([Order Date].[Date], [Ship Date].[Date], DAY)							
Product Name	Sales	Quantity	Discount	Profit	Order On Weekday	Order Processing Time	
Storage shelf, platinum	77,88	2	0	3,894	7 - Sonntag	5	
	6,63	3	0	1,7901	4 - Donnerstag	4	
	5,88	2	0	1,7052	4 - Donnerstag	4	
	5,46	3	0	1,5288	2 - Dienstag	5	
Operated Pencil Sharpener	73,2	5	0	21,228	2 - Dienstag	5	
Book W/Dividers/Space For Phone Numbers, 5 1/4"x8 1/2", 200/Messages	22,72	4	0	10,224	2 - Dienstag	5	
	45,36	7	0	21,7728	7 - Sonntag	3	
	11,34	3	0	5,2164	5 - Freitag	4	
Clear Snap-On Lid, Black Granite	80,3	5	0	20,878	5 - Freitag	4	
	64,74	13	0	30,4278	5 - Freitag	4	
ts	405,64	4	0	12,1692	5 - Freitag	4	

Nachdem die Daten nun vorbereitet sind, kann mit der explorativen Datenanalyse via Visualisierungen begonnen werden.

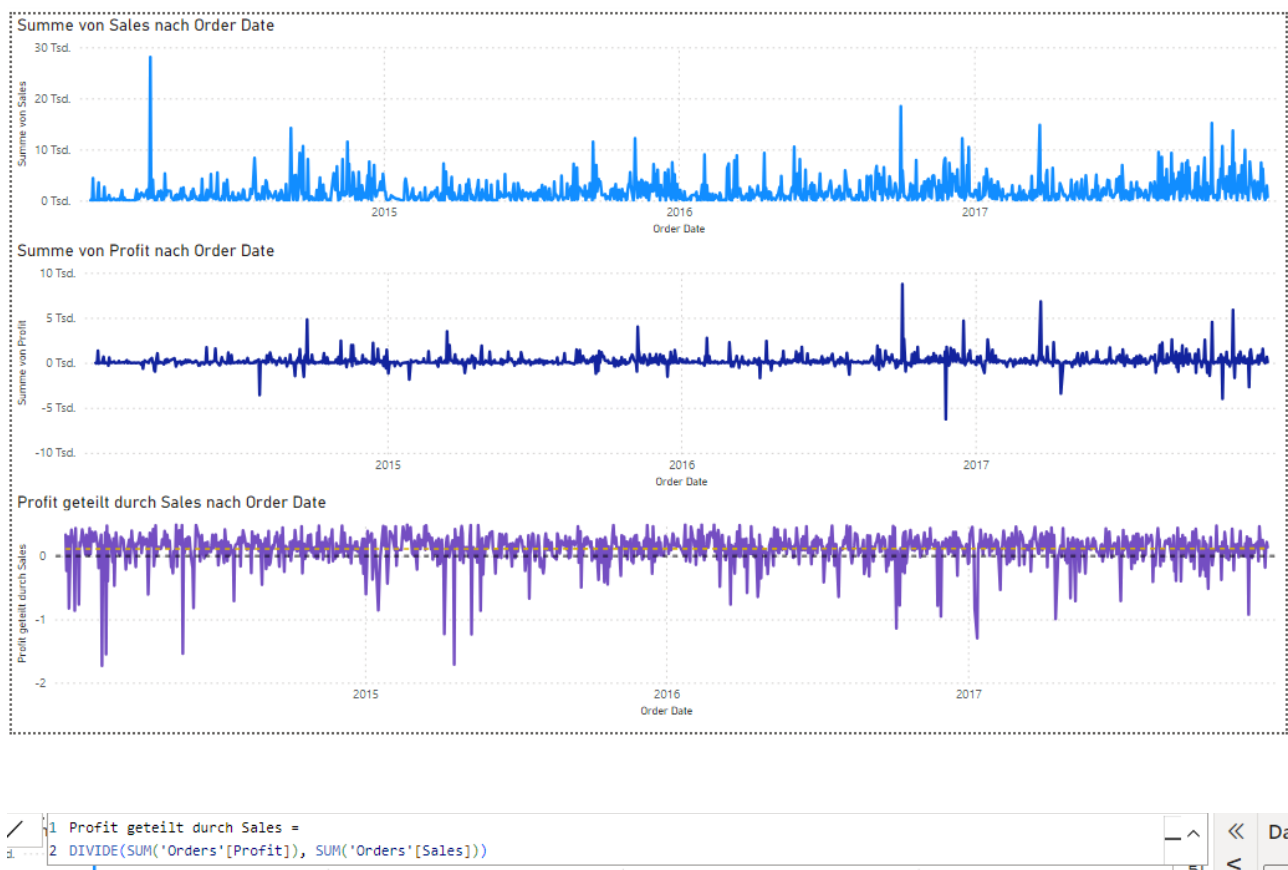
Zuerst wurden die groben Kategorien nach ihrem Profit untersucht, dann das Verkaufsvolumen nach Kategorie und Unterkategorie, sowie der Profit nach Kategorie und Unterkategorie:



Dabei viel sofort der große Verlust durch Tische und Bücherregale auf, der genauer untersucht werden

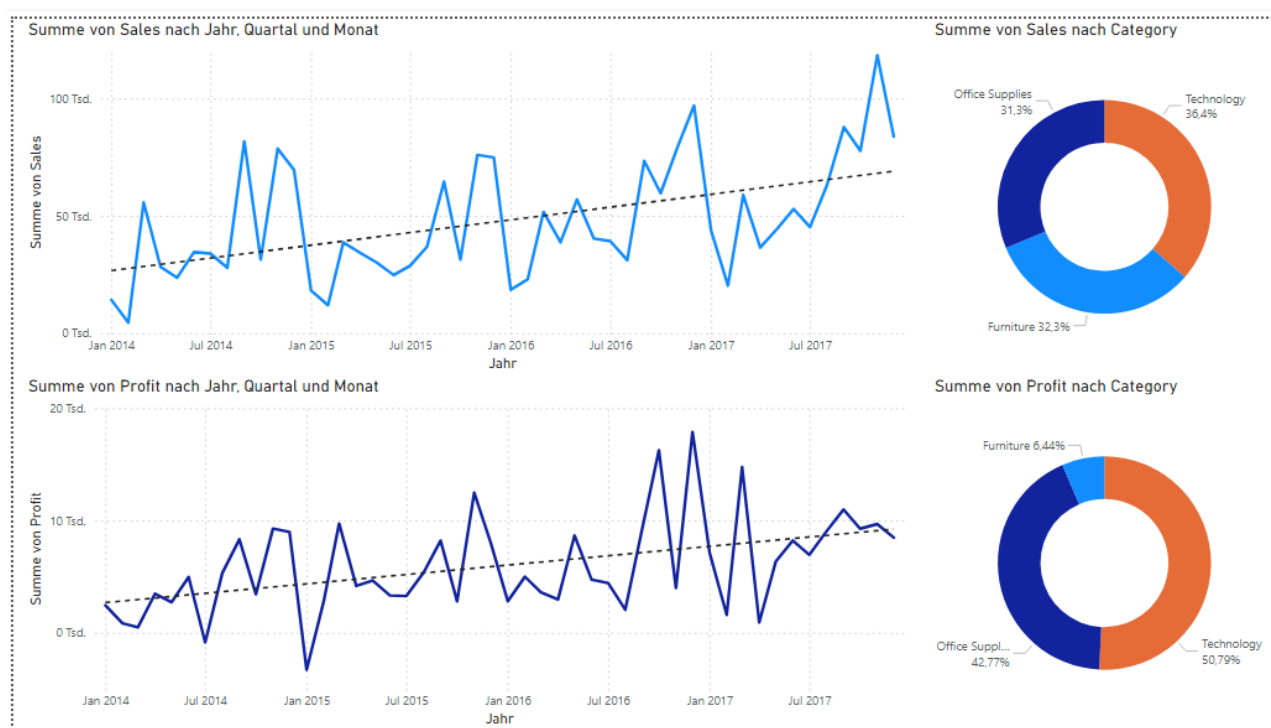
muß.

Zunächst wurden jedoch die Verkäufe noch etwas weiter unter generellen Aspekten untersucht, indem das Verkaufsvolumen und der Profit über die Zeit geplottet wurden. Außerdem wurde über einen „Quickmeasure“ ein berechnetes Feld erstellt, welches den Profit pro Sale repräsentiert.

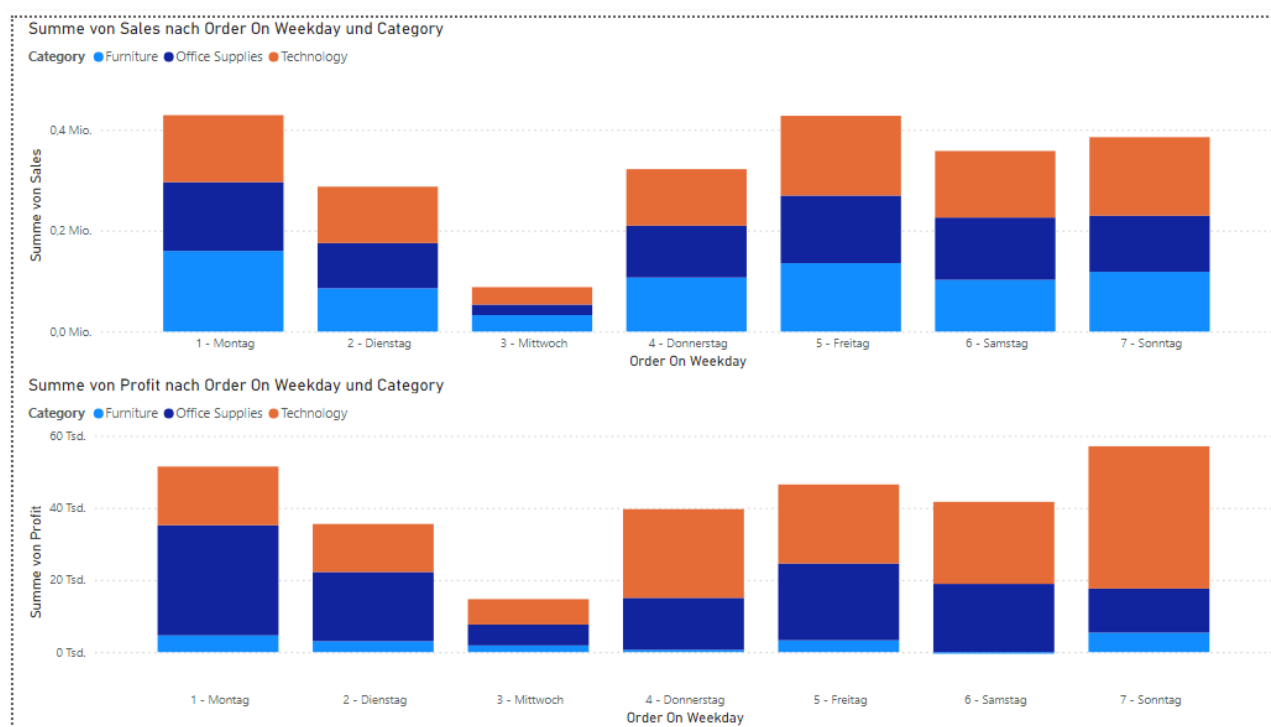


Hier sieht man, dass es viele Verkäufe mit hohem Volumen gibt, die Profite pro Verkauf sehen jedoch recht knapp aus. Das deutet auf ein Geschäftsfeld mit wenig Marge hin. Im untersten Graph, der den Profit pro Verkauf darstellt sieht man, dass der Mittelwert dennoch bei ca. 10-20 Prozent Gewinn liegt. Jedoch gibt es auch starke Ausschläge in den negativen, unprofitablen Bereich.

Untersucht man das Verkaufsvolumen und den Profit jedoch pro Monat über die letzten Jahre, so sieht man einen deutlichen positiven Trend. In den Ring-Diagrammen erkennt man außerdem, dass obwohl die drei Kategorien ähnlich viel Verkaufsvolumen haben, der Beitrag zum Gewinn sehr stark von der Technologie und der Office-Supplies Sparte eingefahren wird. Die Furniture Kategorie, zu welcher auch die Tische und Bücherregale gehören, trägt kaum zum Gewinn bei, wahrscheinlich, weil diese beiden Unterkategorien so viel Verlust ausmachen.



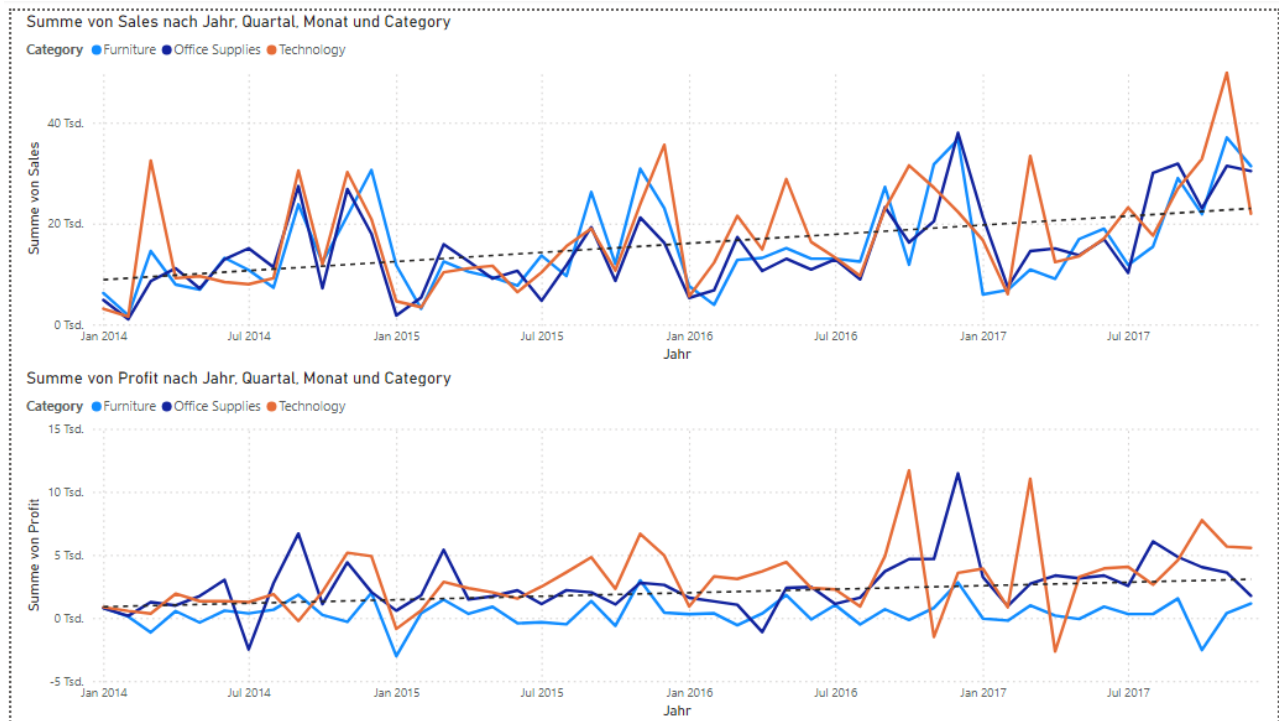
Als nächstes wurde die Bestellverteilung pro Wochentag nach Verkaufsvolumen und Profit untersucht. Hierbei viel auf, dass Mittwoch in allen Kundensegmenten deutlich weniger los ist, als die anderen Tage.



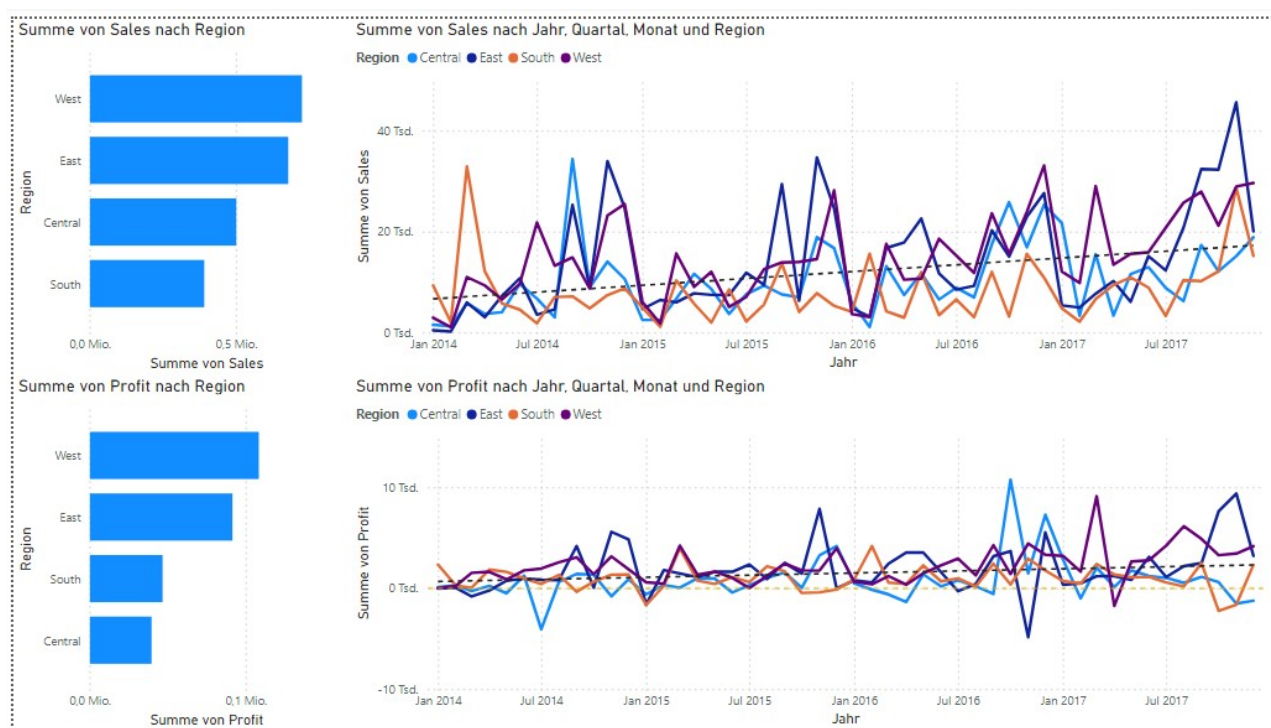
Darauf hin noch einmal eine Trendanalyse von Verkaufsvolumen und Profit und wie sich diese aus den



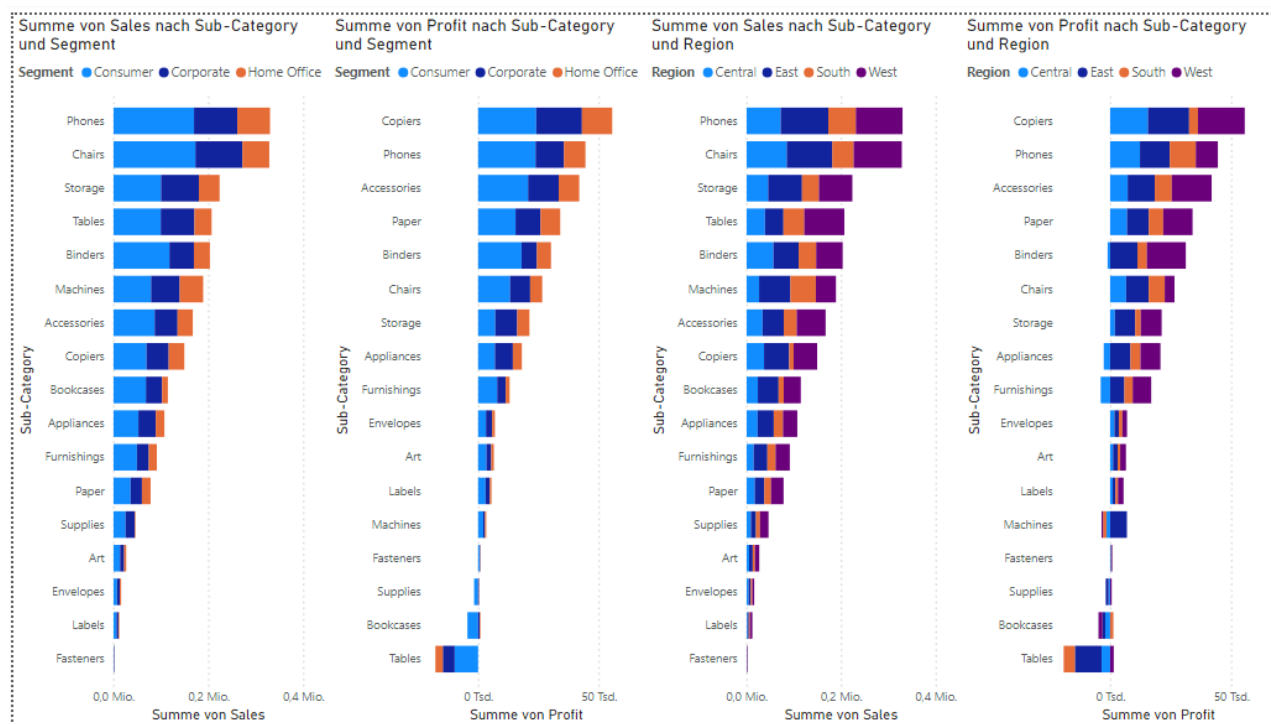
groben Kategorien zusammensetzen. Auch wenn der generelle Trend positiv ist, sieht man, dass die Möbelverkäufe einige Male ins negative ausreißen.



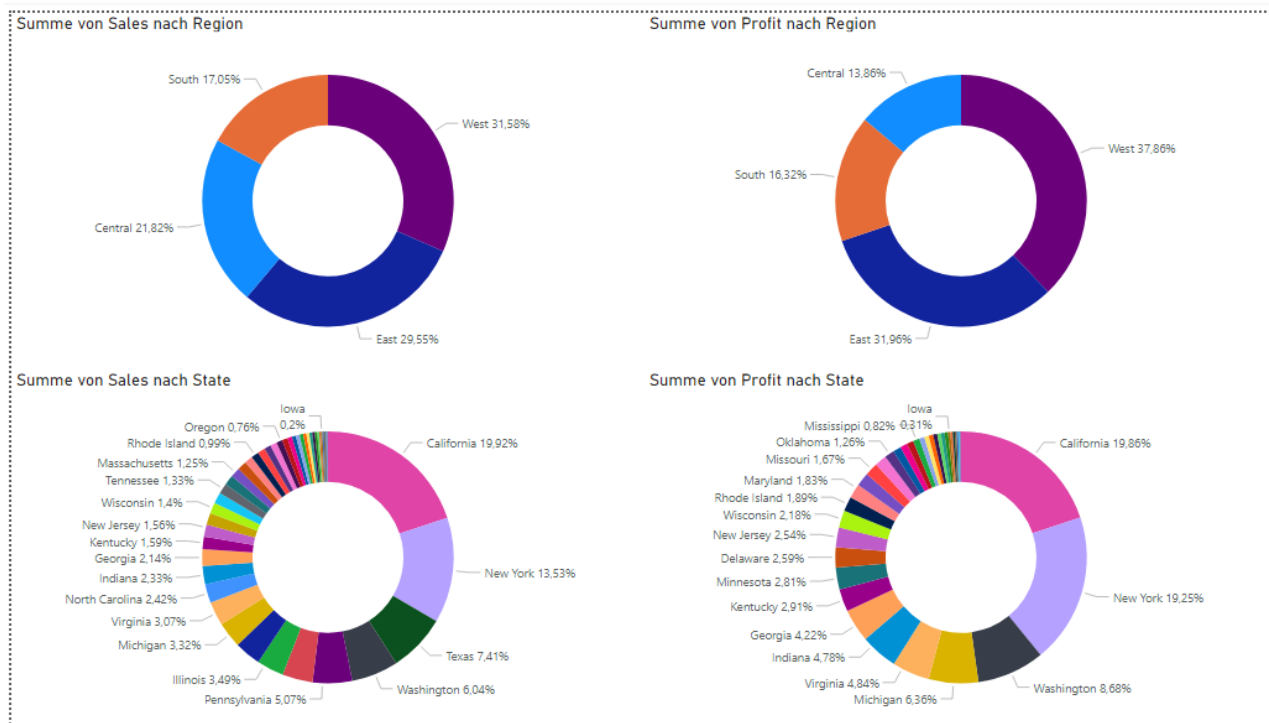
Diese Trendanalyse ist natürlich auch für die verschiedenen Regionen interessant. Dabei fällt auf, dass obwohl die Zentrale ein deutlich höheres Verkaufsvolumen hat, als die Südliche Region, sie in der Profitverteilung deutlich unter dieser liegt.



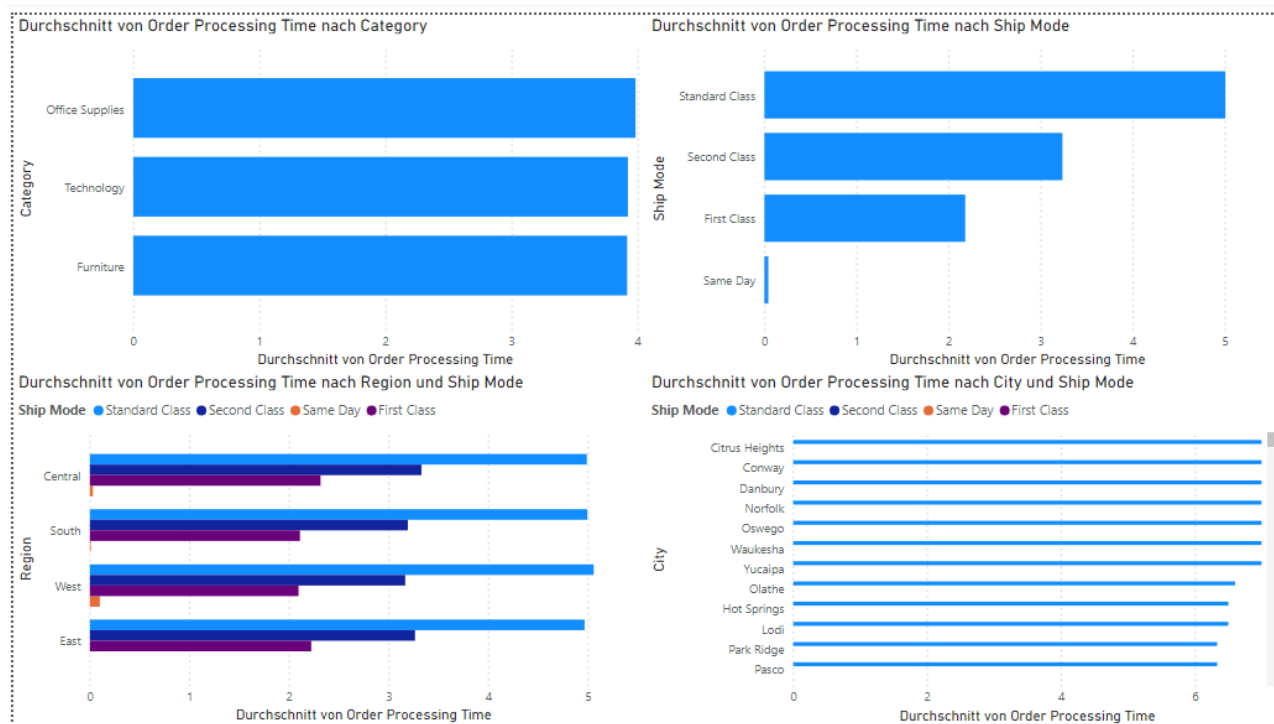
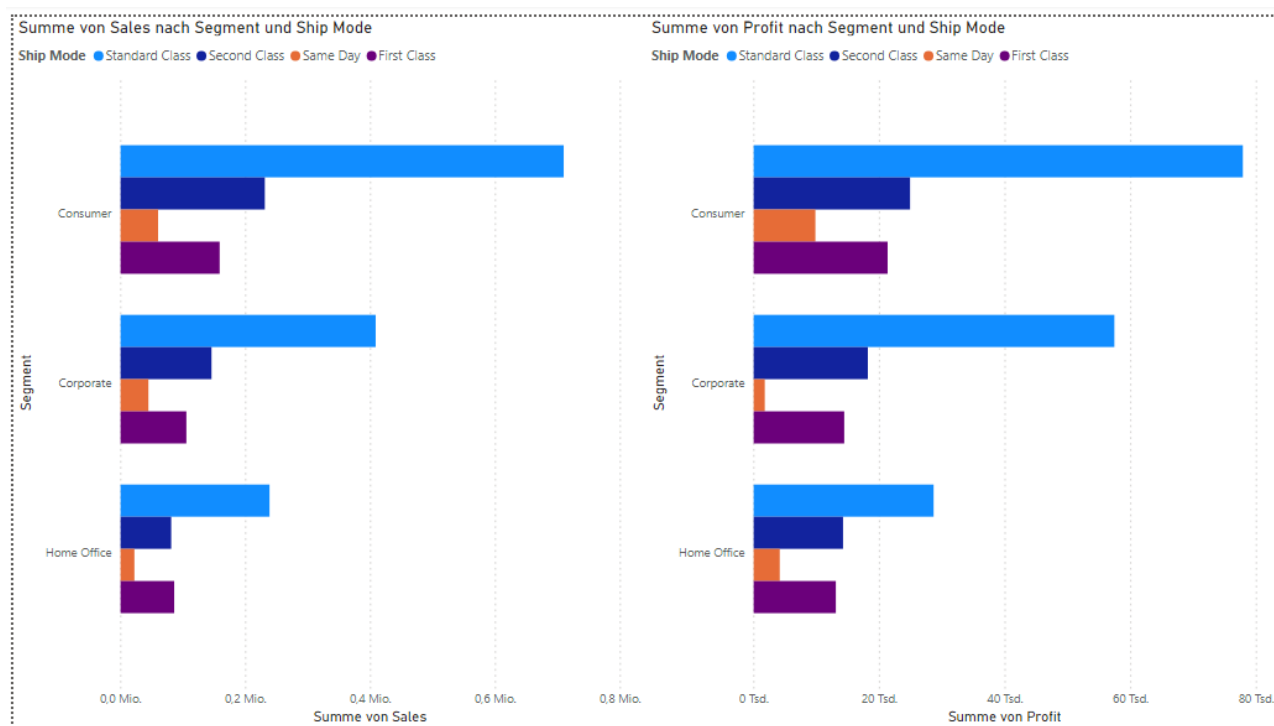
In den nächsten Graphen wurden noch einmal die Unterkategorien nach Verkaufsvolumen und Profit dargestellt, diesmal jedoch unterteilt nach Kundensegment und Region.



Die nächsten Graphen beleuchten, wie sich das Verkaufsvolumen und der Profit auf die Regionen und Staaten verteilen. Man sieht deutlich, dass Kalifornien und New York die größten Umsatzmärkte sind.

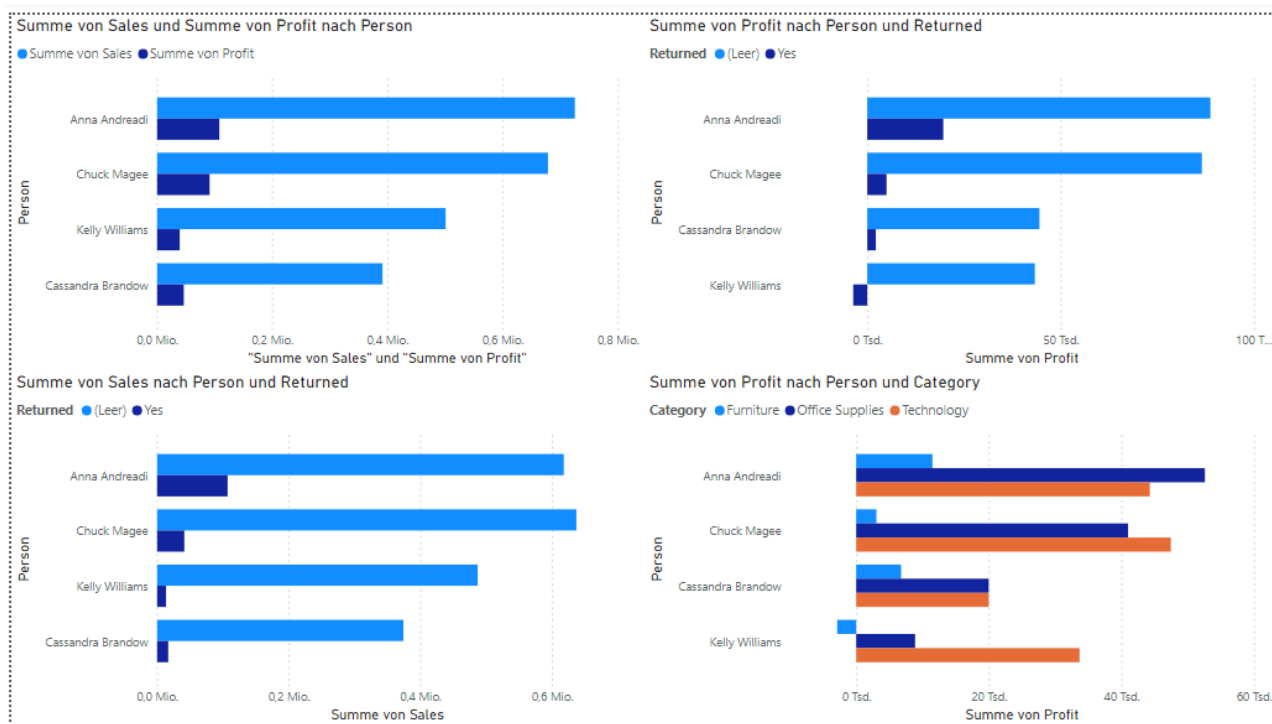


In den folgenden Darstellungen wurden die Versandmethoden genauer angesehen. Der interessanteste Graph ist der „Durchschnitt von Order Processing Time nach City und Ship Mode“. Wenn man diesen nach unten scrollt sieht man, dass die „Standard Class“ einen recht großen Zeitbereich abdeckt und man in manchen Regionen als Kunde geduldiger sein muss als in anderen.



Eine weitere Perspektive auf die Daten sind die Verkäufe und Profite nach Verkaufspersonal. Hier gibt es nur vier Mitarbeiter, wahrscheinlich wurden die Daten anonymisiert und nur als Beispiel aufgeführt. Dennoch sieht man, dass Kelly noch etwas mit Retouren herausholen kann. Diese schneiden aktuell noch in Ihren Profit. Möglicherweise hat sie Kunden Produkte empfohlen, die sie doch nicht wirklich gebraucht haben. Ein

Manager könnte dies mit ihr aufarbeiten.



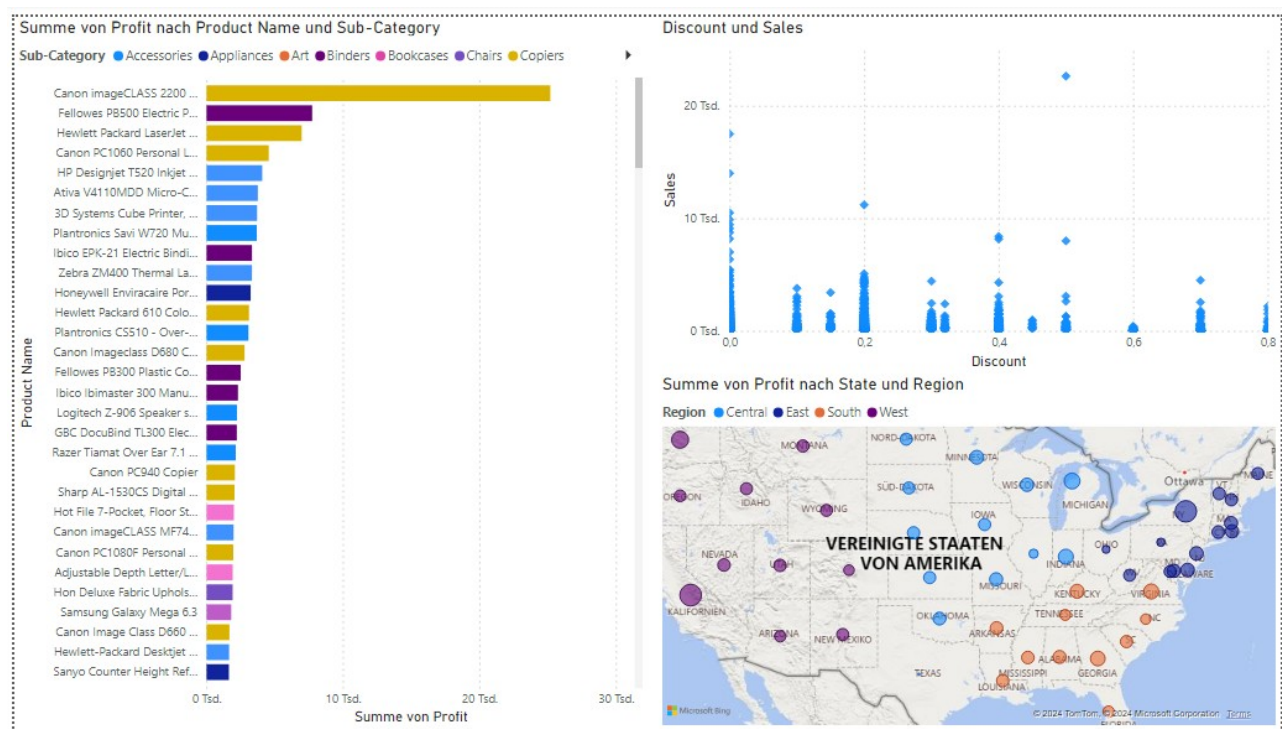
In den folgenden Diagrammen wurden die Bestellungen und Discounts weiter untersucht. Interessant ist, dass alle drei Kundensegmente im Durchschnitt vier Produkte pro Bestellung kaufen.

Stellt man den Discount und den Profit gegenüber, so fällt einem sofort auf, dass die Geschäfte mit einem bestimmten 3D-Drucker die meisten Verluste pro Produkt gemacht haben aber bei normalen Papier-Druckern pro Produkt wiederum den meisten Gewinn machen. Hier liegt möglicherweise eine Menge Potenzial diesen neuen Markt des 3D Druckens noch besser zu bedienen.

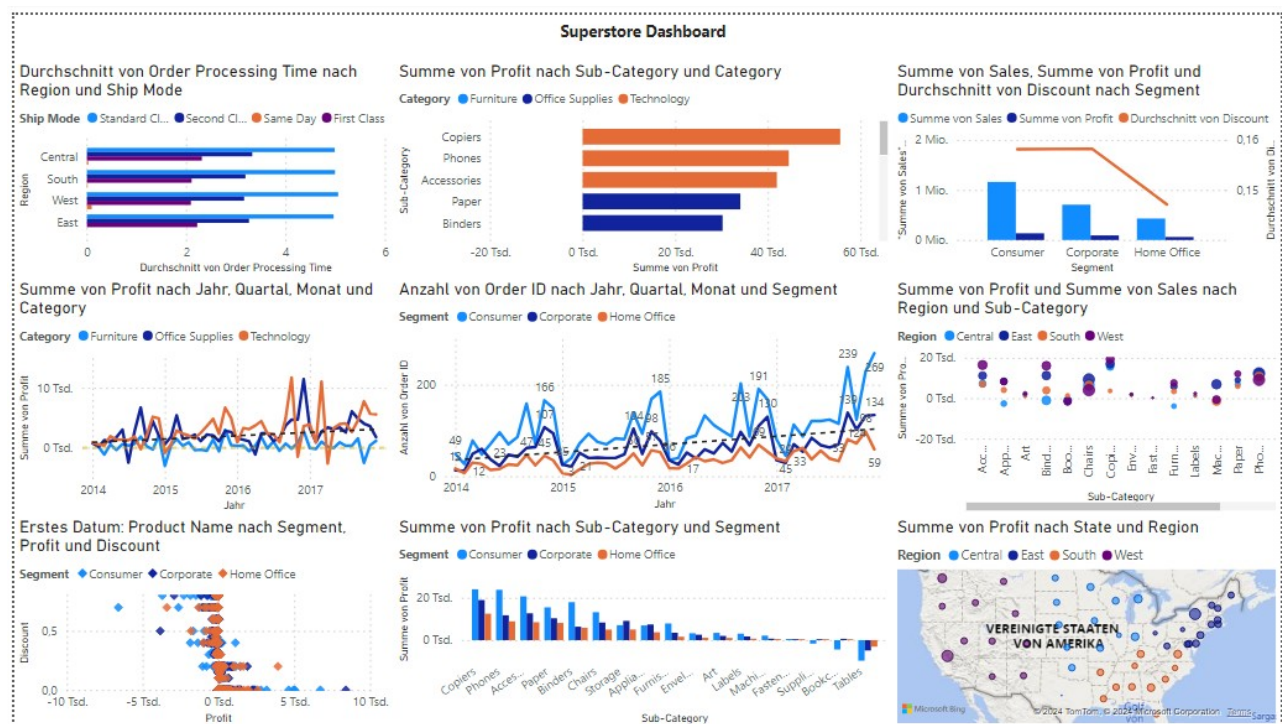


Die letzten Graphen listen noch einmal alle Produkte und deren Profit auf. Die Karte zeigt über die Kreisgröße welche Staaten am meisten zum Profit beigetragen haben. Wie zuvor haben Kalifornien und New York den größten Kreis.





Aus diesen Diagrammen wurden dann die interessantesten ausgewählt und auf einem Dashboard zusammengefasst.



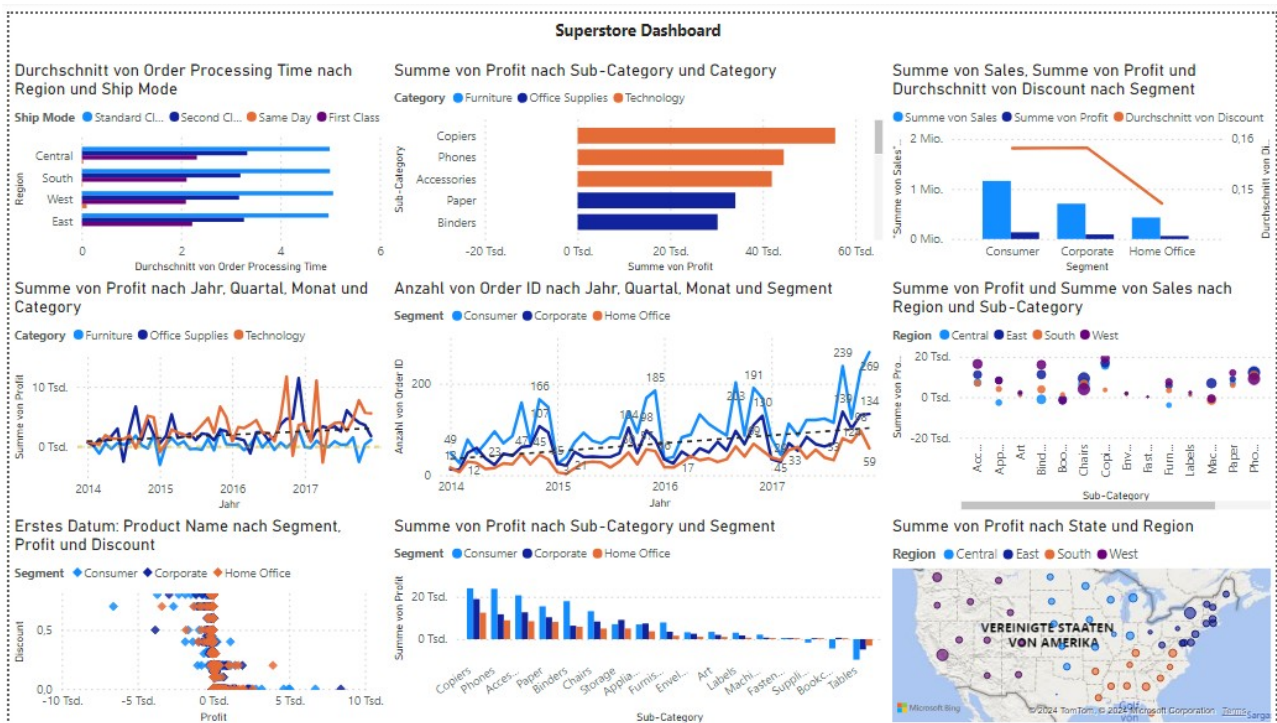
In diesem kann man nun einzelne Daten anklicken um die verschiedenen Ansichten auf diese Daten zu

fokussieren beziehungsweise zu Filtern. Klickt man zum Beispiel auf das „Consumers“ Segment, so sieht man nur noch Daten, die Konsumenten zugeordnet werden können.

## 2.6 Ergebnisse

Zusammenfassung der anschaulichsten Visualisierungen in einem Dashboard.

- „Summe von Profit nach State und Region“: Zeigt über Kreisgröße welche Regionen den meisten Gewinn einfahren.
- „Summe von Profit nach Sub-Category und Segment“: Zeigt, dass „Tables“ in allen Segmenten Verluste einfahren, sowie, dass „Bookcases“ und „Supplies“ im Konsumer-Segment ebenfalls Verluste bringen.
- „Erstes Datum: Product Name nach Segment, Profit und Discount“: Zeigt, dass einige 3D-Drucker nur mit Discount zu starken Verlusten verkauft werden konnten. Filtert man, via klick in einem anderen Diagramm, nach Tischen, so sieht man diese ebenfalls mit hohen Discounts große Verluste einfahren.
- „Summe von Profit nach Jahr, Quartal, Monat und Category“: Zeigt trotzdem einen positiven Profit-Trend aller Kategorien über die letzten vier Jahre.



## 2.7 Ausblick



Die Geschäfte sollten ihr Feingefühl für 3D Drucker und Tische weiter ausbauen oder diese Produkte anderen Wettbewerbern überlassen. Da sie jedoch erstaunlich gut im normalen Drucker-Verkauf sind, können sie vielleicht in den nächsten Jahren ihr Know-How noch besser auf das neue Marktsegment des 3D Druckens anwenden.