

Dokumentation

Informationsvisualisierung



Supply Chain Report

Welf Poser
29.06.2023

1.	EINFÜHRUNG	2
1.1	VERWENDETE METHODIK UND WÜRDIGUNG	2
1.1.1	<i>Das Was-Warum-Wie Modell</i>	2
1.1.2	<i>Vier Ebenen der Validierung</i>	3
1.1.3	<i>Visualization Analysis and Design</i>	4
1.2	DATENSATZ.....	4
2	DOMAIN SITUATION	5
2.1	EINORDNUNG DES ANWENDUNGSGEBIETS IM UNTERNEHMEN.....	5
2.2	ANWENDER.....	6
2.2.1	<i>Anwenderspezifische Fragen</i>	6
3	DATEN- UND AUFGABENABSTRAKTION.....	7
3.1.1	<i>DataCoSupplyChainDataset.csv.....</i>	8
3.1.2	<i>Tokenitized_access_logs.csv.....</i>	10
3.2	AUFGABENABSTRAKTION.....	11
3.2.1	<i>Abstraktion der Domänenspezifischen Fragen</i>	13
4	VORENTWÜRFE.....	15
4.1	BASIC LAYOUT	16
4.2	ERSTER VORENTWURF	16
4.3	ZWEITER VORENTWURF	17
4.4	DRITTER VORENTWURF.....	18
5	DATENVORBEREITUNG.....	19
5.1	BEHANDLUNG FEHLENDER EINTRÄGE	19
5.2	AUSSORTIERUNG UNWICHTIGER VARIABLEN.....	19
5.3	VORBEREITENDE SCHRITTE IN TABLEAU	20
5.4	ERSTELLUNG VON CALCULATED FIELDS IN TABLEAU.....	20
5.4.1	<i>Shipment delay (sched - real)</i>	20
5.4.2	<i>Late Delivery.....</i>	20
5.4.3	<i>Late Delivery Adjusted.....</i>	20
5.4.4	<i>Accuracy Late Delivery.....</i>	21
5.4.5	<i>Daytime</i>	21
6	VISUELLE KODIERUNG	21
6.1	LAYOUT UND DESIGN	22
6.2	FARBPALETTE	23
6.3	VISUALISIERUNGSARTEN.....	24
6.3.1	<i>Liniendiagramme.....</i>	24
6.3.2	<i>Karten.....</i>	25
6.3.3	<i>Balkendiagramme und Histogramme.....</i>	27
6.3.4	<i>Kreisdiagramm</i>	31
6.3.5	<i>Textfelder</i>	32
7	BEWERTUNG DES DASHBOARDS	33
7.1.1	<i>Validierung und Algorithmus.....</i>	33
7.1.2	<i>Fazit.....</i>	33

1. Einführung

Dieses Dokument beschreibt das systematische Vorgehen der Visualisierungs- und Designentscheidungen bei der Erstellung des Supply Chain Report in Tableau. Hierin sollen die Entscheidungen und der Prozess der Report-Erstellung nachvollziehbar dargelegt und wissenschaftlich begründet werden.

1.1 Verwendete Methodik und Würdigung

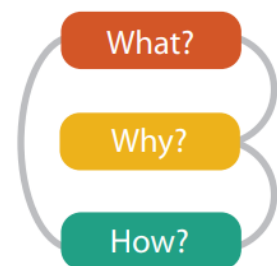
Informationsvisualisierung ist ein leistungsstarkes Werkzeug, um komplexe Datenmengen verständlich und zugänglich zu machen. Eine der einflussreichsten Denkerinnen auf diesem Gebiet ist Tamara Munzner, die ein Drei-Stufen-System für die Gestaltung effektiver Visualisierungen entwickelt hat, bekannt als das "Was-Warum-Wie-System". Das "Was-Warum-Wie-System" wird ergänzt durch das "verschachtelte Modell für Design und Validierung" ebenfalls entwickelt von Tamara Munzner.

Beide Systeme wurden beim Erstellen des Supply-Chain-Reports verwendet und bilden somit auch den Rahmen der Grundstruktur dieses Dokuments. Im Folgenden werden beide Modelle kurz erläutert.

1.1.1 Das Was-Warum-Wie Modell

Das "Was"

Das "Was" bezieht sich auf die Daten, die visualisiert werden sollen. Munzner teilt diese in verschiedene Kategorien ein. Jede Kategorie hat spezifische Eigenschaften, die berücksichtigt werden müssen, wenn diese visualisiert wird. Zum Beispiel haben Datennetze Beziehungen zwischen Knoten, während Felder räumliche Positionen aufweisen. Durch die genaue Definition, was dargestellt werden soll, kann der Designer eine geeignete Visualisierungsstrategie auswählen.



Das "Warum"

Das "Warum" bezieht sich auf die Ziele der Informationsvisualisierung. Was möchte der Betrachter erreichen? Welche Fragen sollen beantwortet werden? Munzner unterteilt diese Ziele in drei Hauptkategorien: "Analyse", "Suche" und "Abfrage". Suche bezieht sich auf das Auffinden spezifischer Datenpunkte oder Muster, während die Analyse darauf abzielt, neue Einsichten zu gewinnen, die Beziehungen und Strukturen in den Daten offenbaren. Mit der Abfrage ist das Identifizieren und Vergleichen von Daten gemeint. Durch das Verständnis der Ziele kann eine geeignete Visualisierungsstrategie entworfen werden, die den Benutzer bei der Erfüllung seiner Ziele unterstützt.

Das "Wie"

Das "Wie" bezieht sich auf die konkrete Umsetzung der Visualisierung. Wie werden die Daten dargestellt? Welche Farben, Formen und Layouts werden verwendet? Wie wird Interaktivität eingesetzt? Munzner bietet eine Reihe von Visualisierungstechniken an, die auf die jeweiligen Daten und Ziele abgestimmt sind. Durch sorgfältige Auswahl und Anwendung dieser Techniken kann der Designer eine wirksame und ansprechende Visualisierung erstellen.

Dieses Modell wird in der unten genannten IEEE-Publikation ausführlich beschrieben:

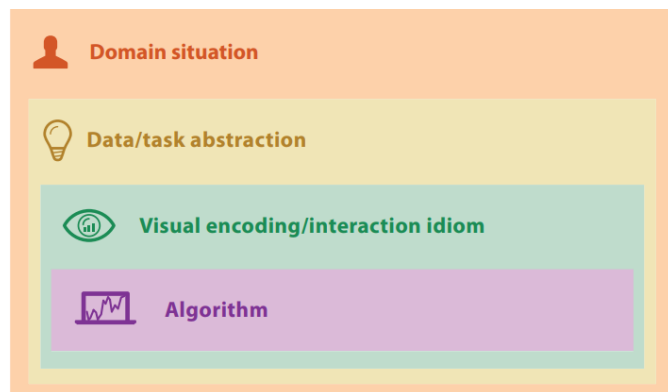
M. Brehmer and T. Munzner, "A Multi-Level Typology of Abstract Visualization Tasks," in IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 12, pp. 2376-2385, Dec. 2013, doi: 10.1109/TVCG.2013.124.

1.1.2 Vier Ebenen der Validierung

Das von Tamara Munzner vorgeschlagene "Verschachtelte Modell für Design und Validierung" umfasst vier Ebenen, die jeweils eine spezifische Rolle im Prozess der Informationsvisualisierung und deren Validierung spielen:

1. Anwendungsbereich:

Die "Domain Situation" stellt die erste Ebene dar und bezieht sich auf das eigentliche Problem oder die Aufgabe, die gelöst oder erreicht werden soll. Sie umfasst die spezifischen Fragen, die beantwortet werden sollen und das Publikum oder die Nutzer, die die Visualisierung verwenden werden.



2. Aufgaben- und Datenabstraktion:

Die Aufgabenabstraktion übersetzt spezifische Ziele aus der Domain Situation in allgemeinere Aufgaben, während die Datenabstraktion rohe Daten in ein abstrakteres Modell umwandelt. Beide dienen dazu, spezifische Anforderungen in ein für die Informationsvisualisierung geeignetes Format zu transformieren. Eine genaue Validierung dieser Abstraktionen ist entscheidend, um die Erfüllung der Domain-Anforderungen sicherzustellen.

3. Visuelle Kodierung:

Hier wird die visuelle Darstellung der Daten erstellt und validiert. Werden die Daten und Ziele korrekt und effektiv durch die gewählten visuellen Elemente (Farben, Formen, Layouts etc.) repräsentiert?

4. Algorithmen-/Technologieebene:

Auf dieser Ebene wird die technische Umsetzung der Visualisierung validiert. Sind die verwendeten Algorithmen und Technologien geeignet, um die visuelle Darstellung effektiv zu erstellen und zu interagieren?

Jede der vier Ebenen in Munznerns Modell erfordert eine separate Validierung. Sollte es auf einer Ebene zu Unstimmigkeiten oder Fehlern kommen, kann dies negative Auswirkungen auf die darüber liegenden Ebenen haben. Durch diesen verschachtelten Validierungsprozess wird gewährleistet, dass die Informationsvisualisierung auf jeder Ebene präzise und wirkungsvoll ausgeführt wird.

Dieses Modell wird in der unten genannten IEEE-Publikation ausführlich beschrieben:

T. Munzner, "A Nested Model for Visualization Design and Validation," in IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 921-928, Nov.-Dec. 2009, doi: 10.1109/TVCG.2009.111

1.1.3 Visualization Analysis and Design

In dem renommierten Werk "Visualization Analysis and Design" präsentiert Tamara Munzner mit bemerkenswerter Detailliertheit die Anwendung ihrer Modelle. Dieses Buch war eine unerlässliche Ressource bei der Erstellung des vorliegenden Berichts und stellt eine reiche Informationsquelle für diejenigen dar, die sich intensiv mit dem Bereich der Informationsvisualisierung auseinandersetzen möchten. Deshalb verdient dieses Buch eine besondere Würdigung und explizite Erwähnung.

MUNZNER, T.: Visualization Analysis and Design: CRC PRESS, 2015 (AK Peters Visualization Series). - ISBN 9781498759717

1.2 Datensatz

Der verwendete Datensatz "DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS" wird auf der Plattform Mendely Data, unter einer Creative Common Lizenz, zur Verfügung gestellt. ([link](#))

Constante, Fabian; Silva, Fernando; Pereira, António (2019), "DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS", Mendeley Data, V5, doi: 10.17632/8gx2fvg2k6.5

Es handelt sich hierbei um einen Datensatz über Lieferketten. Dieser Datensatz wurde von der DataCo Global zu analysezwecken genutzt und daraufhin publiziert.

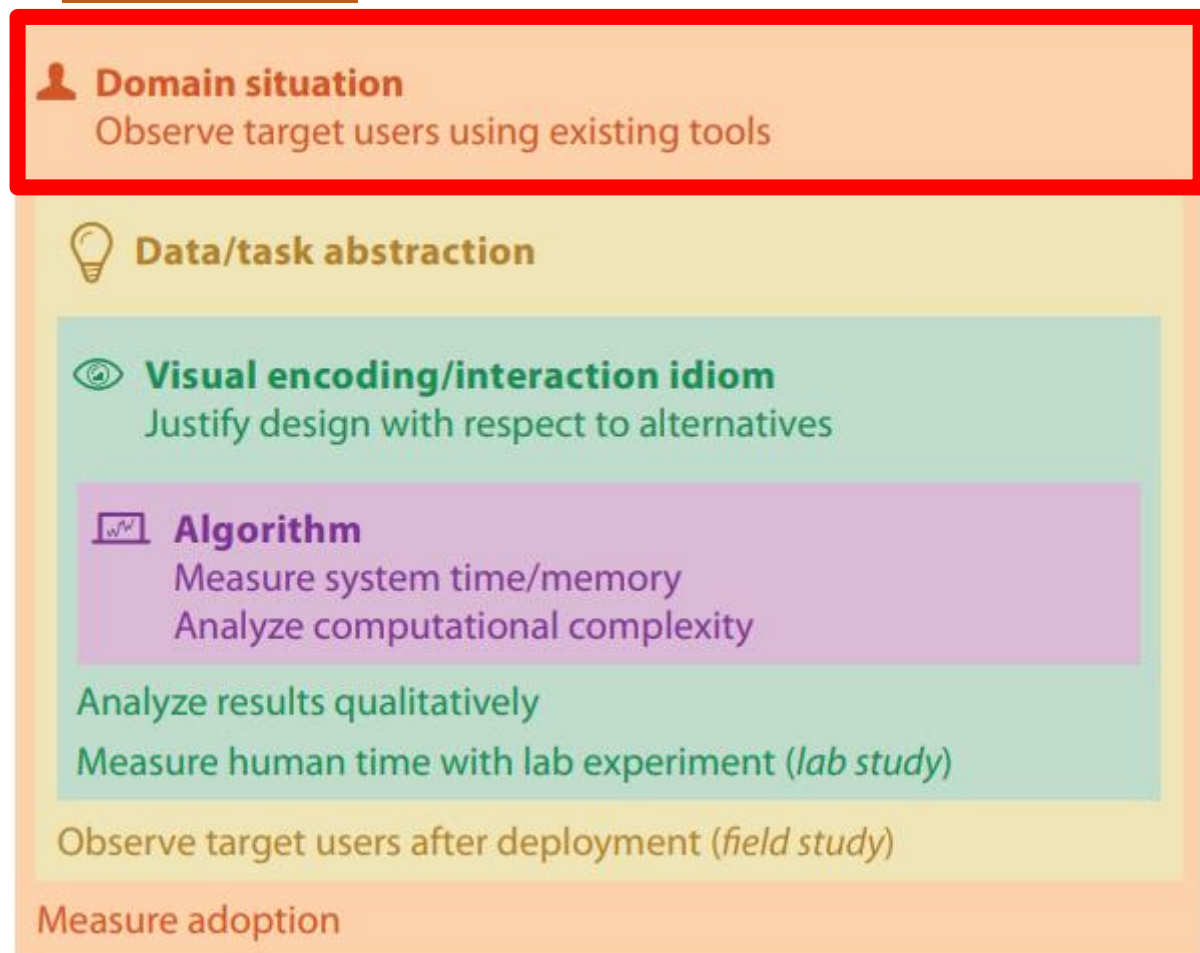
Über den Hintergrund und die Entstehung des Datensatzes ist nichts bekannt. Da DataCo Global ein Unternehmen ist, dass sich auf Analysen und Datenschutz spezialisiert hat, ist anzunehmen, dass der Datensatz für ein früheres Kundenprojekt erstellt und zu Analyse und Trainingsszenarien einer KI verwendet wurde.

Der Datensatz besteht aus drei CSV-Files:

- DataCoSupplyChainDataset.csv
- DescriptionDataCoSupplyChain.csv
- Tokenitized_access_logs.csv

Die genaue Betrachtung der Daten wird im Kapitel 3.1 Datenabstraktion erörtert.

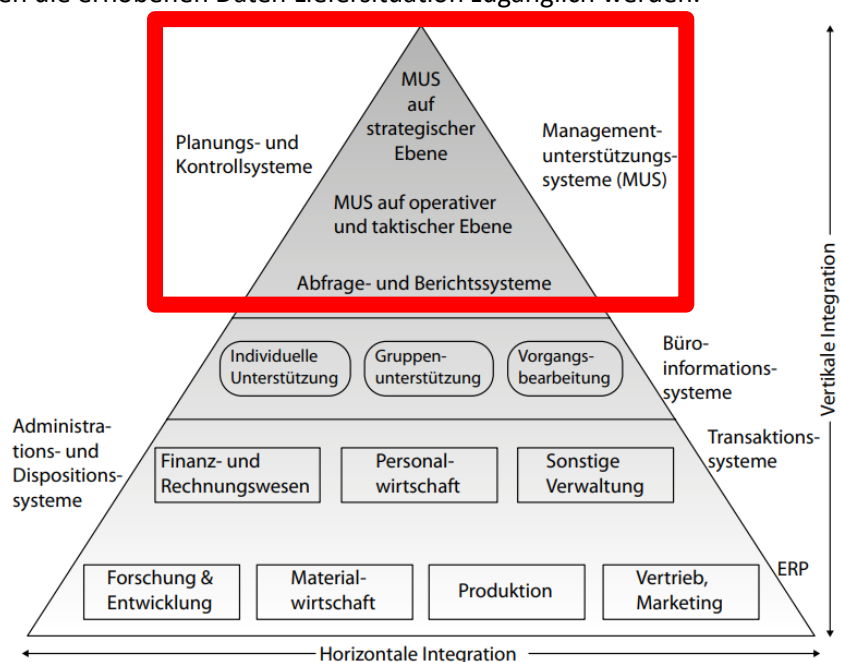
2 Domain Situation



2.1 Einordnung des Anwendungsgebiets im Unternehmen

Der erstellte Report soll als Teil eines Planungs- und Kontrollsystems in einem Unternehmen integriert werden. Durch den Report sollen die erhobenen Daten Liefersituation zugänglich werden.

Man verspricht sich dadurch eine verbesserte Überwachung und Kontrolle der Lieferkette. Zudem soll durch den Report der Überblick des momentanen Geschehens erleichtert werden. Darüber hinaus kann der Report das Management bei wichtigen Fragen und strategischer Unternehmensplanung unterstützen.



2.2 Anwender

Die Anwender haben folgende Rollen im Unternehmen:

- Supply Chain Manager
- Controller
- Logistik Manager
- Vertriebsteams
- CEO oder Geschäftsführer

Die Anwender sind alle einem spezifischen Unternehmen zugeordnet und besitzen dementsprechend ein einheitliches Domänenwissen.

Des Weiteren kann davon ausgegangen werden, dass Führungskräfte auf höherer Managementebene über ausreichende Fachkompetenzen und Fähigkeiten verfügen, um sich effizient in komplexen Berichten orientieren zu können. Jedoch sollte auch beachtet werden, dass gerade Manager oft sehr stark zeitlich beansprucht sind, was dafür spricht, dass der Report möglichst schnell verständlich und intuitiv aufgebaut sein sollte.

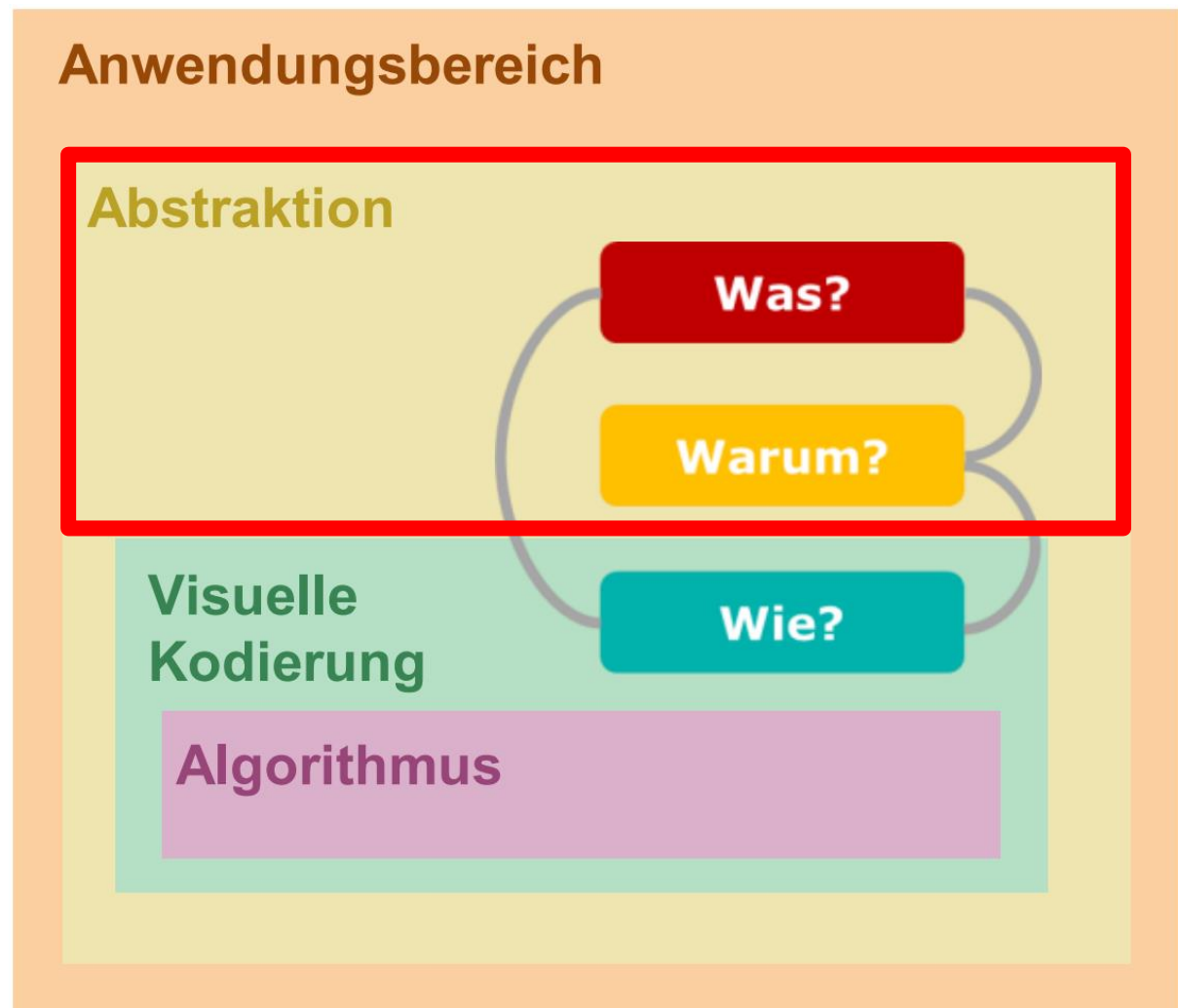
2.2.1 Anwenderspezifische Fragen

Entsprechend des Einsatzgebietes und der Anwender können somit folgende Fragestellungen, in Bezug auf das Supply Chain Management, auftreten:

1. Wie ist die durchschnittliche Lieferzeit? Wie variiert sie zwischen verschiedenen Lieferanten, Produktkategorien oder Zielregionen?
2. Gibt es wiederkehrende Muster oder Trends in den Lieferzeiten? Welche Faktoren könnten diese Trends beeinflussen (z.B. Feiertage, Wetterbedingungen)?
3. Gibt es signifikante Unterschiede in der Qualität oder im Zustand der gelieferten Produkte? Inwiefern korrelieren diese mit bestimmten Lieferanten oder Transportwegen?
4. Wie häufig treten Lieferverzögerungen oder -ausfälle auf? Was sind die häufigsten Ursachen für diese Probleme?
5. Wie genau sind unsere Lieferprognosen im Vergleich zur tatsächlichen Lieferleistung?
6. Gibt es Korrelationen zwischen bestimmten Produkten oder Lieferanten und erhöhten Lieferproblemen?
7. Welche Auswirkungen haben saisonale Schwankungen auf unsere Lieferkette?

Die konkreten Fragestellungen, speziell für diesen Report, werden in dem Kapitel 3.2 Aufgabenabstraktion erstellt.

3 Daten- und Aufgabenabstraktion



Wie in der Einführung angesprochen handelt es sich bei dem verwendeten Datensatz um drei CSV-Dateien:

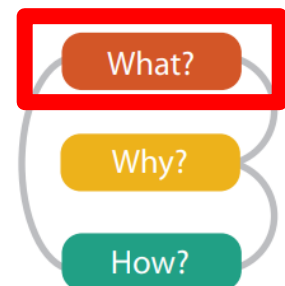
- DataCoSupplyChainDataset.csv
- DescriptionDataCoSupplyChain.csv
- Tokenized_access_logs.csv

Die *DataCoSupplyChainDataset.csv* wird in der Dokumentation von Mendeley Data als strukturierter Datensatz beschrieben, während die *Tokenized_access_logs.csv* als unstrukturierter Datensatz betitelt wird.

Bei beiden CSV's handelt es sich um tabellarische Datenätze.

Die *DescriptionDataCoSupplyChain.csv* beinhaltet die Metadaten, also die Beschreibung der einzelnen Attribute, für die *DataCoSupplyChainDataset.csv*.

Im nachfolgenden Kapitel werden die Datensätze *DataCoSupplyChainDataset.csv* und *Tokenized_access_logs.csv* näher betrachtet.



3.1.1 DataCoSupplyChainDataset.csv

Dieser Datensatz besteht, in seiner Ursprungsform, aus 53 Spalten (Attribute) und 180.520 Zeilen (Items).

Eine Zeile (Item) repräsentiert dabei eine Bestellung eines Kunden.

Die Spalten bzw. Attribute werden im Folgenden erklärt:

Type

Art der getätigten Transaktion (Kategorial)

- Ebenen: 4

Days for shipping (real)

Tatsächliche Versandtage des gekauften Produkts (Quantitativ)

- Wertebereich: 0 – 4

Days for shipment (scheduled)

Tage der geplanten Lieferung des gekauften Produkts (Quantitativ)

- Wertebereich: 0 – 4

Benefit per order

Verdienst pro aufgegebenen Bestellung (Quantitativ)

- Wertebereich: -4.274,98 – 911,80

Sales per customer

Gesamtumsatz pro Kunde (Quantitativ)

- Wertebereich: 7,49 – 1.939,99

Delivery Status

Lieferstatus der Bestellungen (Kategorial)

- Ebenen: 4

Late_delivery_risk

Kategoriale Variable, die angibt, ob die Sendung verspätet (1) oder nicht verspätet ist (0) (Kategorial)

- Ebenen: 2

Category Id

Produktkategorie-Code (Kategorial)

- Ebenen: 51

Category Name

Beschreibung der Produktkategorie (Kategorial)

- Ebenen: 50

Customer City

Stadt, in der der Kunde den Kauf getätigt hat (Kategorial)

- Ebenen: 563

Customer Country

Land, in dem der Kunde den Kauf getätigt hat (Kategorial)

- Ebenen: 2

Customer Email

E-Mail des Kunden (Kategorial)

- Ebenen: 1 (Alle anonymisiert)

Customer Fname

Kundenname (Kategorial)

- Ebenen: 782

Customer Id

Kunden-ID (Kategorial)

- Ebenen: 20.652

Customer Lname

Nachname des Kunden (Kategorial)

- Ebenen: 1.109

Customer Password

Maskierter Kundenschlüssel (Kategorial)

- Ebenen: 1 (Alle unkenntlich)

Customer Segment

Kundentypen (Kategorial)

- Ebenen: 3

Customer State

Bundesland, zu dem der Laden gehört, in dem der Kauf registriert ist (Kategorial)

- Ebenen: 46

Customer Street

Straße, zu der der Laden gehört, in dem der Kauf registriert ist (Kategorial)

- Ebenen: 7.458

Customer Zipcode

Postleitzahl des Kunden (Kategorial)

- Ebenen: 995

Department Id

Abteilungscode des Geschäfts (Kategorial)

- Ebenen: 11

Department Name

Name der Abteilung des Geschäfts (Kategorial)

- Ebenen: 11

Latitude

Breitengrad entsprechend dem Standort des Geschäfts (Quantitativ)

- Wertebereich: -33,94 – 48,78

Longitude

Längengrad entsprechend dem Standort des Geschäfts (Quantitativ)

- Wertebereich: -158,03 – 115,26

Market

Markt, an den die Bestellung geliefert wird (Kategorial)

- Ebenen: 5

Order City

Bestimmungsortstadt der Bestellung (Kategorial)

- Ebenen: 3.597

Order Country

Bestimmungsland der Bestellung (Kategorial)

- Ebenen: 164

Order Customer Id

Kundenbestellcode (Kategorial)

- Ebenen: 20.652

order date (DateOrders)

Datum, an dem die Bestellung aufgegeben wird (Quantitativ)

- Wertebereich:
1/1/2015 0:00 – 31/1/2018 23:38

Order Id

Bestellcode (Kategorial)

- Ebenen: 65.752

Order Item Cardprod Id

Produktcode, der durch den RFID-Leser generiert wird (Kategorial)

- Ebenen: 118

Order Item Discount

Rabattwert des Bestellartikels (Quantitativ)

- Wertebereich: 0,00 – 500,00

Order Item Discount Rate

Rabattprozensatz des Bestellartikels (Quantitativ)

- Wertebereich: 0,00 – 0,25

Order Item Id

Bestellartikelcode (Kategorial)

- Ebenen: 180.519

Order Item Product Price

Preis der Produkte ohne Rabatt (Quantitativ)

- Wertebereich: 9,99 – 1.999,99

Order Item Profit Ratio

Gewinnverhältnis des Bestellartikels (Quantitativ)

- Wertebereich: -2,75 – 0,5

Order Item Quantity

Anzahl der Produkte pro Bestellung (Quantitativ)

- Wertebereich: 1 – 5

Sales

Umsatzwert (Quantitativ)

- Wertebereich: 9,99 – 1.999,99

Order Item Total

Gesamtbetrag pro Bestellung (Quantitativ)

- Wertebereich: 7,49 – 1.939,99

Order Profit Per Order

Gewinn pro Bestellung (Quantitativ)

- Wertebereich: -4.274,98 – 911,80

Order Region

Region der Welt, in die die Bestellung geliefert wird (Kategorial)

- Ebenen: 23

Order State

Staat der Region, in die die Bestellung geliefert wird (Kategorial)

- Ebenen: 1089

Order Status

Bestellstatus (Kategorial)

- Ebenen: 9

Order Zipcode

Postleitzahl (Kategorial)

- Ebenen: 609

Product Card Id

Produktcode (Kategorial)

- Ebenen: 118

Product Category Id

Produktkategoriecode (Kategorial)

- Ebenen: 51

Product Description

Produktbeschreibung (Kategorial)

- Ebenen: 0 (Alles Null-Werte)

Product Image

Link des Besuchs und Kaufs des Produkts
(Kategorial)

- Ebenen: 118

Product Name

Produktname (Kategorial)

- Ebenen: 118

Product Status

Status des Produktlagers (Kategorial)

- Ebenen: 2

Shipping date (DateOrders)

Exaktes Datum und Uhrzeit des Versands
(Quantitativ)

- Wertebereich:
3/1/2015 – 6/2/2018 22:14

Shipping Mode

Versandart (Kategorial)

- Ebenen: 4

3.1.2 Tokenitized_access_logs.csv

Dieser Datensatz besteht aus 8 Spalten (Attributen) und 469.977 Zeilen (Items).

Die Attribute sind dabei:

- Product
- Category
- Date (Keine Werte hinterlegt)
- Month
- Hour
- Department
- Ip
- Url

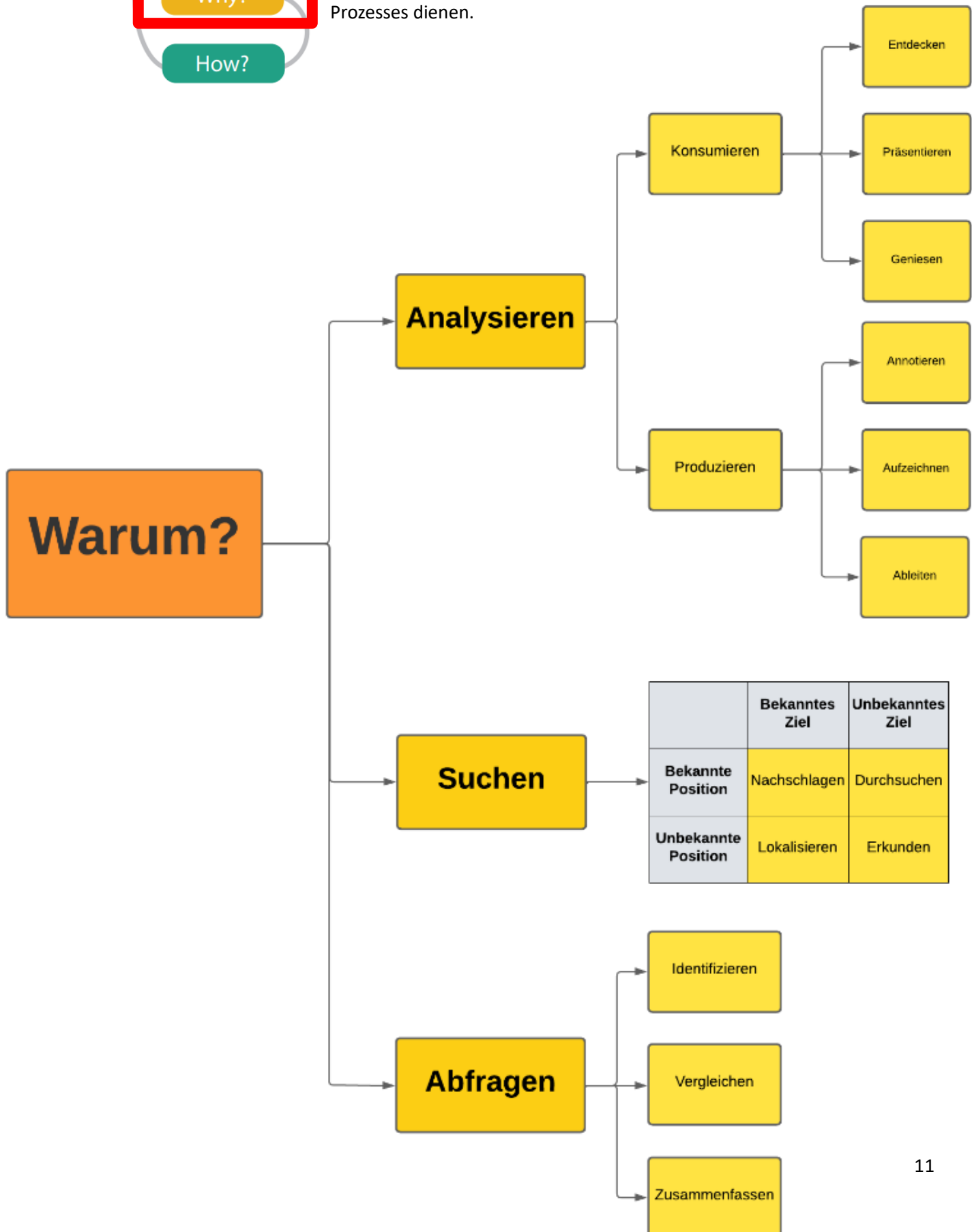
Leider gibt es keine Dokumentation der einzelnen Attribute. Somit ist nicht klar ersichtlich auf sich das Datum und die Zeitangaben beziehen. Darüber hinaus ist in der *DataCoSupplyChainDataset.csv* die Kategorie und das Produkt schon vorhanden und klar einer Kundenbestellung zugeordnet.

Aus den oben genannten Gründen wird dieser Datensatz deshalb nicht weiter berücksichtigt und nicht für den Report verwendet.

3.2 Aufgabenabstraktion



Wie in der Einleitung bereits erklärt, werden in der Aufgabenabstraktion die spezifischen Fragestellungen der Businessdomäne in eine allgemeine Form übersetzt, um den Visualisierungsprozess zu erleichtern. Die folgende Grafik soll eine Orientierung ermöglichen und der Klarstellung dieses Prozesses dienen.



Um die Aufgabenabstraktion effektiv durchzuführen, werden zuerst Domänenspezifische Fragen erhoben. Um gleich zu Beginn eine gewisse Struktur in den Prozess zu bringen, werden nur Fragen berücksichtigt, die mit den hiesigen Daten zu beantworten sind. Zudem werden alle Fragen drei Kategorien zugeordnet:

Versand:

1. Wie verhält sich die geplante mit der realen Versandzeit über den erhobenen Zeitraum?
2. Lassen sich Ursachen für einen verspäteten Versand identifizieren?
 - I. Gibt es Zusammenhänge zwischen dem Lieferort und einem verspäteten Versand?
 - II. Gibt es Filialen, die deutlich häufiger einen verspäteten Versand verursachen?
 - III. Gibt es Zusammenhänge zwischen der Bestellzeit und einem verspäteten Versand?
 - IV. Hängt ein verspäteter Versand mit der Anzahl Artikel zusammen?
3. Wie verhält sich die Liefersituation in Bezug auf die verschiedenen Lieferkategorien?

Lieferung:

1. Wie hoch ist die Genauigkeit für die Vorhersage einer verspäteten Lieferung?
2. Resultiert die Verspätete Lieferung aus einer verspäteten Absendung oder ist der Liefersdienst für eine verspätete Lieferung verantwortlich?
3. Gibt es Zusammenhänge einer verspäteten Lieferung mit dem Zielort der Lieferung?
4. Hängt eine verspätete Lieferung mit der Anzahl Artikel zusammen?

Umsatz, Gewinn und Kundenverhalten:

1. Wie verhält sich der Umsatz und der Profit über den betrachteten Zeitraum?
 - I. Wie hoch ist der durchschnittliche Profit per Bestellung?
 - II. Wie hoch ist der (gesamte) Umsatz?
 - III. Wie hoch ist der (gesamte) Profit?
2. In welche Regionen konzentriert sich der Kundenstamm?
3. Was sind die beliebtesten Produkte?
4. Was sind die bevorzugten Bezahloptionen?

3.2.1 Abstraktion der Domänenspezifischen Fragen

Die domänenspezifischen Fragen werden nun mittels des Frameworks von Tamara Munzner in eine allgemeinere Aufgabenabstraktion überführt. Dieser Prozess hilft dabei, die inhaltliche Komplexität der gestellten Fragen zu reduzieren und sie in handhabbare, analysierbare Einheiten zu zerlegen. Indem wir die benötigten Datentypen und die entsprechenden Aktionen zur Beantwortung der Fragen identifizieren, legen wir den Grundstein für das Design der Datendarstellung. Dieser Schritt ist entscheidend, um zu gewährleisten, dass die geeignetsten Analysetools und Visualisierungstechniken ausgewählt werden, um die gestellten Fragen effektiv und effizient zu beantworten.

Versand:

1. Wie verhält sich die geplante mit der realen Versandzeit über den erhobenen Zeitraum?

- **Datentypen:** **Zeitreihen** für geplante und tatsächliche Versandzeit.
- **Aktionen:** **Entdeckung der Trends und Muster** in geplanten und tatsächlichen Versandzeiten.

2. Lassen sich Ursachen für einen verspäteten Versand identifizieren?

I. Gibt es Zusammenhänge zwischen dem Lieferort und einem verspäteten Versand?

- **Datentypen:** **Geographische Daten** für Lieferorte, **Zeitdaten** für Versand.
- **Aktionen:** **Entdeckung der Korrelationen** zwischen Lieferorten und verspätetem Versand.

II. Gibt es Filialen, die deutlich häufiger einen verspäteten Versand verursachen?

- **Datentypen:** **Kategorische Daten** für Filialen, **Zeitdaten** für Versand.
- **Aktionen:** **Vergleichen der Häufigkeit** von verspätetem Versand je nach Filiale.

III. Gibt es Zusammenhänge zwischen der Bestellzeit und einem verspäteten Versand?

- **Datentypen:** **Zeitdaten** für Bestellungen und Versand.
- **Aktionen:** **Entdeckung der Korrelationen** zwischen Bestellzeit und verspätetem Versand.

IV. Hängt ein verspäteter Versand mit der Anzahl Artikel zusammen?

- **Datentypen:** **Numerische Daten** für Artikelzahlen, **Zeitdaten** für Versand.
- **Aktionen:** **Entdeckung der Korrelationen** zwischen Artikelzahlen und verspätetem Versand.

3. Wie verhält sich die Liefersituation in Bezug auf die verschiedenen Lieferkategorien?

- **Datentypen:** **Klassifikationsdaten** für Lieferkategorien und Versandstatus, **Zeitdaten** für geplante und tatsächliche Versandzeiten.
- **Aktionen:** **Analyse und Vergleichung der Verteilung und Trends** der Versandzeiten über die verschiedenen Lieferkategorien.

Lieferung:

1. Wie hoch ist die Genauigkeit für die Vorhersage einer verspäteten Lieferung?
 - **Datentypen:** **Klassifikationsdaten** für Vorhersagen und tatsächliche Versand- und Lieferzeiten.
 - **Aktionen:** **Auswertung** der **Genauigkeit der Vorhersage** für verspätete Lieferungen.
2. Resultiert die Verspätete Lieferung aus einer verspäteten Absendung oder ist der Lieferdienst für eine verspätete Lieferung verantwortlich?
 - **Datentypen:** **Zeitdaten** für Versand und Lieferung, **Klassifikationsdaten** für Versand- und Lieferstatus.
 - **Aktionen:** **Analyse der Ursachen** für verspätete Lieferungen.
3. Gibt es Zusammenhänge einer verspäteten Lieferung mit dem Zielort der Lieferung?
 - **Datentypen:** **Geographische Daten** für Lieferziele, **Zeitdaten** für Versand und Lieferung.
 - **Aktionen:** **Entdeckung der Korrelationen** zwischen verspäteten Lieferungen und Lieferzielen.
4. Hängt eine verspätete Lieferung mit der Anzahl Artikel zusammen?
 - **Datentypen:** **Numerische Daten** für Artikelzahlen, **Zeitdaten** für Versand und Lieferung.
 - **Aktionen:** **Vergleichung der Korrelationen** zwischen verspäteten Lieferungen und Artikelzahlen.

Umsatz, Gewinn und Kundenverhalten:

1. Wie verhält sich der Umsatz und der Profit über den betrachteten Zeitraum?
 - I. Wie hoch ist der durchschnittliche Profit per Bestellung?
 - **Datentypen:** **Numerische Daten** für Bestellungen und Profit.
 - **Aktionen:** **Berechnung** und Darstellung des **durchschnittlichen Profits** pro Bestellung.
 - II. Wie hoch ist der (gesamte) Umsatz?
 - **Datentypen:** Numerische Daten für Umsatz.
 - **Aktionen:** **Berechnung** und Darstellung des **gesamten Umsatzes**.
 - III. Wie hoch ist der (gesamte) Profit?
 - **Datentypen:** **Numerische Daten** für Profit.
 - **Aktionen:** **Berechnung** und Darstellung des **gesamten Profits**.
2. In welche Regionen konzentriert sich der Kundenstamm?
 - **Datentypen:** **Geographische Daten** für Kundenstandorte.
 - **Aktionen:** **Präsentation der geographischen Verteilung** des Kundenstamms.
3. Was sind die beliebtesten Produkte?
 - **Datentypen:** **Kategorische Daten** für Produkte, **numerische Daten** für Verkaufszahlen.
 - **Aktionen:** **Entdeckung und Präsentation der Verteilung** und Bestellzahlen von Produkten.

4. Was sind die bevorzugten Bezahloptionen?

- **Datentypen:** **Kategorische Daten** für Bezahlmethoden.
- **Aktionen:** **Analyse und Präsentation der Verteilung** und Präferenzen der Bezahlmethoden.

4 Vorentwürfe



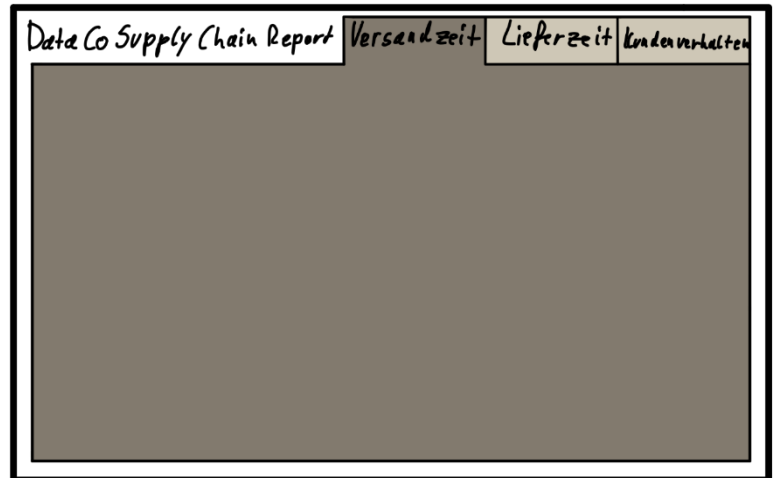
Im folgenden Kapitel werden drei Vorentwürfe präsentiert und miteinander verglichen, die auf den zuvor identifizierten Anforderungen basieren. Jeder Entwurf bietet einen einzigartigen Ansatz, um die gegebenen Informationen zu visualisieren und die gestellten Fragen zu beantworten. Dabei berücksichtigen wir verschiedene Aspekte wie die Auswahl der Visualisierungstechniken, die Art der Datenpräsentation und die Integration der verschiedenen Datenelemente. Das Ziel ist es, die Stärken und Schwächen jedes Entwurfs zu analysieren und zu vergleichen, um letztendlich die effektivste und intuitivste Lösung zu finden.

4.1 Basic Layout

Um einen Rahmen für die folgenden Entwürfe zu bieten, wurde zuerst ein Basic Layout festgelegt:

Dieses Layout bietet im oberen linken Bereich Platz für den Titel des Dashboards und ggf. ein Firmenlogo. Dies ist wichtig, da dieser Bereich typischerweise zuerst betrachtet wird und somit dem User gleich eine Orientierung gibt was zu erwarten ist.

Das Menü im oberen rechten Rand ist durch das Design, dass an einen Fileexplorer oder Browser-Tab erinnert, intuitiv zu verstehen. Zudem bietet es eine zusätzliche Orientierung durch die Farbbelegung der einzelnen Reiter.



4.2 Erster Vorentwurf

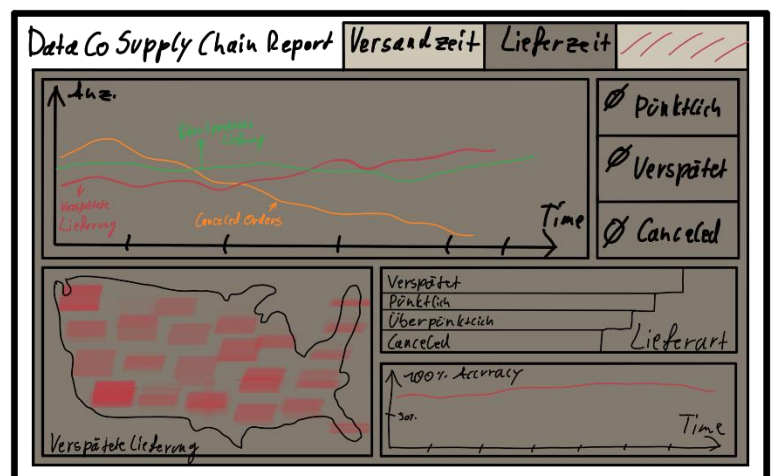
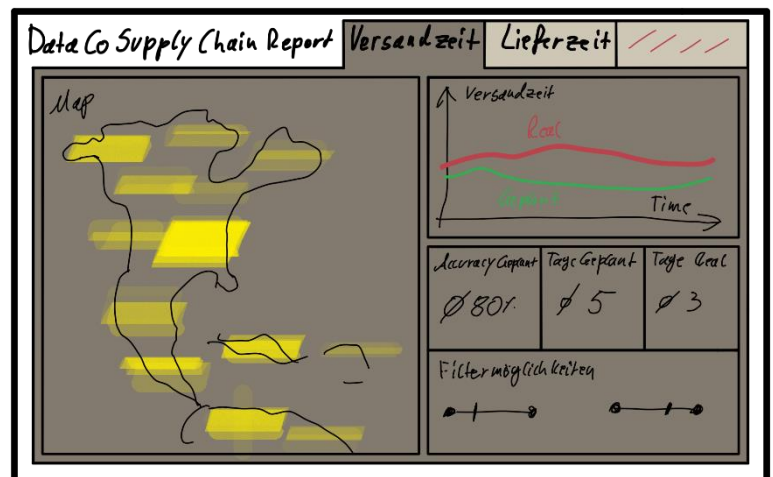
Der erste Entwurf besteht nur aus den Seiten "Versandzeit" und "Lieferzeit". Das hat den Grund, da der Report sich nur auf die wesentlichen Informationen der Lieferkette fokussieren soll.

Auf der Seite "Versandzeit" soll es dem Nutzer ermöglicht werden schnell und effektiv alle nötigen Informationen zu erfassen.

Die große Karte ermöglicht es dem Nutzer, schnell einen Überblick über schlecht performende Stores zu gewinnen. In dem Liniendiagramm gewinnt man einen groben Überblick über das Verhältnis von realer zu geplanter Versandzeit. Die Textfelder mit den wichtigsten Metriken helfen zusätzlich der schnellen Erfassung wichtiger KPI's.

Ausgewählte Filtermöglichkeiten runden das Design ab und ermöglichen bei Bedarf eine flache Analyse.

Die zweite Seite verfolgt den gleichen Ansatz der schnellen Informationsgewinnung. In dem Liniendiagramm kann man die Entwicklungen des Bestellstatus verfolgen. Dieser Chart ist extra groß gewählt, um die Datenaufnahme zu unterstützen. Zudem helfen die Textfelder wieder bei der Erfassung der wichtigen Kennzahlen. Das Balkendiagramm rundet die schnelle Gesamtübersicht über die Liefersituation ab. Um eine gute



Informationsverarbeitung zu gewährleisten, werden die Balken liegend präsentiert. Die Karte gibt Aufschluss über verspätete Lieferungen, farblich markiert in den Staaten. Der Liniendiagramm rechts unten bietet nur eine Zusatzinformation über die Genauigkeit der Vorhersage für eine zu späte Lieferung.

Dieses Design ist geeignet, um sich eine kurze Übersicht zu verschaffen. Es ist also eher ein Überwachungsdashboard. Der Nachteil ist, dass man dadurch kaum die darunter liegenden Daten weiter analysieren kann. Das führt dazu, dass man nur die Ist-Situation bewerten, jedoch für die Analyse unbefriedigender Geschehnisse, keine Gegenmaßnahmen ableiten kann.

Gerade deswegen wurde dieses Design nicht gewählt. Es ist ungeeignet alle Businessfragen zu beantworten.

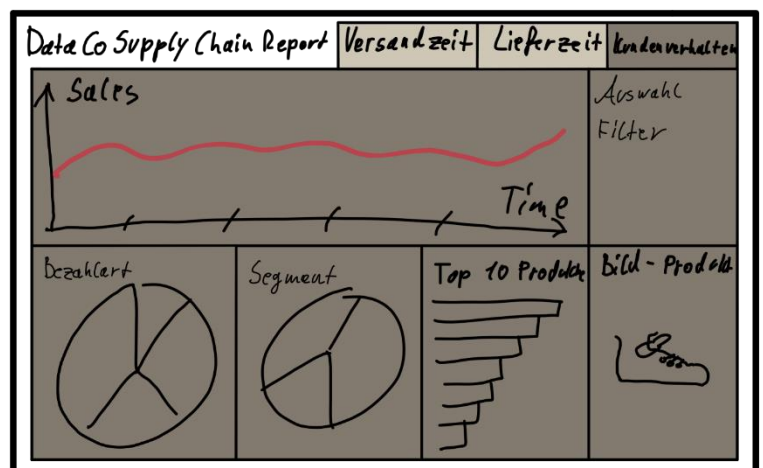
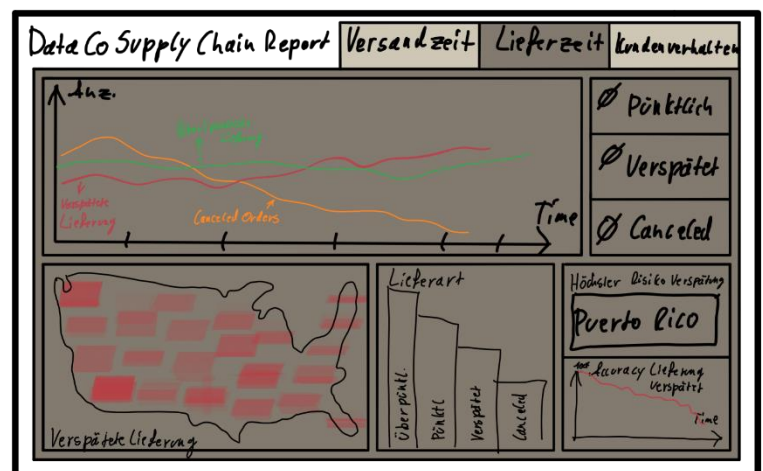
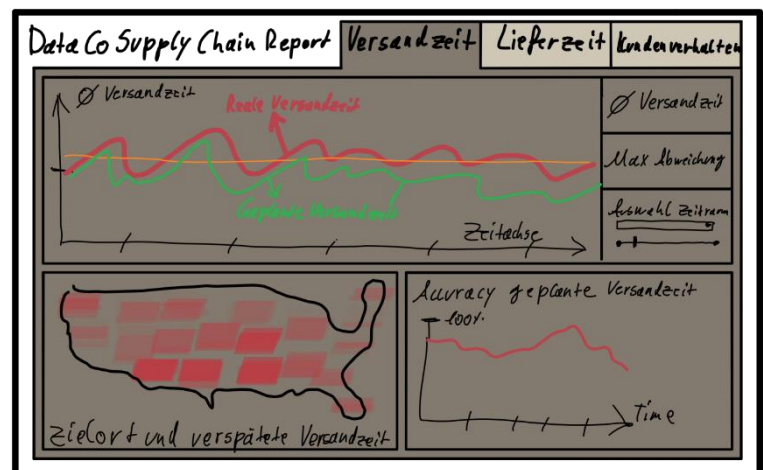
4.3 Zweiter Vorentwurf

Dieser Entwurf ist aus der Verbesserung des ersten Entwurfs entstanden und verfolgt somit den Ansatz dem Nutzer eine tiefere Analyse der Daten zu ermöglichen und darüber hinaus trotzdem noch einen schnellen Einstieg und Überblick der momentanen Versand- und Liefersituationen zu bieten.

Um dies zu realisieren wurde versucht ein möglichst einheitliches Design zu schaffen, um die Orientierung im Report zu vereinfachen. Die Liniendiagramme, die Karten und die Textfelder wurden alle an der gleichen Stelle positioniert.

Zudem wurde der Reiter Kundenverhalten hinzugefügt, um weitere Einblicke in die vorhandenen Daten zu ermöglichen. Die dargestellten Daten in diesem Reiter haben nur sekundär etwas mit dem Liefermanagement zu tun, sind jedoch interessant für Geschäftsführer und Vertriebsteams.

Der Ansatz ein Design zu finden das sowohl als schnelle Übersicht als auch zu analysezwecken genutzt werden kann, geht in die richtige Richtung. Wenn man jedoch die Seiten "Versandzeit" und "Lieferzeit" genauer betrachtet, dann kann man feststellen, dass immer noch zu wenige Visualisierungen vorhanden sind um die Daten geeignet filtern zu können. Somit wurde dieser Entwurf auch verworfen.



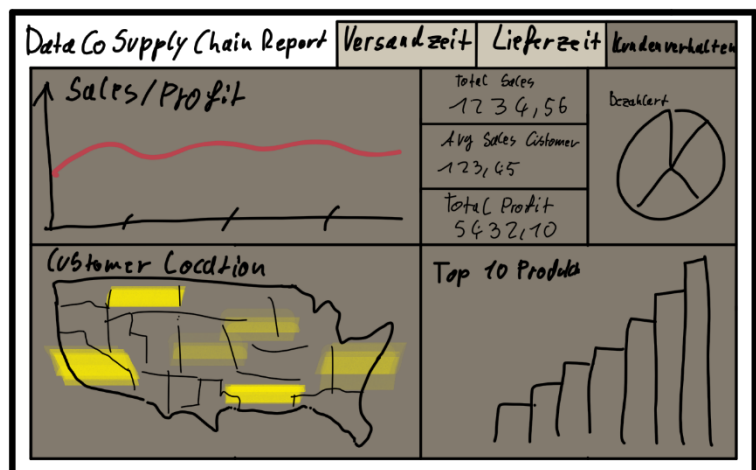
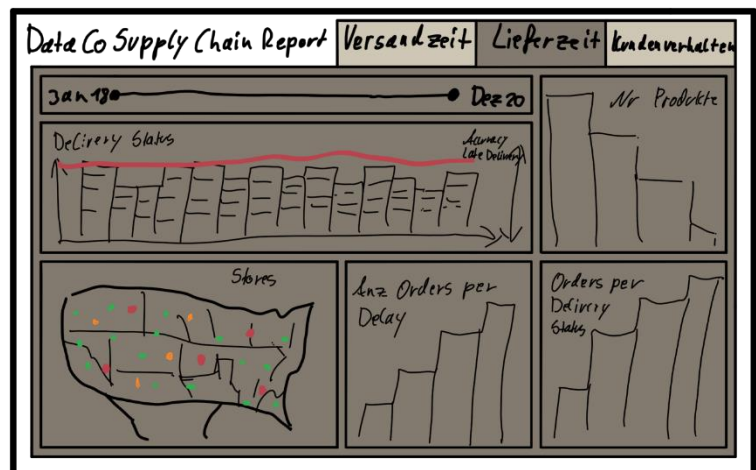
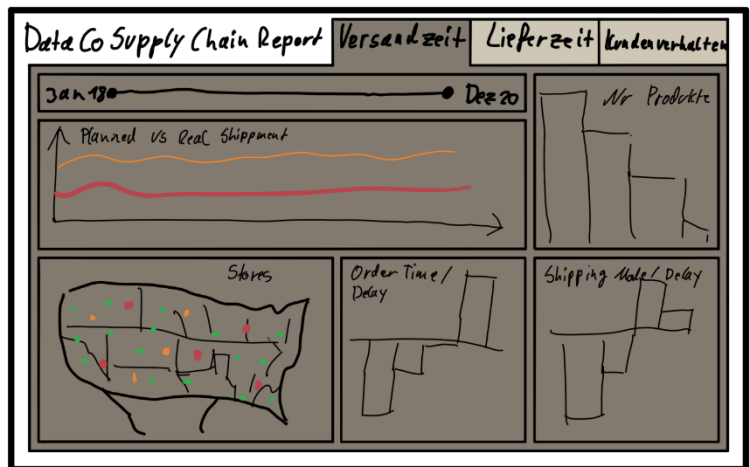
4.4 Dritter Vorentwurf

Im dritten Entwurf wurde das einheitliche Design nochmals optimiert. Die Seiten "Versandzeit" und "Lieferzeit" sind nun vom Layout her gleich. Die Seite Kundenverhalten ist ähnlich, jedoch lässt sich durch die unterschiedlichen Gegebenheiten nur schwer ein absolut einheitliches Layout finden.

Um gute Analysemöglichkeiten zu ermöglichen und um alle Businessfragen abzudecken, wurden zu allen Seiten noch Visualisierungen über wichtige Sachverhalte des Unternehmens, hinzugefügt.

Dies geht ein bisschen zu Lasten der Übersichtlichkeit und der schnellen Auffassung. Jedoch handelt es sich bei den Nutzern, wie in dem Kapitel 2 "Domain Situation" beschrieben, überwiegend um Manager, die in der Lage sein müssen, sich schnell in komplexe Sachverhalte einzuarbeiten und generell ein großes Domainwissen besitzen. Somit ist der Malus der Übersichtlichkeit vertretbar.

Die Feinheiten dieses Layouts werden im Kapitel 6 „Visuelle Kodierung“ genaustens beleuchtet.



5 Datenvorbereitung

Um mit dem Datensatz in Tableau effektiv arbeiten zu können, muss dieser erst aufbereitet werden. Das Vorgehen der Datenvorbereitung wird in diesem Kapitel kurz beschrieben.

Die Datenvorbereitung wurde größtenteils mittels Python durchgeführt. Hierbei wurde die CSV-Datei in ein Dataframe eingelesen und folgende Schritte ausgeführt:

5.1 Behandlung fehlender Einträge

Zuerst wurden alle Variablen auf fehlende Werte untersucht. Dabei wurden bei folgenden Variablen Null-Werte identifiziert.

- Customer Zipcode: 3
- Order Zipcode: 155.679
- Product Description: 180.519

Die fehlenden Einträge in Customer Zipcode wurden durch die Entfernung der gesamten Reihe behoben.

Die Spalten Order Zipcode und Product Description wurden beide gelöscht, da diese durch die überwiegende Anzahl an fehlenden Einträgen unbrauchbar sind.

5.2 Aussortierung unwichtiger Variablen

Im nächsten Schritt wurden die vorhandenen Variablen näher betrachtet und überprüft, ob diese überhaupt relevant für dieses Visualisierungsprojekt sind.

Dabei wurden folgende Variablen aussortiert:

Sensible Informationen über den Kunden:

- Customer Email
- Customer Fname
- Customer Lname
- Customer Password

Nach Betrachtung der unterschiedlichen ID's wurden folgende Spalten für nicht notwendig befunden:

- **Category Id (**
Wird nicht benötigt da Category Name den gleichen identifizierenden Effekt erfüllt
- **Department Id**
Wird nicht benötigt da Department Name den gleichen identifizierenden Effekt erfüllt
- **Order Customer Id**
Wird nicht benötigt da Customer Id, die gleiche ID repräsentiert
- **Order Item Cardprod Id** (Wird zur Visualisierung nicht benötigt)
- **Product Card Id** (Wird zur Visualisierung nicht benötigt)
- **Product Category Id**
Wird nicht benötigt da Category Name den gleichen identifizierenden Effekt erfüllt

Abschließend besteht der Datensatz aus 41 Spalten (Attributen) und 180516 Reihen (Items). Das behandelte Dataframe wurde in eine CSV-Datei überführt und kann nun in Tableau bearbeitet werden.

5.3 Vorbereitende Schritte in Tableau

In diesem Prozess wurde der aufbereitete Datensatz in das Visualisierungsprogramm Tableau eingelesen. Anschließend wurde die erste Reihe des Datensatzes als Header definiert. Für eine korrekte Dateninterpretation wurde das Komma als Delimiter ausgewählt.

Die Bezeichnung "EE.UU." in der Spalte "Customer Country", die auf Spanisch die Vereinigten Staaten bezeichnet, wurde durch den Alias "USA" ersetzt.

Die binären Werte in der Spalte "Late_delivery_risk" wurden in Boolesche Werte konvertiert.

Des Weiteren wurde die Spalte "Product Status" in "Product available" umbenannt und dabei die binären Werte in boolesche Werte umgewandelt und invertiert. Dieser Schritt fördert das intuitive Verständnis dieses Attributs und ermöglicht eine einfachere Dateninterpretation.

Abschließend wurden die Attribute "Customer Id", "Order Id" und "Order Item Id" in die diskrete Kategorie verschoben (Blau), und das Attribut "Type" wurde in "Payment Type" umbenannt, um das intuitive Verständnis weiter zu erhöhen.

5.4 Erstellung von Calculated Fields in Tableau

Calculated Fields in Tableau sind leistungsfähige Werkzeuge, die es ermöglichen, Daten durch benutzerdefinierte Formeln zu manipulieren, um neue Verbindungen oder Einblicke zu generieren. Ihre Anwendung ist von entscheidender Bedeutung, da sie eine tiefere und angepasste Analyse der vorhandenen Daten ermöglicht.

5.4.1 Shipment delay (sched- real)

Um den geplanten mit dem realen Versandtag vergleichen zu können musste die Differenz zwischen den beiden Angaben berechnet werden:

[Days for shipment (scheduled)]-[Days for shipping (real)]

Somit ergibt sich eine Metrik, die eine negative Anzahl an Tagen angibt, wenn der geplante Versandtag verstrichen ist und eine positive Anzahl an Tagen wenn der geplante Versandtag übertroffen wurde.

5.4.2 Late Delivery

Um eine binäre Metrik zu bekommen die eine verspätete Lieferung kennzeichnet wurde folgende Formel verwendet:

IF [Delivery Status]="Late delivery" then 1 ELSE 0 END

Dadurch lässt sich leicht filtern, ob eine Bestellung verspätet geliefert wurde.

5.4.3 Late Delivery Adjusted

Für die Berechnung des Prozentwerts der zu spät gelieferten Bestellungen gesehen auf das gesamte Bestellvolumen wurde folgende Formel erstellt:

SUM([Late Delivery])/COUNT([Order Id])

5.4.4 Accuracy Late Delivery

Zur Evaluation der Genauigkeit der Vorhersage für eine zu Späte Lieferung wurde folgende Metrik ergänzt:

```
if [Late_delivery_risk_bool] == TRUE AND [Delivery Status] == "Late delivery" then 1 ELSEIF  
[Delivery Status] == "Shipping canceled" then NULL  
ELSE 0 EN
```

5.4.5 Daytime

Um eine oberflächliche Analyse zu ermöglichen, wurde eine Aggregation der Bestellzeitpunkte durchgeführt. Dabei wurde die Zeit in vier Intervalle eingeteilt.

```
IF Datepart("hour",[order date (DateOrders)])>=6 and Datepart("hour",[order date (DateOrders)])<12  
then "morning" ELSEIF
```

```
Datepart("hour",[order date (DateOrders)])>=12 and Datepart("hour",[order date (DateOrders)])<18  
then "afternoon" ELSEIF
```

```
Datepart("hour",[order date (DateOrders)])>=18 and Datepart("hour",[order date (DateOrders)])<22  
then "evening" ELSE
```

```
"night" END
```

6 Visuelle Kodierung



6.1 Layout und Design

Im Kapitel 4.1 wurden die Designentscheidungen des Basic Layouts bereits erläutert. In diesem Kapitel soll es nun um die Begründung der Anordnungen der Visualisierungen gehen.

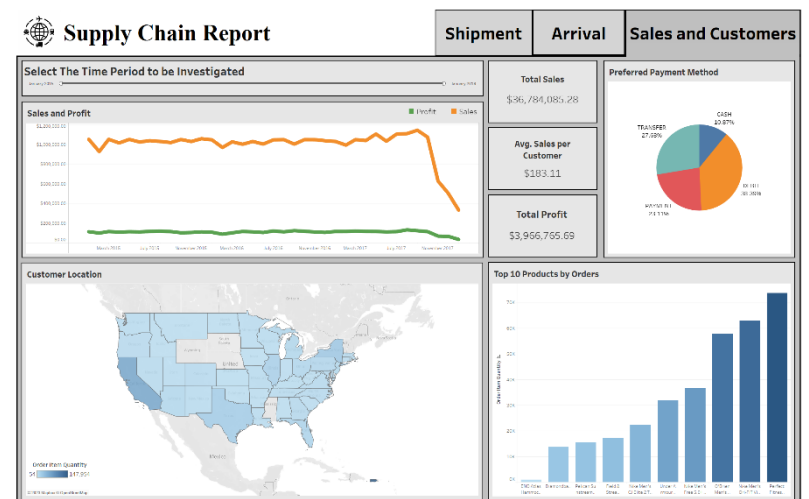
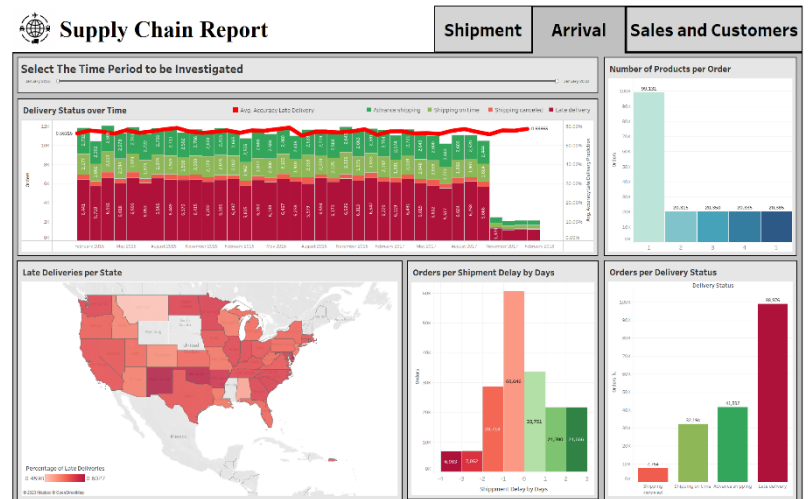
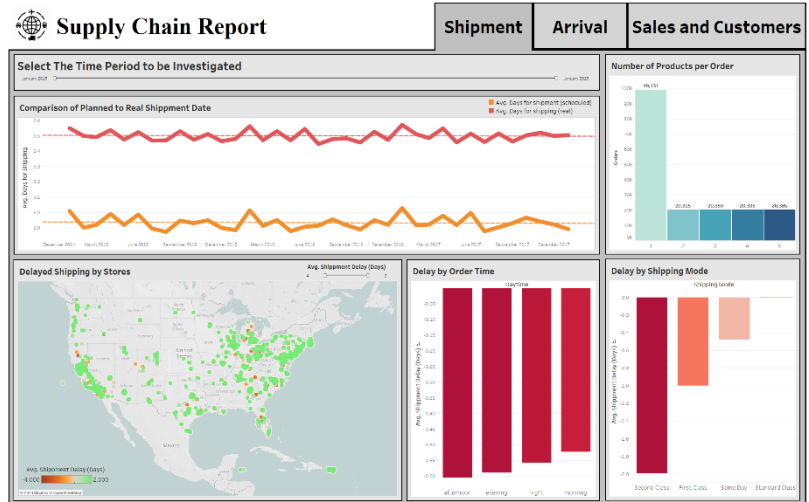
Es wurde darauf geachtet, dass die Visualisierung, die die wichtigsten Sachverhalte darstellt, immer im oberen linken Bereich platziert wurde, da dies oft der als erster fixierte Bereich ist beim erstmaligen Betrachten des Reports.

Die Visualisierungen im rechten Bereich liefern Detailinformationen zum betrachteten Sachverhalt und dienen gleichzeitig als Seitenübergreifende Filtermöglichkeit. Durch diese Visualisierungen ist es möglich Trends und Muster in den Daten intuitiv zu analysieren.

Jede Seite hat zudem eine Karte zusätzliche Informationen in Bezug auf das Seitenthema veranschaulicht. Die Karte gibt außerdem etwas Raum in das Dashboard, so dass es nicht überladen wirkt.

Auf der Seite wurden außerdem noch interaktive Textelemente verwendet um eine schnelle Identifikation von KPI's zu ermöglichen.

Zudem wurde auf ausgefallene Visualisierungen verzichtet, um den Einstieg und die Benutzung des Dashboards zu erleichtern. Darüber hinaus wurde der Report in englischer Sprache angefertigt, da auch die Daten alle in Englisch gehalten wurden.



6.2 Farbpalette

In den Visualisierungen wurde im Grunde mit drei unterschiedlichen Farbvariationen gearbeitet.

Rote Farbtöne sollen dem Benutzer signalisieren, dass es sich um einen schlechten Wert oder eine wichtige Darstellung einer Metrik handelt. Dies ist eine gute Farbwahl, da Rot eine starke Signalwirkung hat und generell eher mit schlechten oder wichtigen Gegebenheiten assoziiert wird.

Grüne Farbtöne repräsentieren genau das Gegenteil der Roten Farbtöne. Sie stellen einen guten Wert oder eine gute Entwicklung dar. Grün ist auch wieder eine Signalfarbe und wurde wegen ihrer positiven Assoziation für diesen Zweck dafür bestimmt.

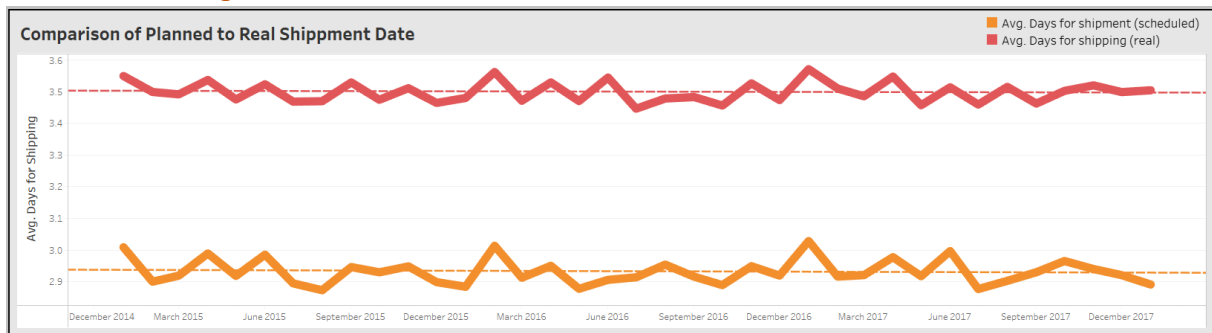
Blaue Farbtöne werden im Report für neutrale Sachverhalte eingesetzt. Die unterschiedlichen Sättigungen werden dafür benutzt, um unterschiedliche Wertebereiche zu verdeutlichen. Neutrale Sachverhalte sind zum Beispiel die Anzahl an Kunden in einem Staat. Dabei wäre eine Farbkodierung der Staaten mit Rot oder Grün irreführend, weil es im Endeffekt neutral zu bewerten ist welcher Staat nun die meisten Kunden beherbergt.

Alle anderen Farbtöne wurden sonst nur dafür benutzt, um Kategorien klar voneinander zu trennen (Siehe Kreisdiagramm).

6.3 Visualisierungsarten

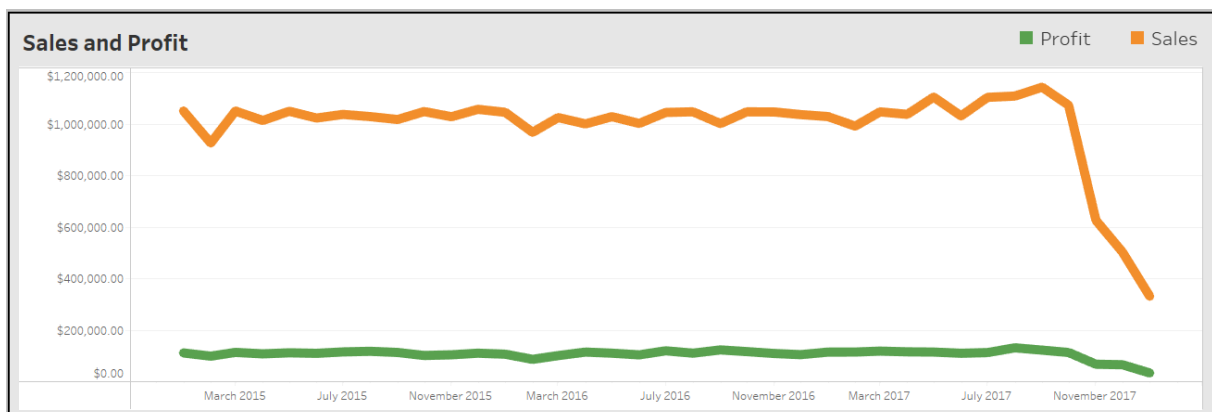
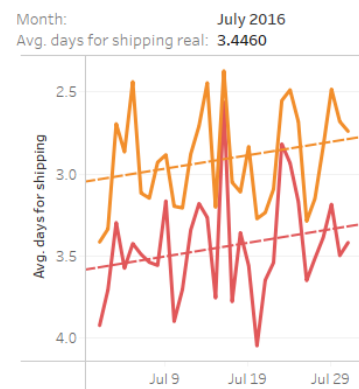
In diesem Kapitel werden die Visualisierungsarten anhand des Frameworks von Tamara Munzner beschrieben. Nach diesem Ansatz werden die Items des Datensatzes als Zeichen verstanden, die von Kanälen ihr Aussehen auf Grundlage von Attributen ändern. Zeichen sind also eine atomare Repräsentation (Punkt, Linie, Fläche) der Werte der Attribute, beeinflusst durch die Kanäle (Position, Form, Größe, Neigung, Farbe).

6.3.1 Liniendiagramme



In diesem Liniendiagramm sind die durchschnittlichen geplanten und realen Versandtage kodiert. Der Verlauf der Tage wird jeweils mittels einer Linie dargestellt. Um die Analyse zu unterstützen, wurde eine Trendlinie hinzugefügt. Da die Aggregation auf der monatlichen Ebene es nicht erlaubt kurzzeitige Spitzen zu identifizieren, wurde dem Tooltip eine detaillierte Tagesansicht hinzugefügt. Diese ist gleich aufgebaut wie das Liniendiagramm auf Monatsebene, nur wird hierbei der gehoverte Monat auf Tagesebene repräsentiert.

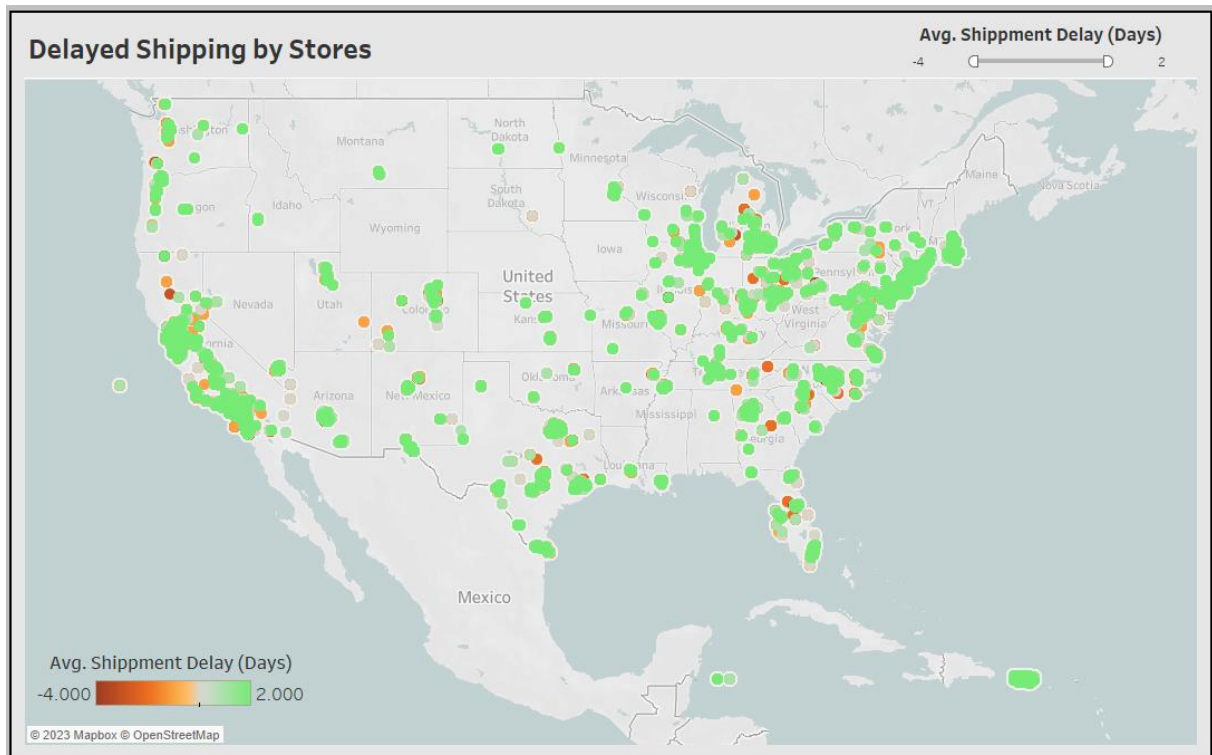
Die Farben der Linien wurden bewusst durch Rottöne gekennzeichnet da es sich um sehr wichtige Kennzahlen handelt.



Dieses Liniendiagramm visualisiert die Entwicklung des Umsatzes und Gewinns des Unternehmens. Der Verlauf der Durchschnittswerte ist wieder durch eine Linie dargestellt. Um die beiden Metriken gut unterscheiden zu können wurden beide Linien mit einer Farbe belegt. Grün für Profit, da dies eine positive und wichtige Metrik für das Unternehmen darstellt. Orange da der Umsatz auch eine Wichtige Kennzahl ist. Auf Rot wurde in diesem Fall verzichtet da sonst der Kontrast zwischen grün und Rot als positive vs. Negativ aufgefasst werden könnte.

Die Linien in beiden Diagrammen wurden zudem vergrößert, um die Lesbarkeit zu erhöhen.

6.3.2 Karten

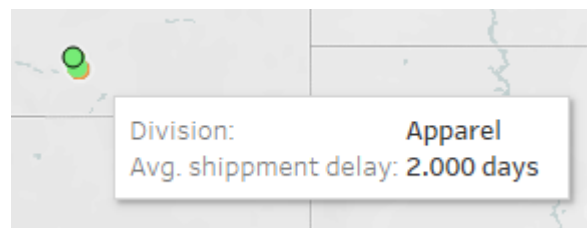


In dieser Karte wird untersucht, ob bzw. welche Filialen im Durchschnitt eine schlechte Versandzeit aufweisen. Dabei wurden die geographischen Daten der Filialen als Punkte auf der Karte dargestellt.

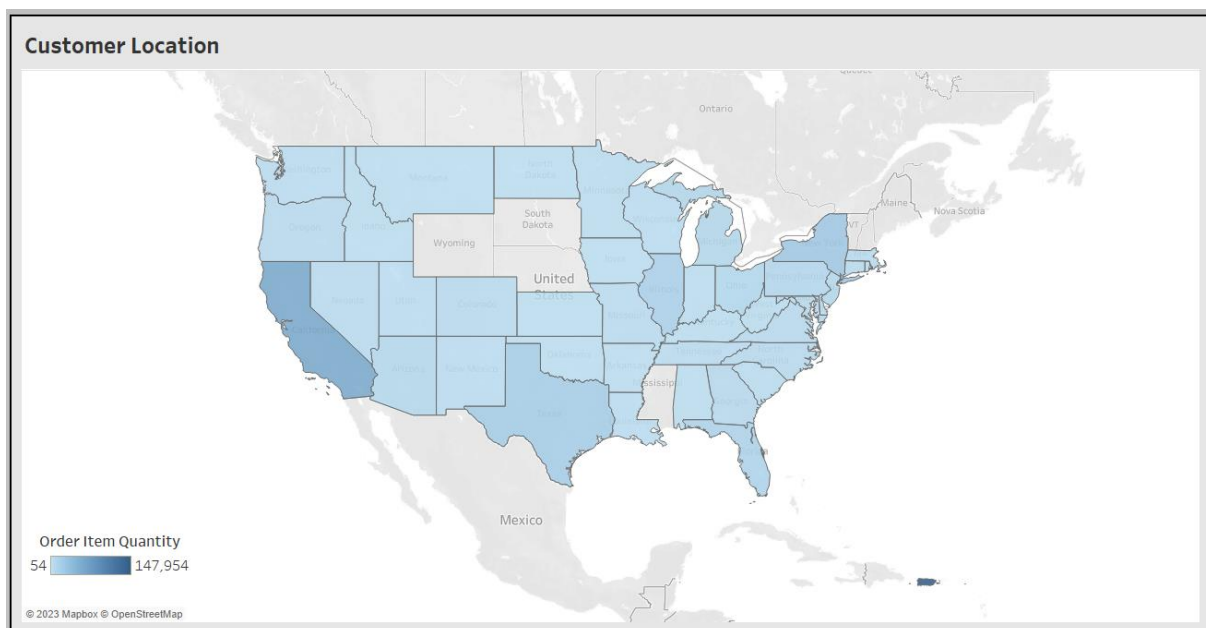
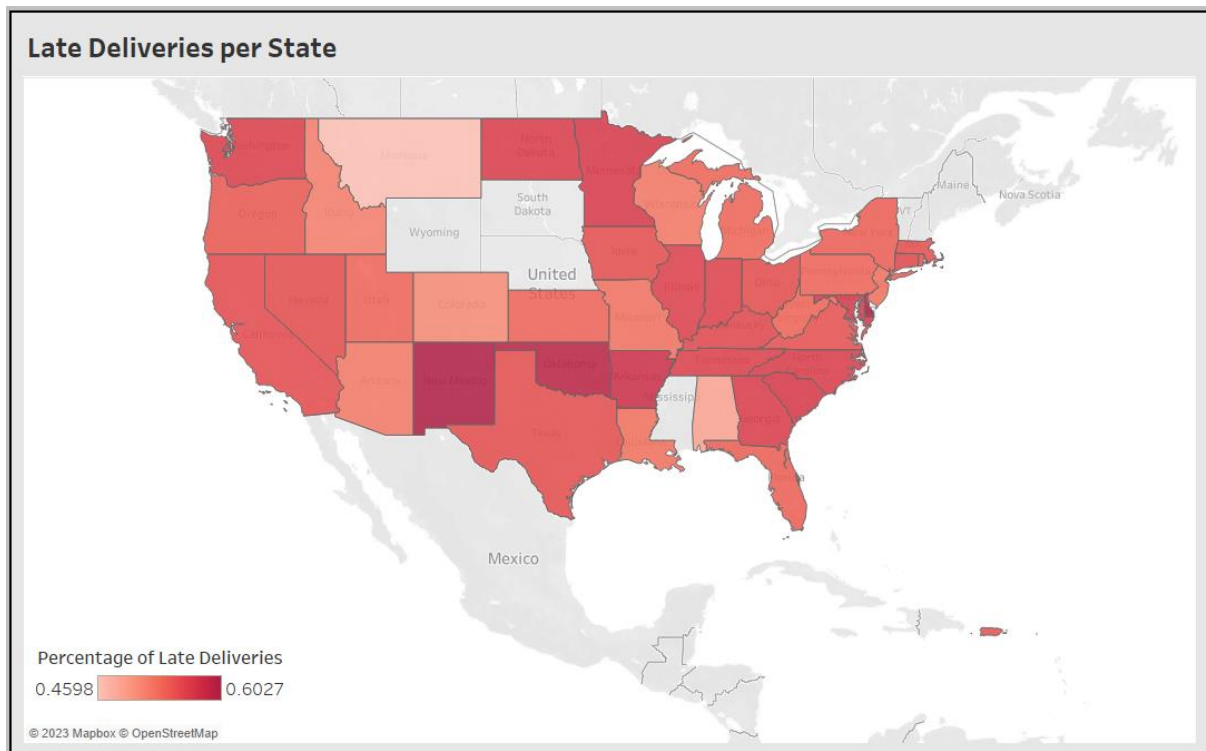
Jeder Punkt ist farblich kodiert, um die Performance der Filiale zu signalisieren. Dabei wurde ein Farbübergang gewählt, der bei dem Nullpunkt ein neutrales Grau zeigt, bei einem durchschnittlichen verspäteten Versand in den roten Farbraum driftet und bei einem verfrühten Versand in den grünen Farbraum. Dies soll eine Erfassung der Lage ermöglichen.

Weitere Informationen wurden dem Tooltip hinzugefügt. In diesem kann man die Sparte der Filiale und die durchschnittlichen Versandtage ermitteln.

Zudem ist es möglich die Karte, mit dem Slider oben rechts, nach den gewünschten durchschnittlichen Versandtagen, zu filtern.



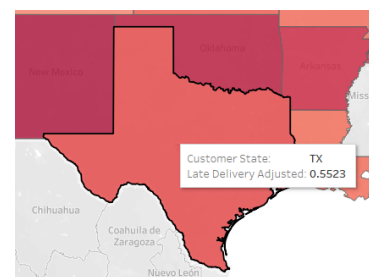
In dieser Karte wurden zudem das Straßennetz und die Stadtnamen mit angezeigt, um eine bessere Orientierung bei der Lokalisierung der Filialen zu bieten.



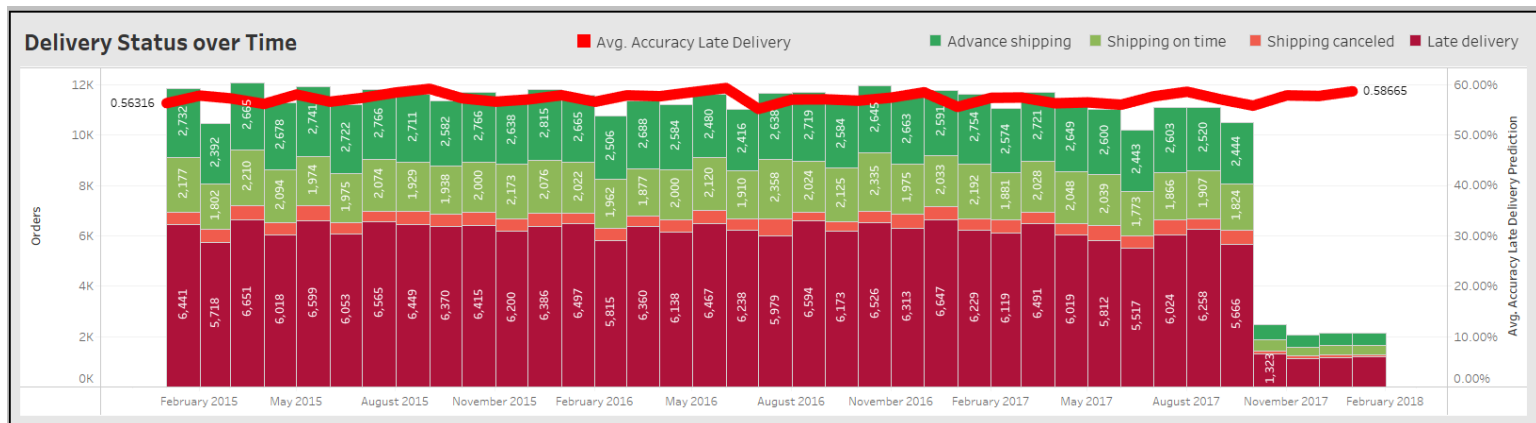
Diese beiden Karten sind relativ gleich aufgebaut. Die erste Karte stellt den Prozentwert der verspäteten Lieferungen im Bezug auf die gesamte Liefermenge dar.

In der zweiten Karte wird die Anzahl der Bestellungen im Kontext der einzelnen Staaten betrachtet.

In beiden Karten wurde jeweils der Sachverhalt durch Farblische die Saturierung der Staaten dargestellt. Dabei wurde in der ersten Karte ein Rotton gewählt, da es sich um einen negativen Sachverhalt handelt. In der zweiten Karte wurde ein Blauton verwendet, da dieser nur die Häufigkeit ausdrücken soll ohne implizite Bewertung. Tooltips wurden wieder zur genauen Betrachtung der einzelnen Werte eingesetzt.



6.3.3 Balkendiagramme und Histogramme

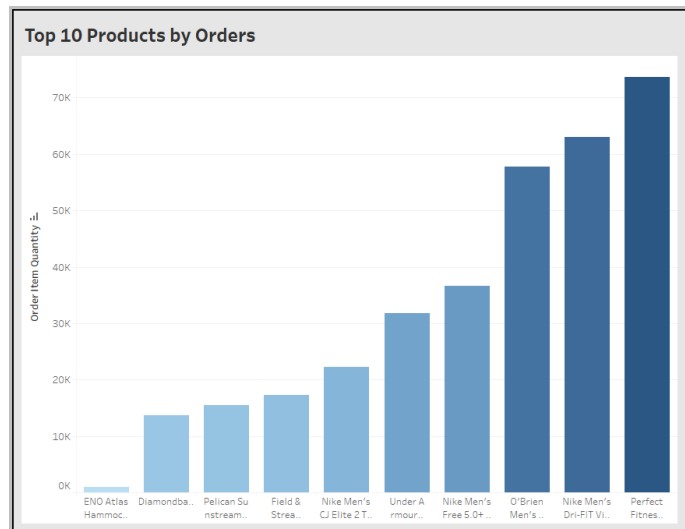


Dieses gestapelte Balkendiagramm, in Kombination mit einem Liniendiagramm, ist bei weitem die Visualisierung mit der höchsten Informationsdichte im ganzen Report.

In dieser Visualisierung wird das Verhältnis der verschiedenen Lieferstatus mithilfe eines gestapelten Balkendiagramms dargestellt. Die Höhe des gesamten Balkens signalisieren dabei die Gesamtzahl der Bestellungen für den betrachteten Monat. Die Höhe der individuellen Balken geben wiederum Aufschluss darüber, wie sich die Gesamtzahl der Bestellungen in die einzelnen Lieferstatus aufschlüsselt. Um die Informationsverarbeitung des Nutzers zu unterstützen, wurden die einzelnen Balken nach dem gängigen Farbschema (rot, schlecht und grün, gut) gekennzeichnet. Um die Analyse und die Informationsaufnahme zu erleichtern, wurde die Bestellanzahl den Balken hinzugefügt.

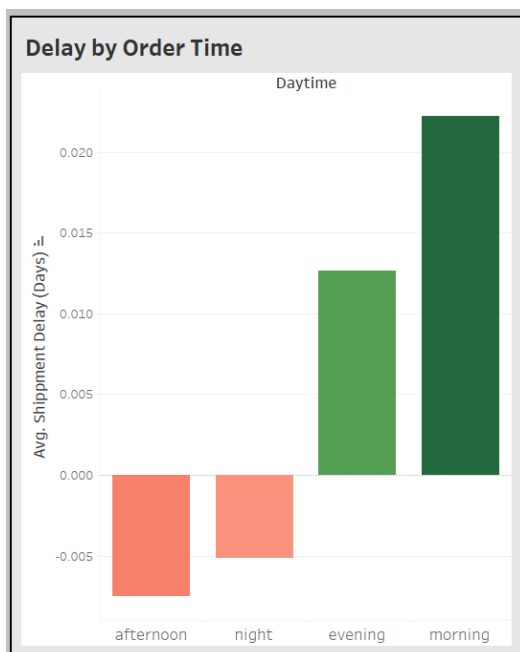
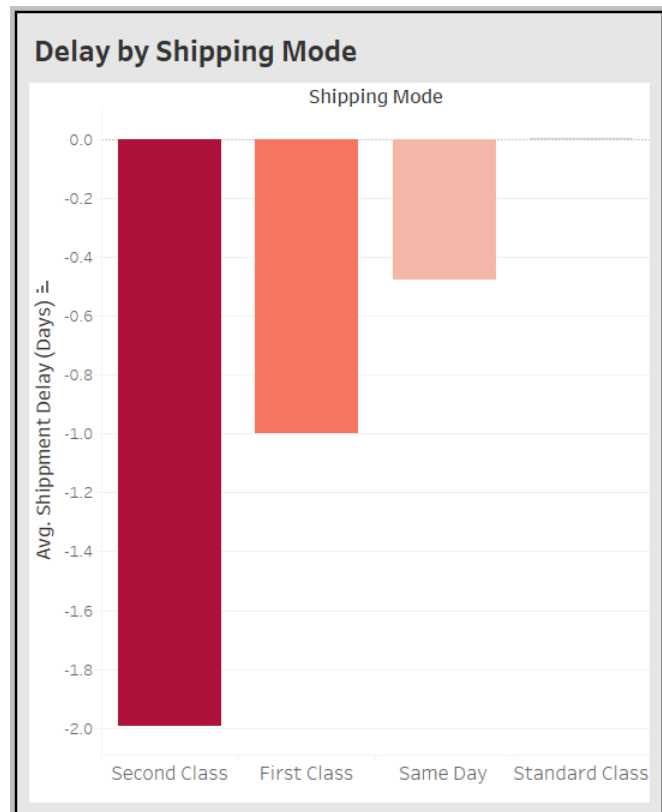
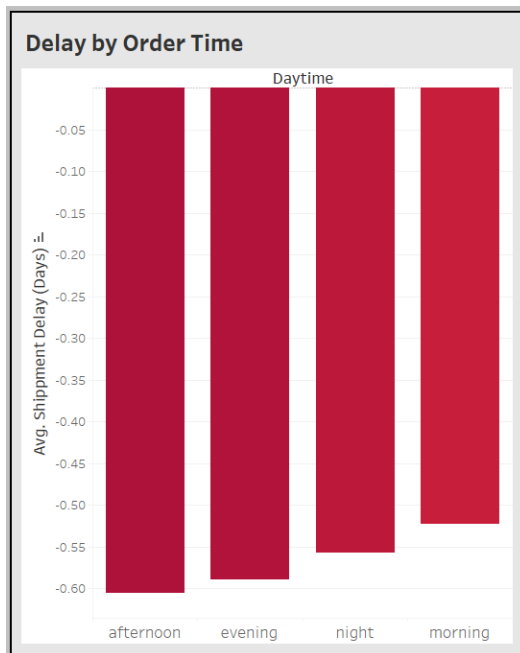
Das Liniendiagramm wiederum stellt die Genauigkeit der Vorhersage einer verspäteten Lieferung dar. Dies wurde in dieser Visualisierung verankert, um ein einheitliches Layout zu ermöglichen, ohne diese wichtige Metrik zu verlieren. Zudem ist das die einzige Visualisierung, wo dies möglich ist, da alle anderen Visualisierungen keinen Bezug auf den Zeitraum nehmen.

Um die Linie von den Balken abzuheben, wurde ein knalliges Rot gewählt. Diese Farbwahl passt auch zur Signalisierung der Wichtigkeit dieser Metrik.



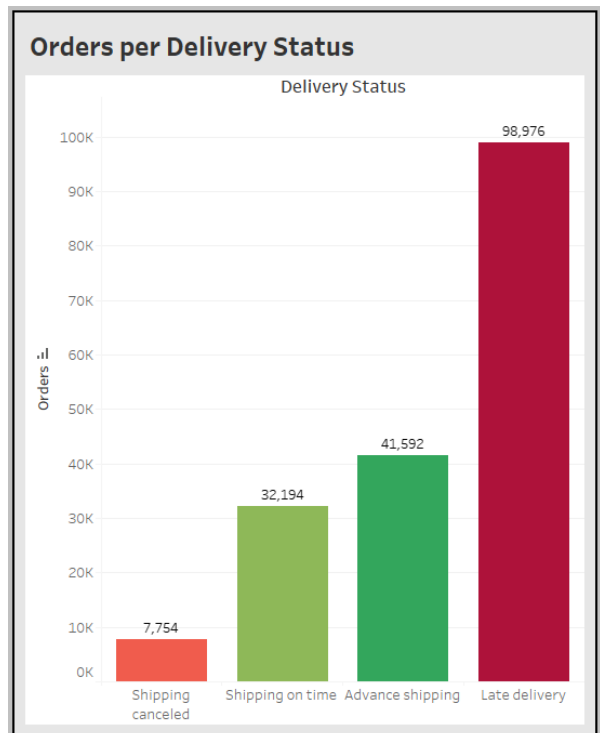
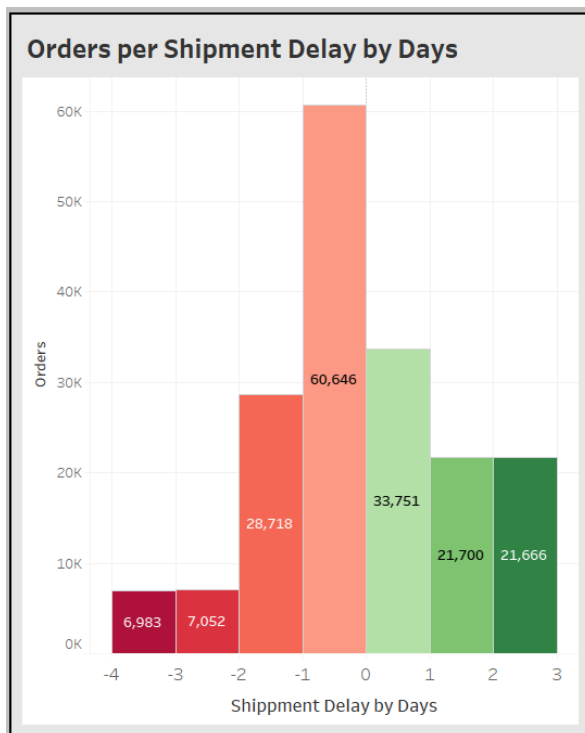
Diese Balkendiagramme sind ähnlich in ihrer Kodierung. Im linken Histogramm wird die Anzahl der Bestellungen mit der Anzahl der jeweiligen Produkte in den Bestellungen visualisiert. Dabei signalisiert die Höhe der Balken die Anzahl der Bestellungen. Um den Fokus auf die Anzahl zu legen wurde die steigende Anzahl mit einem dunkler werdenden Blauton gekennzeichnet. Zur Unterstützung der Informationsaufnahme wurde die Anzahl der jeweiligen Bestellungen über die Balken geschrieben.

Im Balkendiagramm rechts werden die Top 10 Produkte nach Bestellungen visualisiert. Die Höhe der Balken signalisiert die Anzahl der verkauften Produkte. Um dies übersichtlich zu gestalten wurden die Balken aufsteigend geordnet. Zudem wurde ein dunkel werdender Blauton zur doppelten Kodierung des Sachverhaltes genutzt. Dies soll jeweils die Informationsverarbeitung unterstützen.



Diese beiden Balkendiagramme haben wieder viele Gemeinsamkeiten. Da jeweils die Verspätung der Lieferungen im Bezug auf die Lieferart und den aggregierten Zeitpunkt der Bestellung betrachtet wird, können die Balken nach oben, ins positive oder nach unten, ins Negative Schwingen. Die Länge Der Balken repräsentiert dabei wieder die Anzahl der Tage und sorgt für eine intuitive Unterscheidung.

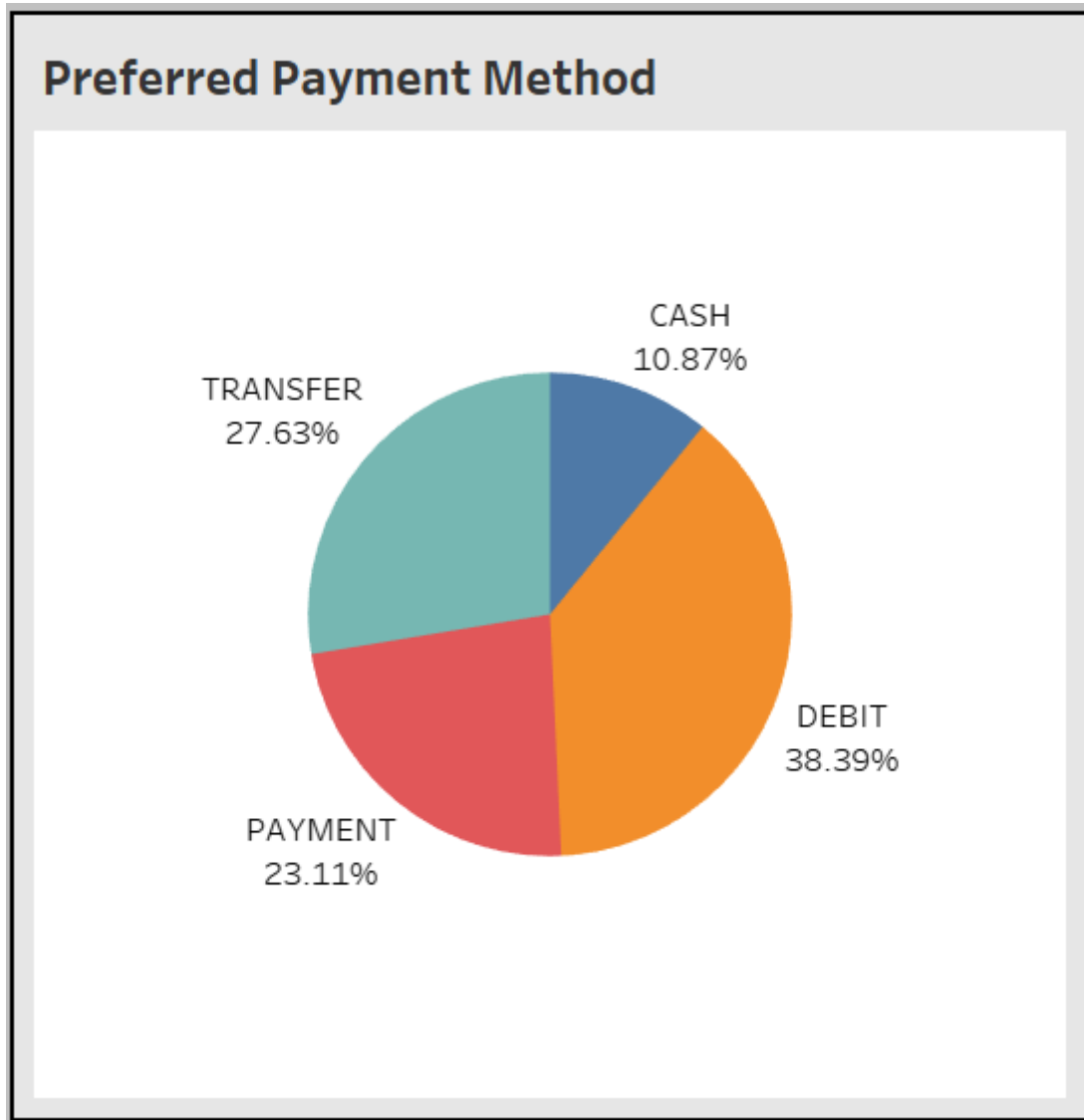
Um den Verzug oder die überpünktliche Lieferung zu signalisieren wurden wieder Farben eingesetzt. Der Verlauf von einem Rotton zu einem Grünton ist dabei wieder passend, da dieser intuitiv als negativ zu positiv verstanden wird.



Genau wie bei den ersten betrachteten Balkendiagrammen dieses Kapitels handelt es sich links wieder um ein Histogramm und rechts um ein Balkendiagramm. Der Unterschied hierbei ist jedoch, dass die dargestellten Sachverhalte in negativ und positiv unterschieden werden können. Um diese Unterscheidung deutlich zu machen, wurde wieder mit einem Farbverlauf von rot zu grün gearbeitet.

Ansonsten sind diese Balkendiagramme genau gleich kodiert wie ihre Vorgänger.

6.3.4 Kreisdiagramm

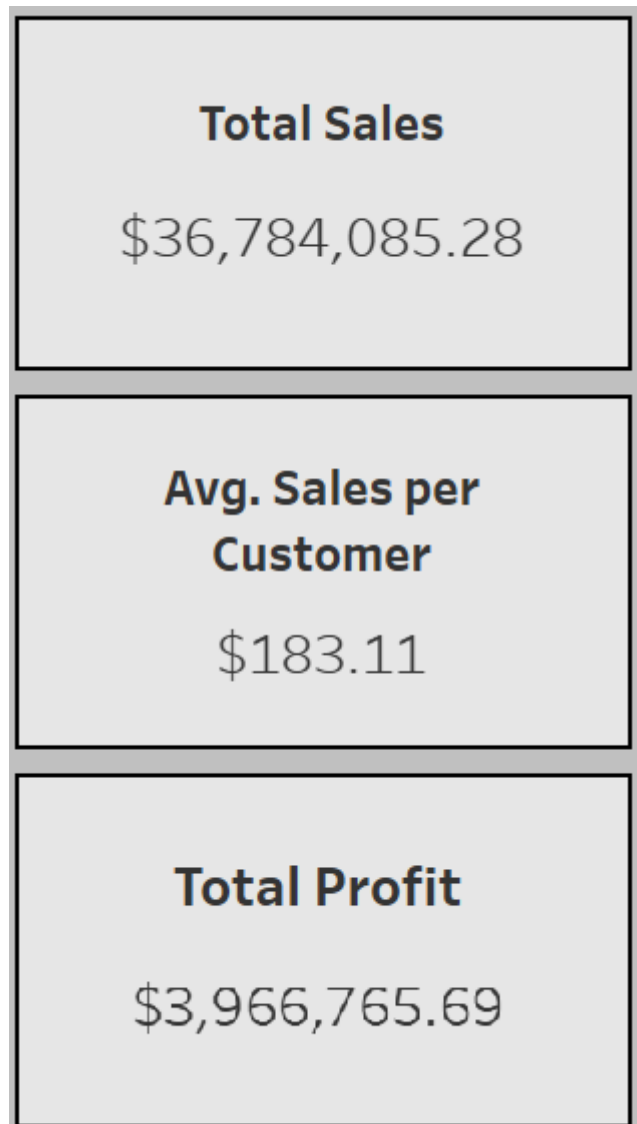


Um einen groben Überblick über die Verteilung der verwendeten Bezahlmethoden zu bekommen wurde ein Kreisdiagramm gewählt. Ein Kreisdiagramm bietet sich an, da nur vier kategorische Variablen dargestellt werden sollen. Dabei wurden mit kräftigen Farben die einzelnen Kategorien unterschieden. Um den Informationsgewinn zu unterstützen, wurde zu der Beschriftung noch der Prozentsatz hinzugefügt.

6.3.5 Textfelder

Um die schnelle Informationsgewinnung zu unterstützen, wurde in der Seite "Sales and Customers" mit Textkarten gearbeitet.

Man hätte diese Werte zwar im Liniendiagramm darstellen können, wenn man sie jedoch noch einmal explizit einzeln darstellt, steigert das die schnelle Informationsgewinnung deutlich. Zudem gibt es keine geeignete Visualisierung, um einen einzelnen Wert darzustellen, weswegen einfach auf einen ganz normalen Text zurückgegriffen wurde.



7 Bewertung des Dashboards



7.1.1 Validierung und Algorithmus

Im letzten Schritt „Algorithmus“ wird untersucht ob die erstellten Visualisierungen performant zu bedienen sind.

Beim Testen des Dashboards in Tableau wurden nur geringe Verzögerungen bemerkt. Es müssten weitere Untersuchungen durchgeführt werden ob diese Verzögerungen durch eine geringe Leistung des Endgeräts stammen oder ob die Datenlage weiter ausgedünnt werden kann, um die Performanz zu erhöhen.

Zur Validierung wurden versucht die zuvor definierten Businessfragen mit dem Tableau-Dashboard zu beantworten. Dies konnte für alle Fragen erfolgreich durchgeführt werden.

7.1.2 Fazit

Durch dieses Dashboard sind die Anwender in der Lage die Versand- und Liefersituation zu überwachen. Bei negativer Entwicklung der Lage kann dieses Dashboard zudem eine große Hilfe in der Ursachenanalyse sein. Es bietet eine Vielzahl von Filtermöglichkeiten und Betrachtungen der Vorgänge auf diversen Ebenen. Zudem kann es bei Bedarf noch deutlich erweitert werden, da die Datenlage noch nicht ausgereizt ist. Unter diesem Gesichtspunkt wurden die Daten im Tableau-Report auch nicht weiter aussortiert.