Group 10 – Dream Five

Group Project

Analytic Tools & Decision Making

Group Members:
1) Devashish Patel
2) Ninad Patil
3) Reshma Thomas
4) Sahiba Gadhok
5) Tanvi Pal

PROPOSAL

Problem Statement

Supply chain management is a complicated and tedious process. It takes a lot of time and effort to keep track of all the activities and other important things of the company especially, if all data is present in a rectangular form only. From this data, it becomes very difficult to understand what your problem areas are and what needs to be improved. But, with our analysis of this data it will become increasingly simpler to understand what the KPI's (Key Performance Indicators) are and how they are faring. This project is aimed at simplifying the work of the entire supply chain management department, help extract valuable insights to act upon and at the same time understand how the supply chain of a particular company works.

Stakeholders that benefit from this analysis

1) The analysis of this data set of supply chain used by DataCo Global, will be beneficial for its Top Management to find out which products are profitable and which products are underperforming.
    - Depending on the above analysis, the top executives can decide some important things like which products' pricings should be changed depending on its current performance, whether a particular product must be discontinued, whether a particular product needs any changes or if the demand for the product is not being met with appropriately.
2) Reviewing the dataset will also help the supply chain department to monitor their distribution network. They can work on finding the flaws in the distribution network and developing innovative ways to overcome those flaws, ensure quality customer service and guarantee customer satisfaction.

Significance of this project

A good supply chain is one that ensures timely delivery of the products or services to the consumers in the least time while maintaining quality of the product. Hence, it is crucial for the success of a company to ensure everything in the chain is working smoothly and all standard operating procedures are being followed to the letter. Every seller's primary vision is customer satisfaction, minimisation of operation costs and improving the financial standing of the company.
The project is important in order to make sure that all the above goals are satisfied efficiently and swiftly. Our analysis can help to identify which country/state's need of goods are unmet, which products are having the highest sales, which area is the most profitable, what is the delivery time for all the cities, and how many orders have been cancelled by the customers? Which customers are potential clients and the ones that are most beneficial, which customers should be targeted separately, which shipping mode has to adapt as per the requirement for quick and safe delivery and what should be the minimum product price which can be offered to customers according to their location and their needs? What is the time at which the customers are most active and what is the highest revenue generating time of the day? This analysis will help other companies as well to enhance their business, achieve customer satisfaction, track customer demands, generate more sales and acquire new customers, and what are the corrective measures that should be taken to do so.

Business Prospects and Objectives of this project

Companies these days are relying much more on their data than ever before to understand and learn from past as well as current data. An efficient supply chain can ensure competitive advantage and better performance, hence ensuring global success.
This will include the following aspects:

1) Forecasting Future Demand
2) Process and Fault Correction

3) Logistics Improvement

1. Based on the performance of particular product(s), the necessary authorities can figure out where to re-focus their efforts by regulating the product release into the market based on the product's relative success. Based on these changes and the future demand, companies will have more information to make well informed decisions with respect to new product releases and any new investments. Balance between supply and demand will ensure that clients are satisfied, and goodwill is maintained in the market for prompt delivery of services with lesser lead times. This will help the company **save millions if not more** in the long run.

2. Using the analysis will help to point at possible **fault correction** and improve them accordingly for example, a particular product's supply chain might not be as efficient as it can be and that could be dealt with correctly by the analysis provided. This will help the company save time and effort into correcting the issues and deal with any and all irregularities that are long existing and or are ones that could come up in the future.

3. Ease in tracking and material handling, movement of inventory can be improved based on this analysis. This will improve customer satisfaction and help in acquiring new customers easily.

Dataset Selection Criteria

Before finding any dataset, we decided on 3 major criteria that our dataset should satisfy:

- There should be at least 2 categorical variables and 3 numeric variables. This is because it would ensure that we have enough variables to work on and to analyse the dataset efficiently and show off our analysis properly. The idea for this sort of criterion was inspired from our project 1 in DAB 501. This was the main criterion for selecting our dataset.
- The dataset should have sufficiently considerable number of observations to show our analysis properly and to deal with inconsistencies in the dataset, if any.
- We should have just enough domain knowledge about the subject of the dataset in order to show our analysis properly and efficiently and at the same time learn a lot of things from it.

We have decided to use **DataCo supply Chain** dataset for our project.

Data Collection Process

From the links provided in the project instructions pdf, the link which consequently led to the discovery of this dataset was the one given below.

https://www.kaggle.com/datasets

Once here, we applied some filters to find only those datasets which are relevant to our needs. The filters used were:

- Size of the dataset (7 MB to 5 GB)
- Most viewed or used datasets.

Finally, we tried to figure out which kind of businesses are handling or working on big data, and then based on some research we zeroed in on supply chain as one of the fields which could be discovered. After looking through some companies, we decided to use the dataset of DataCo Global company named as DataCo Smart Supply Chain for Big Data Analysis.

The link for the same is given below:

https://www.kaggle.com/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis

Data Description

This dataset contains data of supply chains for DataCo Global company. This dataset belongs data of different types of products namely clothing, sports, and electronic supplies. It includes the data of the path of provisioning, production, sales, and commercial distribution methods for these products.

Original Dataset Description of Variables

There are a total of 53 columns in this dataset and the description of each has been given below:

FIELDS	DESCRIPTION
Type	Type of transaction made
Days for shipping (real)	Actual shipping days of the purchased product
Days for shipment (scheduled)	Days of scheduled delivery of the purchased product
Benefit per order	Earnings per order placed
Sales per customer	Total sales per customer made per customer
Delivery Status	Delivery status of ordersAdvance shipping , Late delivery , Shipping canceled , Shipping on time
Late_delivery_risk	Categorical variable that indicates if sending is late (1), it is not late (0).
Category Id	Product category code
Category Name	Description of the product category
Customer City	City where the customer made the purchase
Customer Country	Country where the customer made the purchase
Customer Email	Customer's email
Customer Fname	Customer name
Customer Id	Customer ID
Customer Lname	Customer lastname
Customer Password	Masked customer key
Customer Segment	Types of CustomersConsumer , Corporate , Home Office
Customer State	State to which the store where the purchase is registered belongs
Customer Street	Street to which the store where the purchase is registered belongs
Customer Zipcode	Customer Zipcode
Department Id	Department code of store
Department Name	Department name of store
Latitude	Latitude corresponding to location of store
Longitude	Longitude corresponding to location of store
Market	Market to where the order is delivered Africa , Europe , LATAM , Pacific Asia , USCA
Order City	Destination city of the order
Order Country	Destination country of the order
Order Customer Id	Customer order code
order date (DateOrders)	Date on which the order is made
Order Id	Order code
Order Item Cardprod Id	Product code generated through the RFID reader
Order Item Discount	Order item discount value
Order Item Discount Rate	Order item discount percentage
Order Item Id	Order item code
Order Item Product Price	Price of products without discount
Order Item Profit Ratio	Order Item Profit Ratio
Order Item Quantity	Number of products per order
Sales	Value in sales
Order Item Total	Total amount per order
Order Profit Per Order	Order Profit Per Order
Order Region	Region of the world where the order is delivered: Southeast Asia ,South Asia ,Oceania ,Eastern Asia, West Asia , West of USA , US Center , West Africa, Central Africa ,North Africa ,Western Europe ,Northern , Caribbean , South America ,East Africa ,Southern Europe , East of USA ,Canada ,Southern Africa , Central Asia , Europe , Central America, Eastern Europe , South of USA
Order State	State of the region where the order is delivered
Order Status	Order Status: COMPLETE , PENDING , CLOSED , PENDING_PAYMENT ,CANCELED , PROCESSING ,SUSPECTED_FRAUD ,ON_HOLD ,PAYMENT_REVIEW
Product Card Id	Product code
Product Category Id	Product category code
Product Description	Product Description
Product Image	Link of visit and purchase of the product
Product Name	Product Name
Product Price	Product Price
Product Status	Status of the product stock :If it is 1 not available , 0 the product is available
Shipping date (DateOrders)	Exact date and time of shipment
Shipping Mode	The following shipping modes are presented Standard Class , First Class , Second Class , Same Day

Original Source of this Data

This dataset was taken from the website of Mendeley Data. The link for the same is given below:
https://data.mendeley.com/datasets/8gx2fvg2k6/5

The authors of this dataset are:
Fabian Constante, António Pereira and Fernando Silva

Descriptive Statistics and EDA

Summary Statistics

We have **identified 35 important variables** from the original dataset, and these are the variables that we will be using for our project. There are **1,80,519 rows and 35 columns** in this **dataset**.
The summary statistics like mean, median, standard deviation and IQR are given below along with other statistics. This data is imported from python from a Jupyter notebook.

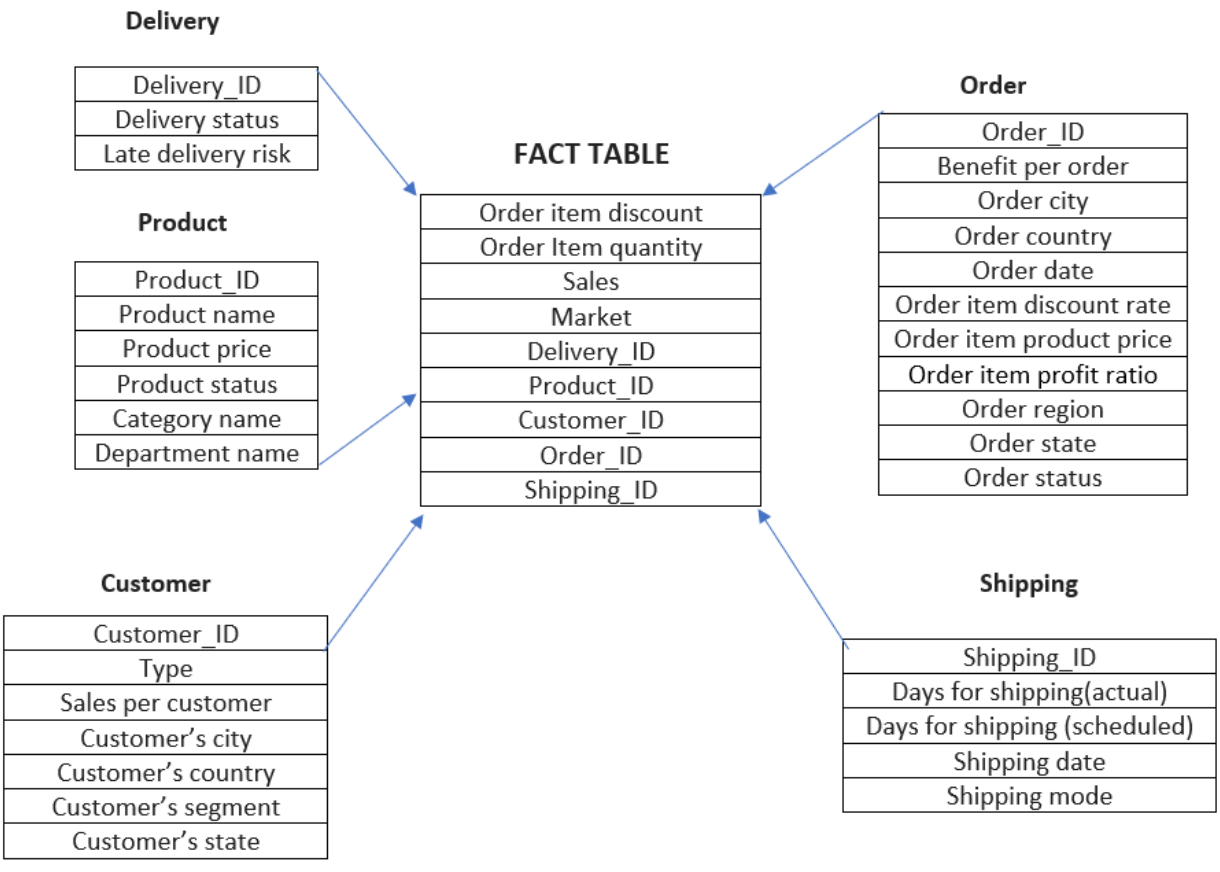
	count	mean	std	min	1st Quartile	Median	3rd Quartile	max	IQR
Days for shipping (real)	180519	3.497654	1.623722	0	2	3	5	6	3
Days for shipment (scheduled)	180519	2.931847	1.374449	0	2	4	4	4	2
Benefit per order	180519	21.974989	104.433526	-4274.97998	7	31.52	64.800003	911.799988	57.800003
Sales per customer	180519	183.107609	120.04367	7.49	104.379997	163.990005	247.399994	1939.98999	143.019997
Late_delivery_risk	180519	0.548291	0.497664	0	0	1	1	1	1
Latitude	180519	29.719955	9.813646	-33.937553	18.265432	33.144863	39.279617	48.781933	21.014185
Longitude	180519	-84.915675	21.433241	-158.025986	-98.446312	-76.847908	-66.370583	115.263077	32.075729
Order Item Discount	180519	20.664741	21.800901	0	5.4	14	29.99	500	24.59
Order Item Discount Rate	180519	0.101668	0.070415	0	0.04	0.1	0.16	0.25	0.12
Order Item Product Price	180519	141.23255	139.732492	9.99	50	59.990002	199.990005	1999.98999	149.990005
Order Item Profit Ratio	180519	0.120647	0.466796	-2.75	0.08	0.27	0.36	0.5	0.28
Order Item Quantity	180519	2.127638	1.453451	1	1	1	3	5	2
Sales	180519	203.772096	132.273077	9.99	119.980003	199.919998	299.950012	1999.98999	179.970009
Product Price	180519	141.23255	139.732492	9.99	50	59.990002	199.990005	1999.98999	149.990005

Type of Variables in the Working Dataset

Given below is the description of the same along with the data type of the variables:

Variable Name	**DESCRIPTION**	**Type of Variable**
Type	Type of transaction made	Categorical: Nominal
Days for shipping (real)	Actual shipping days of the purchased product	Numeric: Discrete
Days for shipment (scheduled)	Days of scheduled delivery of the purchased product	Numeric: Discrete
Benefit per order	Earnings per order placed	Numeric: Continuous
Sales per customer	Total sales per customer made per customer	Numeric: Continuous
Delivery Status	Delivery status of orders: Advance shipping , Late delivery , Shipping canceled , Shipping on time	Categorical: Nominal
Late_delivery_risk	Categorical variable that indicates if sending is late (1), it is not late (0).	Categorical: Binary
Category Name	Description of the product category	Categorical: Nominal
Customer City	City where the customer made the purchase	Categorical: Nominal
Customer Segment	Type of Customer: Consumer , Corporate , Home Office	Categorical: Nominal
Department Name	Department name of store	Categorical: Nominal
Latitude	Latitude corresponding to location of store	Numeric: Continuous
Longitude	Longitude corresponding to location of store	Numeric: Continuous
Market	Market to where the order is delivered Africa , Europe , LATAM , Pacific Asia , USCA	Categorical: Nominal
Order City	Destination city of the order	Categorical: Nominal
Order Country	Destination country of the order	Categorical: Nominal
Order date (DateOrders)	Date on which the order is made	Date
Order time	Time at which the order is made	Time
Order Item Discount	Order item discount value	Numeric: Continuous
Order Item Discount Rate	Order item discount percentage	Numeric: Continuous
Order Item Product Price	Price of products without discount	Numeric: Continuous
Order Item Profit Ratio	Order Item Profit Ratio	Numeric: Continuous
Order Item Quantity	Number of products per order	Numeric: Discrete
Sales	Value in sales	Numeric: Continuous
Order Region	Region of the world where the order is delivered: Southeast Asia ,South Asia ,Oceania ,Eastern Asia, West Asia , West of USA , US Center , West Africa, Central Africa ,North Africa ,Western Europe ,Northern , Caribbean , South America ,East Africa ,Southern Europe , East of USA ,Canada ,Southern Africa , Central Asia , Europe , Central America, Eastern Europe , South of USA	Categorical: Nominal
Order State	State of the region where the order is delivered	Categorical: Nominal
Order Status	Order Status: COMPLETE , PENDING , CLOSED , PENDING_PAYMENT ,CANCELED , PROCESSING ,SUSPECTED_FRAUD ,ON_HOLD ,PAYMENT_REVIEW	Categorical: Nominal
Product Name	Product Name	Categorical: Nominal
Product Price	Product Price	Categorical: Nominal
Shipping date (DateOrders)	Exact date of shipment	Date
Shipping time	Exact time of shipment	Time
Shipping Mode	The following shipping modes are presented Standard Class , First Class , Second Class , Same Day	Categorical: Nominal

Schema:



From the data, we can analyze which **Market** is having a huge demand of specific products for all categories. It may be Health and Science, Clothing, Sports supplies, Electronics or Fitness Accessories, etc. According to that, we will determine the Sales performance, and accordingly understand what the profit margins are for each department, market region, customer segment and customer region.

We are going to analyze the most active **Customer Segments** in a particular market or markets. Customer segment consists of Consumer, Corporate and Home Office. So, we will determine which customer segment is the most revenue generating segment and which products are in high demand.

The **Department** categories comprises of Technology, Apparel, Health and Beauty, Golf, Pet shops, and so on, which will help us analyze the demand for the products coming from those departments. Accordingly figure out which is the best department based on the total sales and which departments need to improve.

We can analyze the difference between **Actual** shipping days and the **Scheduled** shipping days of the purchased product to try and figure out what is the reason for the delay and what needs to be corrected. From this, we can extract information regarding the status of the orders that is whether they are on time or delayed, through which shipping modes are they being shipped and in which customer order regions are always getting late deliveries. This is critical as this affects customer satisfaction and affects the goodwill of the company brand.

These are some of the variables that we have mentioned here, but we have included all the possible analysis plots in the analysis section of this project.

Inconsistencies in the Data (Missing, incomplete or invalid records)

The details of the inconsistencies are given below:

- Product description (data is missing in the entire column)
- Customer Country (Only 2 countries are mentioned even though this is the global data of the company and one of these 2 is EE. UU. which does not make any sense.)
- Date formats are different in both the date columns of the dataset and need to be corrected:
    1. Order Date (Date Orders)
    2. Shipping Date (Date Orders)

Data Segmentation

A majority of our data is segmented into groups in the dataset. The segments are given below:

Customer country → Customer state → Customer city → Customer Street
Customer details are a key factor in the supply chain management system. In our data set, the customer details have been displayed by using details like Customer country, state, city and customer street. They can be considered to be a group of one category of data.

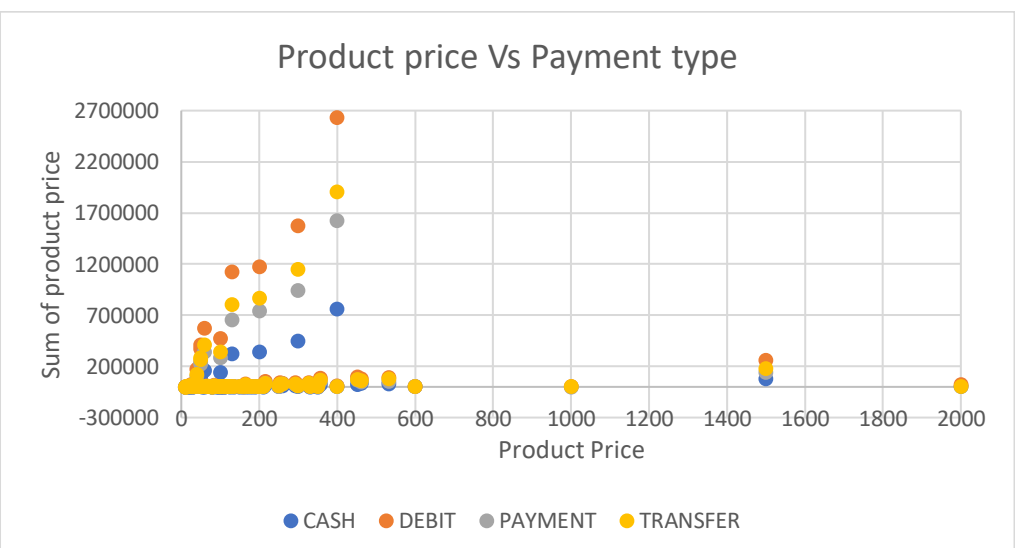
Department Name → Category Name → Product Name.
Each product has been allocated under a different category and each category is handled by different departments.

Market > Order Region > Order country > Order State > Order City
In this dataset, the order details are displayed in the form of market, region, country, state and city in descending order of location.

Correlation between different Variables

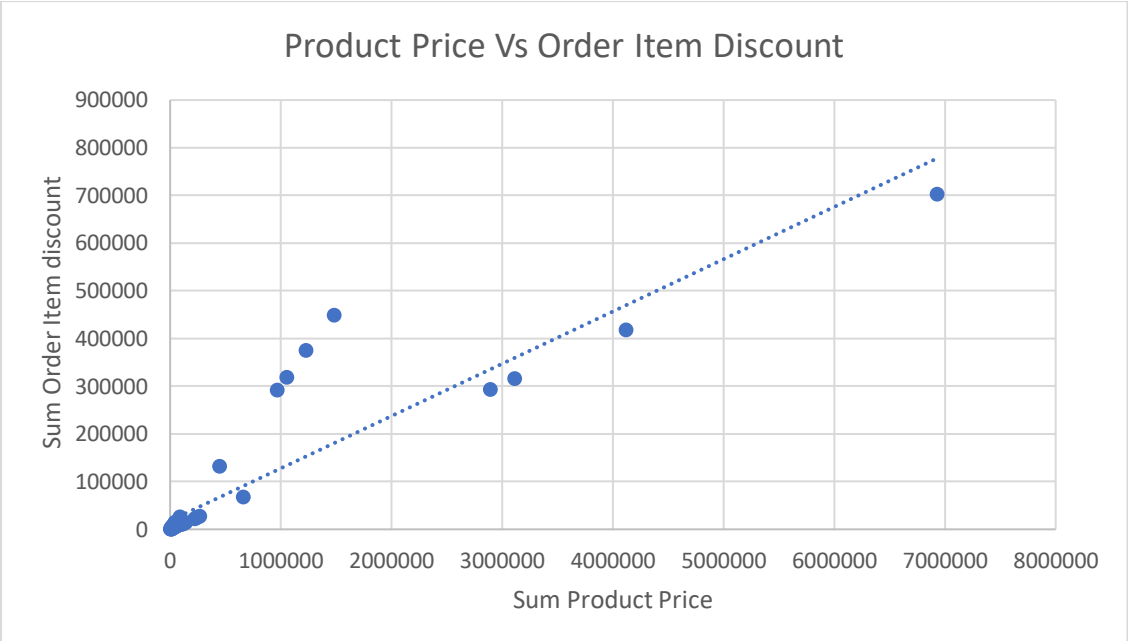
A lot of variables are correlated with each other. Some of them have been shown here with the help of plots:

1) Product Price VS Type of Payment
If the price of the product is high, people prefer Debit and Transfer as mode of payments, on the other hand, if the price is less then they prefer Cash payment or Online Transfer.



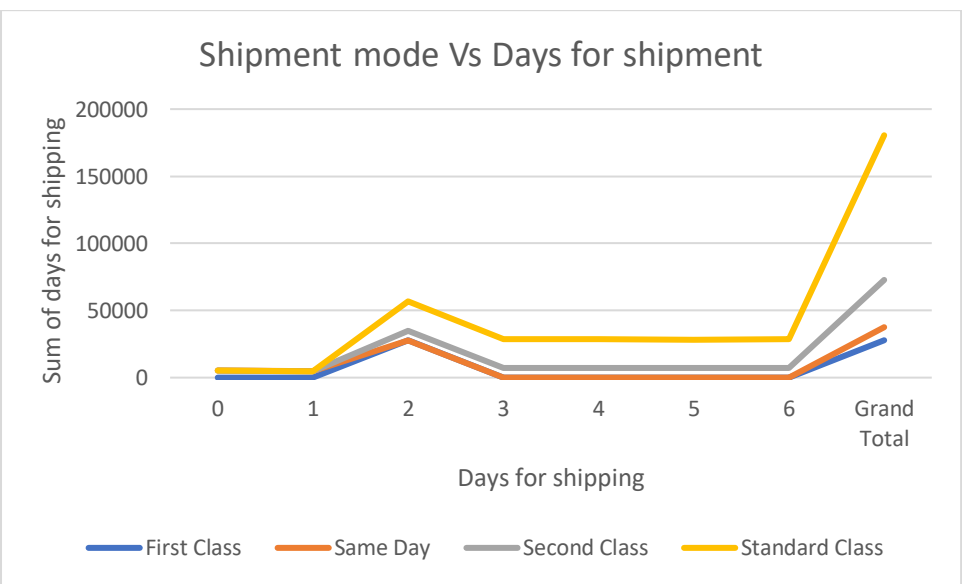
2) *Product Price VS Order Item Discount*

If the price of the product is high, it looks like the order item discount is also high.



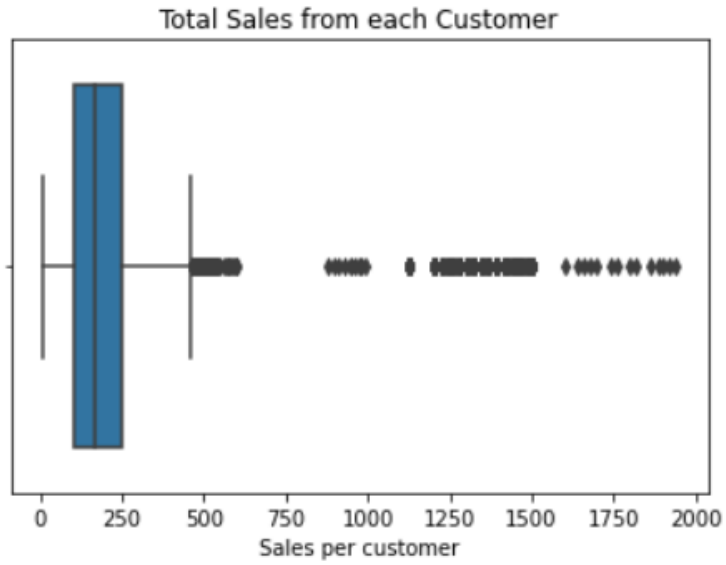
3) *Shipping Mode and Days of Shipping (Real)*

First class and same day shipping mode have an estimated shipping time of 1 day only, and for other shipping modes the estimated shipping days vary from 2 to 6 days.

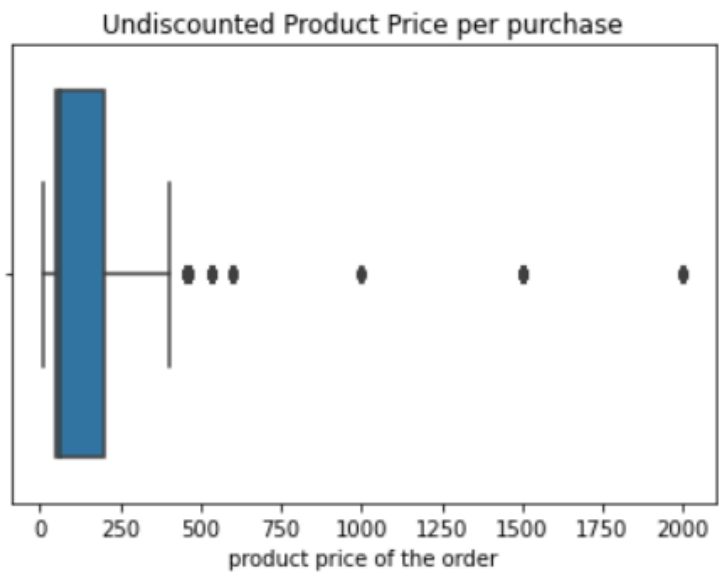


Outliers in the Plots

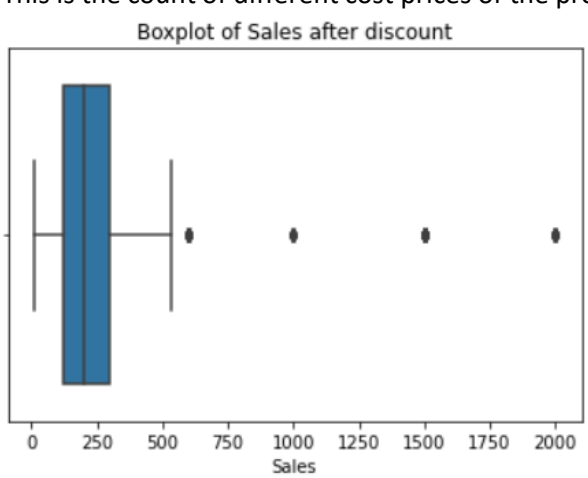
Examples of some of the plots which have outliers are shown below with the help of boxplots.



This is the frequency of sales per customer.



This is the count of different cost prices of the product of the order.

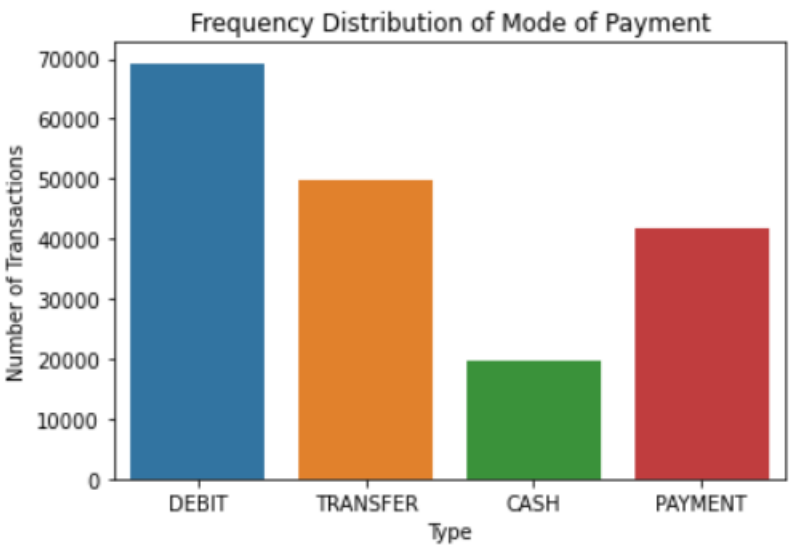


This is the distribution of what are the sales after discount, per order.

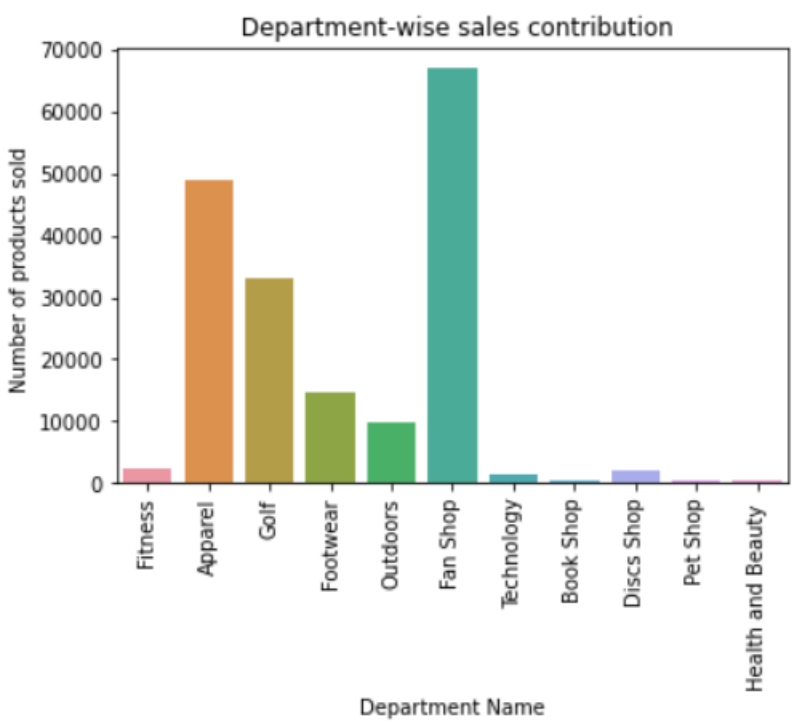
Imbalances in the Data columns

Some of the imbalances in the columns have been shown below with preliminary visualizations with the help of histograms and bar charts. Categorical variables are displayed first followed by numeric variables.

Categorical Variables



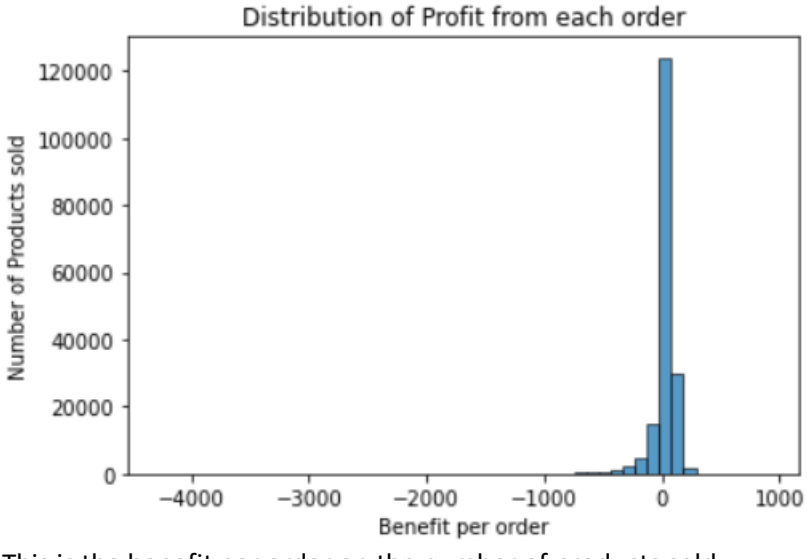
As you can see debit is the most preferred mode of payment followed by transfer and payment and cash being the last choice.



This is the department wise sales contribution where we can clearly see fan shop department being the department to sell the maximum products followed by apparel and so on.

This is the market wise sales distribution where LATAM (Latin America) is the highest revenue generating market followed by Europe and so on.

Numerical Variables



This is the benefit per order on the number of products sold.

Note: Only some of the preliminary visualizations of single variables and two variables are displayed here, to give a brief idea about the correlation of the variables and data imbalances. All the visualizations are displayed in detail in the analysis section of the project.

Data Transformation and Cleaning

1) Product Status has been removed as it contained only 1 value instead of being a binary variable. (This is because all the products are available)
2) Customer Country has been removed as there were only 2 country names “EE.UU” and Puerto Rico. "EE. UU” makes no sense.
3) Shipping Date (Date Orders) has been split into 2 columns namely Shipping Date and Shipping time and the original column has been removed.
4) Order Date (Date Orders) has been split into 2 columns namely Order Date and Order time and the original column has been removed.
5) Customer State has been removed as order region and Market are used as geographical markers.
6) The date and time formats have been made uniform.
7) The country names have been updated properly in English.

8) An extra column namely count of records has been added which works as a serial number column, this was done in order to make a donut chart in Tableau which will be shown below in the analysis section.

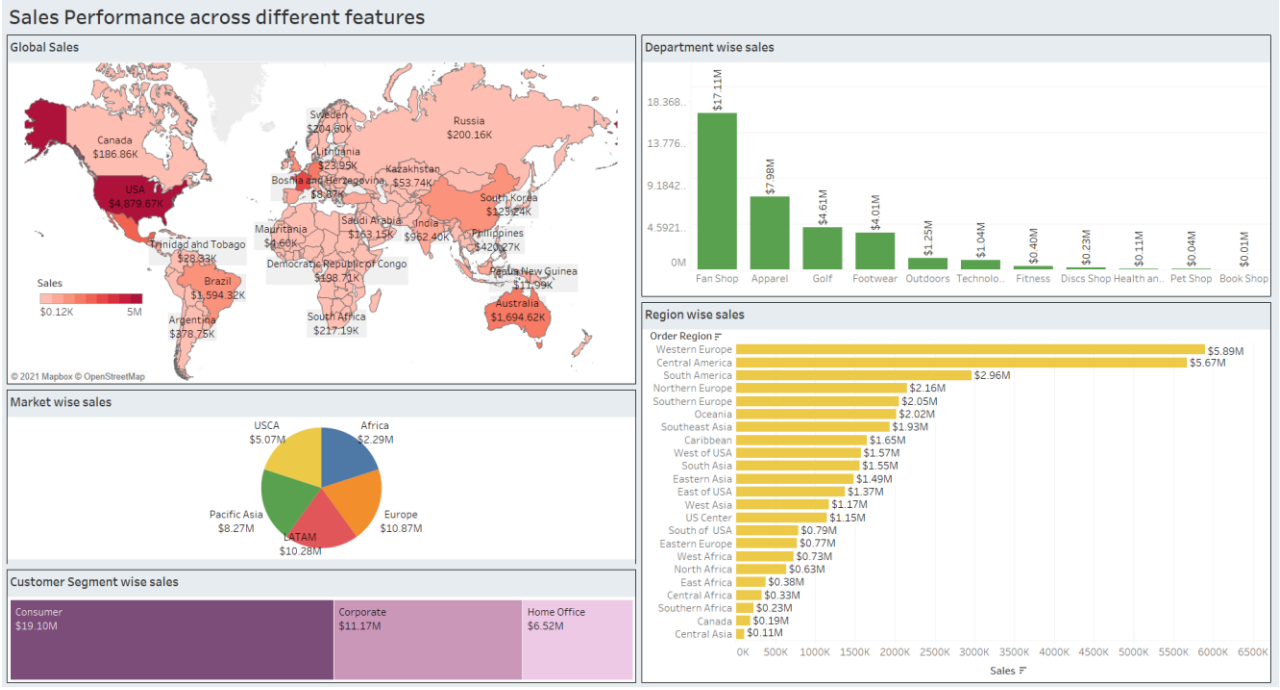
Most of these changes have been implemented using excel directly. The removal of columns mentioned above has been performed in Python which is shown in the source code file.

Data Analysis

General Points to be noted for the entirety of this analysis

1) All capital feature values are represented in Canadian dollars.
2) Shipping delay has been represented in days.
3) All times are in one time zone and no times are adjusted based on any other factors like region.
4) The words features and variables are used interchangeably throughout the analysis below.
5) In all the dashboards, M stands for millions and K stands for thousands.
6) “Payment" in type of payment is used for pending payments and those that have to be reviewed.

We will be attaching the screenshots of the dashboards here and the description of the same will be mentioned below the respective screenshots.



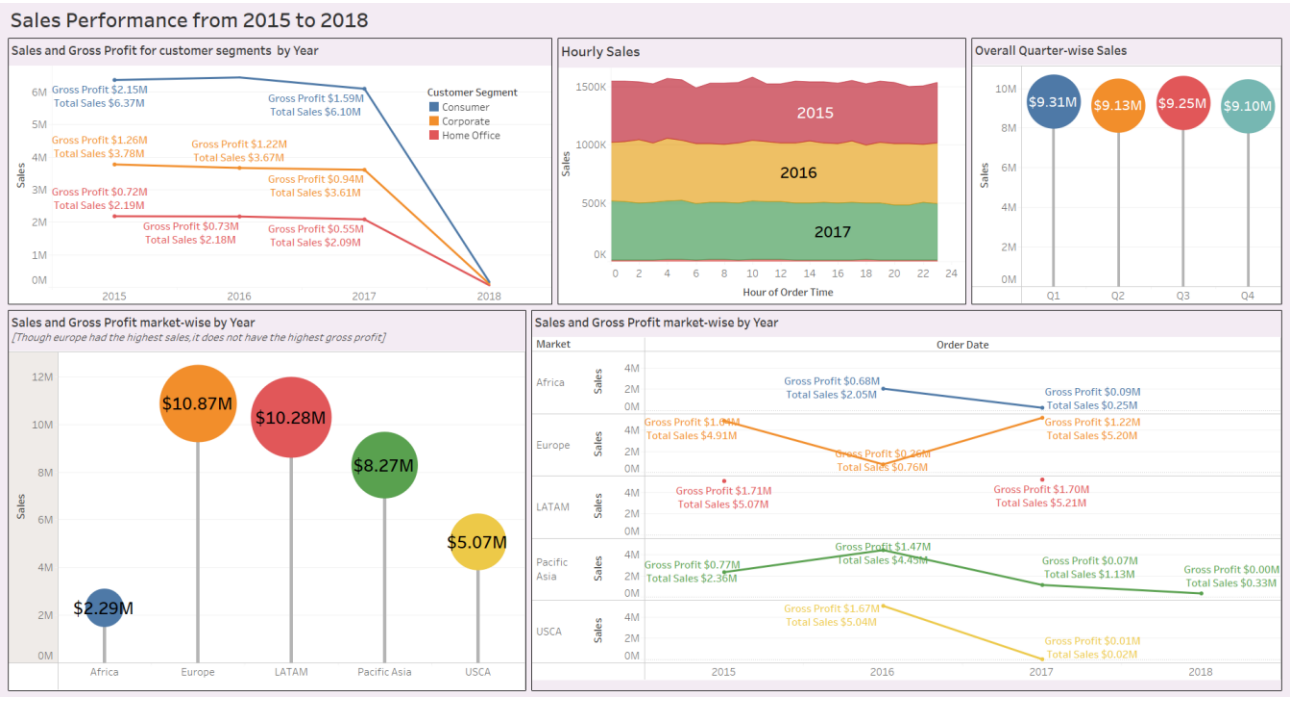
The above dashboard gives us the sales performance of DataCo Global company. The total sales have been represented across different features like geographical location, Market, Department, Order Region and Customer segment.
We can notice the global sales in the map which have been coloured from low to high sales, the lighter shades represent lower sales and the darker shades represent higher sales respectively. USA is the country with highest total sales of approximately 5 million.

The pie chart represents market wise total sales, we can see that Europe has the highest total sales while the lowest total sales is in the African market.

The department wise sales tell us that the Fan shop department has the highest total sales and book shop department has the lowest total sales.

The order region wise sales suggest that Western Europe is the most profitable in terms of total sales followed by Central America and South America. The least profitable is Central Asia region which has a total sale of 0.11 million only.

The Consumer segment is the highest buying segment followed by corporate segment as we can see the same in the Treemap of customer segment wise sales. The consumer segment's total sales are 19.1 million dollars followed by corporate segment with a total sale of 11.17 million dollars.

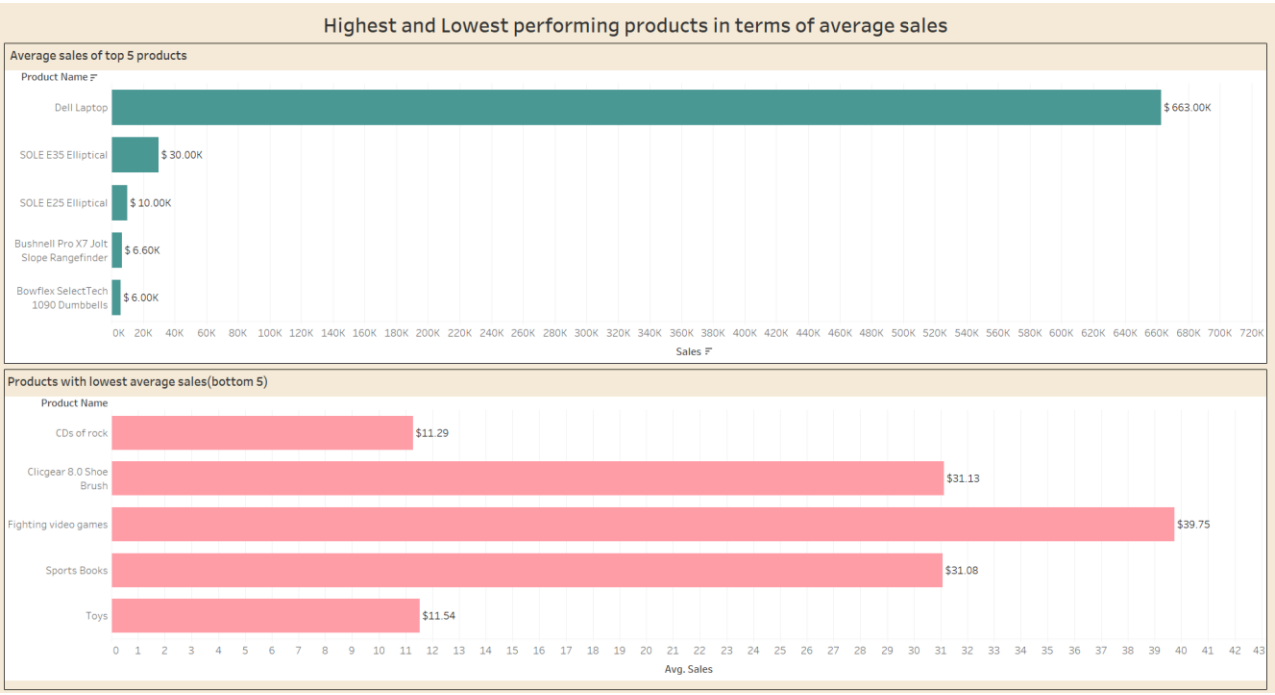


The above dashboard shows the sales performance of the company in the years 2015 to 2018.

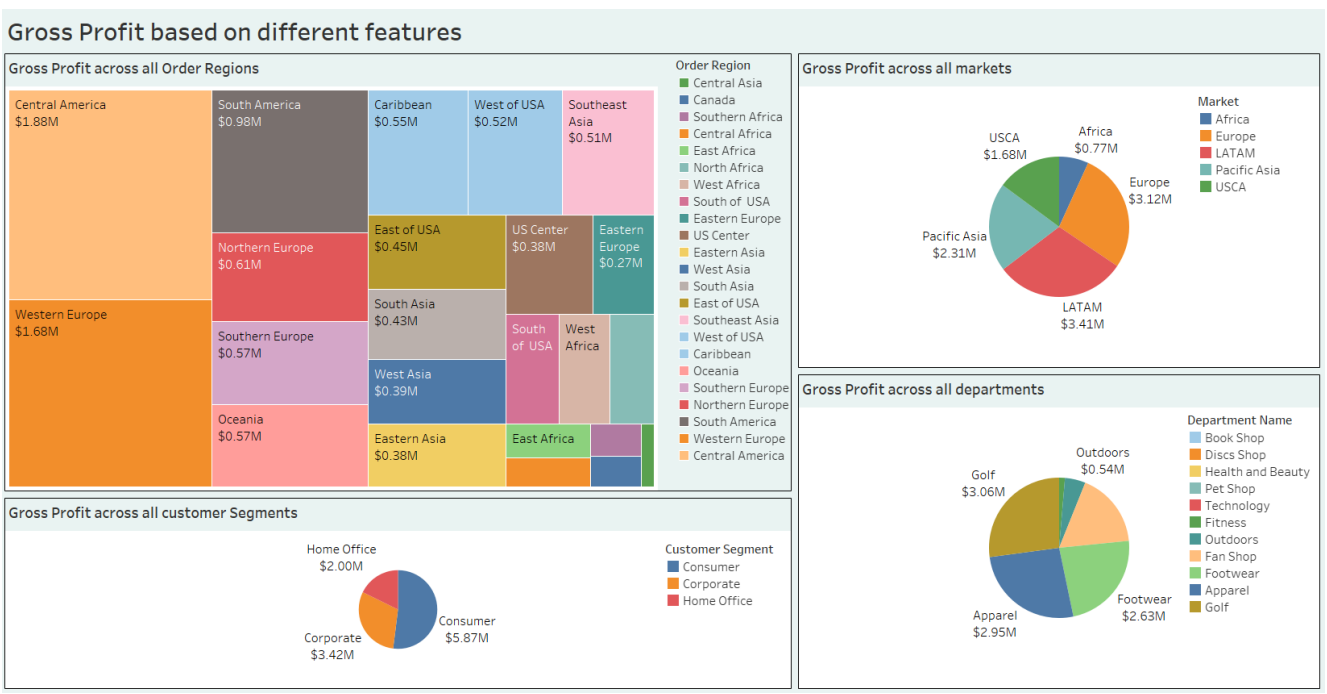
2015 can be seen to be the most profitable year in terms of gross profit as well as total sales. The first sheet is further subdivided into customer segments so that we can find which customer segment is the most profitable. Clearly, consumer segment is the most profitable which is shown in blue with a total sale of 6.37 million and gross profit of 2.15 million dollars in 2015.

We have also used order time in order to find out which time are the customers most and least active and hence which times can be focused on in order to improve the growth of the company. 10 am in the data collectors time zone seems to be the most profitable in terms of total sales. It has a total sale of 0.51 million dollars in total in 2015.

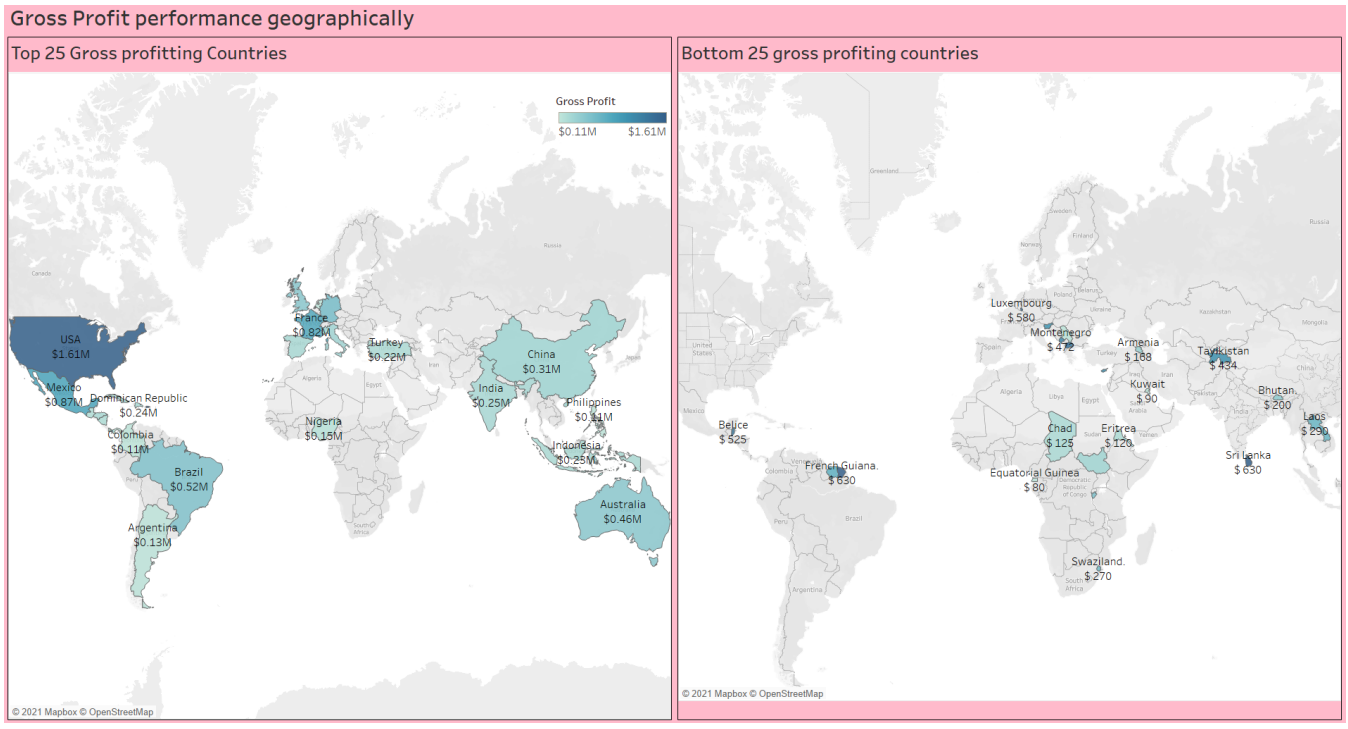
As shown by the sheets above, Europe continues to be the most profitable market in terms of total sales. However, Europe is not the most profitable when it comes to gross profit, Latin America or LATAM as shown in this dataset is the most profitable in terms of gross profit visible by the size of the circle of LATAM as compared to Europe in the bottom left sheet in the dashboard. The first quarter in all these years is most profitable in terms of total sales.



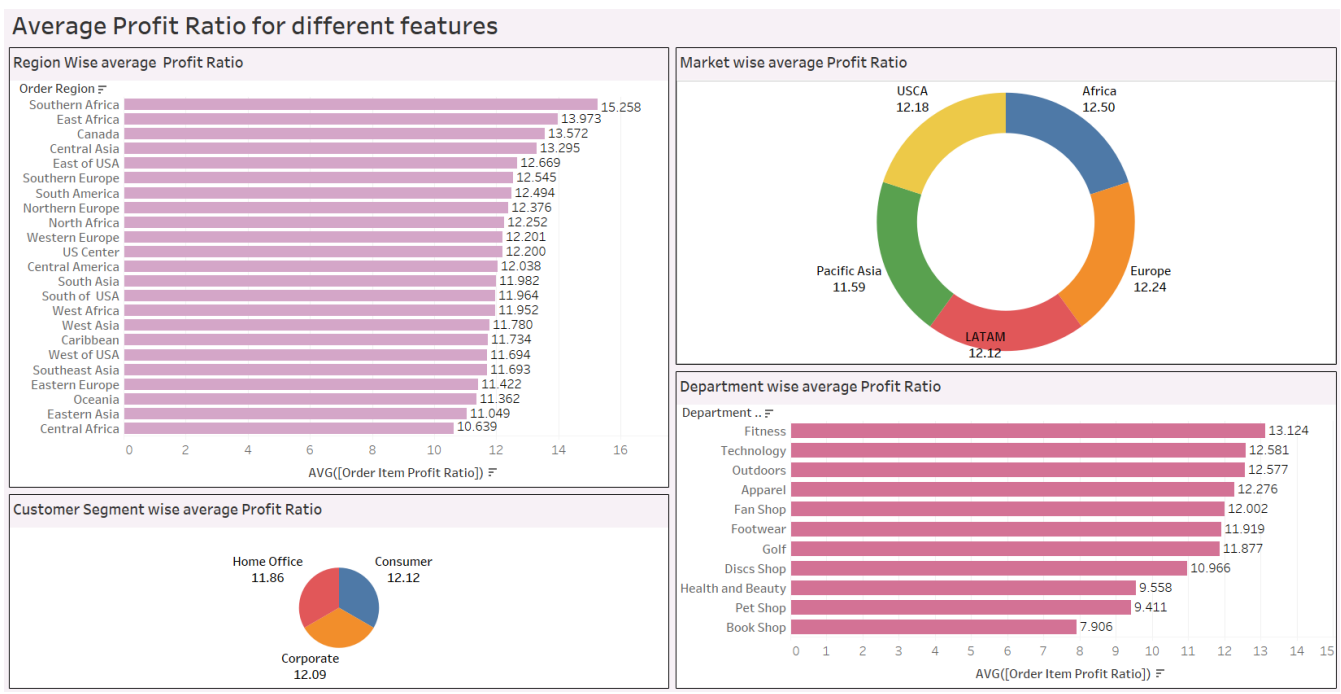
The above dashboard shows us the top 5 and the bottom 5 performing products in terms of average sales. On average the highest selling product is Dell Laptop and the least selling product are CDs of Rock. This maybe because the usage of CD's has become very rare in this era.



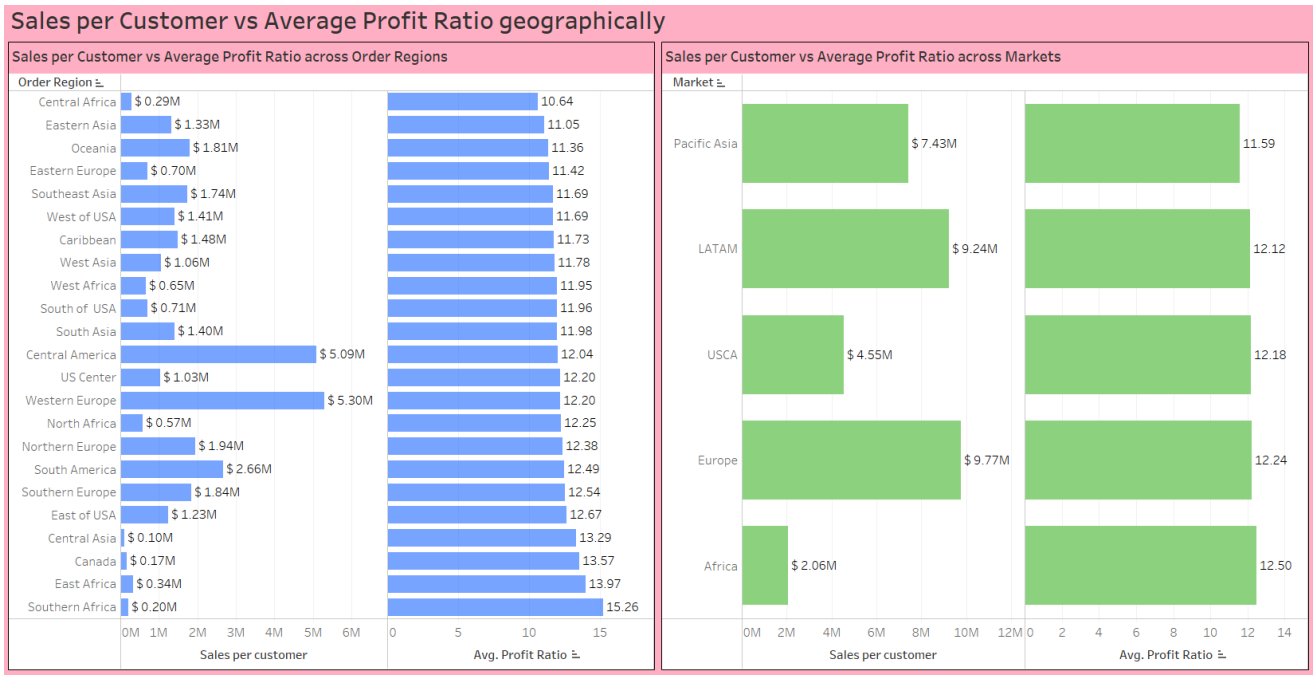
The above plot shows gross profit for different features in the dataset. We can see the gross profit for all order regions, customer segments, Markets and all departments.
As we can see, central America order region has the highest gross profit of 1.88 million dollars followed by western Europe which has the highest gross profit of 1.68 million dollars. Consumer customer segment has the highest gross profit and golf department has the highest gross profit in departments of 3.06 million dollars.



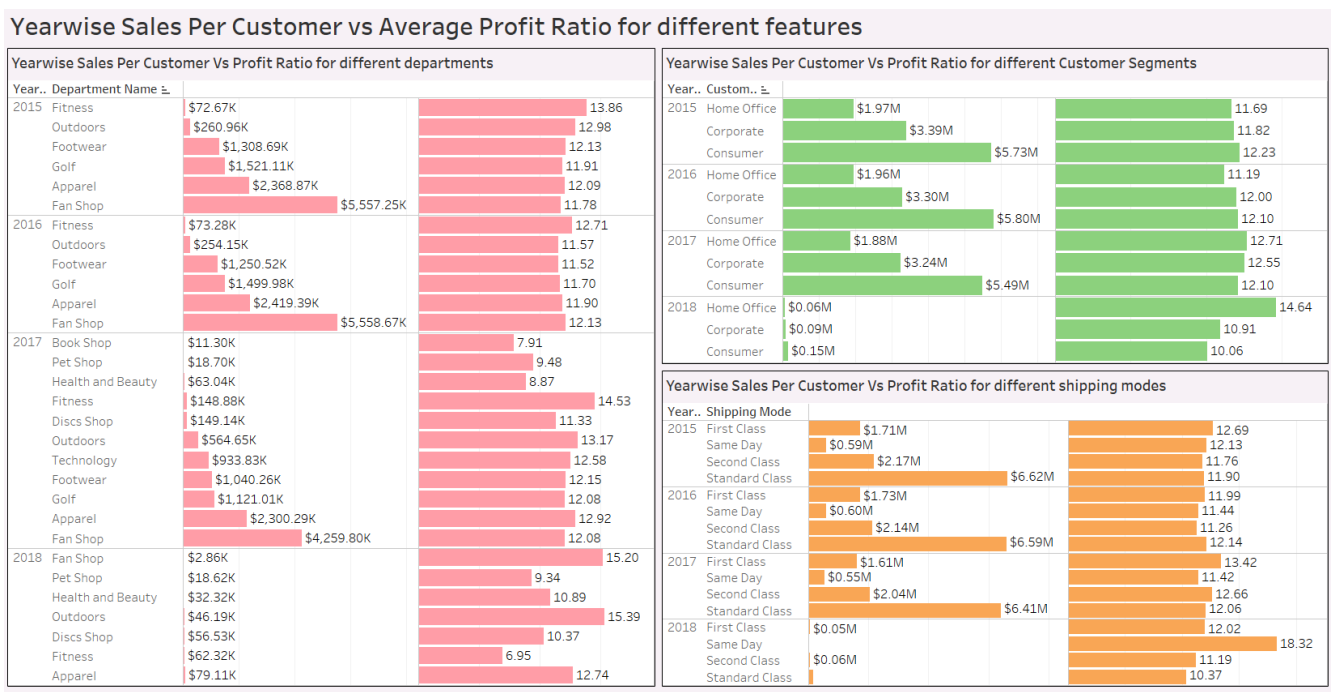
The above dashboard shows us the geographical performance of the countries in terms of gross profit. It shows us the top 25 and the bottom 25 countries with their gross profit values. We can see that USA has the highest gross profit of 1.61 million dollars followed by Mexico with 0.87 million dollars. The country with the least gross profit is Equatorial Guinea with a gross profit of 80 dollars only.



The above plot shows us the average profit ratio for different features like order region, markets, customer segments and departments.
We can see that Southern Africa is the order region which has the highest average profit ratio of 15.25 while consumer segment is the winner for customer segments. Fitness department has the highest average profit ratio and Africa is the market with the highest average profit ratio of 12.5.

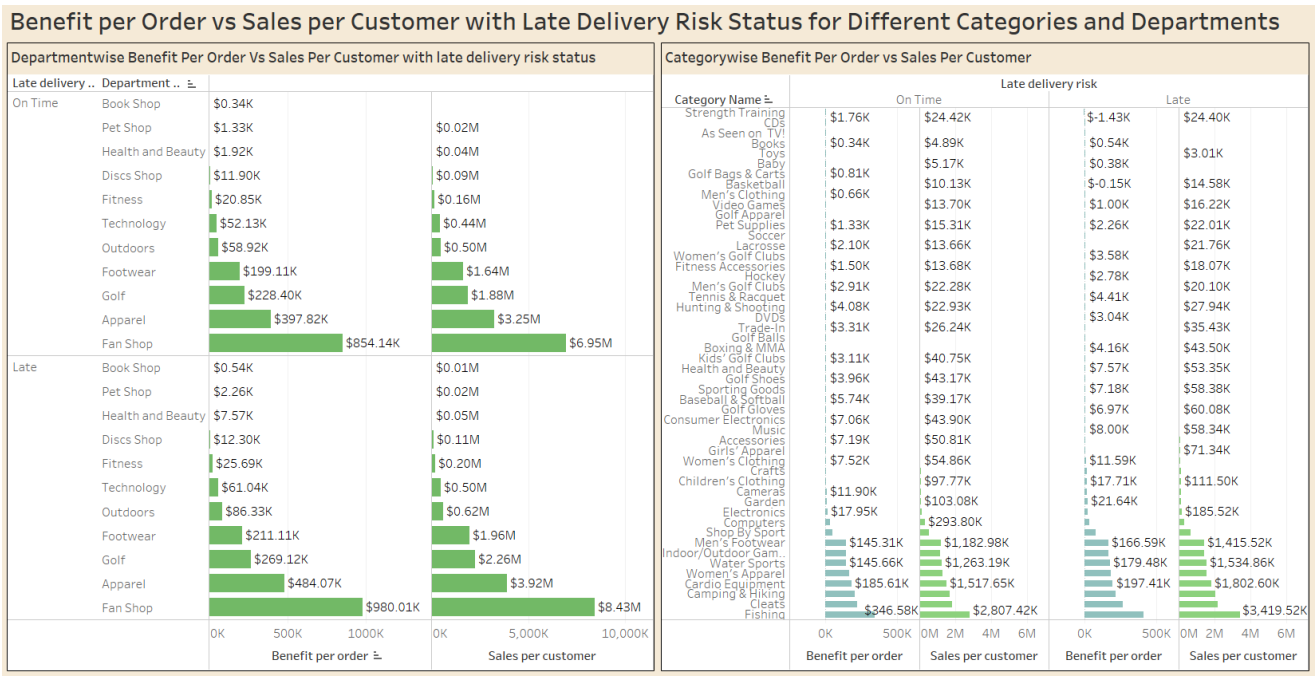


There is a very important insight that this dashboard provides that is even though Southern Africa has the highest average profit ratio, it's total sales per customer are the lowest across the regions. This can be worked on in order to increase the growth of the company rapidly. A similar insight can be interpreted from the other sheet which shows Africa market as the highest average profit ratio market but with lowest total sales per customer.

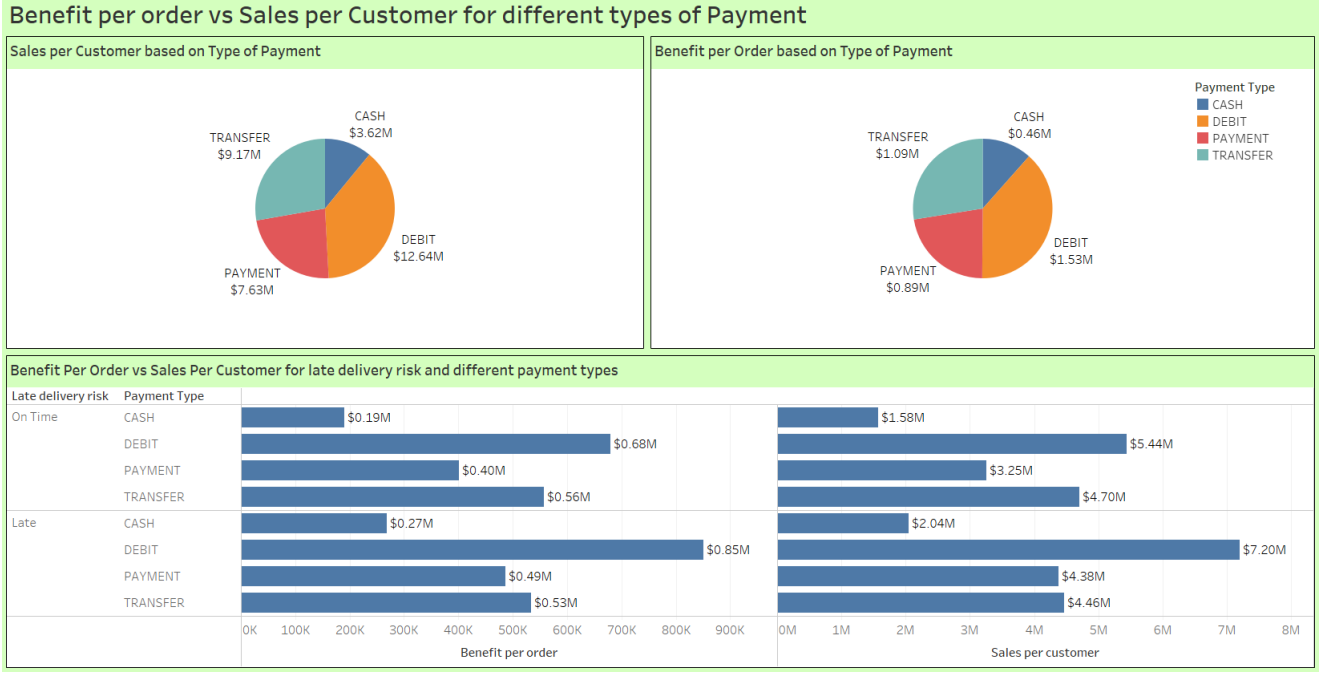


We can see here the total sales per customer vs average profit ratio for different features like shipping mode, departments and customer segments.

As one can interpret from the previous descriptions, the fan shop department is the most profitable in terms of total sales per customer and also has a good average profit ratio across all the years. The standard class shipping mode is more benefitable in terms of sales per customer as compared to same day or first class. This can be a potential problem as first class and same day modes lose their significance if this continues. Again, the consumer segment is the one with the highest total sales per customer and relatively a good average profit ratio.

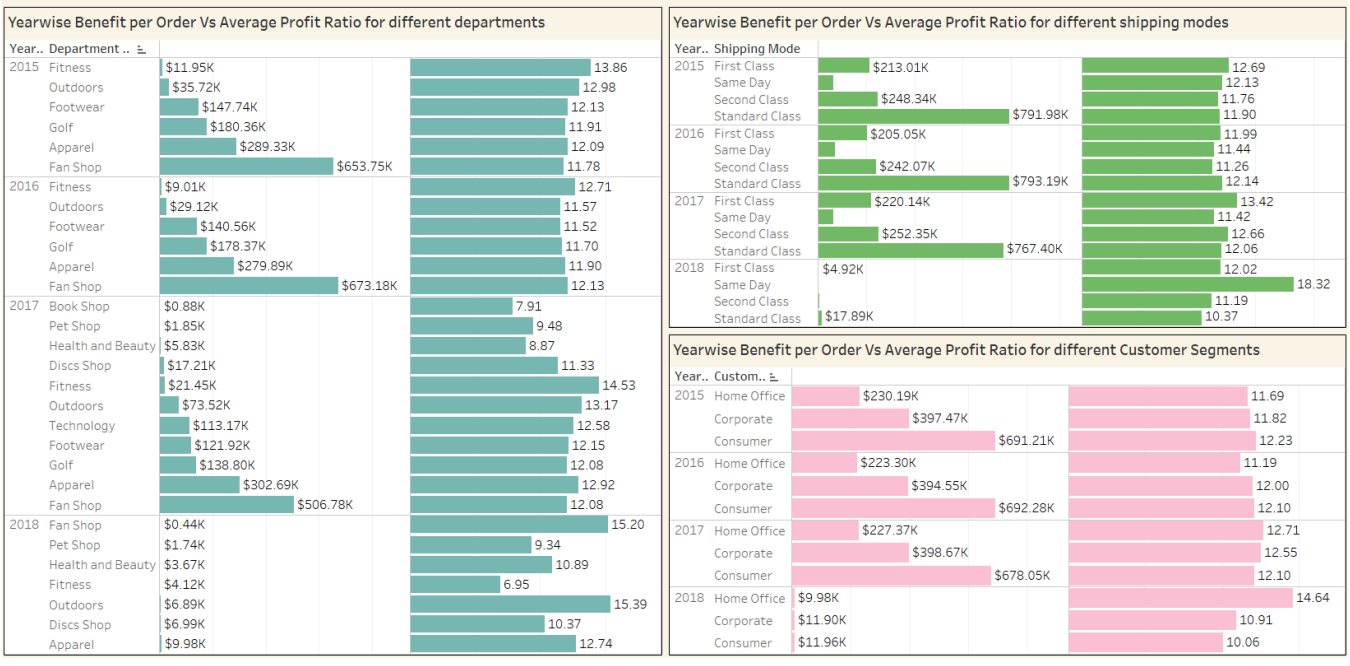


The above dashboard shows us the Total Benefit per order vs Total Sales per Customer for departments and categories.
It shows us that regardless of whether the delivery of the product from fan shop department is late or on time, it has the highest total benefit per order as well as the highest total sales per customer. It also shows us that fishing category is the winner when it comes to categories for both highest total benefit per order as well as the total sales per customer.



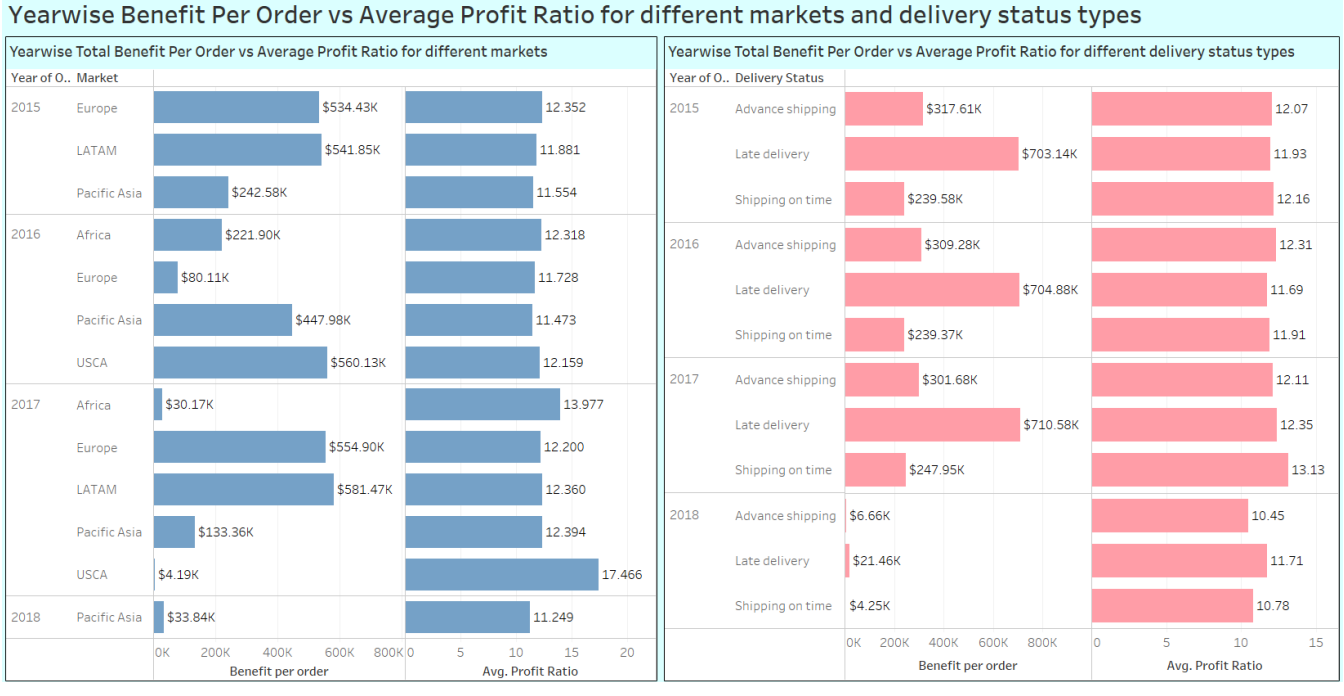
We can see from above that the debit payment type is the most beneficial or the preferred choice for customers as the highest total sales as well as highest total benefit per order are for this payment type which is 12.64 million dollars and 1.53 million dollars respectively.
A peculiar observation here is that the late delivered products which were paid by debit have higher total sales per customer as well as higher total benefit per order that is 7.20 million dollars and 0.85 million dollars respectively. This is highly peculiar as the products that were delivered late are more profitable as compared to the ones that were delivered on time.

Yearwise Benefit per Order vs Average Profit Ratio for different features



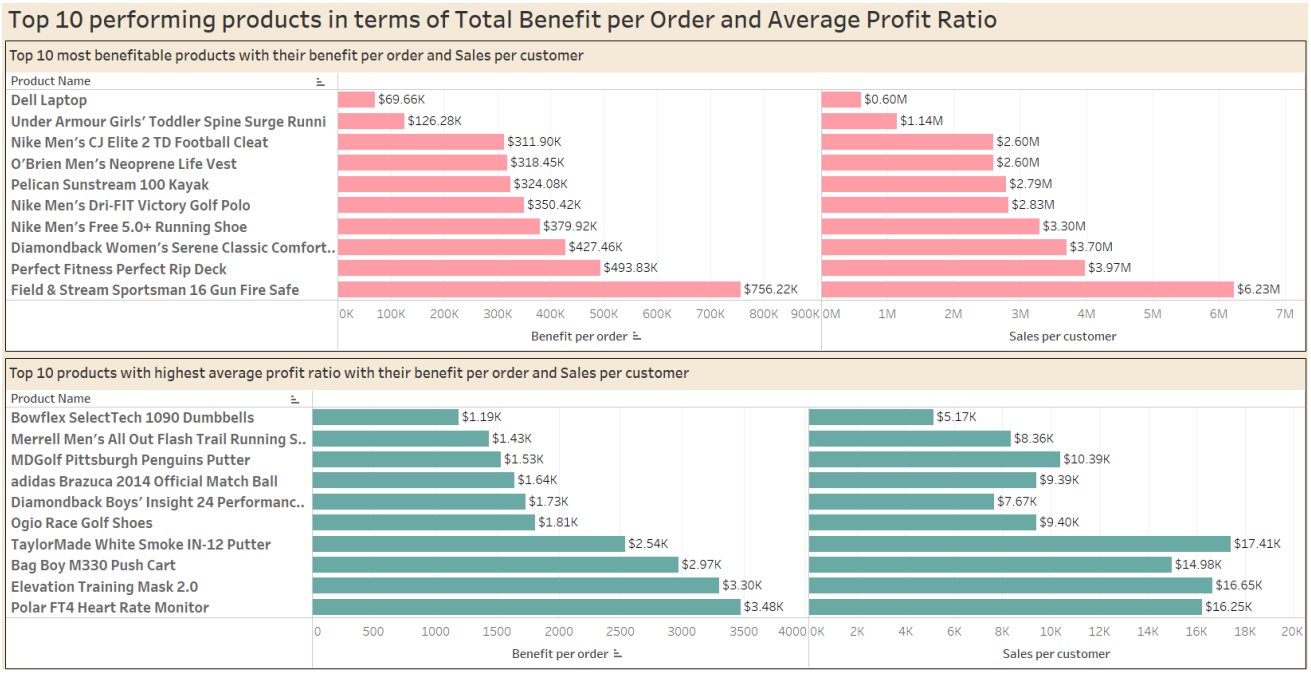
The above dashboard shows us the yearly total benefit per order vs average profit ratio for different features. The fan shop department has the highest total benefit per order regardless of the year. The standard class shipping mode again has a higher total benefit per order. Also, as one could understand from the above dashboards and analysis, consumer segment is the most profitable when it comes to total benefit per order.

Yearwise Benefit Per Order vs Average Profit Ratio for different markets and delivery status types

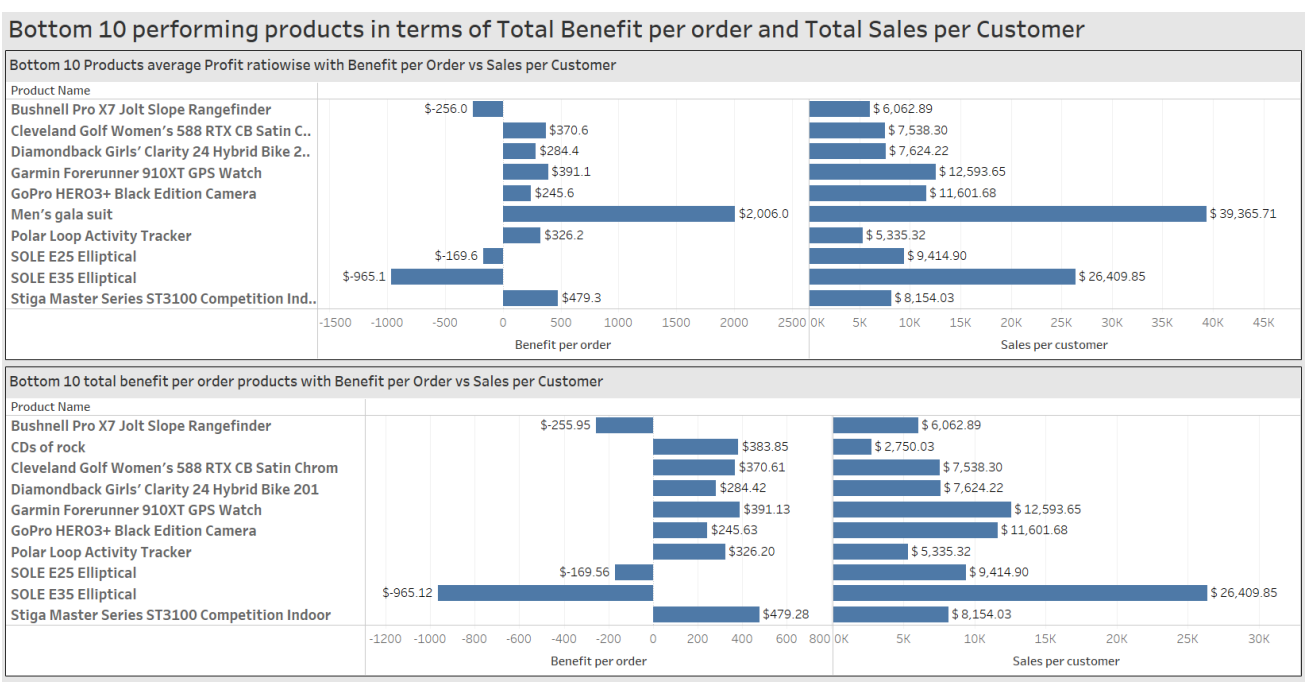


The above dashboard shows us that Europe market has the highest total benefit per order of 534.43 thousand dollars and relatively a very good average profit ratio of 12.35 in 2015. In 2016, the story is very different as there was a massive fall in total benefit per order to a mere 80 thousand dollars approximately. Also, USCA market was the winner when it came to the highest total benefit per order which was 560.13 thousand dollars in this year. Europe recovered tremendously in 2017 and in fact it even topped its 2015 value of total benefit per order by approximately 20 thousand dollars. The highest total benefit per order was for Latin America market in 2017 while the previous year's top contender for the highest total benefit per order USCA was the last in 2017.

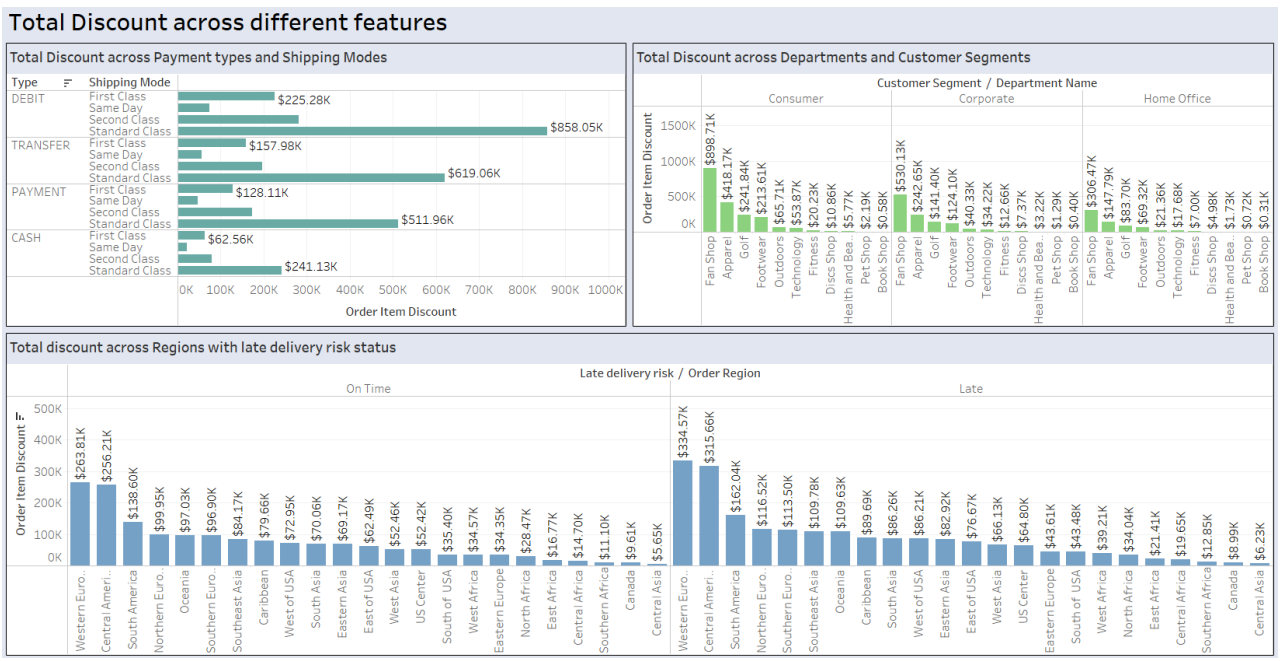
For the shipping modes, the late delivered products show the highest total benefit per order regardless of the year.



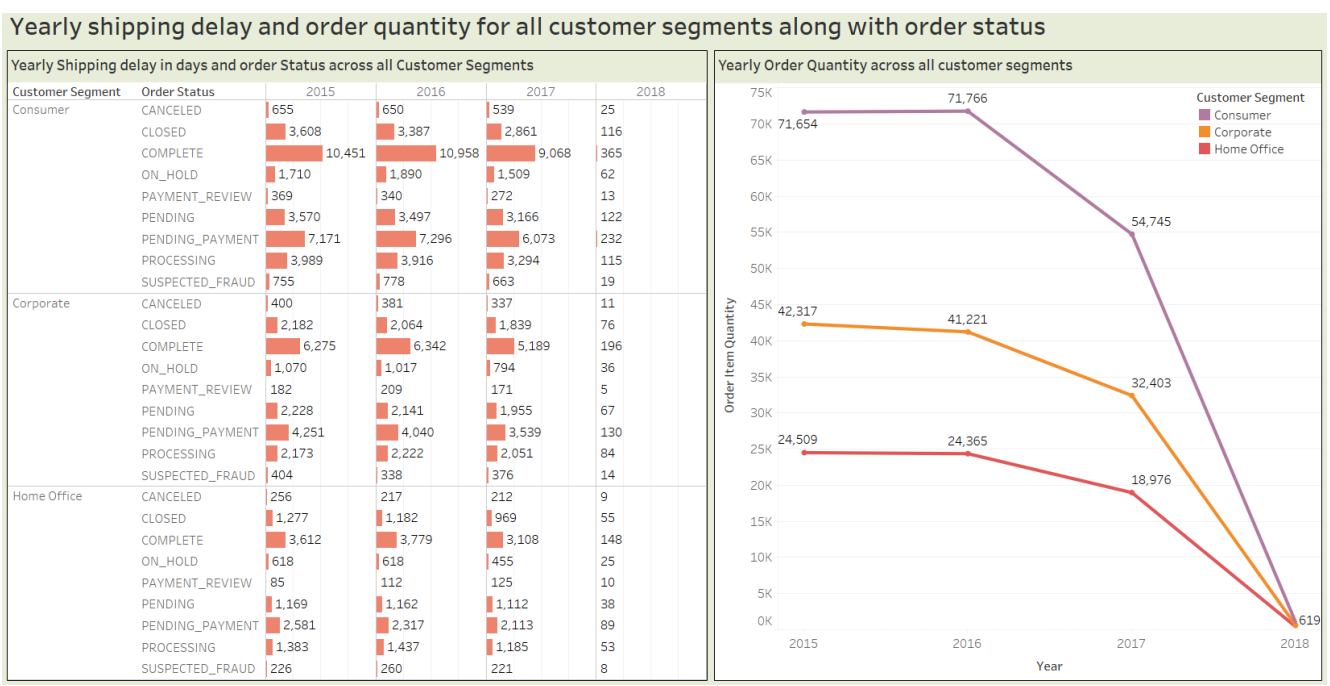
The above dashboard shows us the top 10 products in terms of benefit per order as well as highest average profit ratio.
The Field & Stream sportsman 16 Gun Fire Safe is the most benefitable product in terms of benefit per order the company as shown with a total benefit per order of 756.22 thousand dollars and 6.23 million dollars of total sales per customer.
The Polar FT4 Heart Rate Monitor is the best product in terms of average profit ratio with a total benefit per order of 3.48 thousand dollars and 16.25 thousand dollars of total sales per customer. This means that the products with the highest average profit ratio do not even come close to selling the most as seen above. As we can see not one of the top 10 in the bottom graph of the dashboard are present in the top 10 products with highest benefit per order. This is alarming as the most profitable products are not being sold as much as they should.



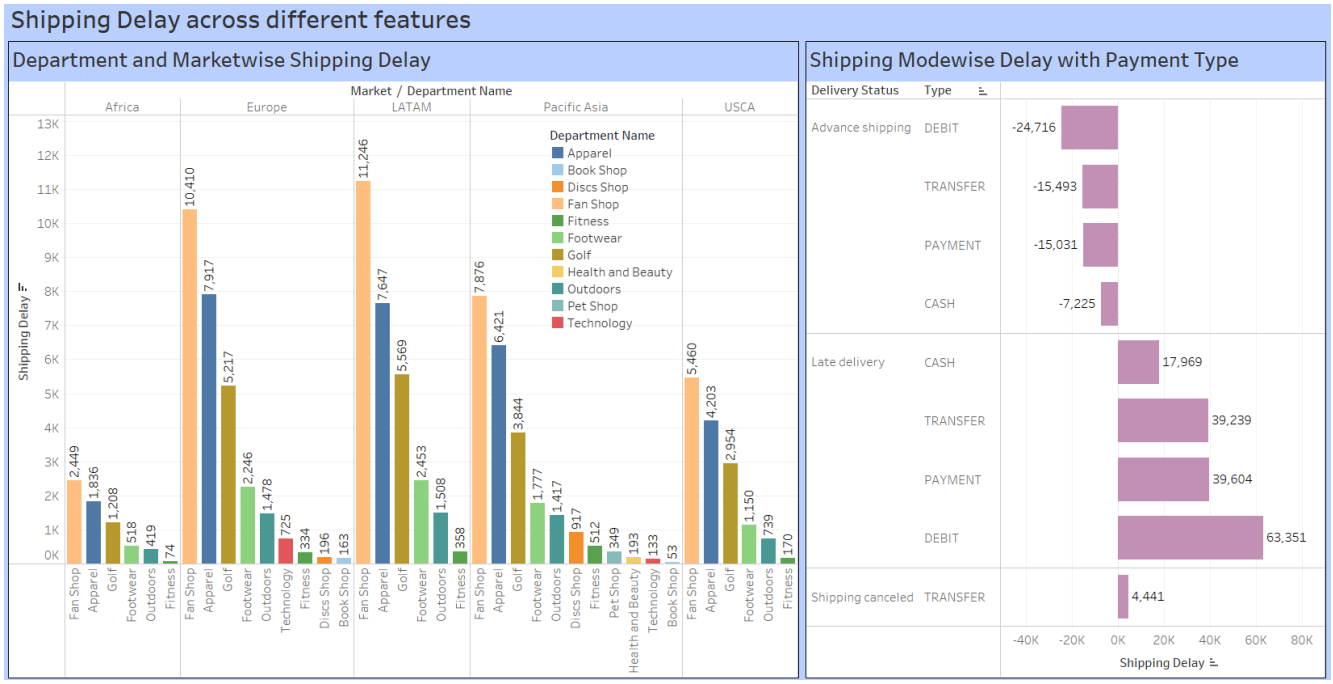
We can see the worst performing products in terms of average profit ratio as well as total benefit per order above in the dashboard respectively.
These products should also be focused on in order to bring them to a considerate level of benefit per order as well as average profit ratio.



The standard class shipping mode has the highest total discount provided regardless of the type of payment method. We can also see that fan shop department has the highest total discount provided whether the customer belongs to consumer, corporate or home office segment.
We can see that Western Europe is the region with the highest total discount provided regardless of whether the product delivered on time or not.

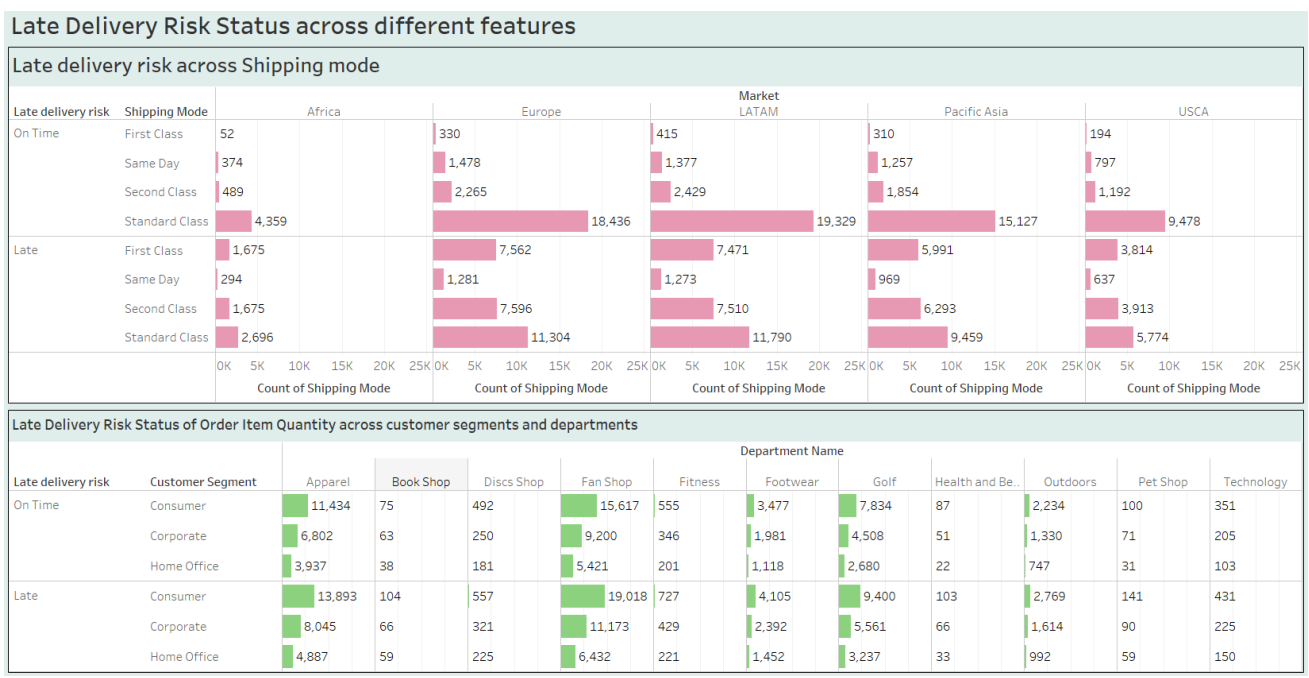


We can see the yearly shipping delay in days vs the order quantity for all customer segments along with the order status. We can see that the complete orders have the highest shipping delays regardless of the customer segment. The consumer segment orders the maximum products from this company with their highest being 71,654 order quantity in 2015.



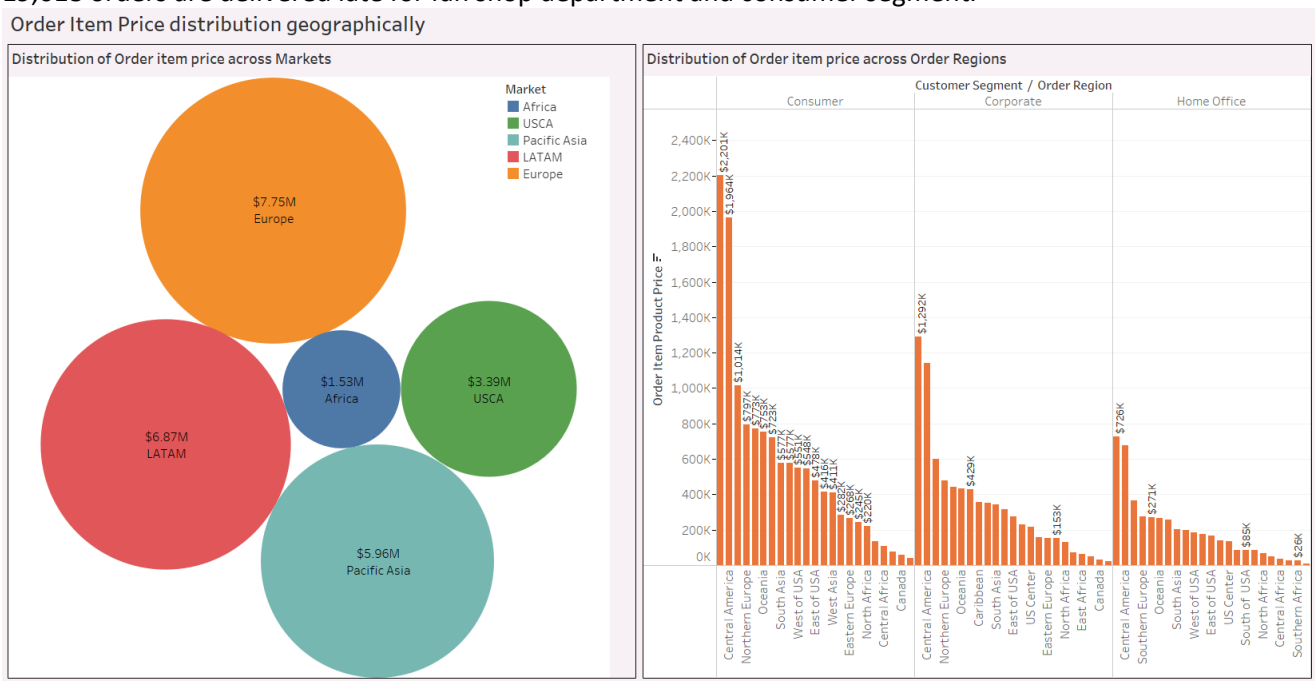
We can see the shipping delay for the shipping modes as well as payment types with the different departments in all the markets.

The fan shop department has the highest shipping delay across all markets and as one would expect the late delivery has the highest shipping delay but the shipping delay is maximum for debit card payment type that is a total of 63,351 days of shipping delay across 2015 to 2018 for all their orders and products delivered.

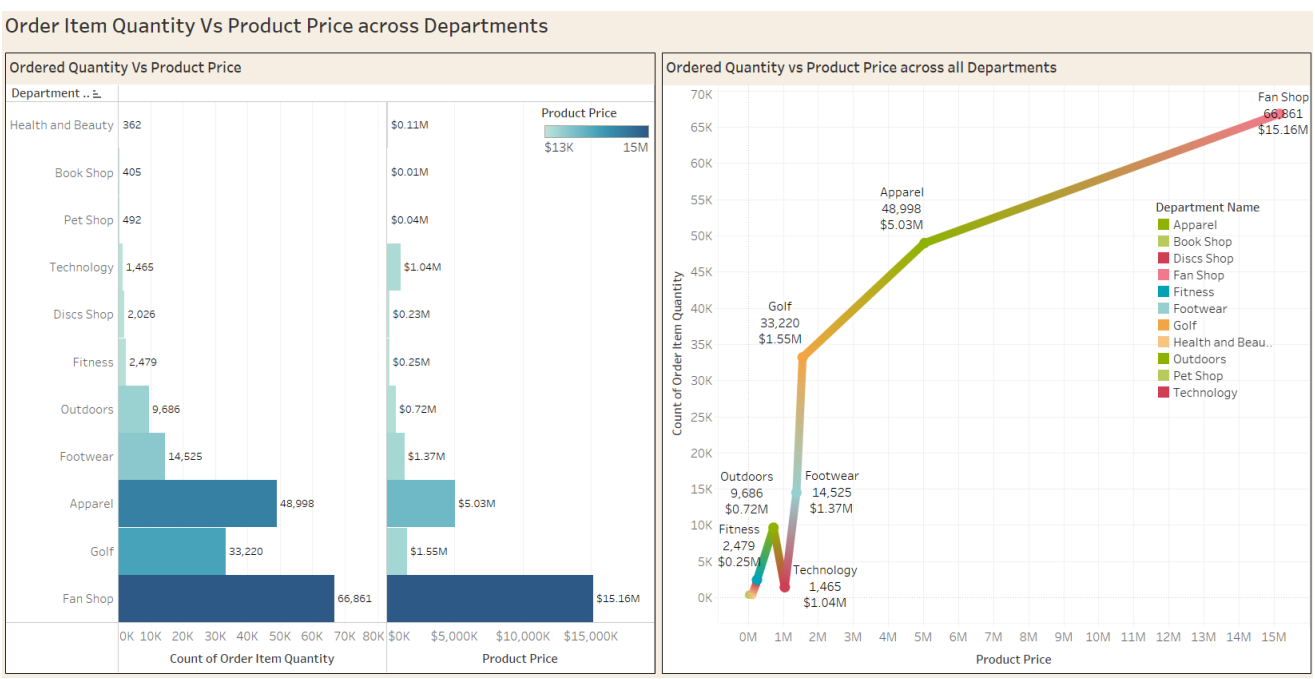


The above dashboard shows us the late delivery risk status across different features namely customer segments, departments and shipping modes.
The standard class has the highest late delivered products regardless of market. One other thing to be noticed here is that standard class also has the highest delivery of products delivered on time which is a positive.

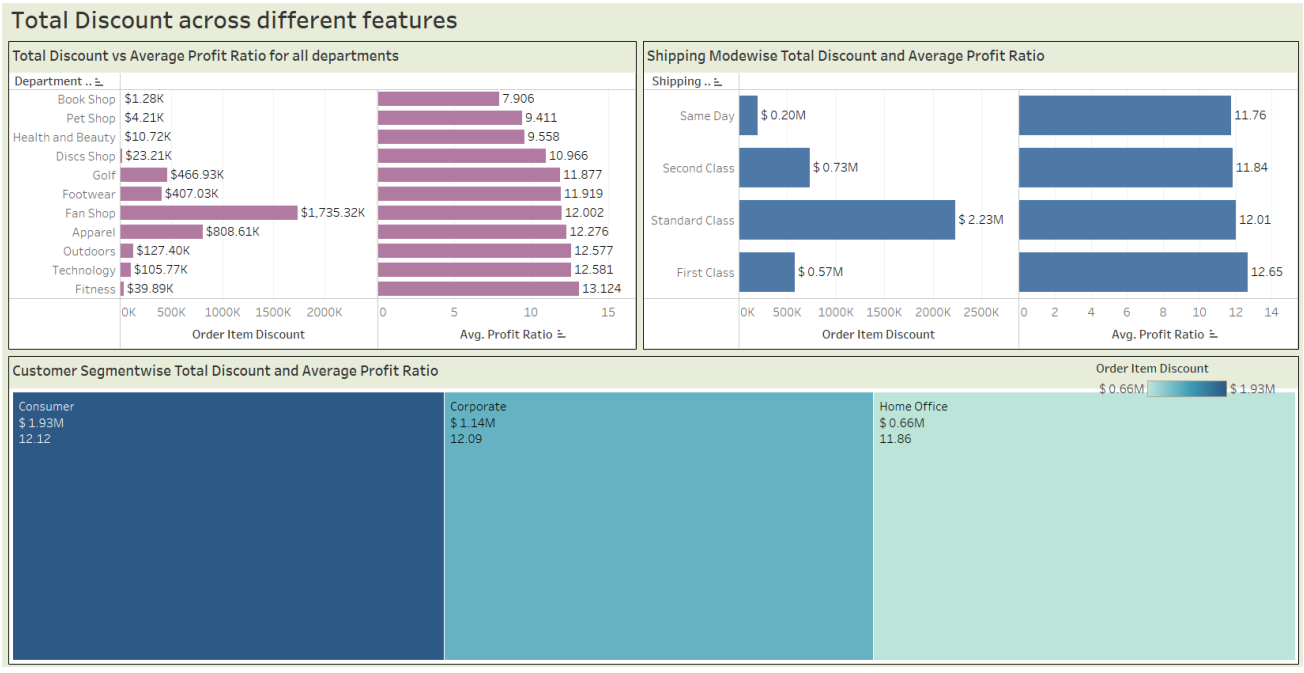
The consumer segment customers who order from the fan shop department have the highest deliveries regardless of whether they are on time or late. 15,617 orders are delivered on time and 19,018 orders are delivered late for fan shop department and consumer segment.



The total order item price is highest for Europe market which is 7.75 million dollars and western Europe region specifically contributes the highest to the above insight regardless of their customer segments.



The above dashboard shows us the order item quantity vs total product price across all departments. We can see that fan shop has the highest order item quantity of 66,861 with a total product price of 15.16 million dollars. The line chart shows the relationship between distribution of order item quantity vs product price. The colours represent the departments. As we can see shown by the other sheet, fan shop has the highest number of quantities as well as highest total product price as shown in dark red on the top of the graph on the right.



This dashboard shows us the total discount provided across different features like customer segment, shipping modes and departments.
While fan shop has the highest total discount provision it does not have the highest average profit ratio. Similarly, standard class has the highest total discount provision but it does not have the highest average profit ratio.
Consumer segment has the highest total discount provision and highest average profit ratio.

Summary of Data Analysis

Europe is the market with the highest total sales of 10.87 million dollars and the western Europe order region is the region with the highest total sales of 5.89 million dollars and finally USA is the country with the highest total sales of 4.89 million dollars. This might look peculiar as Europe is the continent with the higher total sales but Europe has many countries while USA is only a single country.
USA is solely contributing almost 44% of what Europe is contributing to the total sales of the company even though the company is originally based in Europe.

The geographic distribution of the gross profit from small to big in terms of land mass is as follows:
USA is the country with the highest total gross profit of 1.61 million dollars with Central America being the most profitable order region with the highest total gross profit of 1.88 million dollars and the Latin America market with the highest total gross profit of 3.41 million dollars being the most profitable.
This means that the 2 American continents are jointly responsible for majority of the total gross profit value of the company even though the company is originally based in Europe.

The golf department is the department with the highest total gross profit even though the department with the highest total sales per customer and highest total benefit per order is fan shop department.
This means that even with lesser total sales from the golf department it has a higher gross profit value than the fan shop department.

Europe has the highest total benefit per order in terms of market of 1.17 million dollars followed by Latin America with 1.12 million dollars in close second. Western Europe is the order region with the highest total benefit per order of 0.63 million dollars for order regions and USA is the country with highest total benefit per order amongst all the countries with 0.54 million dollars.
Like total sales, USA is contributing almost half of Europe's total value of benefit per order to the company's total benefit per order.

The standard class shipping mode has the highest total benefit per order of 2.37 million dollars and the late delivered orders have a higher total benefit per order of 2.14 million dollars as compared to orders that are on time.

After seeing the dashboards, we have noticed that the performance of the company had reduced gradually from 2015 to 2017 and from 2017 to 2018 it took a massive nosedive and they were reduced to almost no sale in that year.

Our Comments and Suggestions

1) The top 10 products with the highest average profit ratio are not even in the list of the top 10 products with highest benefit per order and highest sales per customer. This means that these products could be advertised differently or the approach for their entry into the market could be changed in order to extract the highest benefits from these products.
2) The standard class shipping mode has the highest total benefit per order and sales per customer as compared to same day or first class shipping mode. While this could also be based on the customer's personal choices, we think that this is an area which needs to be improved in order to maintain the significance of same day and first class shipping modes.
3) The late delivered orders have higher total sales and higher total gross profit as compared to the orders that are on time, this needs to be improved as this could prove to be very harmful to the growth of the company in the long run.
4) The standard class shipping mode has no total shipping delay in fact they are delivered before the scheduled delivery date as their total shipping delay is negative while the shipping modes first class and same day have considerable delays. Again, this needs to be corrected as quickly as possible.
5) Africa is the market with the highest average profit ratio while its total sales are the lowest. This is another area where the company could improve and thereby drastically improve their benefits.
6) Fitness department is the most profitable in terms of average profit ratio while its total discount and total sales are the lowest. This is the department which needs to change its strategy and work towards using this potential golden opportunity to reap maximum benefits from it.
7) Shipping mode standard class has the highest provision of total discount even though it does not have the highest average profit ratio amongst all shipping modes. This needs to be corrected.
8) In 2018, the performance was the worst in the four years from 2015 to 2018. This is a highly drastic fall and concerning matter which should have been rectified instantly.