Ministry of Communications and Information Technology

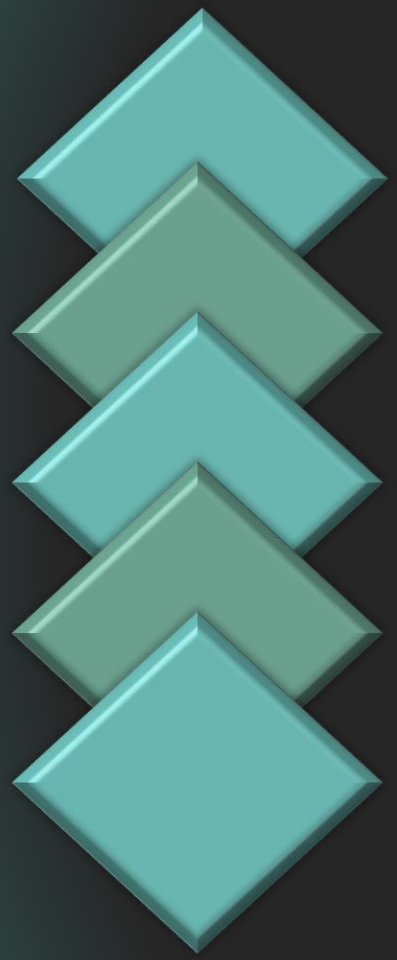
رواد مصر الرقمية

Sales & Supply Chain Analysis Report

Technical Provider: AST

Instructor: Ahmed Abdellatif

Meet our team



Mohamed Shehata

Youssef Tarek

Hassan Hashish

Ibrahim Mohamed

Khaled Hossam

Agenda


Objective &
Data Overview
Cleaning &
Modelling
EDA with Python
Blueprint &
Dashboards
Analysis Report
Recommendations

Objective

The goal of this project is to analyze the sales and supply chain performance of Data Co company to identify insights and key requirements by analyzing the performance over the past few years and providing decisions and recommendations for improvement to the company.

Data Overview

The dataset contains 180,519 rows and 53 columns of data related to sales and supply chain operations, including product details, delivery times, shipping methods, customer details, sales information, and department details. The dataset contains only one fact table, so it needs to be normalized to reduce the data redundancy and make it suitable for data modeling. And there some cleaning and nulls to handle.

Cleaning & Modelling

Cleaning & Modelling

First, we imported libraries like Pandas for cleaning, Matplotlib & Seaborn for visualization, and HTML for downloading the file after we finished cleaning. Then, we uploaded the file and checked the number of rows and columns, and there was 180,519 rows and 53 columns.

Imported Libraries

[2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from IPython.display import HTML
```

Overview

[4]:
```
df = pd.read_csv(r'C:\Users\nopog\OneDrive\Documents\DEPI\Datasets\Data Co\DataCoSupplyChainDataset.csv', encoding = 'unicode_escape')
```

[6]:
```
df.shape
```

[6]:
```
(180519, 53)
```

Cleaning & Modelling

Second, we checked each column for its data type and the number of null values. Then, we examined how many duplicates were in the data and reviewed the format of each column and deleted two columns that contained nearly 100% null values.

```
[8]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180519 entries, 0 to 180518
Data columns (total 53 columns):
 #   Column                         Non-Null Count   Dtype
---  ------                         --------------   -----
 0   Type                           180519 non-null  object
 1   Days for shipping (real)       180519 non-null  int64
 2   Days for shipment (scheduled)  180519 non-null  int64
 3   Benefit per order              180519 non-null  float64
 48  Product Name                   180519 non-null  object
 49  Product Price                  180519 non-null  float64
 50  Product Status                 180519 non-null  int64
 51  shipping date (DateOrders)     180519 non-null  object
 52  Shipping Mode                  180519 non-null  object
dtypes: float64(15), int64(14), object(24)
memory usage: 73.0+ MB

[7]: df.duplicated().sum()

[7]: 0

[8]: df.head(10)

[9]: df.drop(['Product Description', 'Order Zipcode'], axis = 1, inplace = True)
```

Cleaning & Modelling

Third, we created the customer table, which contained all the customer information. Then, we deleted duplicates from the table and ensured there were no duplicates in the customer ID, confirming that there were no multivalued attributes.

[18]:
```
dfcus = df[['Customer Id', 'Customer Fname', 'Customer Lname', 'Customer Email', 'Customer Password', 'Customer Country', 'Customer State', 'Customer Cit
```

[22]:
```
dfcus.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180519 entries, 0 to 180518
Data columns (total 11 columns):
 #   Column     Non-Null Count   Dtype  
---  ------     --------------   -----  
 0   ID         180519 non-null  int64  
 1   Fname      180519 non-null  object 
 2   Lname      180511 non-null  object 
 3   Email      180519 non-null  object 
 4   Password   180519 non-null  object 
 5   Country    180519 non-null  object 
 6   State      180519 non-null  object 
 7   City       180519 non-null  object 
 8   Street     180519 non-null  object 
 9   Zipcode    180516 non-null  float64
 10  Segment    180519 non-null  object 
dtypes: float64(1), int64(1), object(9)
memory usage: 15.1+ MB
```

[24]:
```
dfcus.duplicated().sum()
```

[24]:
```
159867
```

[26]:
```
dfcus = dfcus.drop_duplicates()
```

[32]:
```
dfcus['ID'].duplicated().sum()
```

[32]:
```
0
```

Cleaning & Modelling

Fourth, we checked the format and values in each column in the table, replacing any necessary values, whether they were garbage values or nulls values, and changed the column names if necessary.

[36]:
```
dfcus.head()
```

[36]:

	ID	Fname	Lname	Email	Password	Country	State	City	Street	Zipcode	Segment
0	20755	Cally	Holloway	XXXXXXXXX	XXXXXXXXX	Puerto Rico	PR	Caguas	5365 Noble Nectar Island	725.0	Consumer
1	19492	Irene	Luna	XXXXXXXXX	XXXXXXXXX	Puerto Rico	PR	Caguas	2679 Rustic Loop	725.0	Consumer
2	19491	Gillian	Maldonado	XXXXXXXXX	XXXXXXXXX	EE. UU.	CA	San Jose	8510 Round Bear Gate	95125.0	Consumer
3	19490	Tana	Tate	XXXXXXXXX	XXXXXXXXX	EE. UU.	CA	Los Angeles	3200 Amber Bend	90027.0	Home Office
4	19489	Orli	Hendricks	XXXXXXXXX	XXXXXXXXX	Puerto Rico	PR	Caguas	8671 Iron Anchor Corners	725.0	Corporate

[38]:
```
dfcus['Country'].unique()
```

[38]:
```
array(['Puerto Rico', 'EE. UU.'], dtype=object)
```

[40]:
```
dfcus['Country'] = 'United States'
```

Cleaning & Modelling

Fifth, we deleted these columns from the fact table and once again ensured there were no duplicates in the fact table. Then, we downloaded the table as a CSV file.

[54]:
```
df.drop(['Customer Fname', 'Customer Lname', 'Customer Email', 'Customer Password', 'Customer Country', 'Customer State', 'Customer City', 'Customer Stre
```

[56]:
```
df.duplicated().sum()
```

[56]:
```
0
```

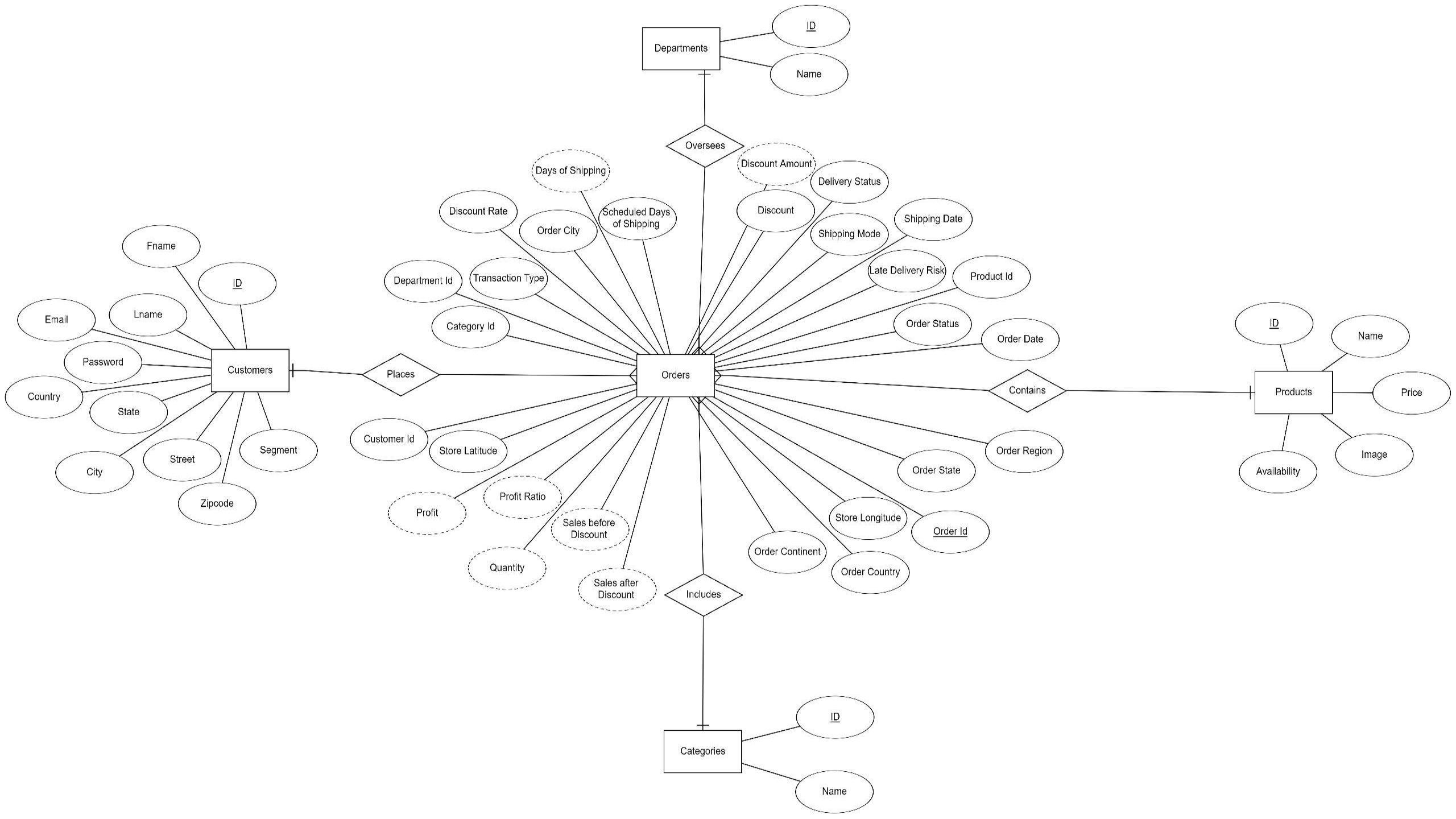
[60]:
```
dfcus.to_csv("Customers.csv", index = False)
HTML('<a href="Customers.csv" download="Customers.csv">Click here to download the CSV</a>')
```

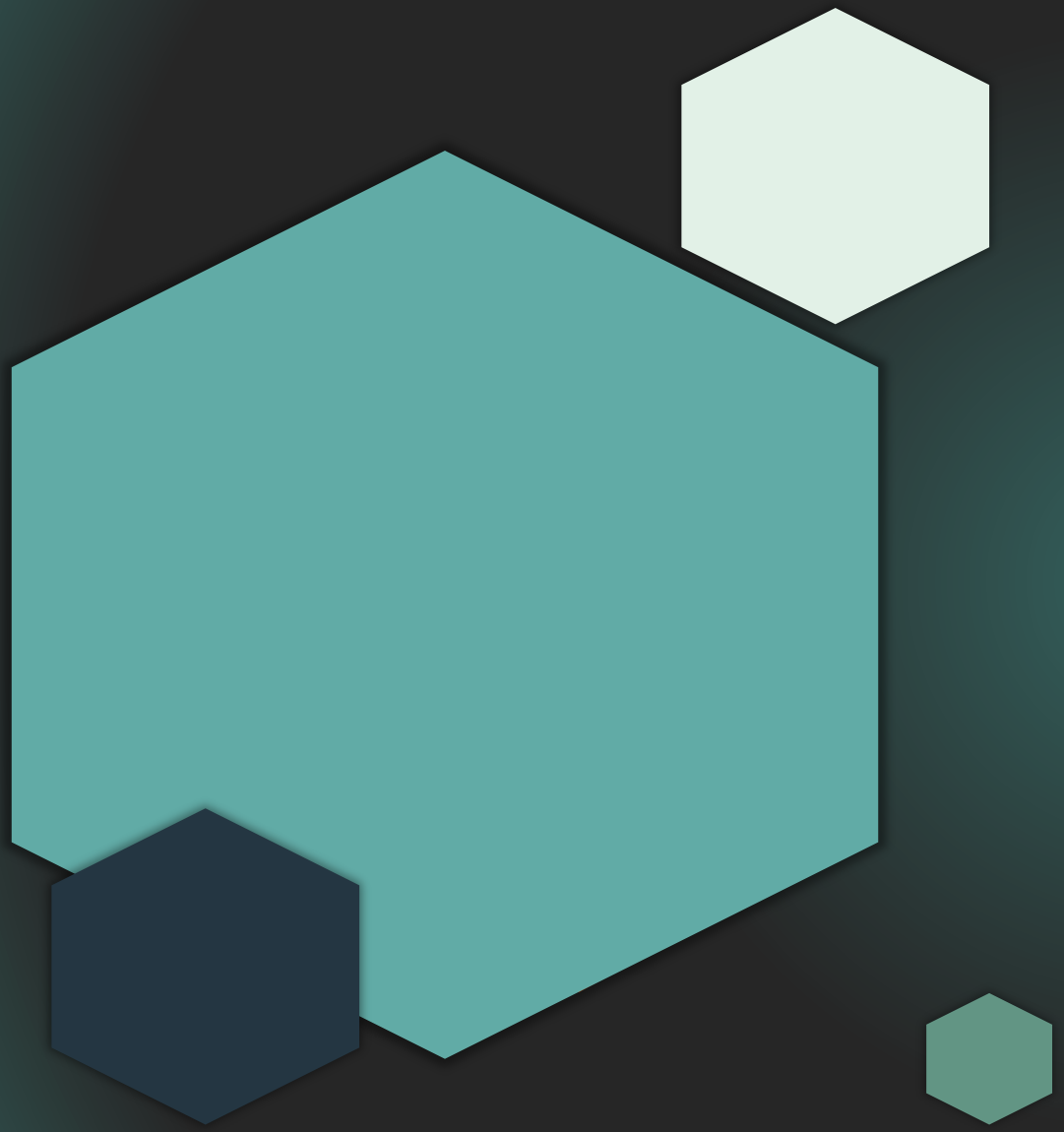
[60]: Click here to download the CSV

Cleaning & Modelling

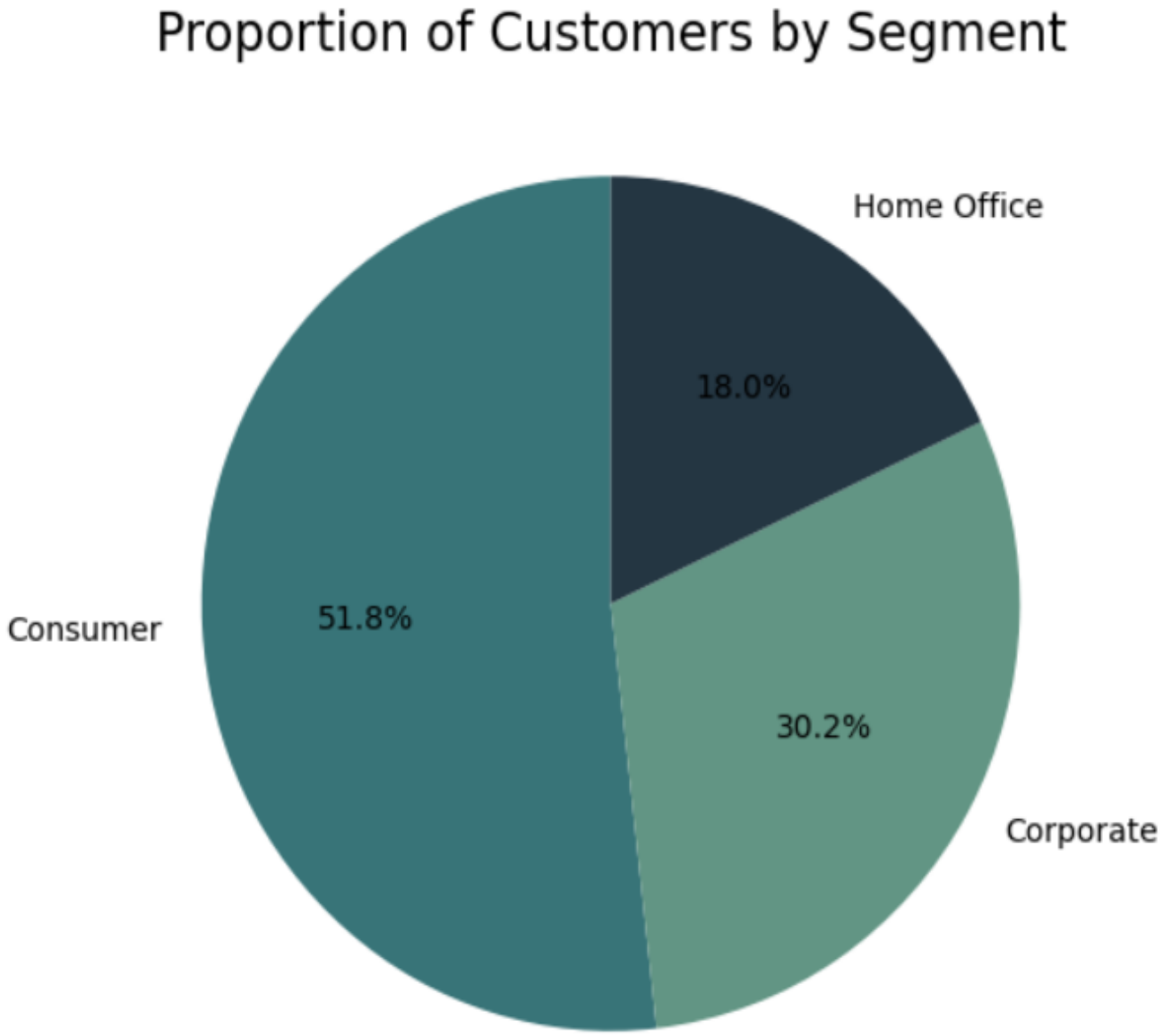
After applying all the previous steps to each table, we ended up with one fact table, which was the order table, and dimension tables, which included the customer table, department table, products table, and category table.

1. Order Table: 180,519 Row and 28 Column.
2. Customer Table: 20,652 Row and 11 Column.
3. Product Table: 118 Row and 5 Column.
4. Category Table: 51 Row and 2 Column.
5. Department Table: 11 Row and 2 Column.

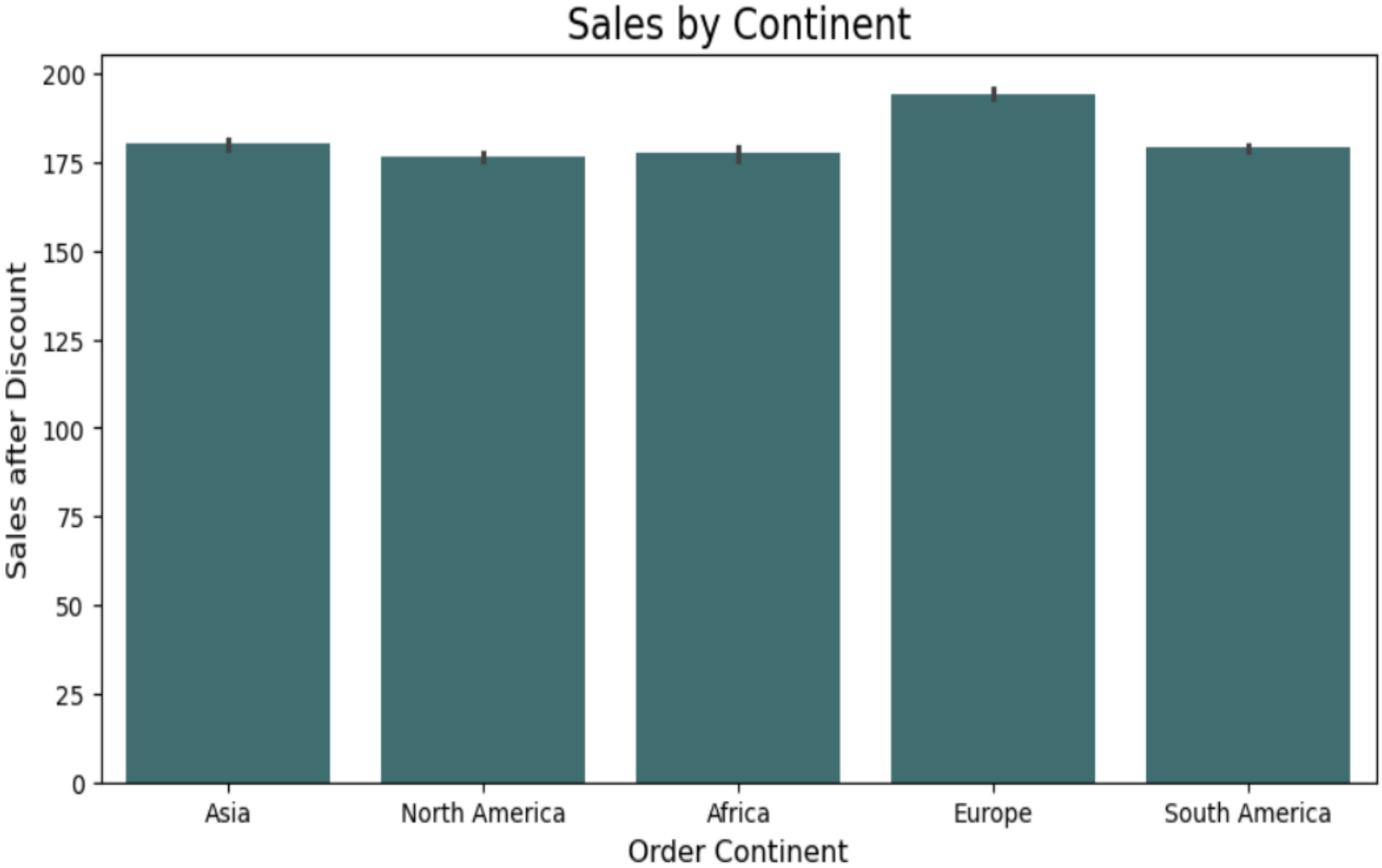

Departments
ID
Name
Oversees
Days of Shipping
Discount Amount
Delivery Status
Discount Rate
Scheduled Days of Shipping
Discount
Shipping Date
Order City
Shipping Mode
Fname
Late Delivery Risk
Product Id
Department Id
Transaction Type
ID
Email
Lname
Order Status
Order Date
Category Id
Password
Customers
Places
Orders
Contains
Products
ID
Name
Price
Country
State
Customer Id
Store Latitude
Order Region
Segment
Image
City
Street
Order State
Availability
Profit Ratio
Profit
Store Longitude
Order Id
Sales before Discount
Order Continent
Zipcode
Quantity
Order Country
Sales after Discount
Includes
ID
Categories
Name



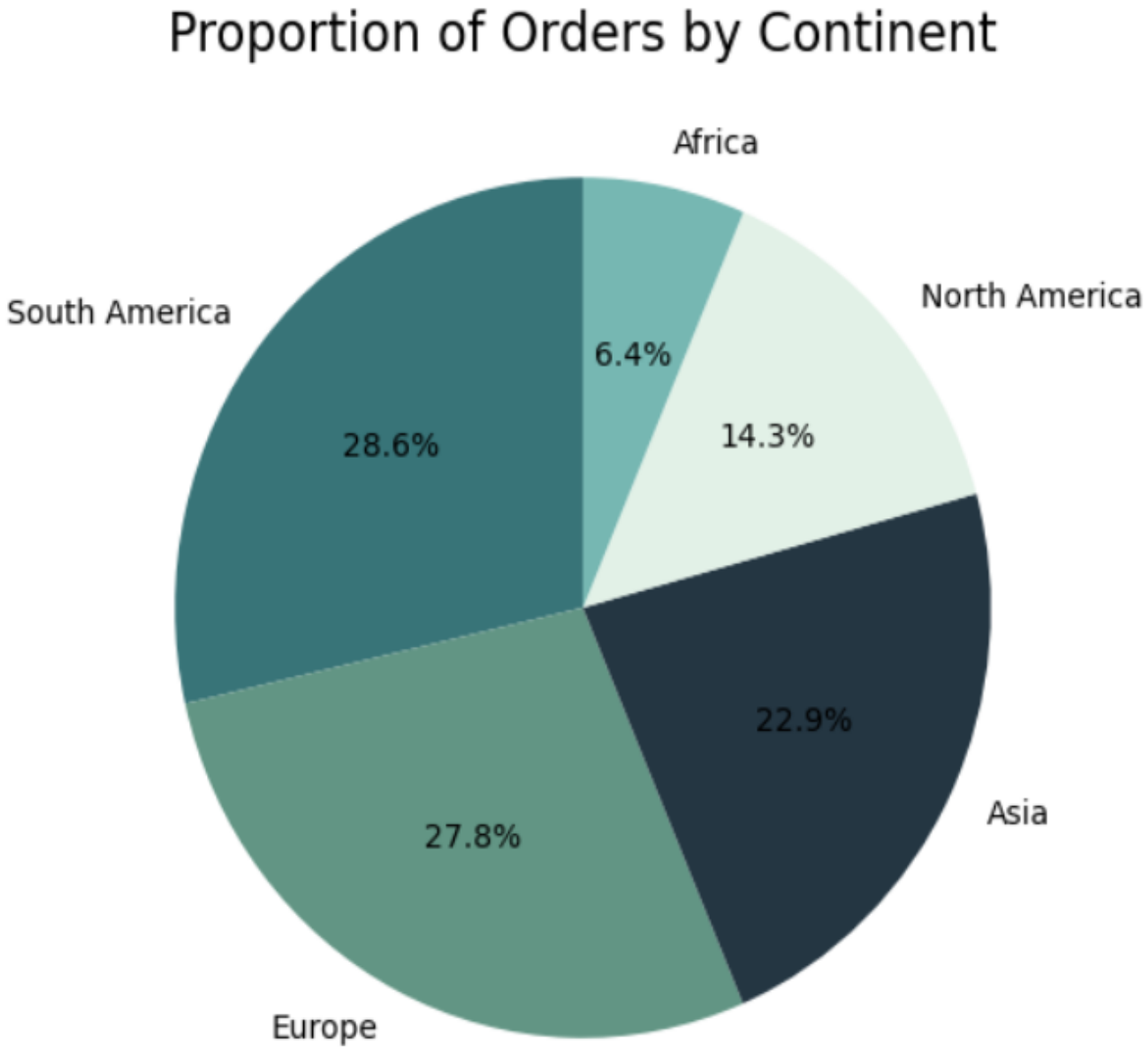
EDA with Python



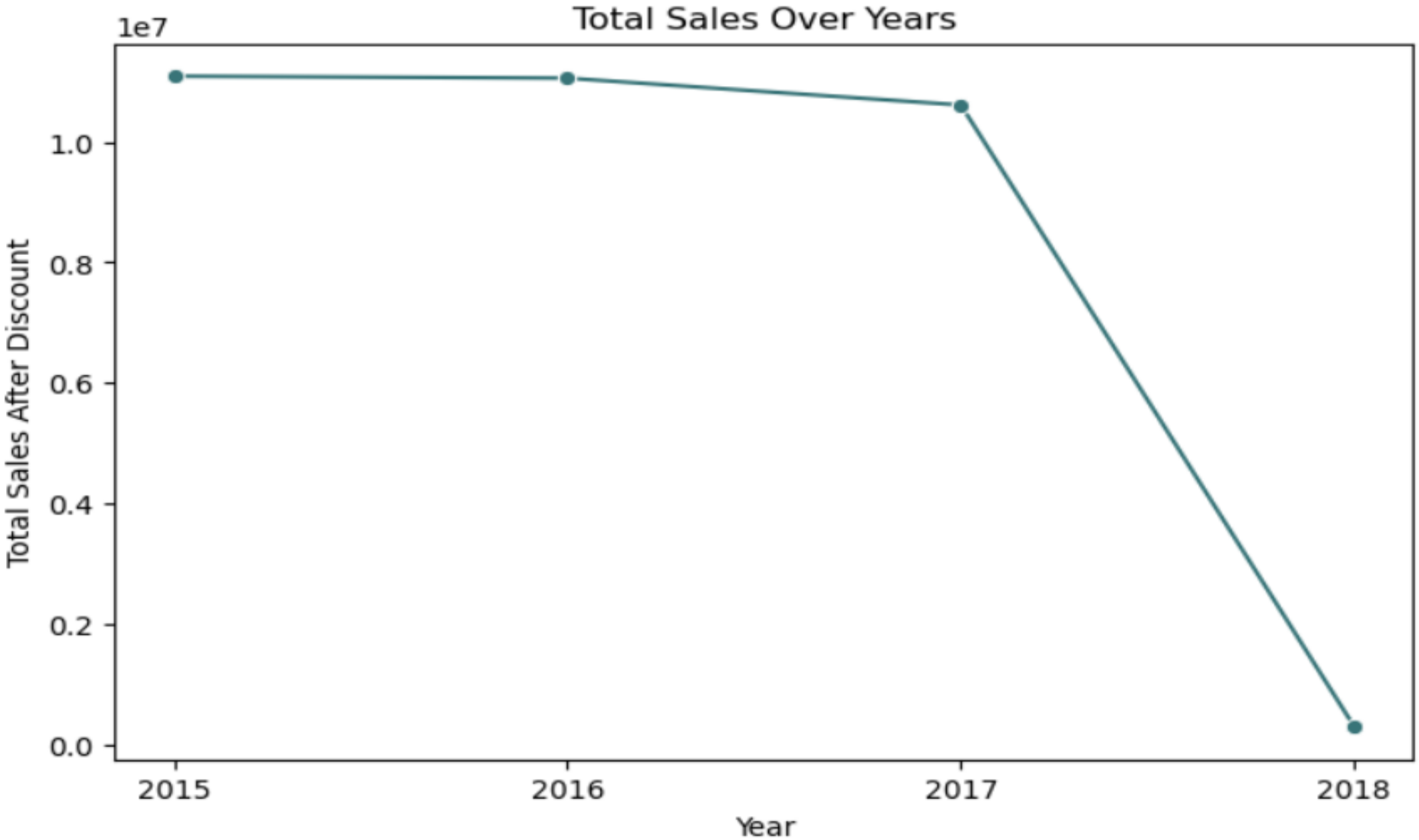
- The pie chart Proportion of Customers by Segment shows the data into Consumer, Corporate, and Home Office across the number of customers.



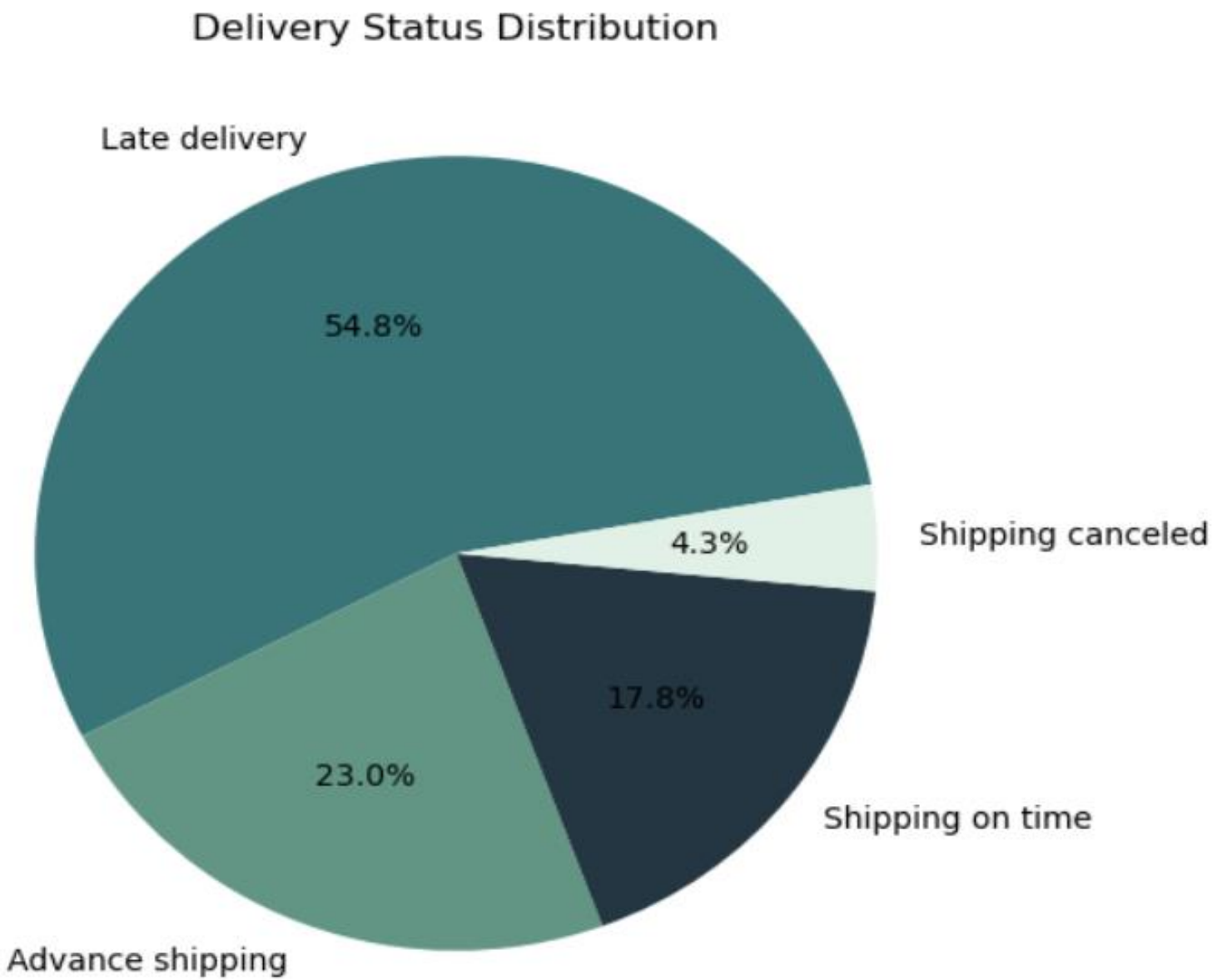
- The bar plot Sales by Continent shows the data into continents like North America, Europe, and Asia with its sales.



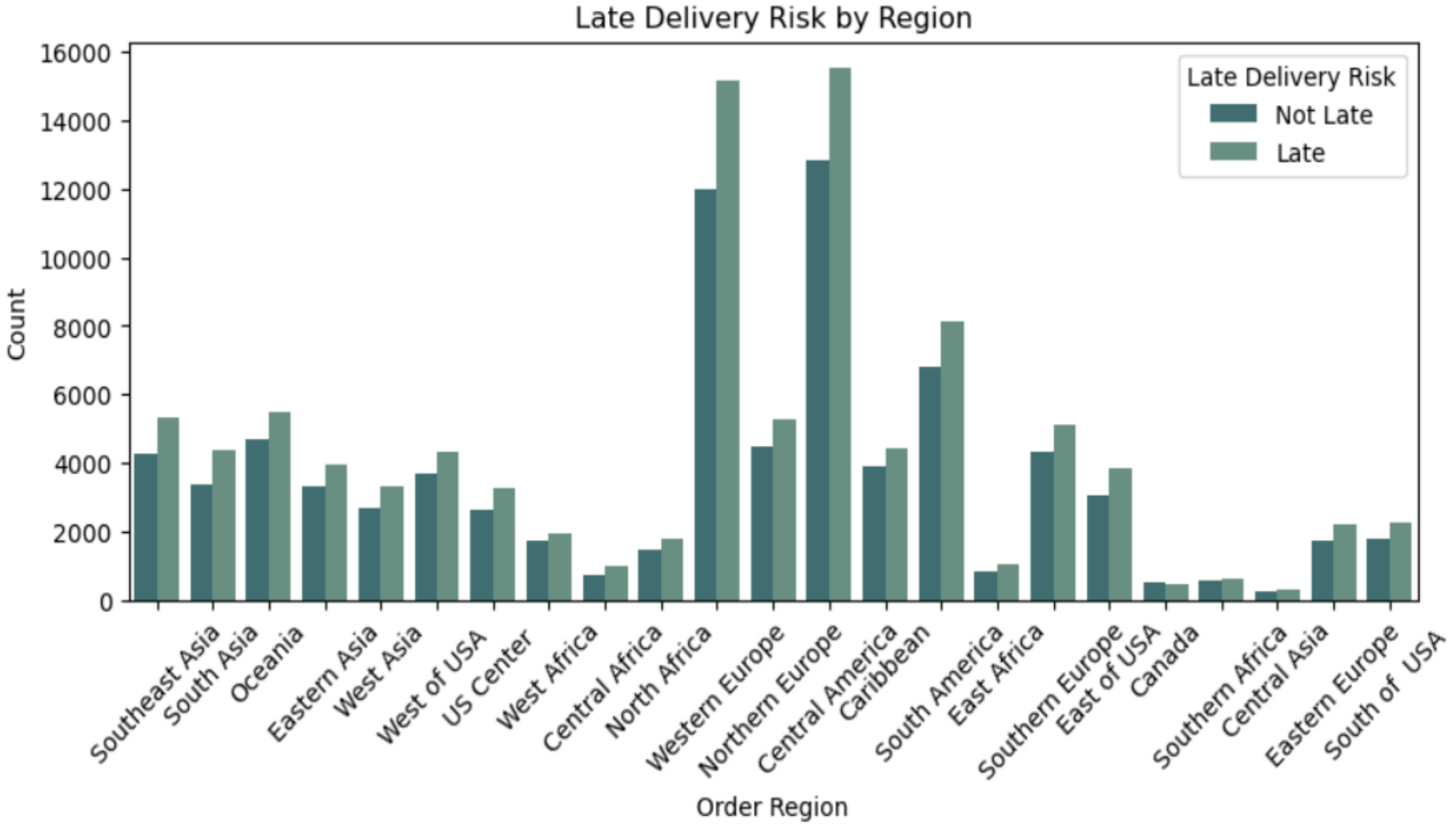
- The pie chart Proportion of Orders by Continent shows the data into North America, Europe, Asia, Africa, and South America across the number of orders.



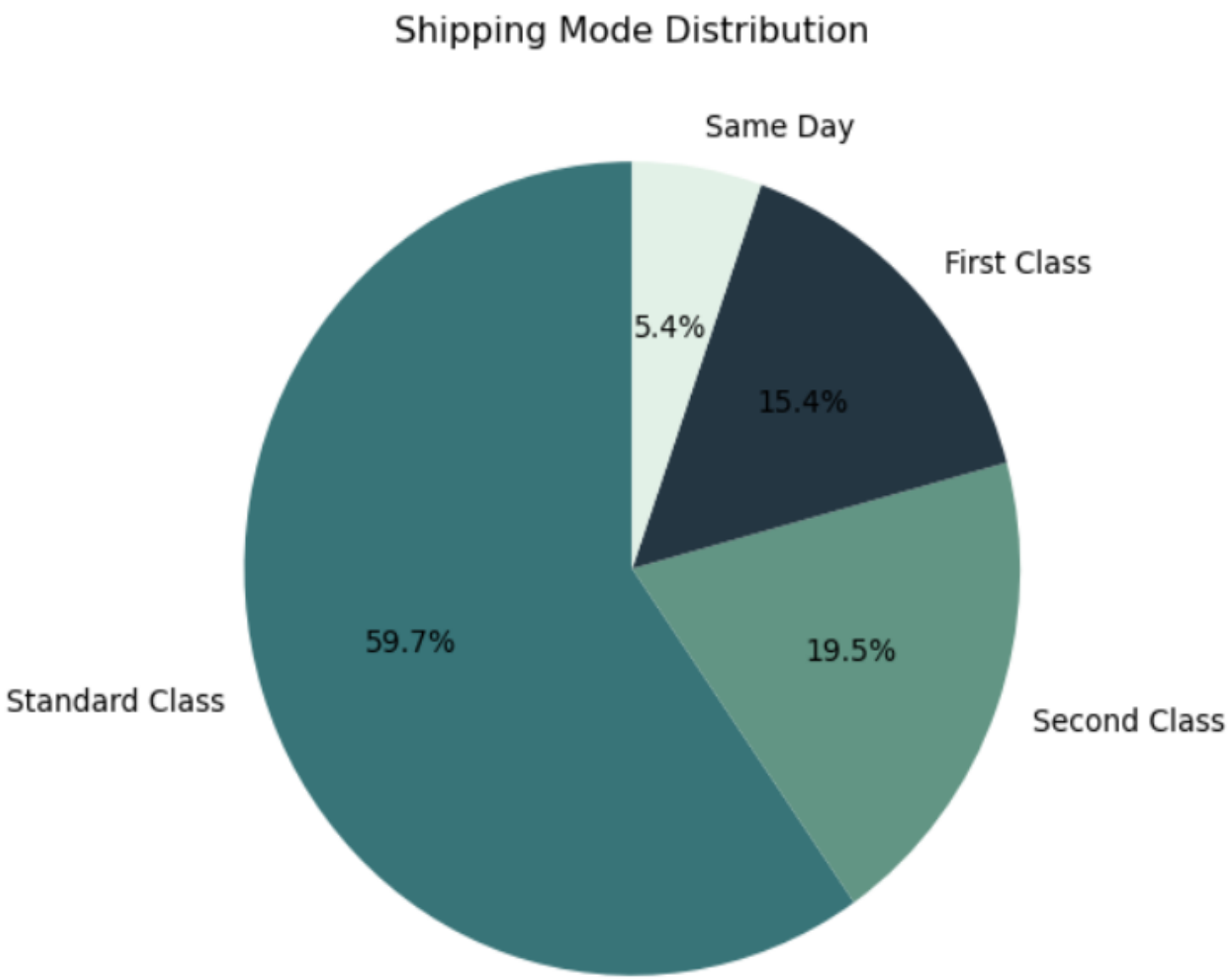
- The line plot Total Sales Over Time shows the sales by year.



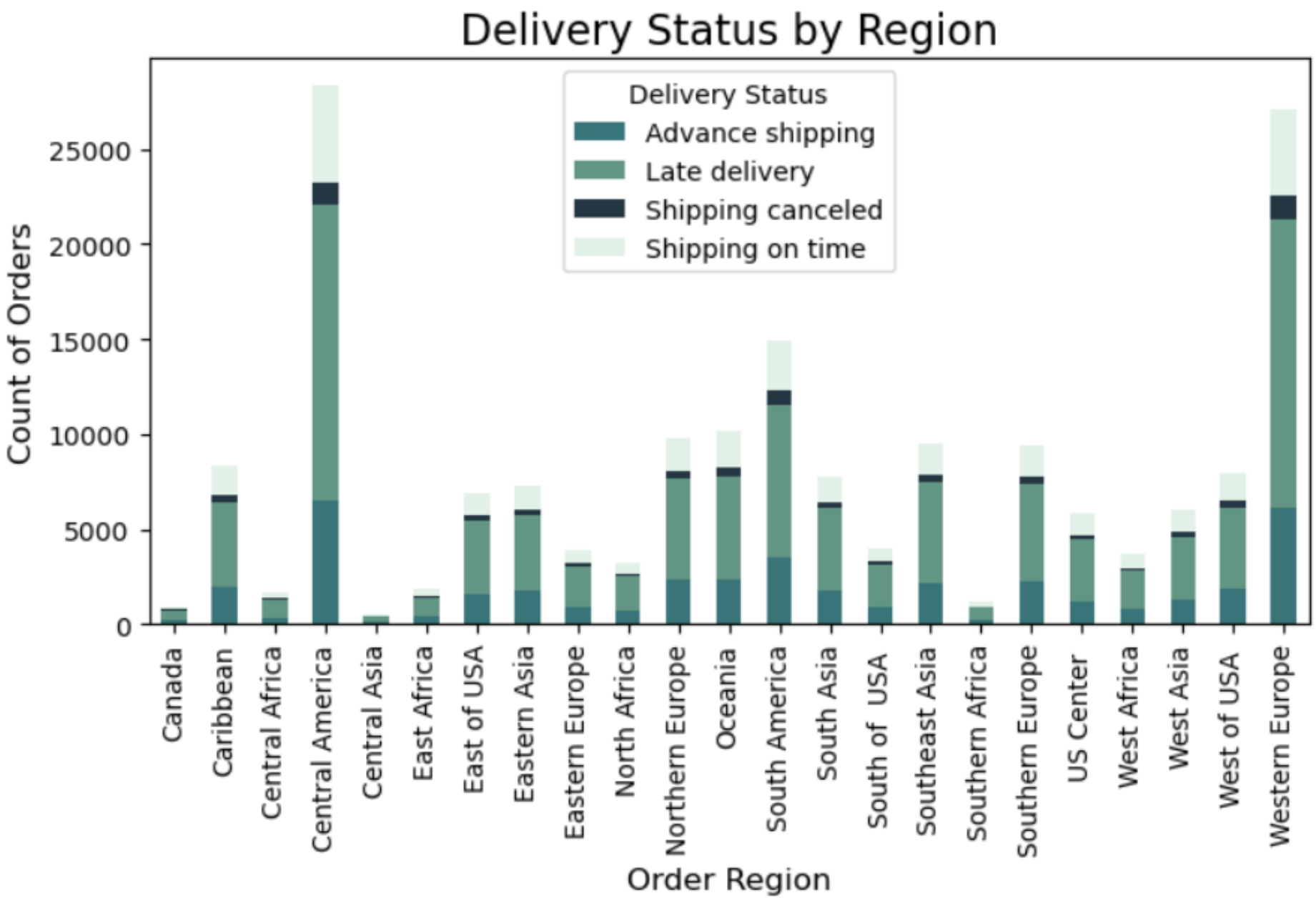
- The pie chart Delivery Status Distribution shows the data into On Time, Late, Advanced, and Canceled.



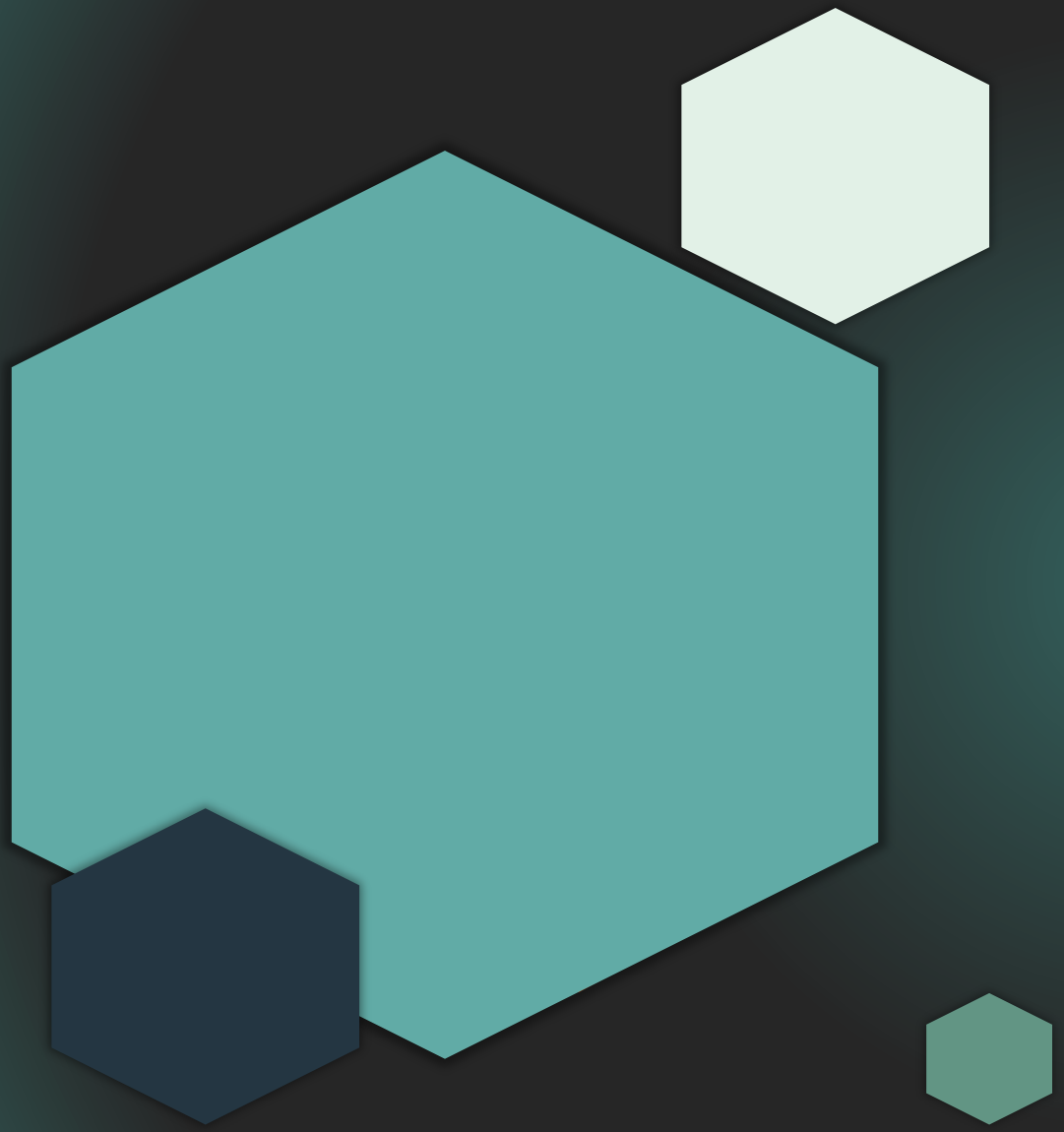
- The bar plot Late Delivery Risk by Region shows the data into regions such as North Aftica, Western Europe, and Central Asia with the late delivery risk.



- The pie chart Shipping Mode Distribution shows the data into Standard Shipping, First class, Second class, and Same day across the number of orders.



- The stacked bar plot Delivery Status by Region shows the data into Advanced, Late, On Time, and Cancelled orders across different regions.



Blueprint

Dashboard Name

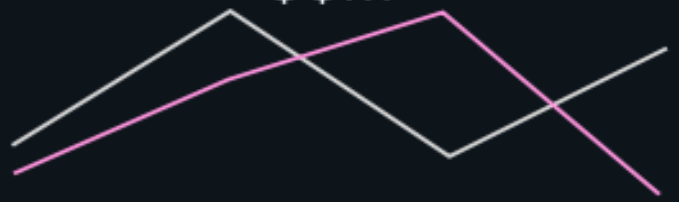
Pages

Filters

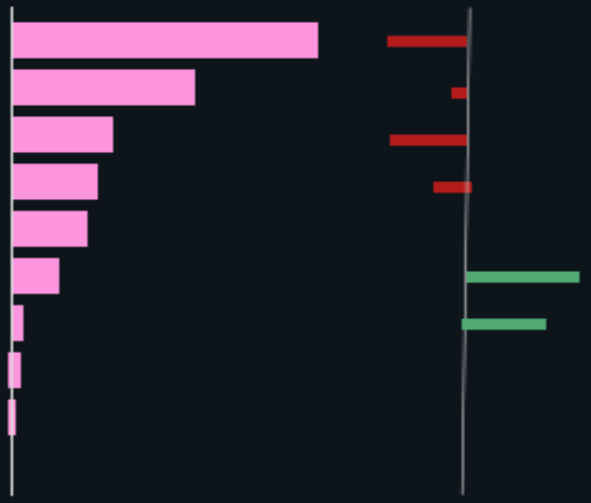
Cards

Sales KPI

$$M



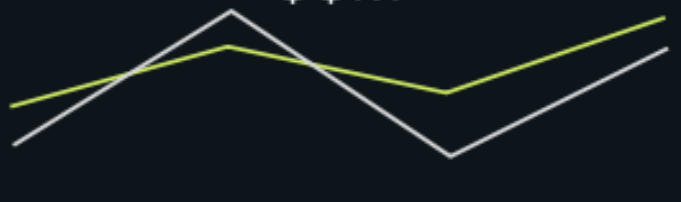
Department



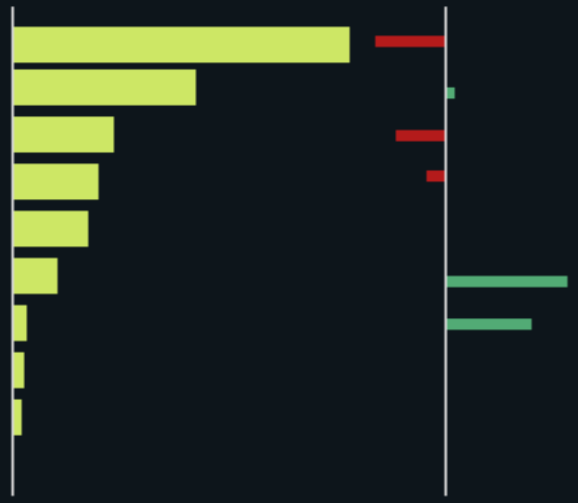
Top 5 Products

Profit KPI

$$M



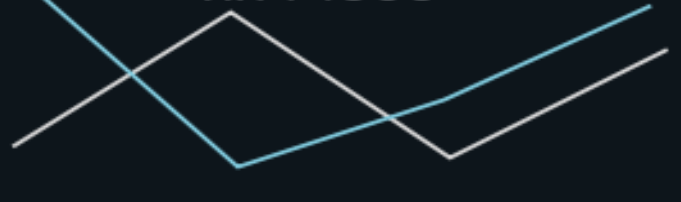
Department



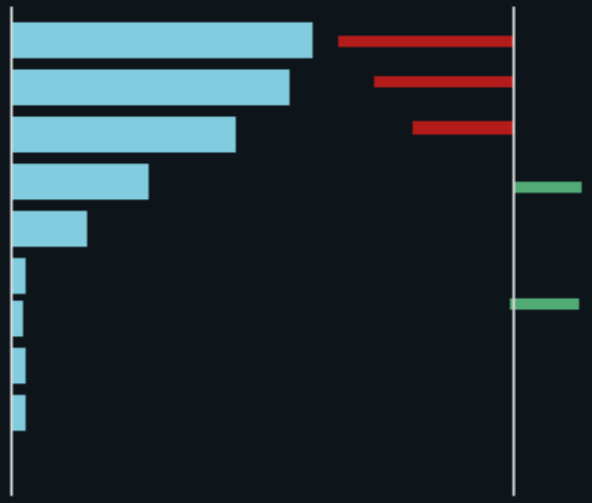
Top 5 Products

Quantity KPI

xx Piece



Department

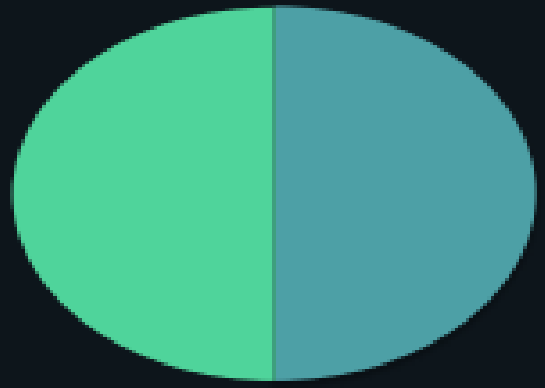


Top 5 Products

Dashboard Name Pages Filters

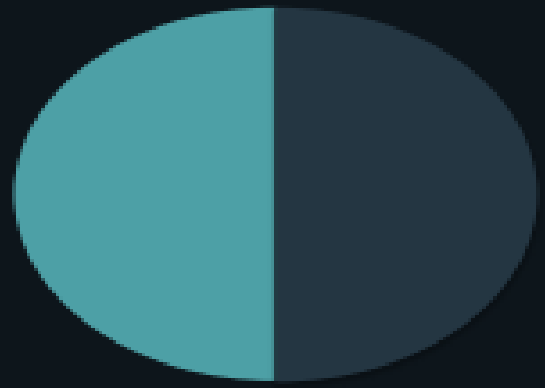
Cards

Orders by Shipping Mode

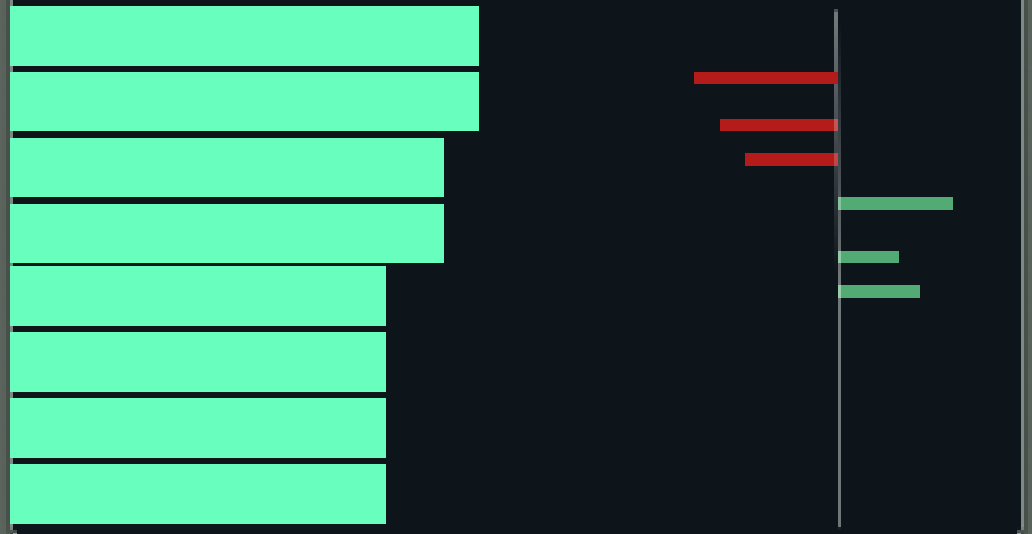


Orders by Delivery Risk

Orders by Delivery Status



Delivery Days & Scheduled by Department



Average Delivery Days by Country

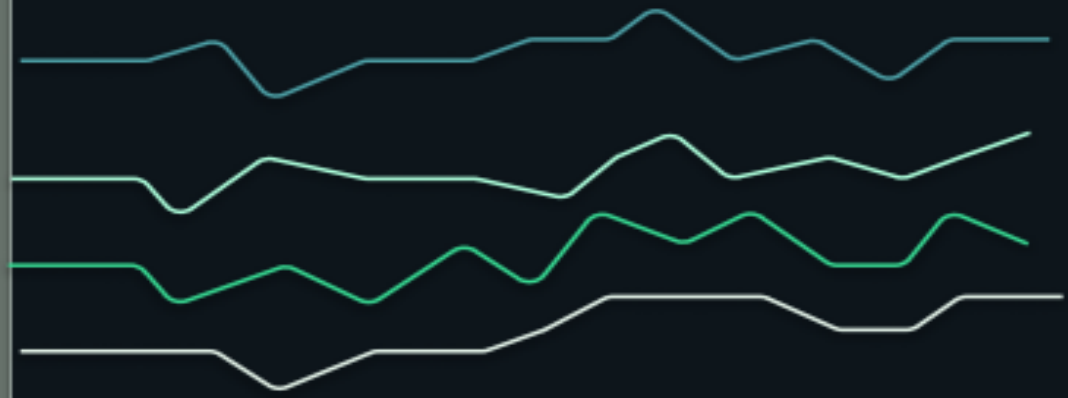
Dashboard Name

Pages

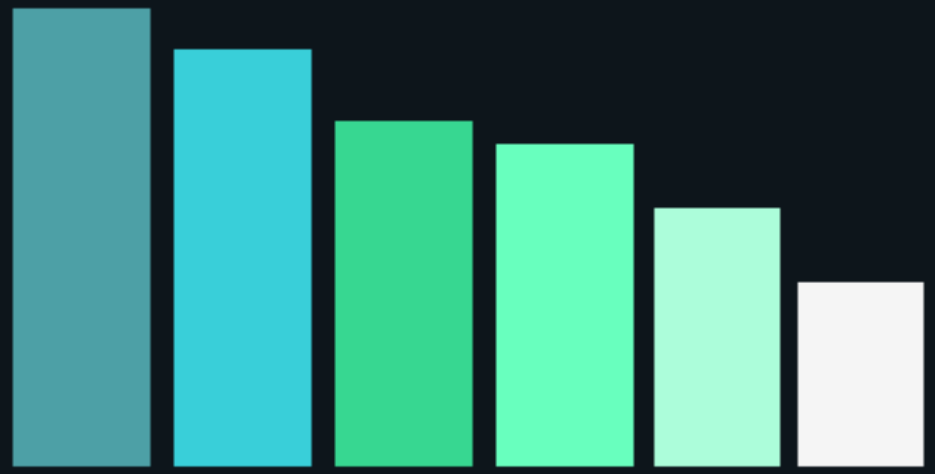
Cards

Number of Customers by Segment

Purchase Hours by Transaction Type

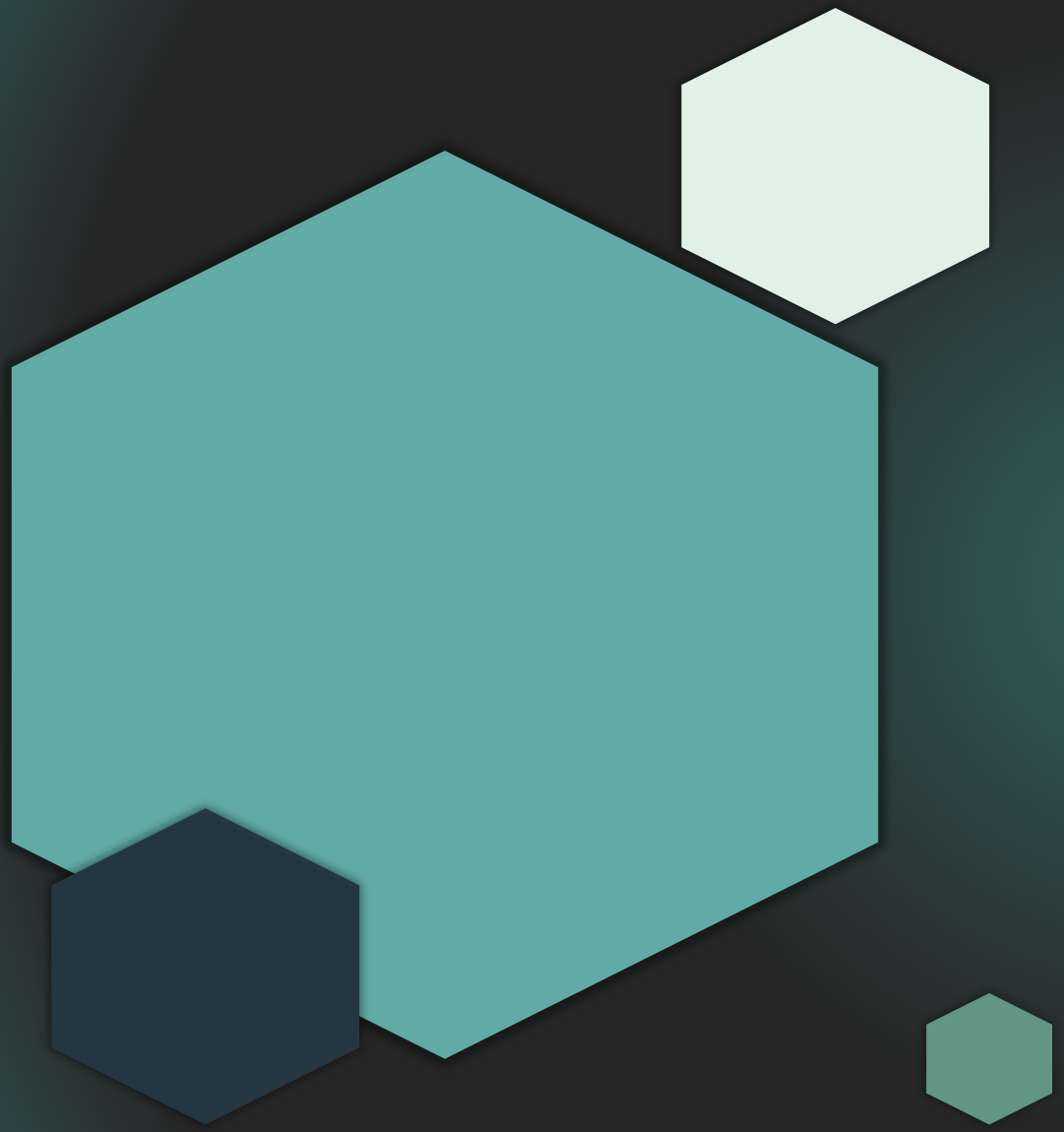


Customer Distribution by No. of Orders



Top 10 Customers by Profit

Customers	Sales	Profit	Quantity	# of Orders



Dashboards

Sales Dashboard | 2017

Sales | Delivery | Customers

$33.1M Total Sales

$4.0M Total Profit

181K Total Orders

384K Total Quantity

20.7K Total Customers

11 Total Departments

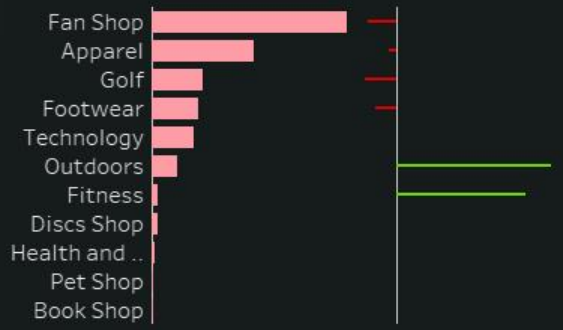
118 Total Products

Sales KPI

$ 10.65M

▼ -3.73% vs. PY

Departments



Top 5 Products

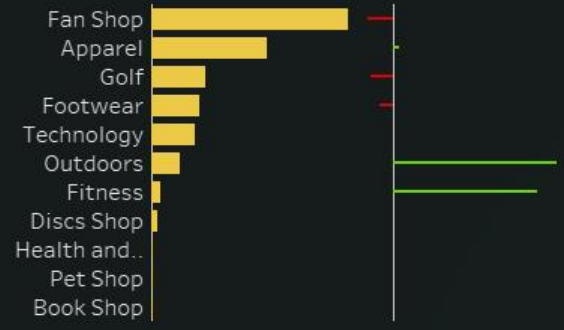
Field & Stream Sportsman 16 ..	$1,735K
Perfect Fitness Perfect Rip Deck	$1,081K
Diamondback Women's Serene..	$1,055K
Nike Men's Free 5.0+ Running ..	$919K
Nike Men's Dri-FIT Victory Golf..	$785K

Profit KPI

$1.31M

▼ -0.36% vs. PY

Departments



Top 5 Products

Field & Stream Sportsman 16 ..	$219K
Perfect Fitness Perfect Rip Deck	$149K
Nike Men's Free 5.0+ Running ..	$114K
Diamondback Women's Serene..	$106K
Nike Men's Dri-FIT Victory Golf..	$104K

Quantity KPI

107.2K Piece

▼ -22.05% vs. PY

Departments

Top 5 Products

Perfect Fitness Perfect Rip Deck	20.1K
Nike Men's Dri-FIT Victory Golf..	17.5K
O'Brien Men's Neoprene Life V..	15.7K
Nike Men's Free 5.0+ Running ..	10.2K
Under Armour Girls' Toddler S..	8.8K

Delivery Dashboard

Sales | Delivery | Customers

$33.1M
Total Sales

$4.0M
Total Profit

181K
Total Orders

384K
Total Quantity

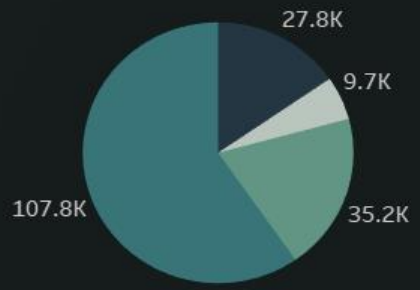
20.7K
Total Customers

11
Total Departments

118
Total Products

Orders by Shipping Mode

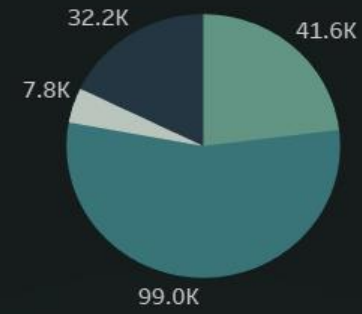
Same Day | First Class | Second Class | Standard



Orders by Delivery Risk

Orders by Delivery Status

Canceled | Late | On Time | Advanced



Delivery Days & Scheduled by Department

Average Delivery Days by Country

Customers Dashboard

Sales | Delivery | Customers

$33.1M
Total Sales

$4.0M
Total Profit

181K
Total Orders

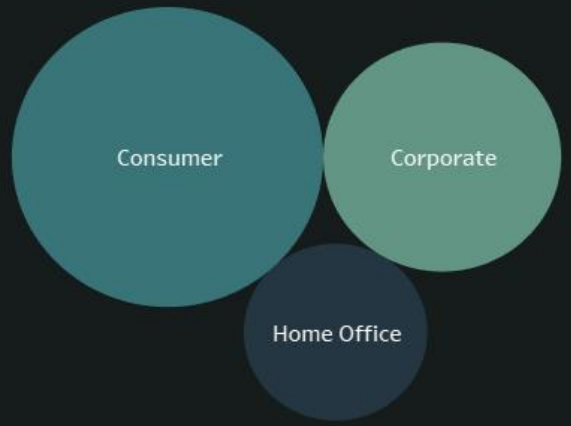
384K
Total Quantity

20.7K
Total Customers

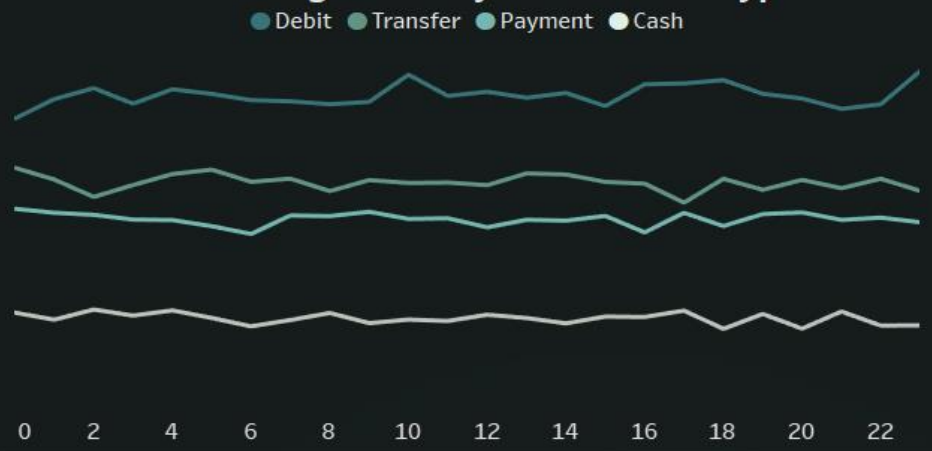
11
Total Departments

118
Total Products

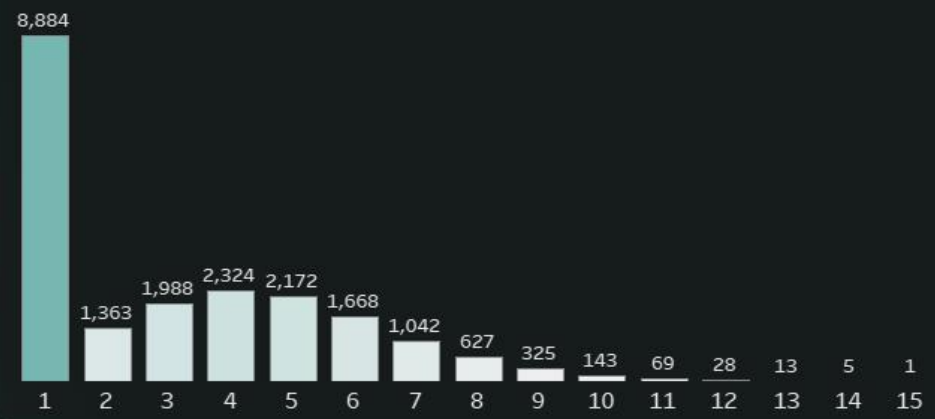
Number of Customers by Segment



Purchasing Hours by Transaction Type



Customer Distribution by Orders



Top 10 Customers by Profit

Customer ID	Customers	Sales	Profit	Quantity	Orders
2641	Betty Spears	$8.22K	$2.44K	90	43
1657	Betty Phillips	$8.17K	$2.20K	111	42
9833	Jacob Smith	$5.39K	$1.94K	48	24
2626	Laura Smith	$5.48K	$1.93K	55	27
5004	Dorothy Friedman	$7.50K	$1.92K	114	45
3735	Mary Lewis	$5.53K	$1.91K	61	27
749	Jesse Matthews	$6.95K	$1.86K	64	38
5560	Mary Rodriguez	$6.12K	$1.83K	74	29
10967	Alexander Cunningham	$5.18K	$1.82K	64	27
5053	Mary Smith	$6.70K	$1.81K	67	33

Analysis Report

Sales Dashboard

- The dashboard shows our business performance for every selected year, focusing on sales, profit, and quantity sold. We made $10.65M in sales in 2017, but that's a decline of 3.73% from last year, which is a bit worrying. The Fan Shop and Apparel departments did the best. However, they did better last year.

- For profit, we earned $1.31M in 2017, showing a slight drop of 0.36%. This is similar to the sales trend, meaning our profits depend on how well we sell. The same departments lead in profit, but areas like Book shop and Pet shop are struggling. Additionally, we sold 22.05% fewer products, which raises concerns about customer demand or stock issues. This decline suggests we need to work harder to keep our customers.

- To improve, we should find out why the quantity sold has dropped so much. We should focus on our successful departments like Fan Shop and Apparel. Focusing on our best departments can also help reduce this declines and improve the business. With these changes, we can aim for better growth in the future.

Delivery Dashboard

- The Delivery Dashboard provides a clear view of our delivery operations and reveals several important findings. One major issue is that nearly 60% of our orders are delivered late, which can frustrate customers and hurt our business. While a large number of deliveries are on time, we need to focus on improving our late deliveries. This could involve better planning and partnerships with shipping companies to ensure items arrive when promised.

- Another finding is that most customers use standard shipping, which is the cheapest option. However, this can lead to longer delivery times. We should promote faster shipping options, like First Class or Same Day delivery, to improve customer satisfaction. Offering these methods could encourage more customers to choose them, helping to reduce delays.

- Additionally, the dashboard shows that average delivery time is always less than scheduled. By identifying and addressing the reasons for these delays, we can improve efficiency. The geographical breakdown also indicates that delivery times can vary based on location. Targeting improvements in areas with longer delivery times can help us serve our customers better. Overall, these insights can guide us in making better decisions to enhance our delivery performance and boost customer satisfaction.

Customer Dashboard

- The Customer Dashboard gives us important information about our customers and their buying habits. One key finding is that many customers only make one purchase. This shows we need to work on getting them to come back for more. Building customer loyalty is crucial because it's often easier and cheaper to keep existing customers than to find new ones.

- Another important point is that a small number of customers make up an important part of our profits and reputation. This means we should pay special attention to these customers to keep them happy. By offering them exclusive deals or personalized services, we can encourage them to shop with us again and again, which will help our business grow.

- Lastly, the Home Office segment is much smaller than the Consumer and Corporate segments. Overall, the dashboard shows us where we can improve, especially in keeping customers, supporting our best ones, and reaching out to new segments.

Recommendations

Recommendations

- After conducting the analysis, we recommend that the company focus on a specific value proposition, such as fitness, rather than expanding into unrelated departments like book shop, pet shop, Discs shop, Health and Beauty, and technology. These additions have made the company's value proposition unclear, resulting in a loss of sales and customers. The departments that should be prioritized are apparel, footwear, golf, fitness, outdoors, and potentially the fan shop, while the rest should be closed. Then, we need to hire a delivery agency because the company takes more time than scheduled to deliver to the customers.

- We also need to rebrand the company to match its new focus. The rebranding should highlight the main areas: apparel, footwear, golf, fitness, and the outdoors, while removing the other, less relevant departments. The goal is to create a clear and simple brand that connects with customers and helps the company grow in these key areas.

- Lastly, we need to focus our marketing on our existing customers who align with our value proposition and segments. At the same time, we should target new customer segments through strategies like social media marketing. This will help us strengthen relationships with loyal customers, attract new ones who fit our brand, and build a trusted, well-known brand in the market.

Ministry of Communications and Information Technology

رواد مصر الرقمية

Thank you

Technical Provider: AST

Instructor: Ahmed Abdellatif