Super Store Analysis
Project Presentation

Prepared by/ Mohamed Ali Gawiesh

Table of contents

01

About Data

The Data

Data source

Data was obtained from [Kaggle, Superstore Dataset, CSV file]

What is the Data?

The dataset used for this project is from a large retail store that includes a wide variety of sales and customer data. It contains information on orders, products, customers, shipping details, and financial metrics.

Columns description

Sales & Profit

Columns for **Sales** and **Profit** measure transaction value and profitability

Customer & Product Info

Columns like **Customer ID**, **Product Category**, and **Sub-Category** provide details about buyers and the types of products sold.

Order & Shipping

Order Date, **Ship Date**, and **Region** help track order timelines and geographic distribution.

02

Objective of the project

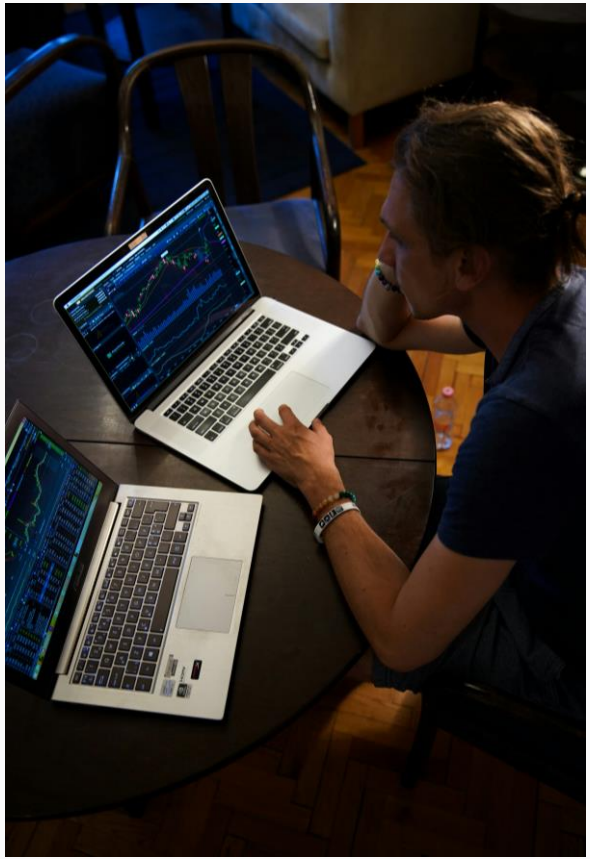
Objectives

Our aim

The primary aim of the project is to analyze Superstore's sales data to identify trends and patterns that can help improve business performance and customer satisfaction. The analysis focuses on understanding customer behavior, optimizing sales strategies, and improving profitability.

The goal

Customer Behavior Analysis: Analyze customer purchasing patterns to understand preferences and identify the most profitable customer segments.
Sales Performance Optimization: Examine product sales performance across different regions and categories to identify opportunities for growth and areas requiring improvement.
Profitability Insights: Determine key factors contributing to high or low profit margins and suggest strategies to enhance profitability.
Operational Efficiency: Assess shipping methods and order fulfillment times to improve delivery efficiency and customer experience.



03

Data Cleaning

Data Cleaning

1

Handling Missing & duplicated Values

Total Records: 9994 rows
Missing Records: 0
Duplicated values: 0

2

Data Formatting

Data types for columns such as **order** and **shipping date** were corrected for easier analysis.

3

Dropped Columns

the following columns were dropped as they did not contribute to the analysis or provided redundant information:

- Row ID
- Order ID
- Customer ID
- Country
- Product ID
- Postal Code

Cleaning Processes

We use Python to clean the data.

1

```
[ ]  df.duplicated().sum()
0
```

```
[ ]  df.isna().sum()
0
```

2

```
[ ]  df['Order Date'] = pd.to_datetime(df['Order Date'], format="%m/%d/%Y", errors='coerce')
     df['Ship Date'] = pd.to_datetime(df['Ship Date'], format="%m/%d/%Y", errors='coerce')
```

3

```
[ ]  columns_to_drop = ['Row ID', 'Order ID', 'Customer ID', 'Country', 'Product ID', 'Postal Code']
     df.drop(columns=columns_to_drop, axis=1, inplace=True, errors='ignore')
```

04

Exploratory Data Analysis (EDA)

$2,297,201

Total sales of the Superstore

37,873 units

Total quantity sold

15.6%

Average Discount

$286,397

Total profit of the superstore

Descriptive Statistics

$ 0.444

Minimum Sales

$ 22,638.48

Maximum Sales

$ 229.86

Average Sales

Additional information

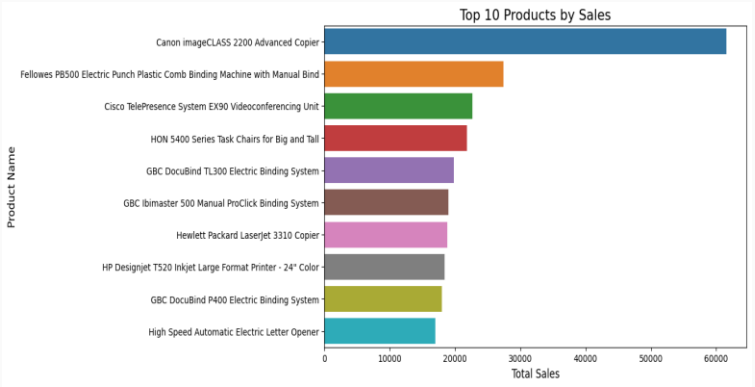
There are 1546 products achieve profit.

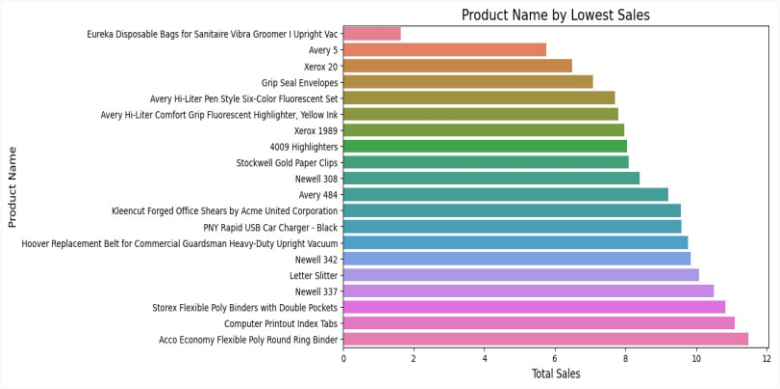
There are 3 products without Profit or loss (at BOP):

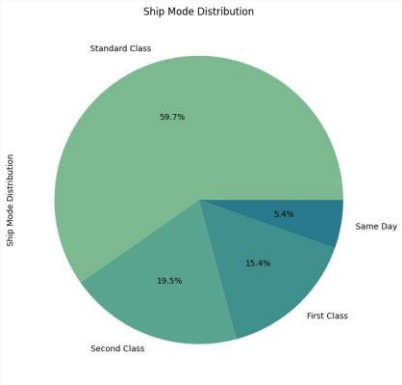
- Alliance Big Bands Rubber Bands, 12/Pack.
- Avery Trapezoid Extra Heavy Duty 4" Binders.
- Belkin OmniView SE Rackmount Kit.

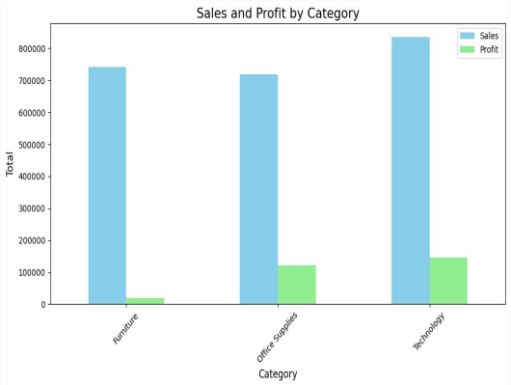
There are 301 products achieve Losses.

Initial visual analysis









Insights from initial visual analysis

The top-selling category was Technology, and it is also the highest profit.
The lowest selling category was office supplies, but the lowest profit is Furniture.

The most frequently used ship mode is standard class.
The least frequently used ship mode in the same day.

In the West the Sales are the Highest (31.6% of total sales).
In the South the sales are least (17.1% of total sales).

New York city was the top-selling city, and California was the top-selling state.
San Antonio is the lowest selling city, and North Dakota is the lowest selling state.

05

Analytics

We divided the analysis into three sections:

1. Customer Behavior Analysis
2. Sales Performance Analysis
3. Shipping and Delivery Analysis

Customer Behavior Analysis

In this section, we analyzed various aspects of customer interactions, including buying patterns, preferred product categories, and frequency of purchases. This helped us identify customer segments, understand their needs, and develop strategies to enhance customer engagement and retention.

We identified key questions related to customer behavior, provided detailed answers, and offered recommendations to improve customer engagement and experience.

Customer Behavior Analysis – Questions and Answers

1

What is the most active customer segment?

The **consumer** segment is the most active, making up **52%** of total sales.

2

Who is the top customer by sales?

Tamara Chand is the top customer by sales.

3

What is the distribution of customers by state?

Wyoming has the least number of customers, while **California** has the most.

Customer Behavior Analysis – Questions and Answers

4

Which cities have the most customers?

New York City has the highest number of customers.

5

What are the busiest months in terms of customer activity?

September, **November**, and **December** are the most active months, while **January** and **February** are the least active.

Recommendations

1. Enhance engagement with the consumer segment through exclusive deals and loyalty programs.
2. Offer tailored promotions to retain top customers like Tamara Chand.
3. Develop marketing strategies for low-performing states like Wyoming.
4. Focus advertising efforts in high-customer cities like New York City
5. Maximize revenue during peak months with holiday promotions and boost engagement in quieter months.

Sales Performance Analysis

We examined sales trends over time, highlighting peak periods, best-selling products, and regions with the highest sales. This analysis allowed us to assess profitability, monitor the impact of discounts and promotions, and identify areas for potential sales growth.

We identified key questions related to sales performance, provided detailed answers, and offered recommendations to improve sales performance.

Sales Performance Analysis – Questions and Answers

1

What is the top product by sales?

The top product is the **Canon imageCLASS 2200 Advanced Copier**.

2

Which region has the highest sales?

The **West** region has the highest sales figures.

3

Which state has the lowest sales?

North Dakota has the lowest sales figures among the states.

Sales Performance Analysis – Questions and Answers

4

What category generates the most sales?

The **Technology** category generates the most sales overall.

5

What is the trend in sales performance?

Sales are generally **increasing** but **fluctuate** seasonally.

6

What category has the least sales?

The **Office Supplies** category has the least sales compared to others.

Recommendations

1. Enhance marketing efforts for the Canon imageCLASS 2200 Advanced Copier to boost visibility.
2. Increase promotional activities in the West region to maximize growth potential.
3. Develop strategies to improve sales in North Dakota by identifying customer needs.
4. Invest in the Technology category to expand product lines and promotions.
5. Monitor sales trends closely to anticipate fluctuations and adjust strategies.
6. Boost Office Supplies sales through attractive bundling offers.

Shipping and Delivery Analysis

This part focused on the efficiency of the shipping and delivery process. We evaluated shipping times, delivery methods, and the impact of delayed deliveries on customer satisfaction. By identifying bottlenecks and areas for improvement, we aimed to enhance operational efficiency and reduce delivery costs.

We identified key questions related to Shipping and delivery, provided detailed answers, and offered recommendations to improve shipping.

Shipping and Delivery Analysis – Questions and Answers

1

What is the latest shipping method?

Standard Class is the Most time consuming shipping method, taking an average of **5 days**.

2

Which region has the most average shipping days?

The **Central** region has the highest average shipping days.

3

What trend is observed in average shipping days?

Average shipping days are nearly **4**, but they increased to **5** days after the **4th quarter of 2017**.

4

Which state has the most and the least shipping days?

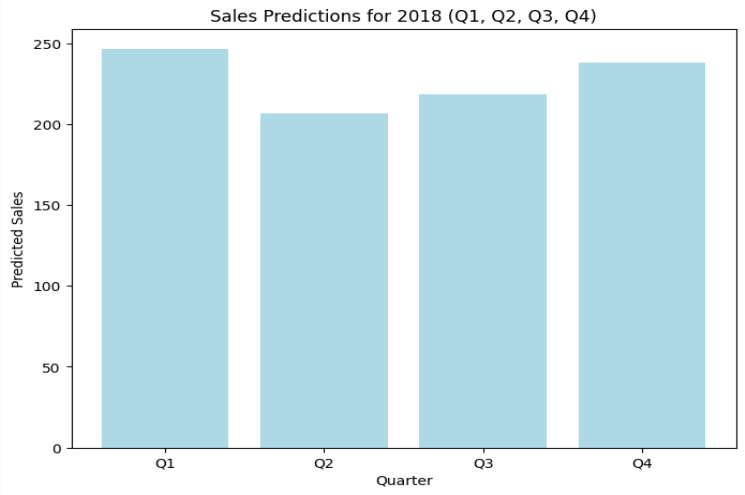
The **District of Columbia** has the most shipping days, while **North Dakota** has the least.

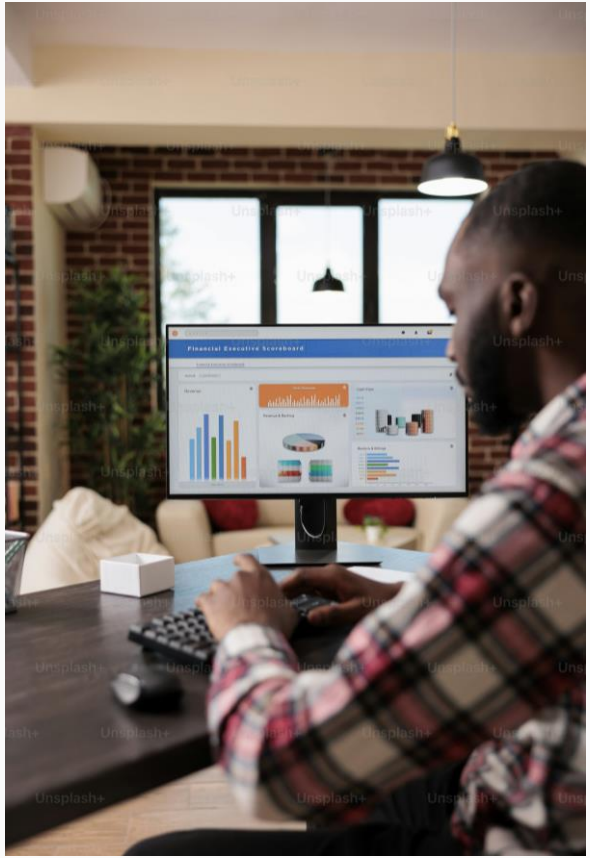
Recommendations

1. Optimize shipping processes to reduce delivery times in the Central region.
2. Enhance shipping options to improve service in the District of Columbia.
3. Investigate the increase in shipping days post-Q4 2017 for potential improvements.
4. Monitor shipping performance regularly to ensure timely deliveries across all regions.

Forecasting Phase

In our project, we implemented machine learning techniques using Python to help us forecast the sales for the upcoming year, 2018. By analyzing historical data and trends, we trained a model to predict the sales for each quarter of 2018.





06

Dashboards

Dashboard 1: Customer Behavior Analysis Dashboard

Description: The Customer Behavior Analysis Dashboard provides a comprehensive view of customer purchasing patterns, preferences, and engagement metrics. It analyzes key factors such as product categories, demographics, and seasonal trends to understand how customers interact with our offerings. By visualizing customer behavior, this dashboard enables us to tailor marketing strategies, enhance customer satisfaction, and improve retention rates.

Customer Behavior Dshboard

Sales Performance | Shipping and Delivery

No. of customers	Total Sales	Total Profit
793	2,297,201	286,397

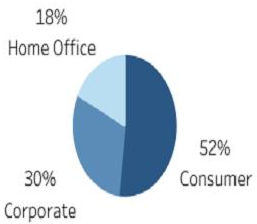
Segment

All

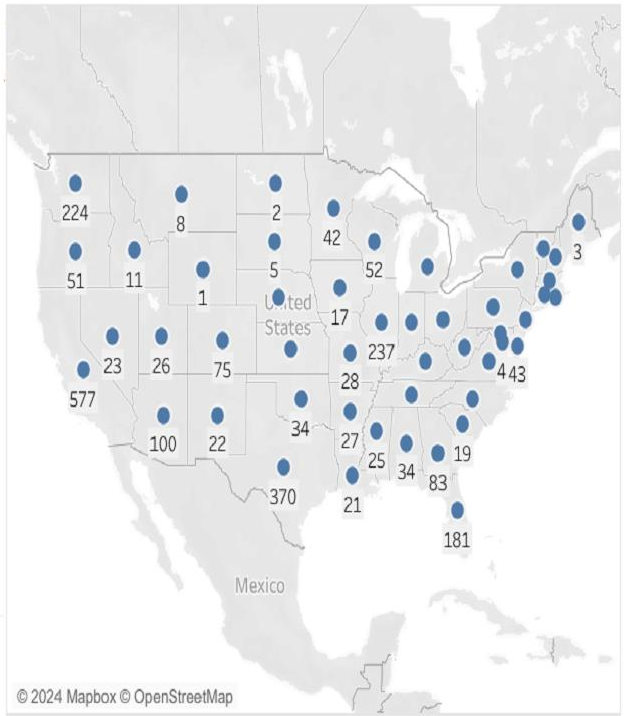
State

All

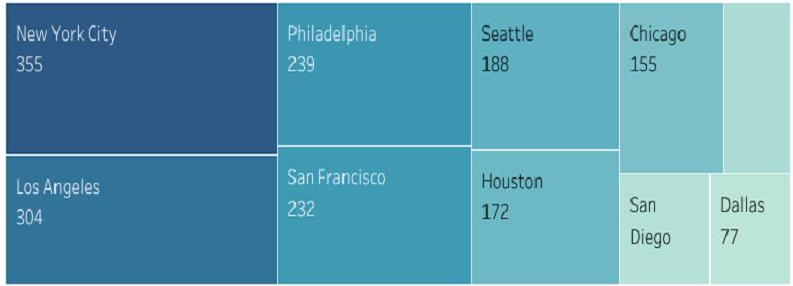
Most active segment



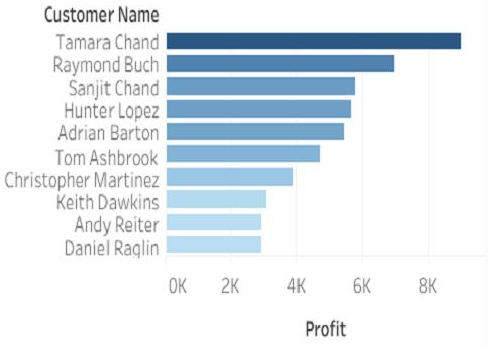
No. of customers by state



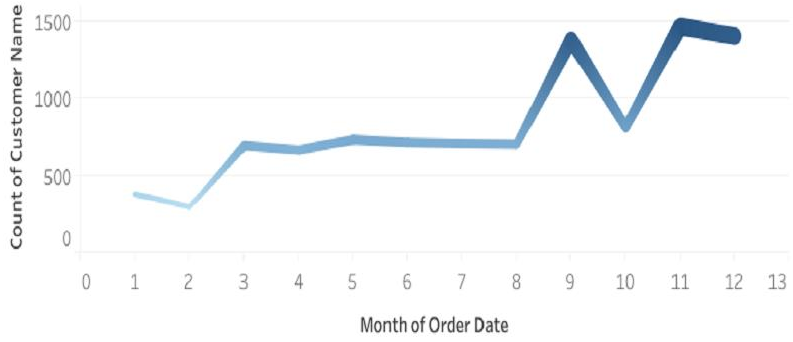
Top 10 cities by customers



top 10 Customers by profit



Customers by month



Dashboard 2: Sales Performance Dashboard

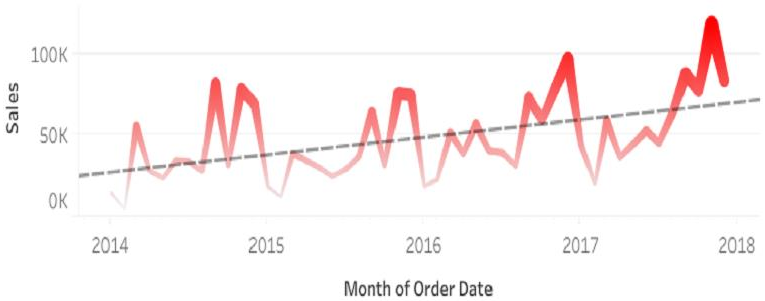
Description: The Sales Performance Dashboard focuses on the overall sales effectiveness of the organization. It highlights key performance indicators (KPIs) such as total sales, sales by region, top-selling products, and sales trends over time. This dashboard serves as a vital tool for sales teams to identify successful strategies, forecast future sales, and make data-driven decisions to drive revenue growth.

Sales Performance Dashboard

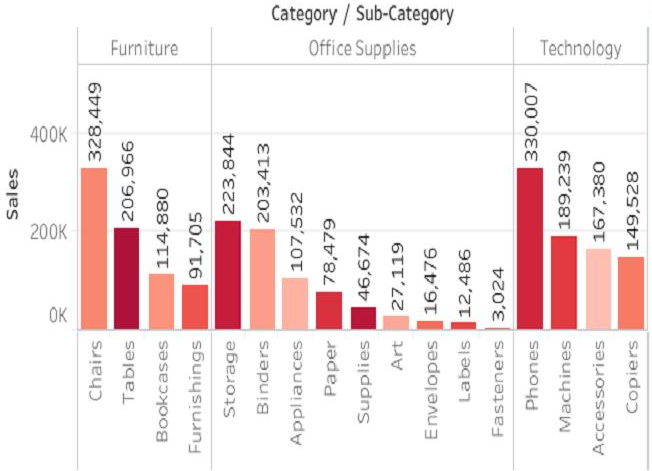
Customer Behavior | Shipping and Delivery

Total Sales	Total Profit	Total Quantity
2,297,201	286,397	37,873

Sales Trend Over Time

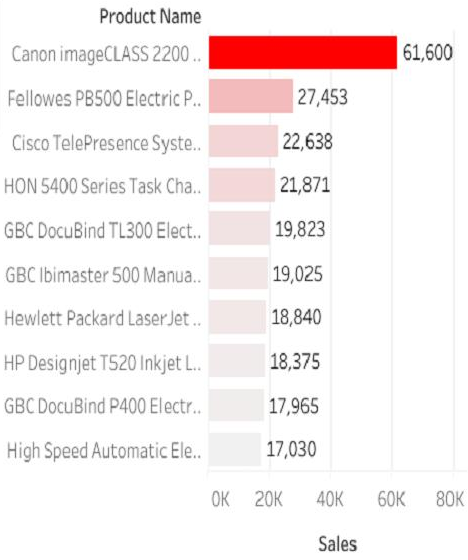


Sales Across Categories and Sub-Categories

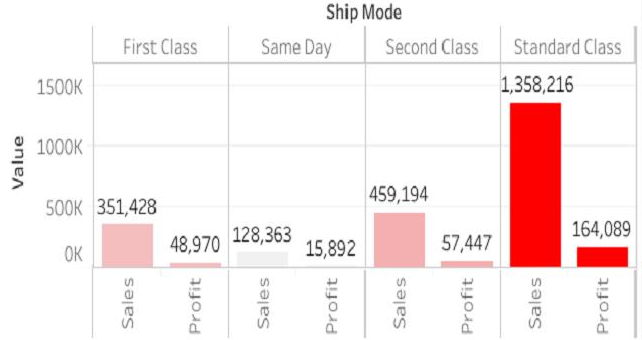


Profitability Analysis by Region

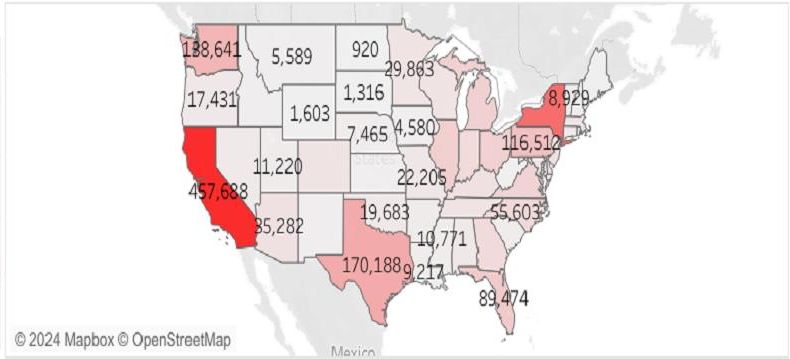
Top Products by Sales



Affect Ship mode on sales & profit



Sales by state



Dashboard 3: Shipping and Delivery Analysis Dashboard

Description: The Shipping and Delivery Analysis Dashboard provides insights into the logistics and delivery performance of the organization. It monitors average shipping days, regional delivery efficiency, and identifies any delays in the shipping process. This dashboard aims to enhance operational efficiency, reduce shipping costs, and improve customer satisfaction by addressing potential issues in the delivery workflow.

Shipping and Delivery Dashboard

Sales Performance | Customer Behavior

Avg. shipping days	Total Sales
3.958	2,297,201

Total Profit	Total quantity
286,397	37,873

Avg. Shipping days

To Null

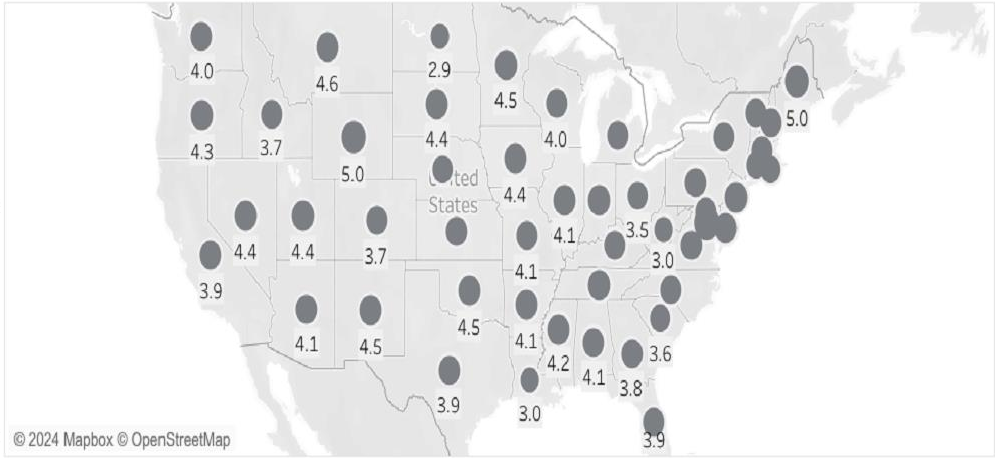
Ship Mode

- ✓ First Class
- ✓ Same Day
- ✓ Second Class
- ✓ Standard Class

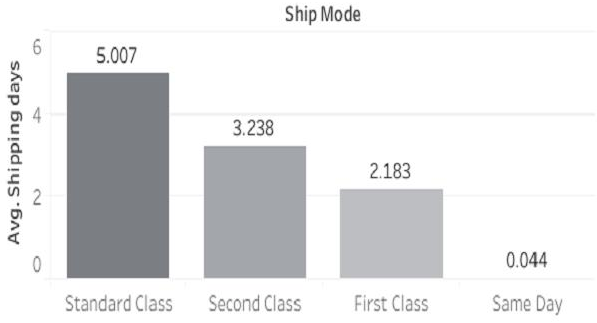
Region

- ✓ Central
- ✓ East
- ✓ South
- ✓ West

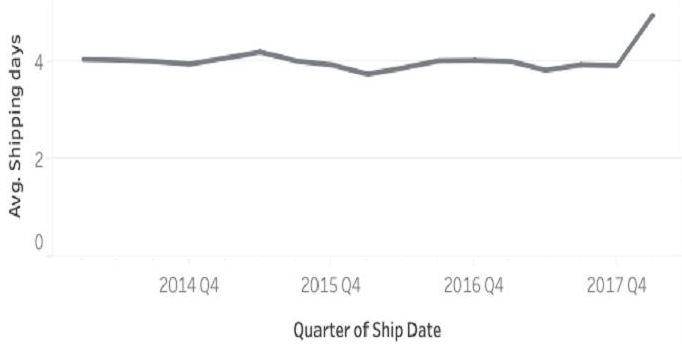
Average shippind days by state



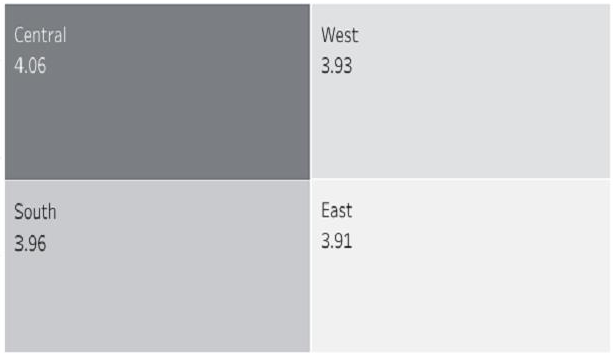
Shipping days by shipping method



Timeline of shipping days



Region by avg. shipping days



Explore the Complete Notebook and View the Dynamic Dashboards:

1. Notebook (Colab)
2. Dashboards (Tableau Public)

Links

Thanks!

Team members
Mohamed Ali
Abdalrahman Soliman
Omar Youssef
Mahmoud Hamdi
Marwan Mahmoud