

The background is a collage of various business and technology-related graphics. At the top, a pair of hands is shown typing on a laptop keyboard. To the right, there are two circular progress indicators: the top one is labeled 'CPU' with '68%' and '8000' below it, and the bottom one is labeled 'MEMORY' with '56%' and '8192 KB' below it. The left side features several bar and line charts. The bottom section contains a grid of six pie charts, some of which are partially filled. There are also various geometric shapes, including triangles and rectangles, in shades of orange, white, and grey, creating a modern, layered effect.

Supply Chain Management System

Made by
Darshanesh Patil

Index

- Importing Library and Dataset
- Dataset Overview
- Data Preprocessing
- Post-Processed Dataset Overview
- Visualization using Tableau
- Hypothesis Testing
- Customer Segmentation
- ABC Analysis
- Machine learning models for Predicting Late Delivery



Importing Basic Library and Dataset

Basic Library:-

- Pandas: Data manipulation with DataFrames.
- NumPy: Array math and computations.
- Matplotlib: Data visualization plotting.
- Seaborn: Stylish statistical plots.

Dataset:-

Uploading Dataset on Google drive and mounting it in Google Colab to enhances workflow and collaboration.

Dataset Overview

Entries: 180,519

Columns: 53

- Float64: 15 columns
- Int64: 14 columns
- Object: 24 columns
- Memory Usage: 73.0+ MB

Columns Name:-

['Type', 'Days for shipping (real)', 'Days for shipment (scheduled)', 'Benefit per order', 'Sales per customer', 'Delivery Status', 'Late_delivery_risk', 'Category Id', 'Category Name', 'Customer City', 'Customer Country', 'Customer Email', 'Customer Fname', 'Customer Id', 'Customer Lname', 'Customer Password', 'Customer Segment', 'Customer State', 'Customer Street', 'Customer Zipcode', 'Department Id', 'Department Name', 'Latitude', 'Longitude', 'Market', 'Order City', 'Order Country', 'Order Customer Id', 'order date (DateOrders)', 'Order Id', 'Order Item Cardprod Id', 'Order Item Discount', 'Order Item Discount Rate', 'Order Item Id', 'Order Item Product Price', 'Order Item Profit Ratio', 'Order Item Quantity', 'Sales', 'Order Item Total', 'Order Profit Per Order', 'Order Region', 'Order State', 'Order Status', 'Order Zipcode', 'Product Card Id', 'Product Category Id', 'Product Description', 'Product Image', 'Product Name', 'Product Price', 'Product Status', 'shipping date (DateOrders)', 'Shipping Mode']

Data Preprocessing

- **Null data**

Null data was found in four columns, Order Zipcode (86% null values, i.e. a total of 155679 records), Product description (100% null values, i.e. a total of 180519 records) and customer zipcode & customer Last Name (<0.1% is say a total of 3 records and 8 record respectively).

- **Removing Redundancies**

- ~~Benefit per order~~ and Order Profit Per Order are the same
- Sales per customer and ~~Order Item Total~~ are the same
- ~~Category Id~~ and Product Category Id are the same
- ~~Customer Id~~ and Order Customer Id are the same
- ~~Order Item Cardprod Id~~ and Product Card Id are the same
- ~~Order Item Product Price~~ and Product Price are the same

Data Preprocessing

- **State Abbreviations**

States in the United States are commonly represented using two-letter abbreviations.

Example: California → CA, New York → NY, Texas → TX

Converted standardized state abbreviations to State Name for consistency and compatibility.

- **Grouping Multiple Regions**

When dealing with large datasets, Grouped multiple regions of North America into North America.

- **Country Name Translation**

Multilingual datasets country names to be translated to English for compatibility.

Data Preprocessing

- **Group infrequent values into an "Others" category**

“Categories Name” having count less than 500.

“Product Name” having count less than 1500.

- **Irrelevant information for the objective of the Study**

"Customer Email", "Customer Password", "Product Image", "Customer Fname", "Customer Lname", "Product Description", "Order Zipcode", "Customer Zipcode", "Customer Country" (Multilingual), "Order City"(Multilingual), "Customer Street", "Order State"(Multilingual).

Post-Processed Dataset Overview

Entries: 180,516

Columns: 34

- Float64: 9 columns
- Int64: 10 columns
- Object: 16 columns
- Memory Usage: 49.6+ MB

Columns Name:-

['Type Of Payments', 'Days for shipping (real)', 'Days for shipment (scheduled)', 'Sales per customer', 'Delivery Status', 'Late_delivery_risk', 'Category Name', 'Customer City', 'Customer Segment', 'Customer State', 'Department Id', 'Department Name', 'Latitude', 'Longitude', 'Market', 'Order Country', 'Order Customer Id', 'order date (DateOrders)', 'Order Id', 'Order Item Discount', 'Order Item Discount Rate', 'Order Item Id', 'Order Item Profit Ratio', 'Order Item Quantity', 'Sales', 'Order Profit Per Order', 'Order Region', 'Order Status', 'Product Card Id', 'Product Category Id', 'Product Name', 'Product Price', 'Product Status', 'shipping date (DateOrders)', 'Shipping Mode']

Visualization using Tableau

[Link](#)

Sales Dashboard

Total Sales
\$36.78M

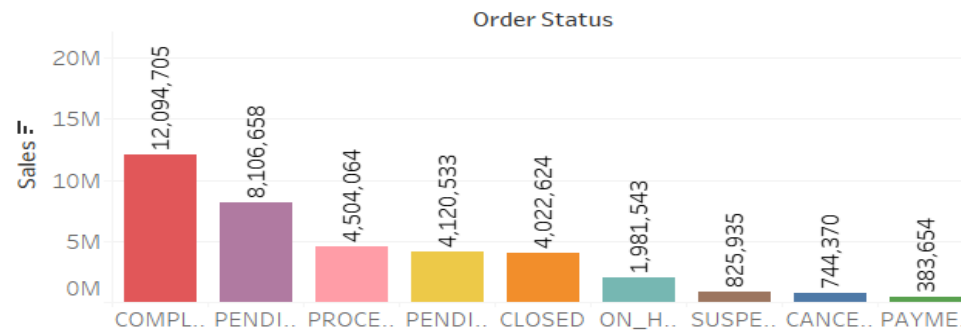
Average Sales Per Customer
\$183.11

Total Profit
\$3.97M

Total Orders
180,516

Average Order Profit
\$21.97

Sales by Order Status



Category Name

(All)

Shipping Mode

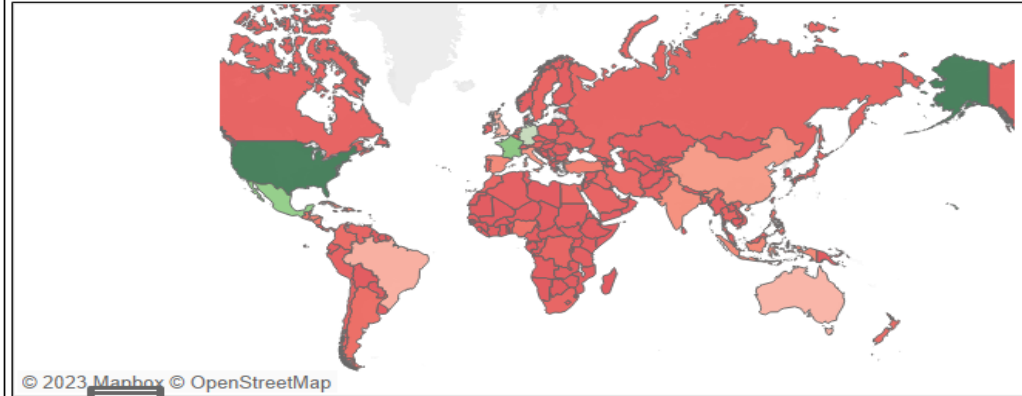
(All)

Market

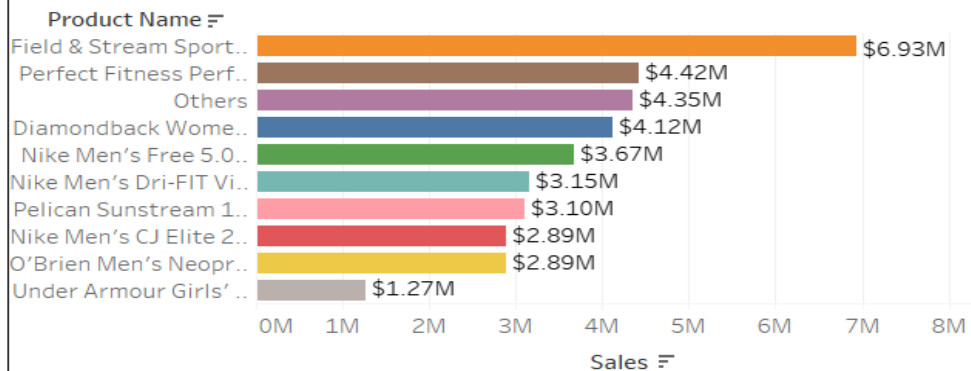
(All)

Year of order date (Date)

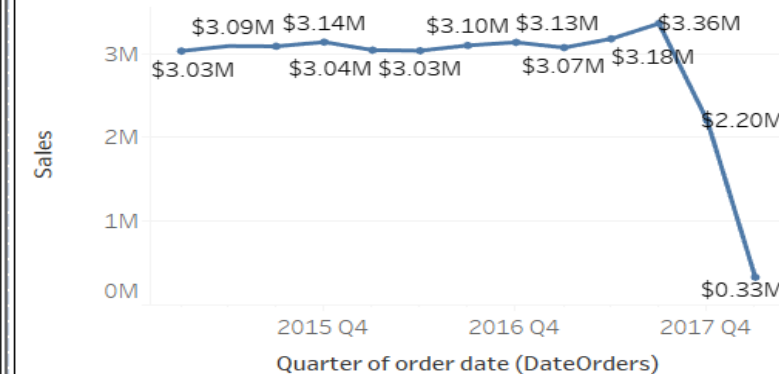
(All)



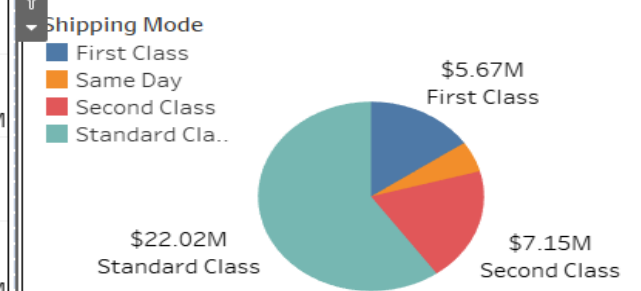
Top 10 Selling Products



Total Sales Trend Chart



Sales by Shipping Mode



Hypothesis Testing

Co-relation Matrix

Days for shipping (real)	1	0.5	0.002	0.4	-0.002	-0.004	0.004	0.003	-0.002	0.002	0.001	-0.002	-0.005	-0.0008	0.002	-0.005	-0.0009	-0.0003	0.002
Days for shipment (scheduled)	0.5	1	0.006	-0.4	-0.0002	-0.005	0.004	0.0009	-0.002	0.003	8e-05	-0.002	-0.002	-0.003	0.006	-0.0002	-0.0004	-0.0004	0.007
Sales per customer	0.002	0.006	1	-0.004	0.2	-0.0002	0.001	0.06	0.08	0.5	-0.1	0.08	-0.001	0.1	1	0.1	0.2	0.2	0.8
Late_delivery_risk	0.4	-0.4	-0.004	1	0.001	0.0007	-0.002	0.002	-0.001	-0.0008	0.0004	-0.001	-0.002	-0.0001	-0.004	-0.004	0.002	0.002	-0.002
Department Id	-0.002	-0.0002	0.2	0.001	1	0.002	-0.002	0.1	0.1	0.1	0.0002	0.1	-0.003	-0.3	0.2	0.03	0.9	0.9	0.4
Latitude	-0.004	-0.005	-0.0002	0.0007	0.002	1	-0.5	0.002	-0.003	-0.003	-0.004	-0.003	-9e-05	-0.002	-0.0007	0.0003	0.002	0.002	0.0005
Longitude	0.004	0.004	0.001	-0.002	-0.002	-0.5	1	-0.005	0.003	0.002	0.0005	0.003	-0.004	0.004	0.002	-0.003	-0.003	-0.002	-0.0009
Order Customer Id	0.003	0.0009	0.06	0.002	0.1	0.002	-0.005	1	0.2	0.04	0.002	0.2	-0.0003	-0.09	0.06	0.009	0.2	0.3	0.1
Order Id	-0.002	-0.002	0.08	-0.001	0.1	-0.003	0.003	0.2	1	0.05	0.0005	1	0.003	-0.09	0.08	0.01	0.2	0.2	0.1
Order Item Discount	0.002	0.003	0.5	-0.0008	0.1	-0.003	0.002	0.04	0.05	1	0.7	0.05	-0.003	0.07	0.6	0.06	0.1	0.1	0.5
Order Item Discount Rate	0.001	8e-05	-0.1	0.0004	0.0002	-0.004	0.0005	0.002	0.0005	0.7	1	0.0005	-0.003	-2e-05	0.0003	-0.02	0.0007	0.0006	0.0003
Order Item Id	-0.002	-0.002	0.08	-0.001	0.1	-0.003	0.003	0.2	1	0.05	0.0005	1	0.003	-0.08	0.08	0.01	0.2	0.2	0.1
Order Item Profit Ratio	-0.005	-0.002	-0.001	-0.002	-0.003	-9e-05	-0.004	-0.0003	0.003	-0.003	-0.003	0.003	1	0.001	-0.002	0.8	-0.002	-0.002	-0.002
Order Item Quantity	-0.0008	-0.003	0.1	-0.0001	-0.3	-0.002	0.004	-0.09	-0.09	0.07	-2e-05	-0.08	0.001	1	0.1	0.02	-0.3	-0.3	-0.5
Sales	0.002	0.006	1	-0.004	0.2	-0.0007	0.002	0.06	0.08	0.6	0.0003	0.08	-0.002	0.1	1	0.1	0.2	0.2	0.8
Order Profit Per Order	-0.005	-0.0002	0.1	-0.004	0.03	0.0003	-0.003	0.009	0.01	0.06	-0.02	0.01	0.8	0.02	0.1	1	0.03	0.03	0.1
Product Card Id	-0.0009	-0.0004	0.2	0.002	0.9	0.002	-0.003	0.2	0.2	0.1	0.0007	0.2	-0.002	-0.3	0.2	0.03	1	1	0.5
Product Category Id	-0.0003	-0.0004	0.2	0.002	0.9	0.002	-0.002	0.3	0.2	0.1	0.0006	0.2	-0.002	-0.3	0.2	0.03	1	1	0.5
Product Price	0.002	0.007	0.8	-0.002	0.4	0.0005	-0.0009	0.1	0.1	0.5	0.0003	0.1	-0.002	-0.5	0.8	0.1	0.5	0.5	1
	Days for shipping (real)	Days for shipment (scheduled)	Sales per customer	Late_delivery_risk	Department Id	Latitude	Longitude	Order Customer Id	Order Id	Order Item Discount	Order Item Discount Rate	Order Item Id	Order Item Profit Ratio	Order Item Quantity	Sales	Order Profit Per Order	Product Card Id	Product Category Id	Product Price

Hypothesis Testing

Hypothesis 1: Impact of delivery time on order profitability

Null hypothesis: There is no significant difference in order profitability between orders with shorter actual delivery time and those with longer delivery time.

Alternative hypothesis: Orders with a shorter actual delivery time are more profitable than those with a longer delivery time.

Hypothesis 1: Impact of delivery time on order profitability

t-statistic: 1.1405142299533106

p-value: 0.25407583425621577

Fail to reject the null hypothesis

Hypothesis Testing

Hypothesis 2: Relationship between sales per customer and delivery status

Null hypothesis: There is no significant difference in sales per customer between orders delivered on time and those that are not.

Alternative hypothesis: Customers who receive their orders on time have higher sales per customer than those who do not.

Hypothesis 2: Relationship between sales per customer and delivery status

t-statistic: 1.8831012266781173

p-value: 0.05968849992613582

Fail to reject the null hypothesis

Hypothesis Testing

Hypothesis 3: Impact of late delivery risk on order profit

Null hypothesis: There is no significant difference in order profitability between orders with high and low late delivery risk.

Alternative hypothesis: Orders with a higher late delivery risk are less profitable than those with a lower late delivery risk.

```
Hypothesis 3: Impact of late delivery risk on order profit  
t-statistic: -1.5871340032452448  
p-value: 0.11248410145210566  
Fail to reject the null hypothesis
```


Hypothesis Testing

Hypothesis 4: Relationship between product price and sales:

Null Hypothesis (H_0): There is no significant relationship between product price and sales.

Alternative Hypothesis (H_1): There is a significant relationship between product price and sales.

Hypothesis: Relationship between product price and sales

Correlation coefficient: 0.7899460191767189

p-value: 0.0

Reject the null hypothesis

Hypothesis Testing

Hypothesis 5: Impact of shipping mode on order profitability

Null hypothesis: There is no significant difference in order profitability between different shipping modes.

Alternative hypothesis: Certain shipping modes are more profitable than others.

Hypothesis 5: Impact of shipping mode on order profitability

F-statistic: 1.9760400987761573

p-value: 0.11516423806216304

Fail to reject the null hypothesis

Hypothesis Testing

Hypothesis 6: Relationship between product category and order profitability

Null hypothesis: There is no significant difference in order profitability between different product categories.

Alternative hypothesis: Certain product categories are more profitable than others.

Hypothesis 6: Relationship between product category and order profitability

F-statistic: 50.74877365263413

p-value: 0.0

Reject the null hypothesis

Customer Segmentation

Customer segmentation is the practice of dividing a company's customer base into distinct groups based on certain shared characteristics and behaviours.

Purpose: It helps businesses better understand their customers, tailor marketing strategies, and provide personalized experiences.

Types of Segmentation:

Demographic:

Geographic:

Psychographic:

Behavioural:

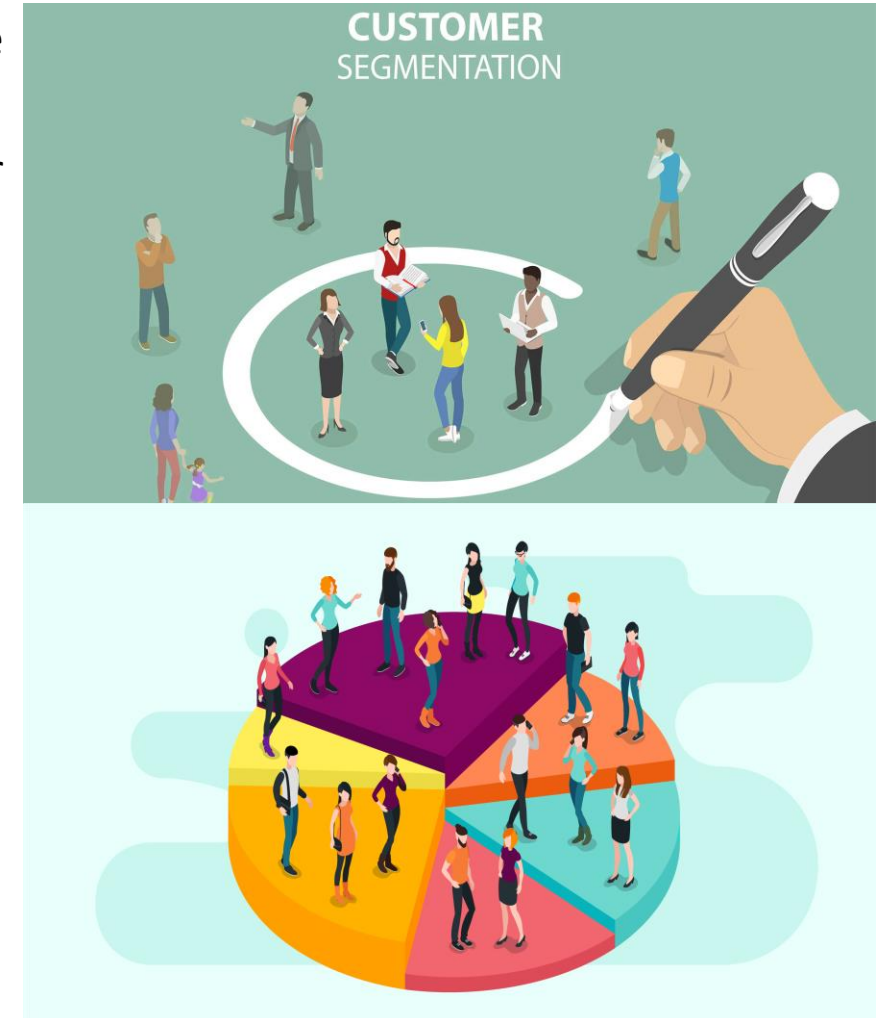
Benefits:

Improved Marketing

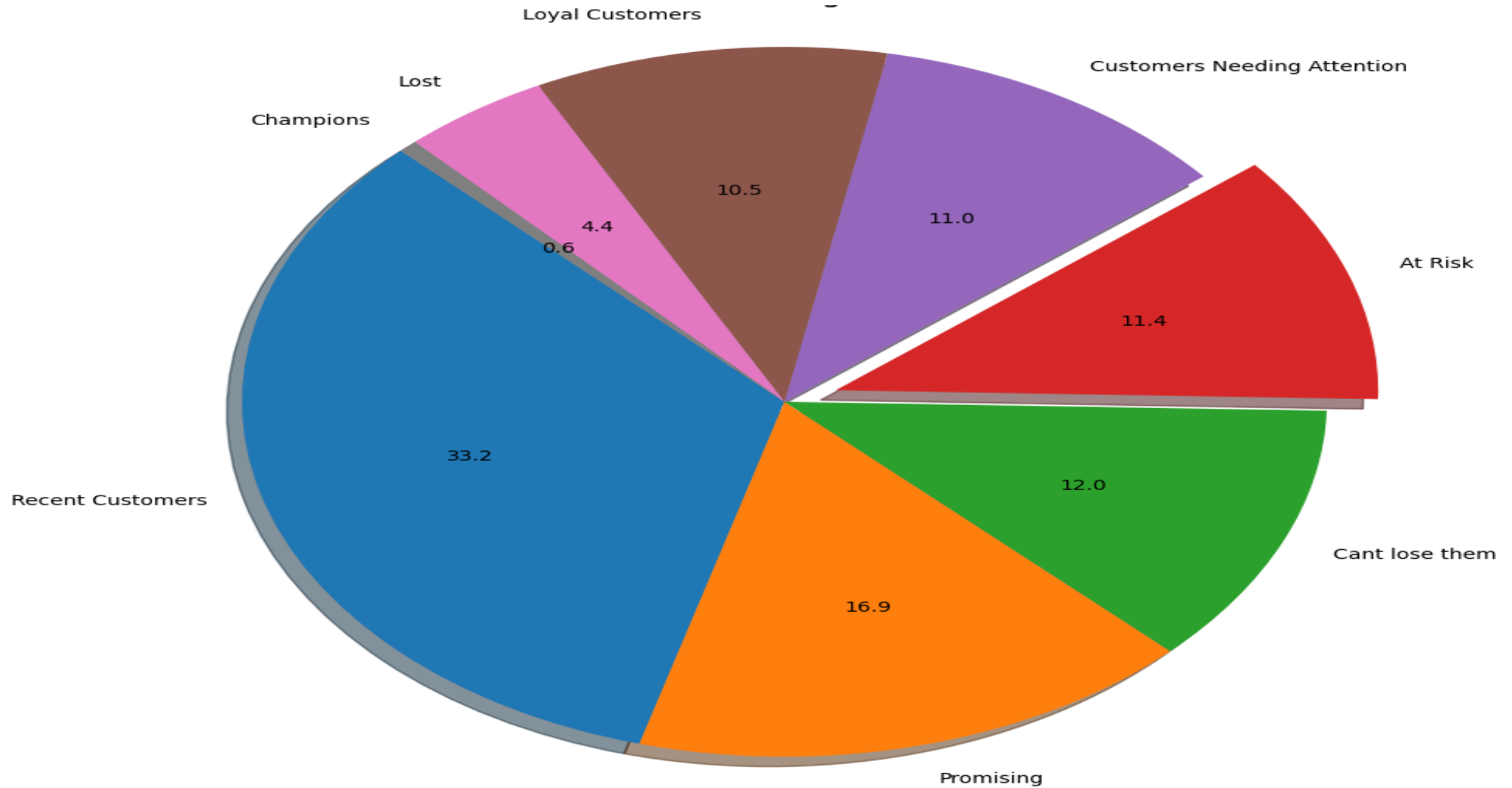
Enhanced Customer Experience:

Efficient Resource Allocation:

Better Product Development:



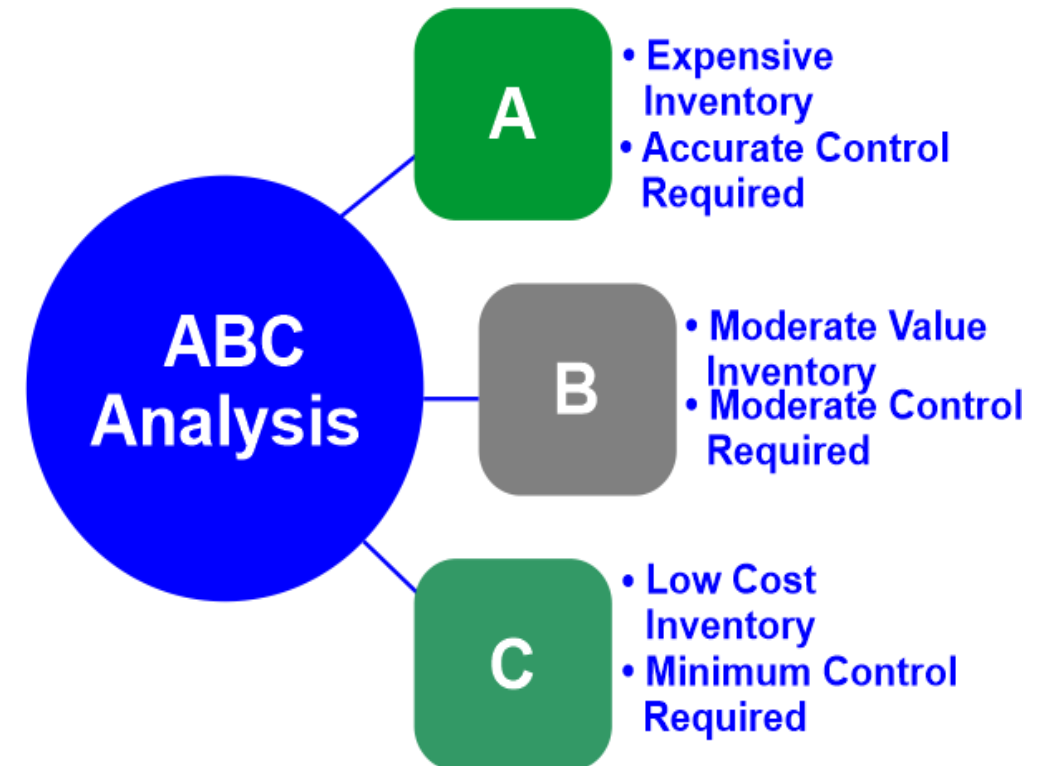
Customer Segmentation



ABC Analysis

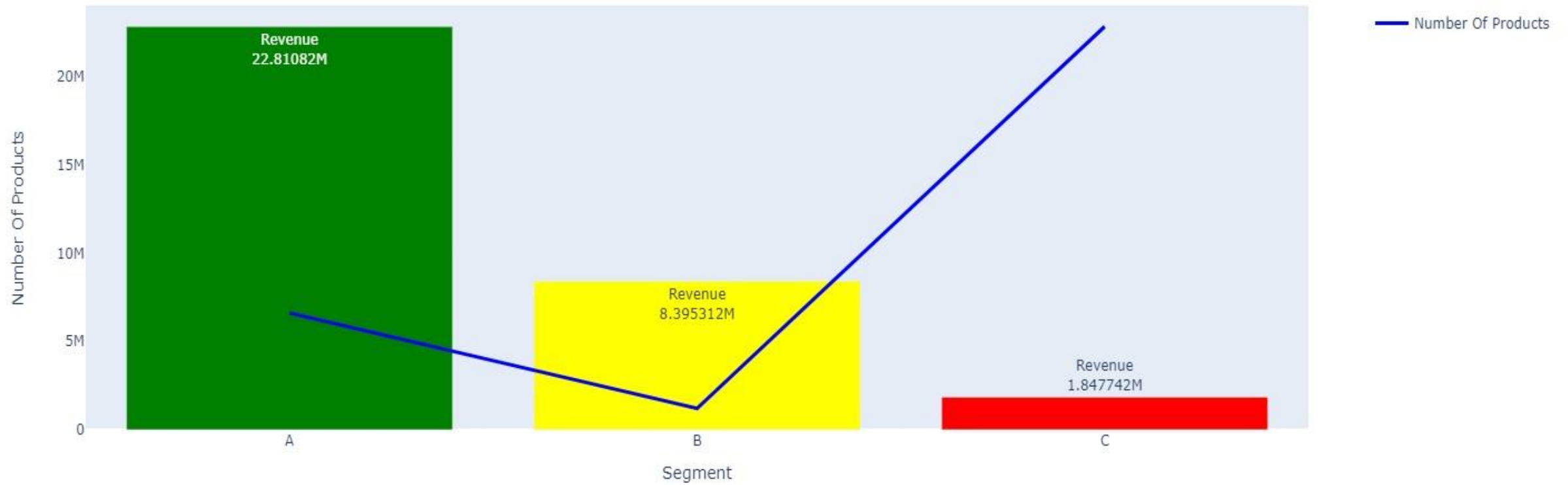
ABC analysis, also known as Pareto analysis or the 80/20 rule, is a technique often used in inventory management and supply chain optimization. It categorizes items into three groups based on their value or importance. The pareto rule which states that 80% of outcome comes from 20% of causes. Similarly here, 80% of the revenue is generated by 20% of all products. High care should be taken that these 20% of the products are always in-stock.

Product Class	Revenue Percentage Range	Description
Class A	0% - 75%	Highest revenue-generating products
Class B	75.01% - 95%	Next significant revenue-generating products
Class C	95.01% - 100%	Long tail products contributing less revenue



ABC Analysis

Revenue Generated By Products in Different ABC Segments



Machine learning models for Predicting Late Delivery

The classification models used in this project are Logistic Regression, Linear Discriminant Analysis, Gaussian Naive Bayes, Support Vector Machines, and Random Forest classification to predict "**Late Delivery**" based on accuracy, recall and F1 score metrics.

Metrics:

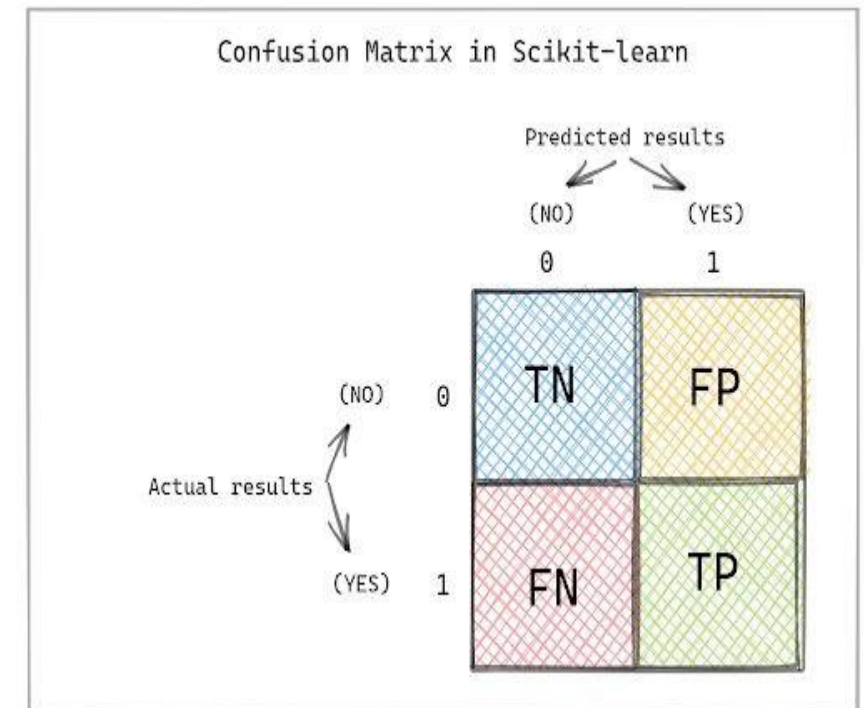
TP: The model correctly predicts that "Late Delivery" will occur.

TN: The model correctly predicts that the "Late Delivery" will not occur.

FP: The model incorrectly predicts that the "Late Delivery" will occur.

FN: The model incorrectly predicts that the "Late Delivery" will not occur.

The model with the highest score of F1 is selected



Machine learning models for Predicting Late Delivery

RandomForestClassifier Confusion Matrix - Late Delivery Prediction

Actual \ Predicted	Not Late	Late
Not Late	True Negative 24403	False Positive 60
Late	False Negative 0	True Positive 29692

Machine learning models for Predicting Late Delivery

LinearSVC Confusion Matrix - Late Delivery Prediction

Actual \ Predicted	Not Late	Late
Not Late	True Negative 23571	False Positive 892
Late	False Negative 0	True Positive 29692

Machine learning models for Predicting Late Delivery

LogisticRegression Confusion Matrix - Late Delivery Prediction

Actual \ Predicted	Not Late	Late
Not Late	True Negative 23571	False Positive 892
Late	False Negative 1	True Positive 29691

Machine learning models for Predicting Late Delivery

Linear Discriminant Analysis Confusion Matrix - Late Delivery Prediction

Actual \ Predicted	Not Late	Late
Not Late	True Negative 23409	False Positive 1054
Late	False Negative 771	True Positive 28921

Machine learning models for Predicting Late Delivery

Gaussian Naive Bayes Confusion Matrix - Late Delivery Prediction

Actual \ Predicted	Not Late	Late
Not Late	True Negative 21606	False Positive 2857
Late	False Negative 17	True Positive 29675

Machine learning models for Predicting Late Delivery

Models Evaluation

Algorithms	F1 Score	Accuracy	Recall	True Negative	False Positive	False Negative	True Positive
Random Forest Classification	99.8991	99.8892	99.7983	24403	60	0	29692
Support Vector Machines	98.5201	98.3529	97.0834	23571	892	0	29692
Logistic Classification Model	98.5185	98.351	97.0833	23571	892	1	29691
Linear Discriminant Analysis	96.9414	96.63	96.4837	23409	1054	771	28921
Gaussian Naive Bayes Model	95.3812	94.693	91.2179	21606	2857	17	29675



THANK YOU