

2024

SALES

DATA

ANALYSIS



PROJECT REPORT



PREPARED BY RAHUL BADOLA

FOR AFAME TECHNOLOGIES

Table of Contents

Introduction	1
Python Libraries Used	2
Basic Composition of Dataset	4-5
Data Cleaning	5
Order Analysis	6-10
Geographical Analysis	11-14
Product Analysis	14-16
Profit Analysis	17-19
Shipping Analysis	18-19
Market Analysis	19-20
Conclusion and Recommendation	21-23

Introduction

In today's dynamic business landscape, understanding and leveraging data is crucial for making informed decisions and staying competitive. Sales data, in particular, offers valuable insights into customer behavior, market trends, and the effectiveness of business strategies. In this project, we delve into an exploratory analysis of a comprehensive sales dataset, aiming to extract actionable insights that can drive business growth and optimization.

The dataset under examination encompasses a plethora of information related to sales transactions, including order details, product attributes, customer demographics, and geographical information. Through this analysis, we seek to uncover hidden patterns, trends, and correlations within the data, shedding light on critical aspects of the sales process.

By employing advanced data analysis techniques and visualization tools, we aim to address key questions pertaining to sales performance, customer segmentation, product profitability, and the impact of various factors such as discounts and shipping costs. The insights gained from this analysis will not only provide a deeper understanding of the underlying dynamics driving sales but also serve as a foundation for strategic decision-making and targeted interventions to enhance overall business performance.

Throughout this report, we will walk through our methodology, explore key findings, and draw meaningful conclusions that can inform and guide stakeholders in their pursuit of business excellence in the realm of sales and marketing. Let us embark on this journey of discovery and unlock the potential hidden within the vast troves of sales data at our disposal.

Python Libraries Used

1. NumPy (Numerical Python):

NumPy is a fundamental library for numerical computing in Python. It offers powerful tools for handling large arrays and matrices, along with a vast collection of mathematical functions to perform various numerical operations efficiently. We utilized NumPy extensively for tasks such as data aggregation, calculation of statistical metrics, and handling numerical computations underlying our analysis.

2. pandas:

pandas is a versatile data manipulation library that provides intuitive data structures and functions for working with structured data. It excels in data ingestion, cleaning, transformation, and analysis, making it indispensable for handling tabular data like our sales dataset. With pandas, we effortlessly managed dataframes, performed data preprocessing tasks, and conducted exploratory data analysis to extract meaningful insights.

3. seaborn:

seaborn is a powerful data visualization library built on top of matplotlib, offering a high-level interface for creating attractive and informative statistical graphics. It simplifies the process of generating complex visualizations, including histograms, scatter plots, box plots, and heatmaps, thereby enabling us to visualize relationships, distributions, and patterns within the sales dataset. seaborn's rich visualization capabilities played a vital role in conveying our findings effectively through insightful plots and charts.

4. matplotlib:

matplotlib is a comprehensive plotting library widely used for creating static, animated, and interactive visualizations in Python. It provides a flexible toolkit for generating a wide range of plots and charts, allowing for detailed customization and fine-tuning of visual elements. We leveraged matplotlib alongside seaborn to craft visually appealing plots, annotate data points, and present our analysis results in a clear and concise manner.

Basic Composition

24
Rows

51290
Columns

Data Types

Row ID	int64
Order ID	object
Order Date	datetime64[ns]
Ship Date	datetime64[ns]
Ship Mode	object
Customer ID	object
Customer Name	object
Segment	object
City	object
State	object
Country	object
Postal Code	float64
Market	object
Region	object
Product ID	object
Category	object
Sub-Category	object
Product Name	object
Sales	float64
Quantity	int64
Discount	float64
Profit	float64
Shipping Cost	float64
Order Priority	object
dtype:	object

Null Values

Row ID	0
Order ID	0
Order Date	0
Ship Date	0
Ship Mode	0
Customer ID	0
Customer Name	0
Segment	0
City	0
State	0
Country	0
Postal Code	41296
Market	0
Region	0
Product ID	0
Category	0
Sub-Category	0
Product Name	0
Sales	0
Quantity	0
Discount	0
Profit	0
Shipping Cost	0
Order Priority	0
dtype:	int64

Description of Data

	Row ID	Order Date	Ship Date	Postal Code	Sales	Quantity	Discount	Profit	Shipping Cost
count	51290.00000	51290	51290	9994.000000	51290.000000	51290.000000	51290.000000	51290.000000	51290.000000
mean	25645.50000	2013-05-11 21:26:49.155781120	2013-05-15 20:42:42.745174528	55190.379428	246.490581	3.476545	0.142908	28.610982	26.375818
min	1.00000	2011-01-01 00:00:00	2011-01-03 00:00:00	1040.000000	0.444000	1.000000	0.000000	-6599.978000	0.002000
25%	12823.25000	2012-06-19 00:00:00	2012-06-23 00:00:00	23223.000000	30.758625	2.000000	0.000000	0.000000	2.610000
50%	25645.50000	2013-07-08 00:00:00	2013-07-12 00:00:00	56430.500000	85.053000	3.000000	0.000000	9.240000	7.790000
75%	38467.75000	2014-05-22 00:00:00	2014-05-26 00:00:00	90008.000000	251.053200	5.000000	0.200000	36.810000	24.450000
max	51290.00000	2014-12-31 00:00:00	2015-01-07 00:00:00	99301.000000	22638.480000	14.000000	0.850000	8399.976000	933.570000
std	14806.29199	NaN	NaN	32063.693350	487.565361	2.278766	0.212280	174.340972	57.296810

Data Cleaning

Dropping Null Values

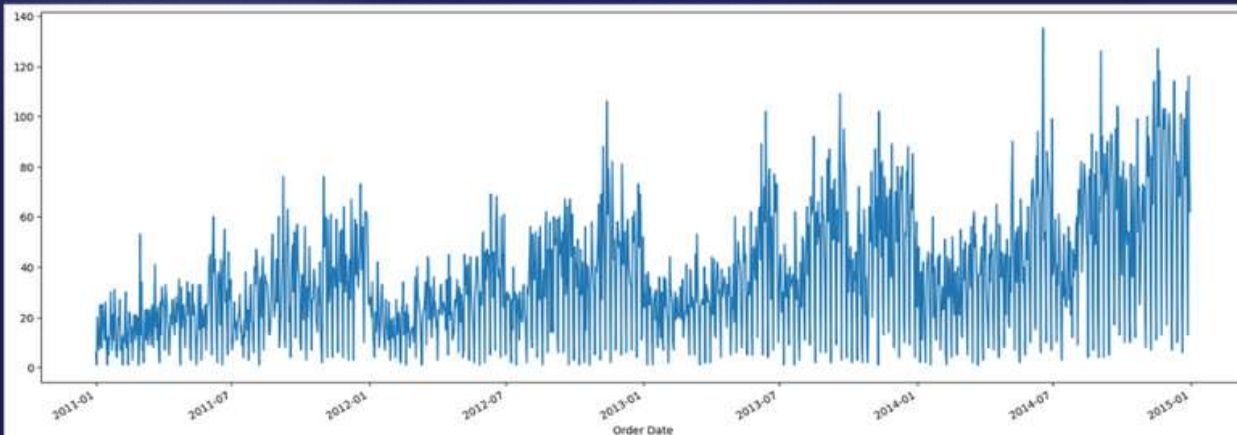
```
# Dropping 'Postal code' column because most of the values are null in that  
df.drop('Postal Code',axis=1,inplace=True)
```

Dropping Duplicates

```
# Dropping Duplicates  
df.drop_duplicates(inplace=True)
```

Order Analysis

Order Frequency Overtime



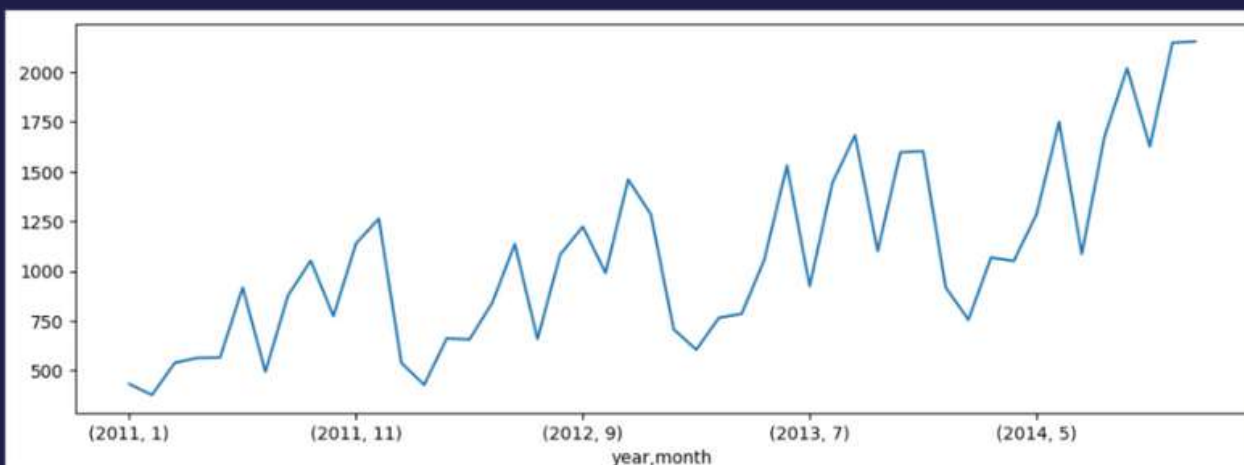
Max Orders - 135
On 2014-06-18

135

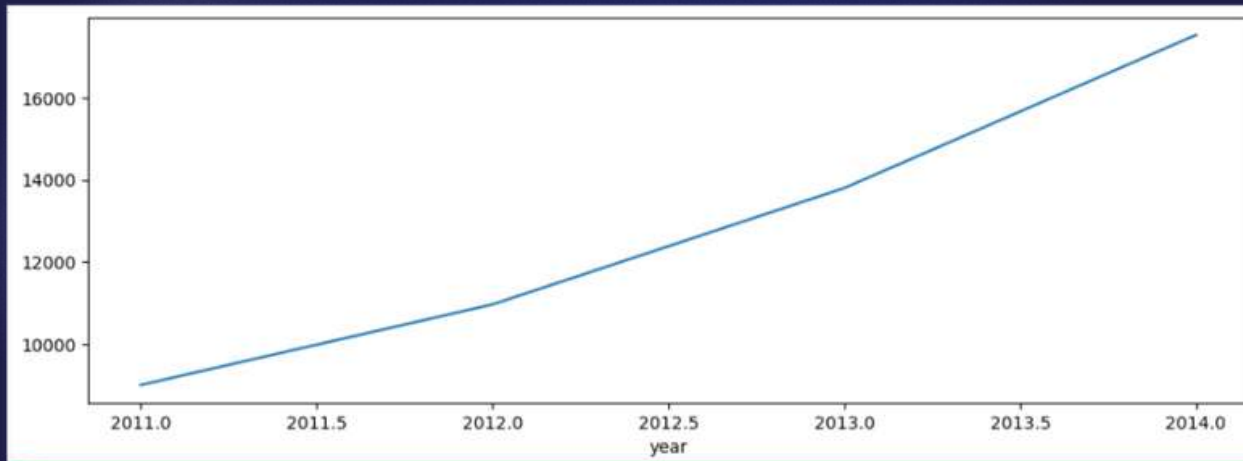
Min Orders - 1
On 2011-02-06

135

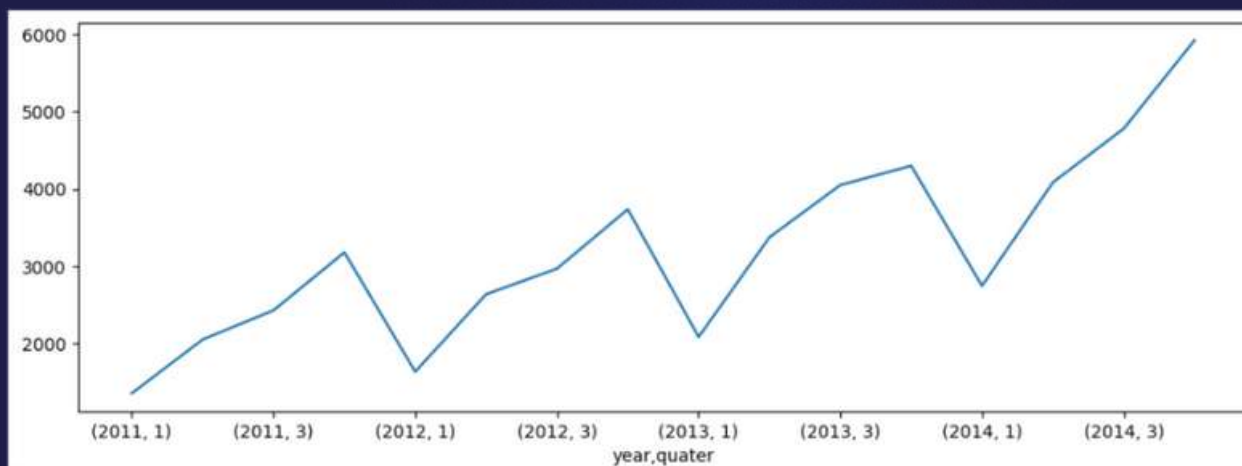
Distribution of Orders by month



Distribution of Order by Year



Distribution of Order by Quarterly

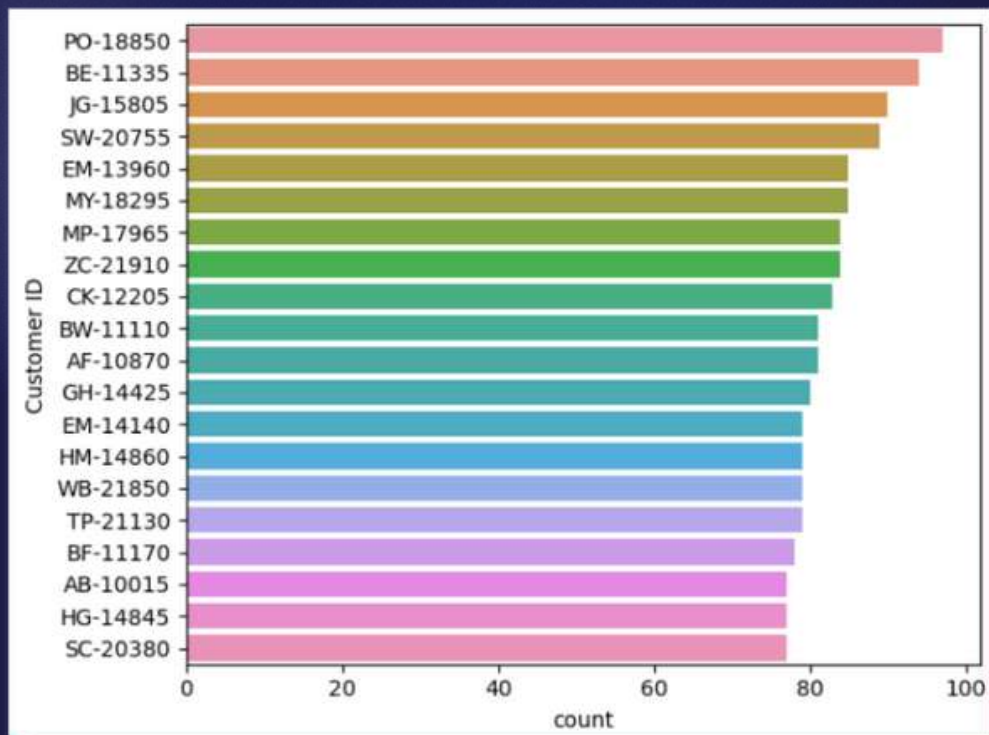


**Customer
Return - 99.6%**

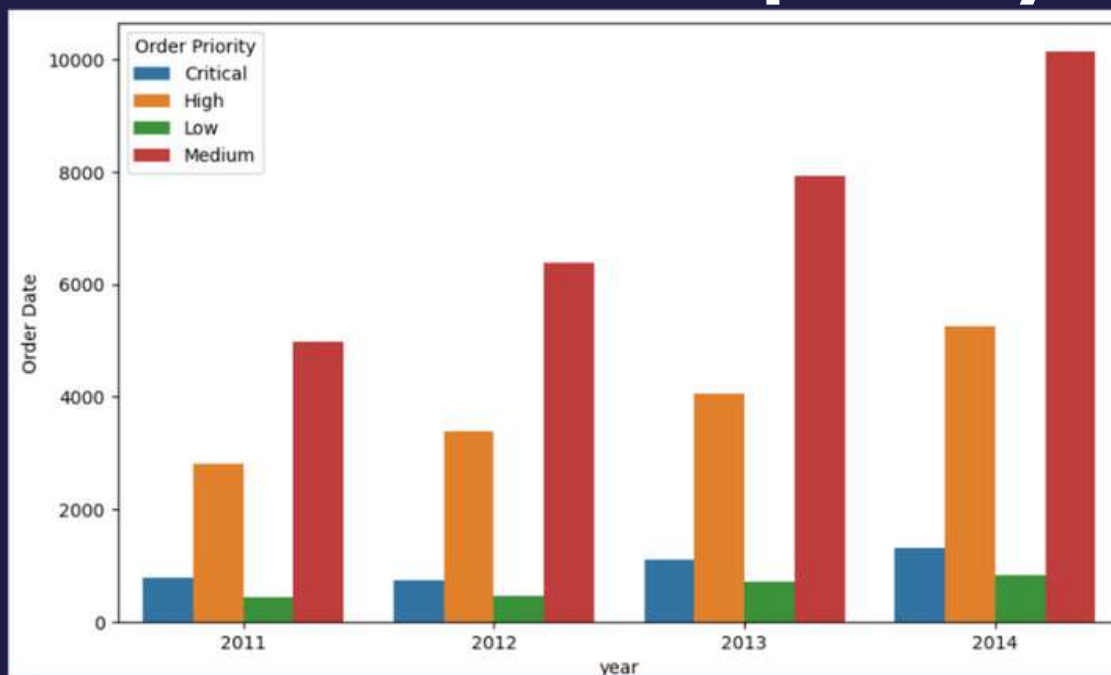
**Customer not
Returned - 0.4%**

135

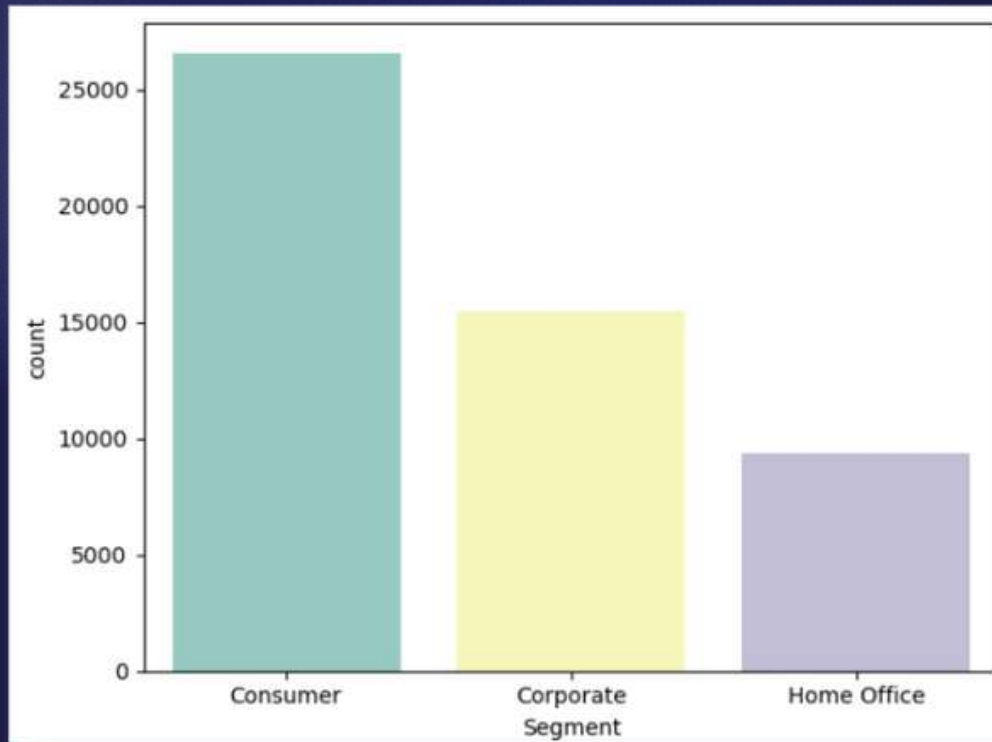
Highest no. of Purchase



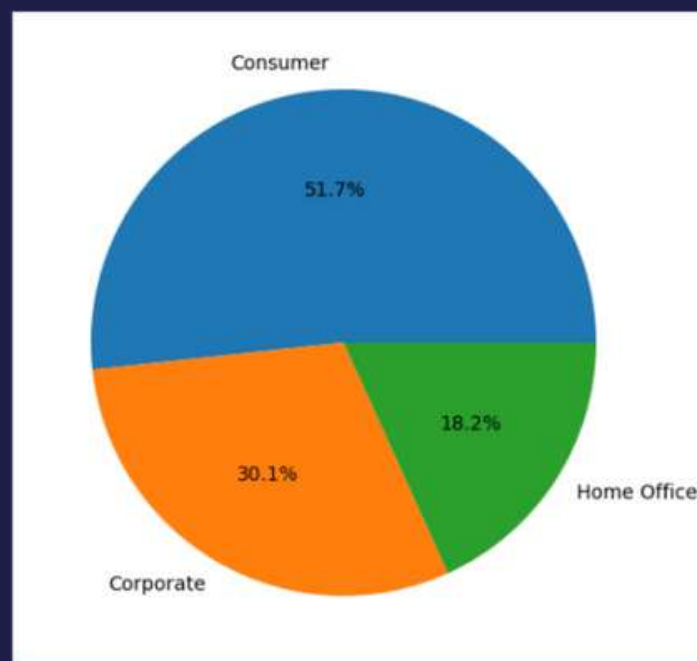
Orders based on priority



Most popular categories(bar)



most popular categories(pie)



Average Customer Retention

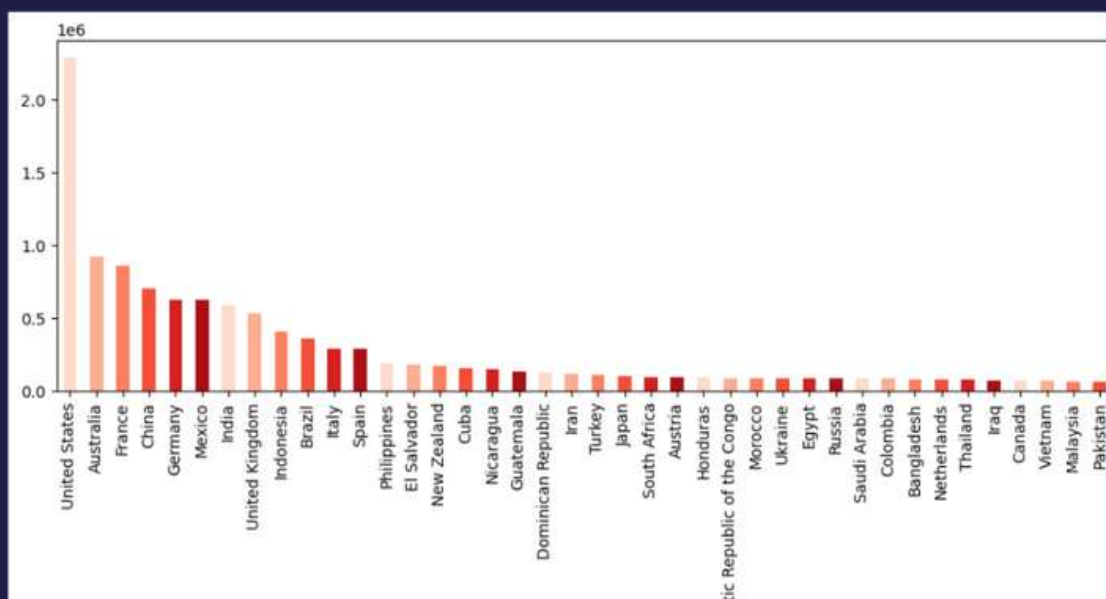
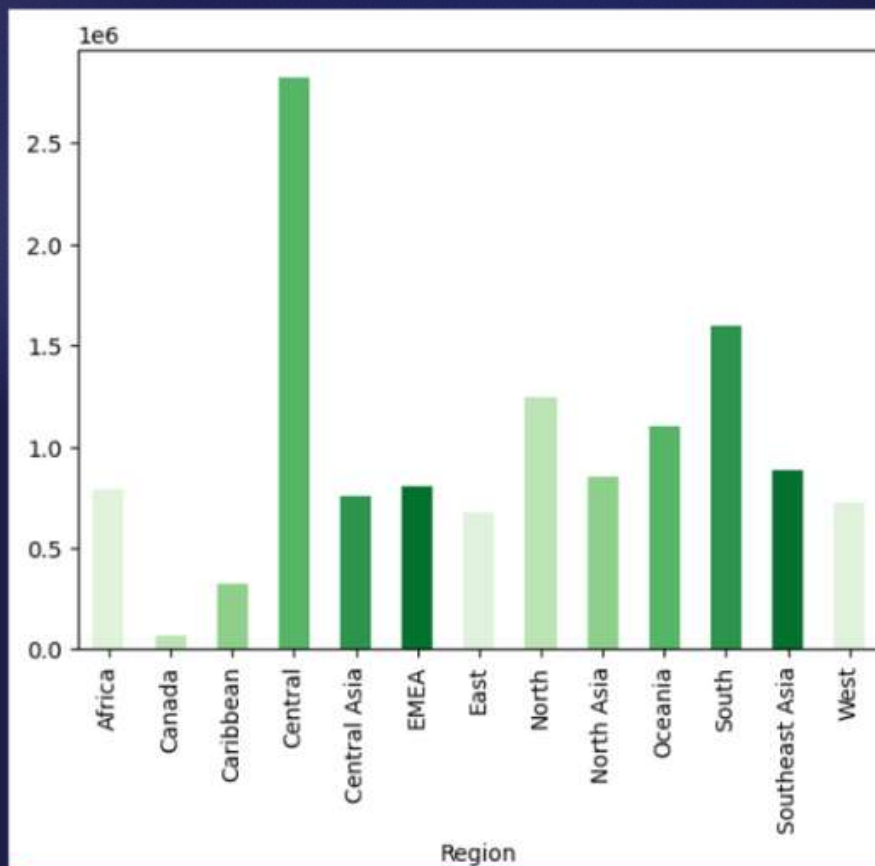
	customer_id	year	month	end_year	end_month	total_days
0	PO-18850	2011	12	2014	8	974
1	BE-11335	2011	4	2014	6	1157
2	JG-15805	2011	8	2014	12	1218
3	SW-20755	2011	6	2014	11	1249
4	EM-13960	2011	8	2014	3	943
...
1585	MG-7650	2013	8	2013	8	0
1586	ME-8010	2013	12	2013	12	0
1587	RC-9825	2012	12	2012	12	0
1588	MG-7890	2013	12	2013	12	0
1589	ZC-11910	2014	6	2014	6	0

1590 rows × 6 columns

**Average Customer
Retention - 988 Days**

Geographical Analysis

Sales By Region & Country



Sales Identifier Program

```
sales_identify()
```

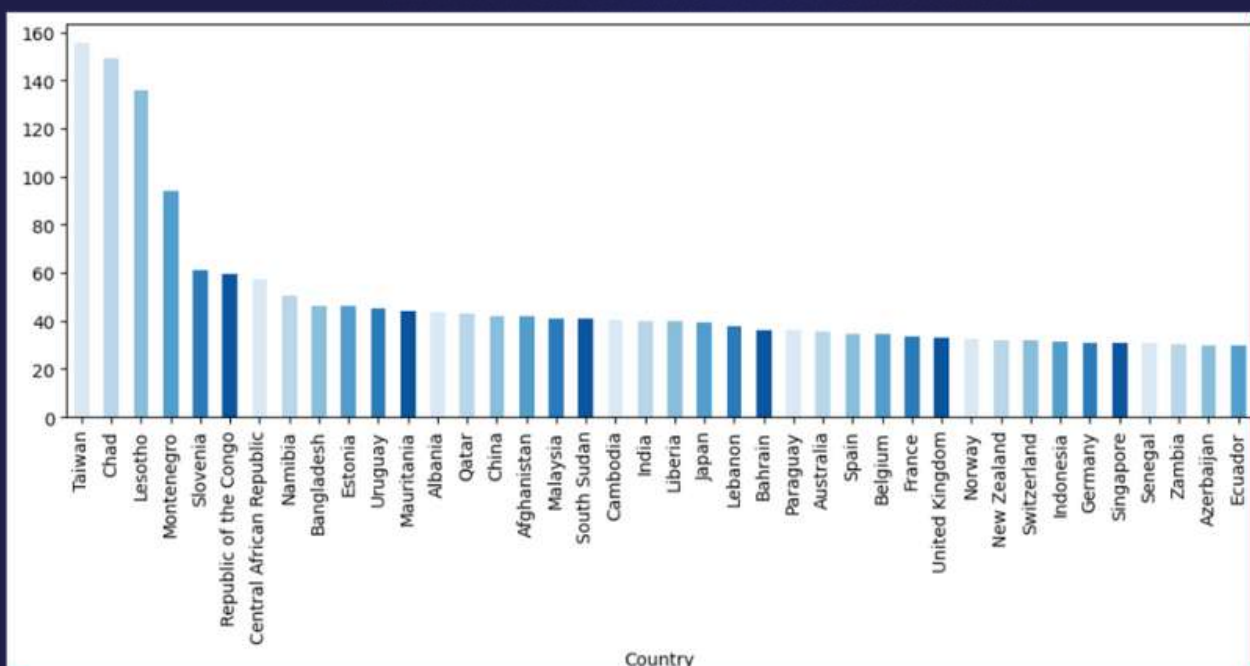
Enter the Country Name: India ←..... Enter Country Name

Enter the name of State: Gujarat ←..... Enter State Name

Enter City Name: Surat ←..... Enter City Name

Total Sales 12216 ←..... Get output of total sales in that particular city

Avg shipping Cost by Location

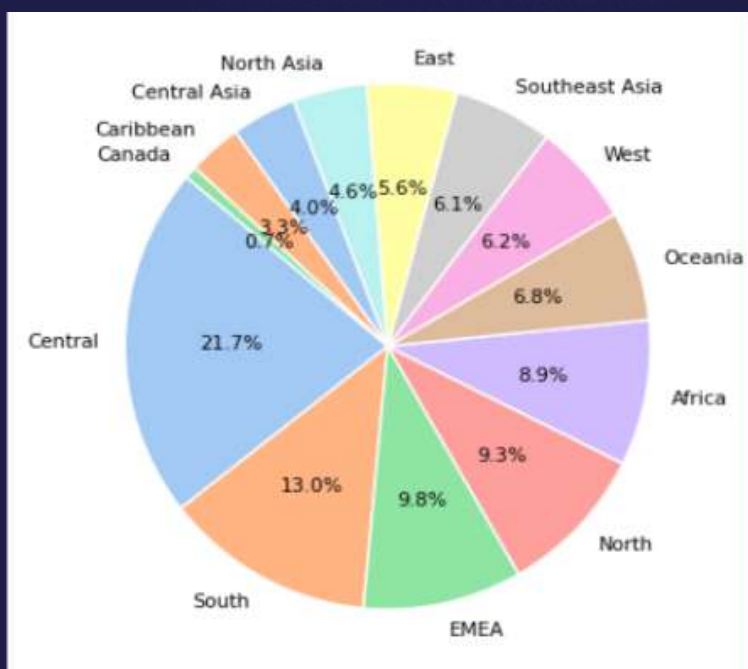


Shipping Cost Identifier Program

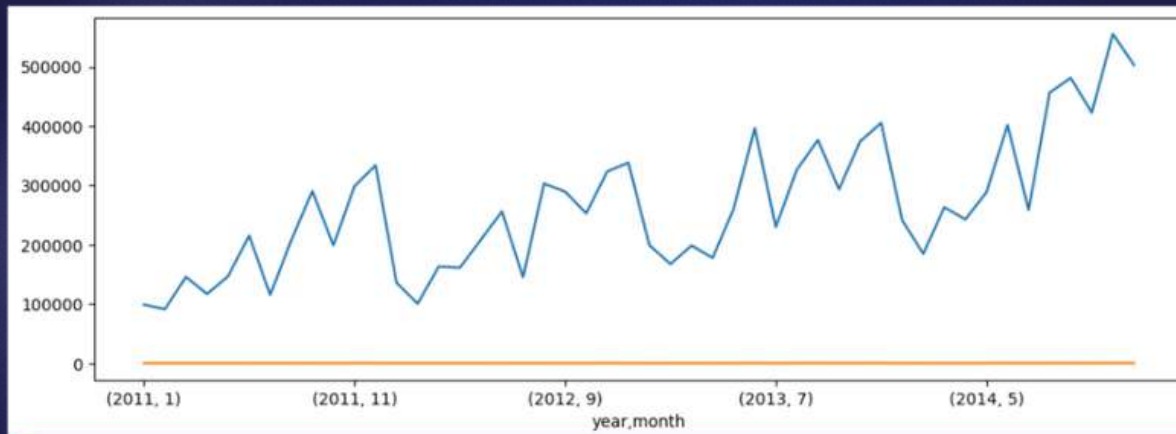
```
shipping_cost_identify()
```

Enter the Country Name: United States ← Enter Country Name
Enter the name of State: New York ← Enter State Name
Enter City Name: New York City ← Enter City Name
Total Shipping Cost: 29 ← Get output of shipping cost in that particular city

Market Share obtained by Region

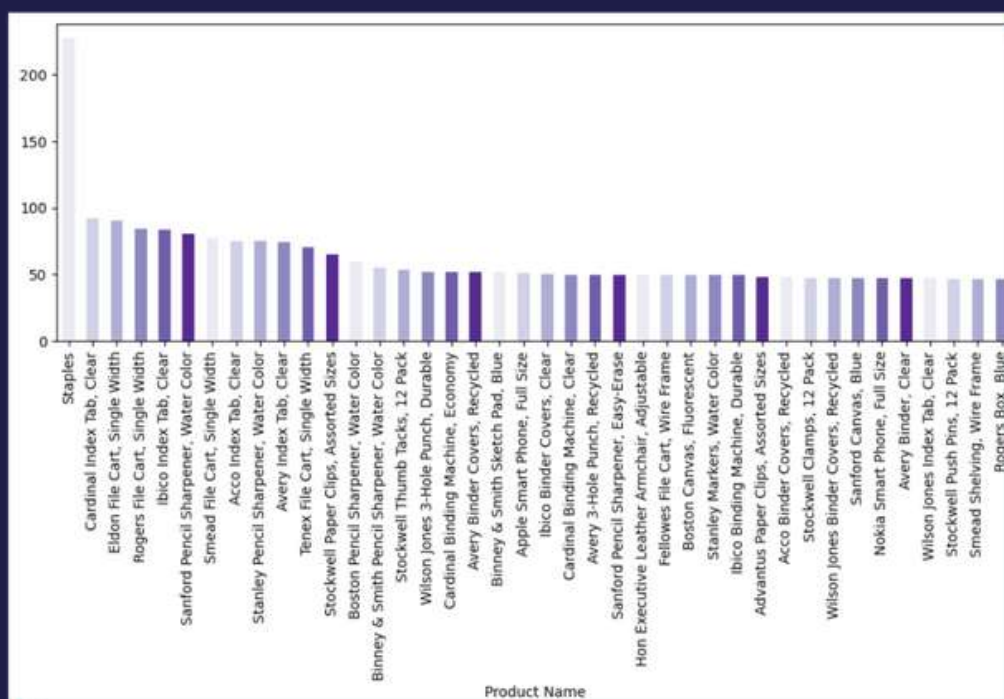


Impact of Discount on Sales

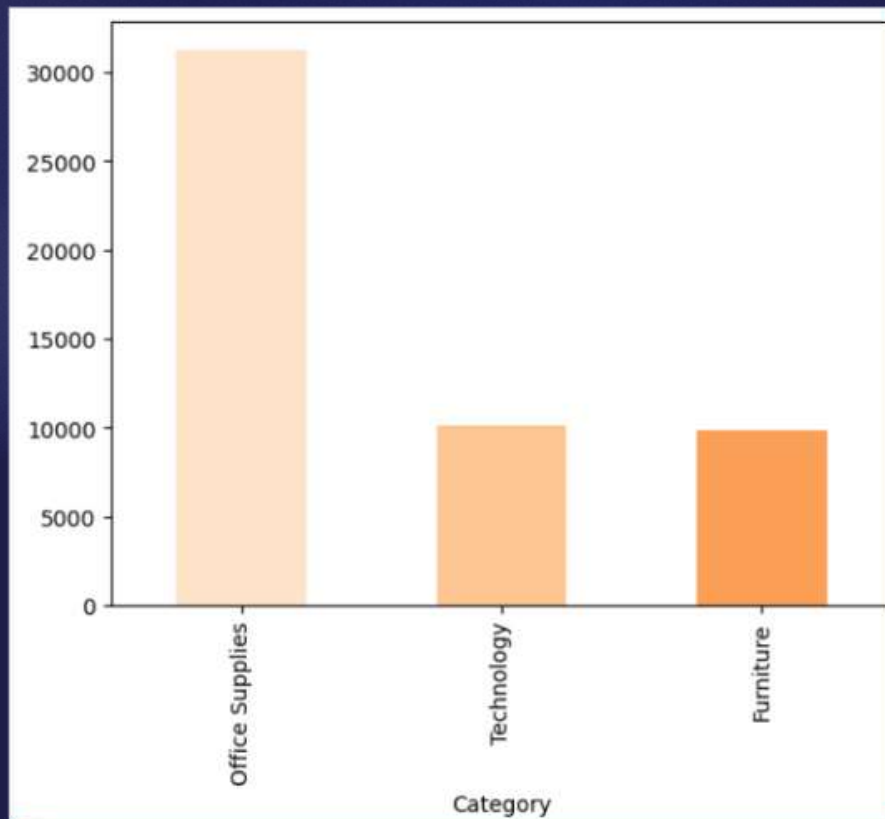


Product Analysis

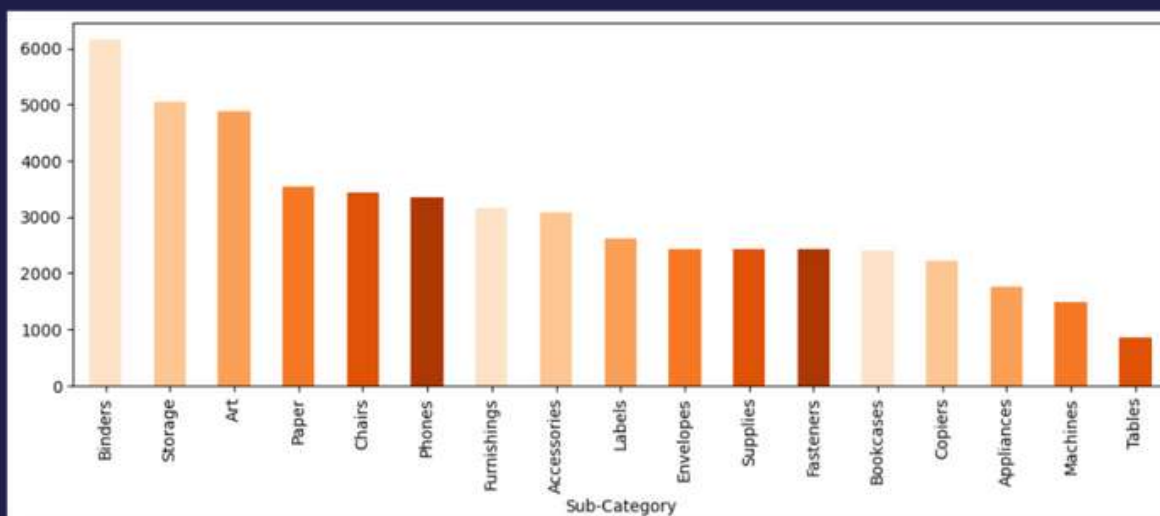
Top Selling Products



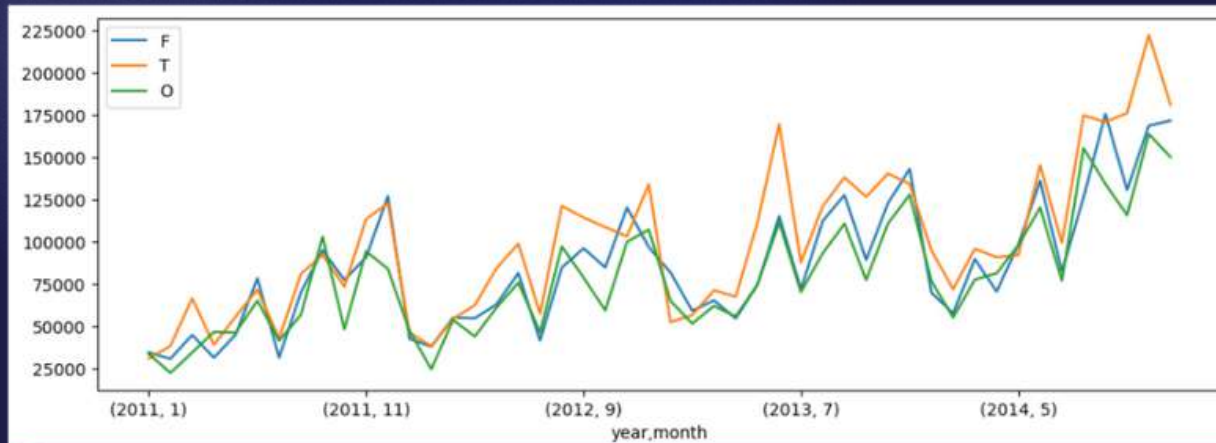
Most Popular Category



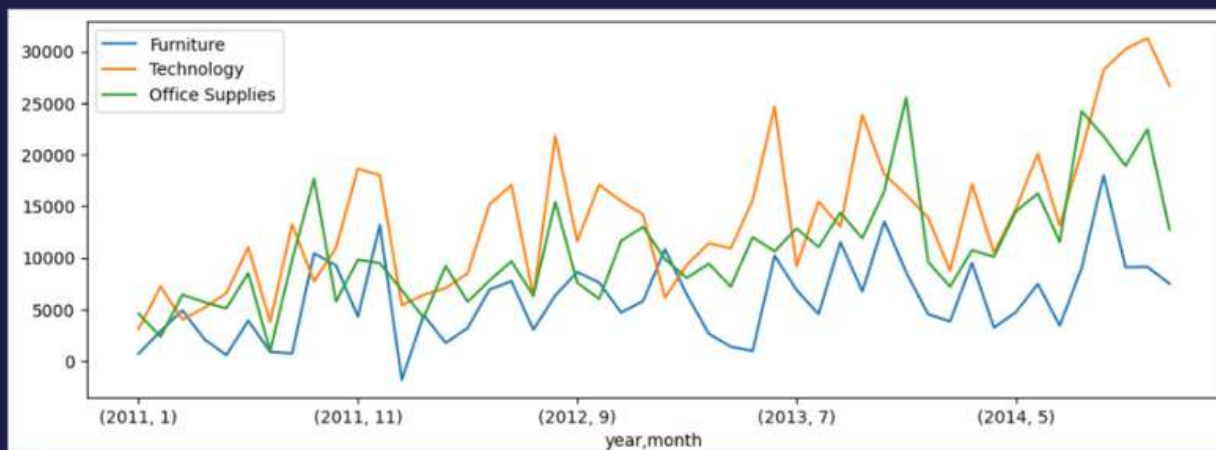
Most Popular Sub-Category



Sales trend by Product Category

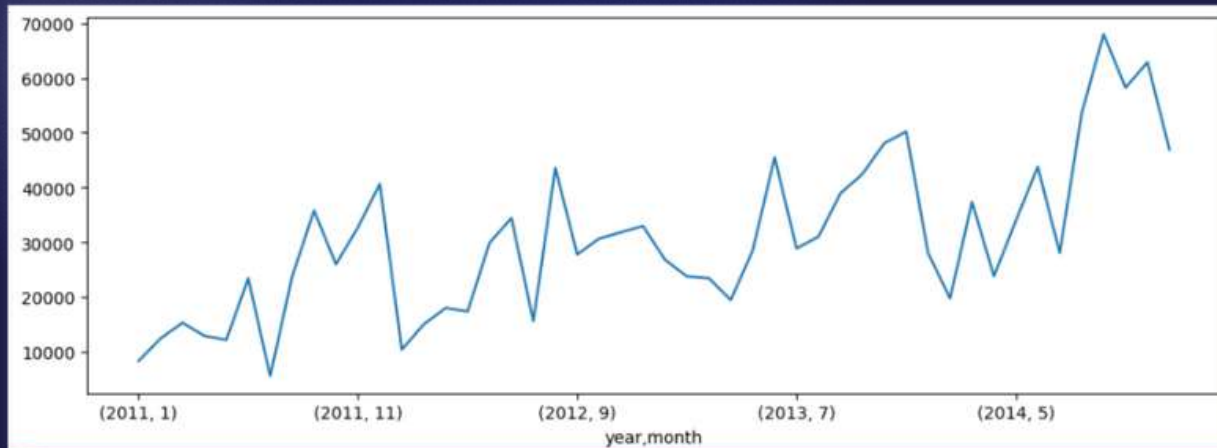


Profitability for Different Categories overtime

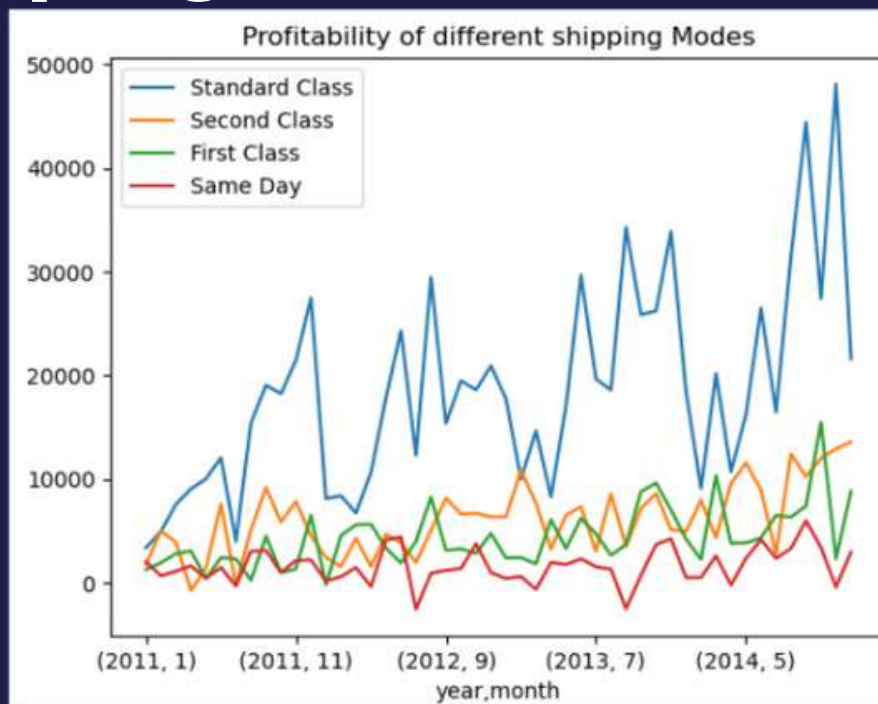


Profit Analysis

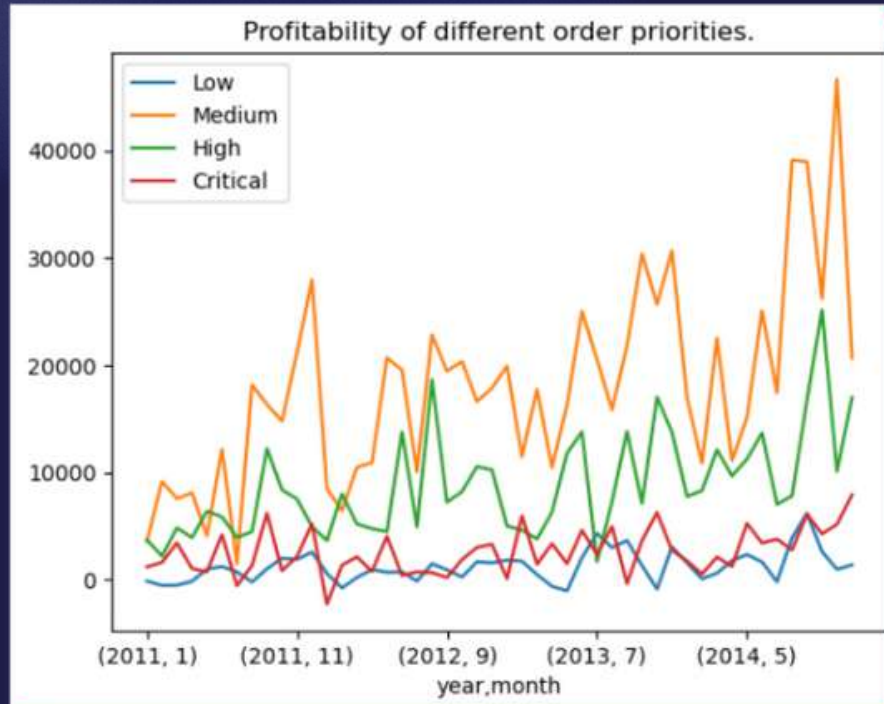
Profit Trend OverTime



Profitability in different shipping Modes

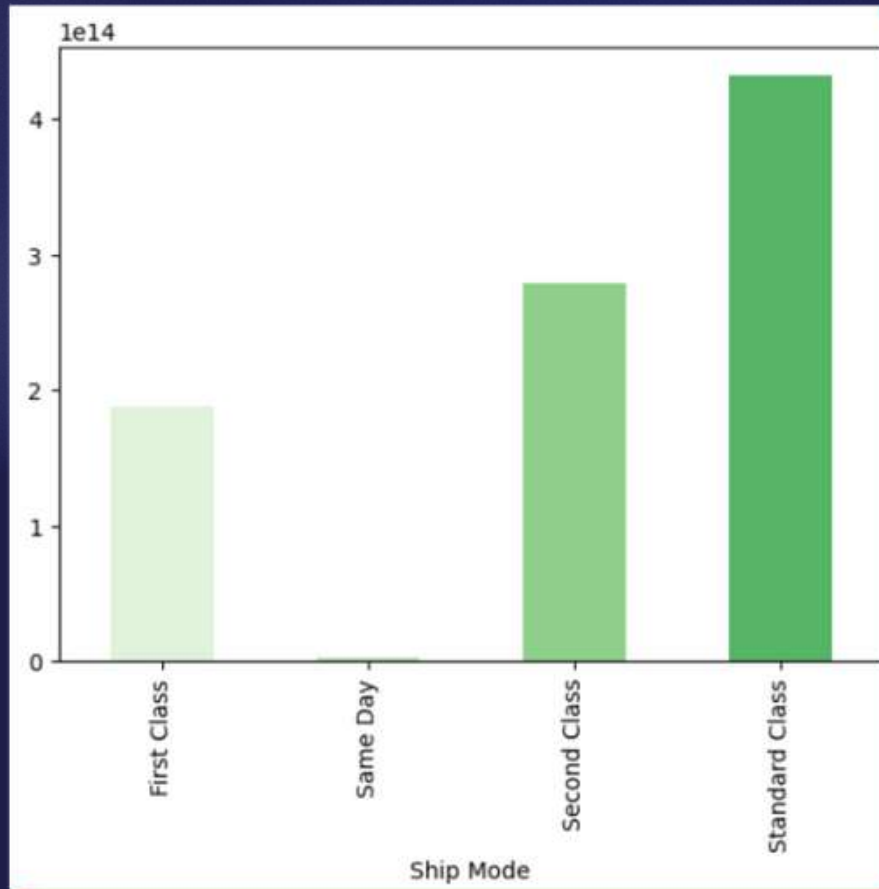


Profitability of different order priorities

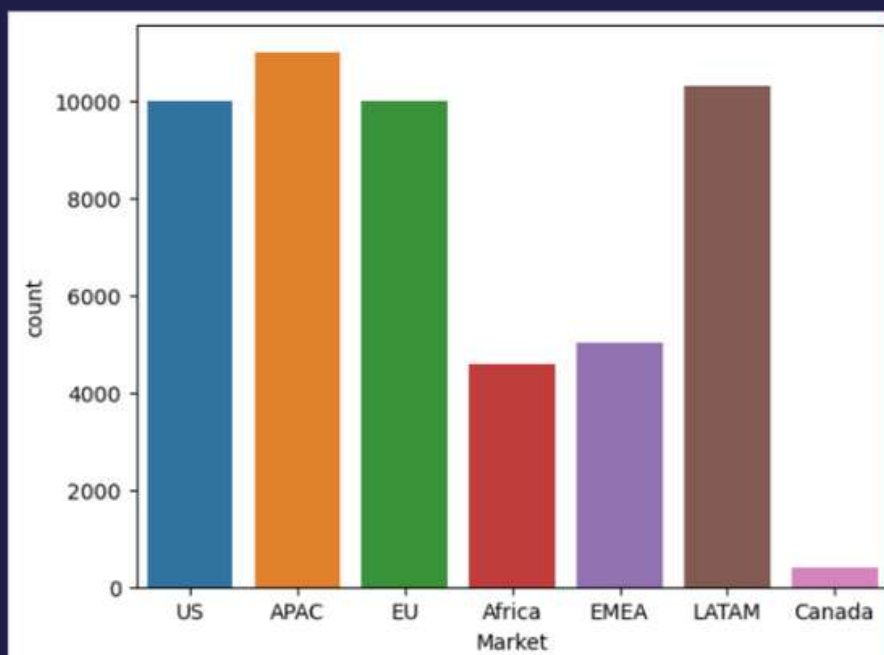


Shipping Analysis

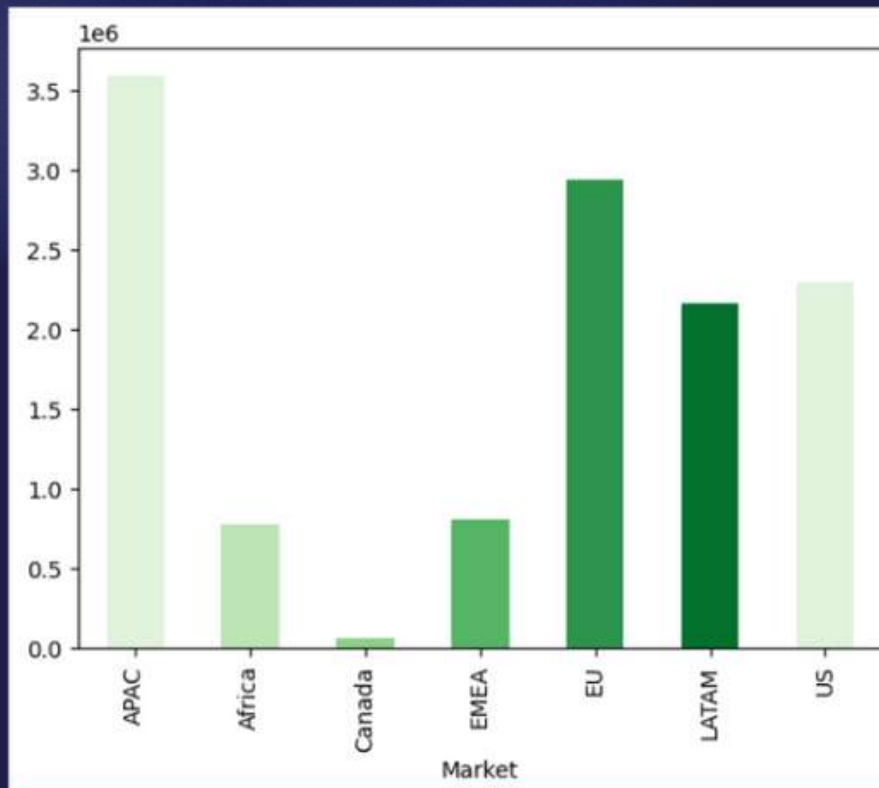
Efficiency of different shipping modes



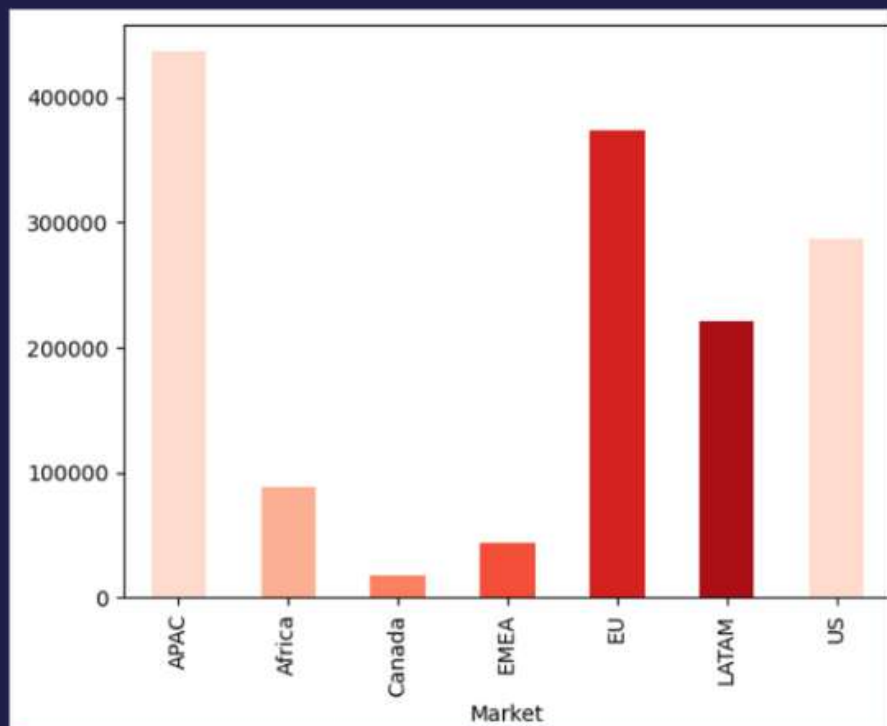
Market Analysis



Market by sales



Profit Obtained from Market



Conclusion

In conclusion, the exploratory data analysis (EDA) conducted on the sales dataset has provided valuable insights into various aspects of the sales process. Through meticulous examination and visualization of the data, we have gained a deeper understanding of sales trends, customer behavior, and factors influencing profitability.

The analysis revealed several key findings:

- 1. Sales Trends:** We observed fluctuating sales patterns over time, with certain months or seasons exhibiting higher sales volumes. Understanding these trends can help businesses optimize inventory management and marketing strategies to capitalize on peak periods.
- 2. Customer Segmentation:** By analyzing customer segments based on demographics and purchasing behavior, we identified distinct customer profiles with varying preferences and purchasing power. Tailoring marketing campaigns and product offerings to specific customer segments can enhance customer engagement and loyalty.
- 3. Product Performance:** We identified top-selling products and categories, as well as products with low sales volumes. This insight can guide inventory management decisions, such as reallocating resources to promote high-performing products or optimizing pricing strategies for underperforming ones.
- 4. Discount Impact:** The analysis revealed the impact of discounts on sales and profitability. While discounts may stimulate short-term sales growth, they can also erode profit margins if not carefully managed. Balancing discount offers with profitability goals is crucial for sustainable business growth.
- 5. Shipping Costs:** We examined the distribution of shipping costs and their impact on profit margins. Optimizing shipping strategies, such as

negotiating rates with carriers or implementing cost-effective shipping options, can mitigate expenses and improve overall profitability.

Overall, the insights derived from this analysis serve as valuable inputs for strategic decision-making and business optimization. By leveraging data-driven insights, businesses can identify growth opportunities, mitigate risks, and enhance operational efficiency in the competitive marketplace.

As we conclude this project, it is essential to acknowledge the limitations of our analysis, such as data quality issues or the absence of external factors that may influence sales performance. Moving forward, continued data collection, analysis, and refinement of analytical techniques will be essential for maintaining a competitive edge and driving sustainable business growth in the dynamic landscape of sales and marketing.

Recommendations

Based on the findings of the exploratory data analysis (EDA), several recommendations emerge to guide strategic decision-making and optimize business performance:

- 1. Personalized Marketing Strategies:** Tailor marketing campaigns and promotions to specific customer segments identified through the analysis. Leverage demographic and behavioral insights to deliver personalized offers, product recommendations, and targeted communications that resonate with each customer segment.
- 2. Dynamic Pricing Strategies:** Implement dynamic pricing strategies based on product performance, customer demand, and market trends. Utilize data-driven pricing models to adjust prices in real-time, optimize profit margins, and remain competitive in the market.
- 3. Optimized Inventory Management:** Utilize sales trends and product performance metrics to optimize inventory management practices. Implement demand forecasting models to anticipate customer demand,

minimize stockouts, and reduce excess inventory holding costs.

4. **Strategic Discounting:** Evaluate the effectiveness of discount offers in driving sales and profitability. Implement targeted discounting strategies for specific products, customer segments, or promotional periods to maximize sales while maintaining profit margins.

5. **Efficient Shipping Practices:** Analyze shipping costs and optimize shipping strategies to minimize expenses and improve overall profitability. Explore opportunities to negotiate favorable shipping rates with carriers, optimize packaging processes, and leverage cost-effective shipping options.

6. **Investment in Customer Experience:** Prioritize investments in enhancing the overall customer experience across all touchpoints. Focus on providing seamless shopping experiences, responsive customer support, and convenient shipping options to foster customer satisfaction and loyalty.

7. **Continuous Monitoring and Adaptation:** Establish a framework for continuous monitoring of sales performance metrics and market dynamics. Regularly review and adapt strategies based on emerging trends, customer feedback, and competitive intelligence to stay agile and responsive in the ever-evolving marketplace.

8. **Investment in Data Analytics Capabilities:** Strengthen data analytics capabilities by investing in advanced analytics tools, talent development, and infrastructure enhancements. Leverage emerging technologies such as machine learning and predictive analytics to unlock deeper insights and drive innovation in sales and marketing strategies.

By implementing these recommendations, businesses can leverage data-driven insights to optimize sales performance, enhance customer satisfaction, and drive sustainable growth in today's competitive marketplace. Continued focus on data-driven decision-making and agility in strategy execution will be key to maintaining a competitive edge and achieving long-term success.

☀ Thank you!

Project Created by Rahul badola

