

CSV Matrix

Unlocking Insights from Data



Team Details

Team Name: Joyguru

Team Leader: Somesh Ghosh (2022UGEC013)

Team Member 1: Siddhant Dash (2022UGEC100)

Team Member 2: Subhomoy Maji (2022UGEC011)

Note:

This content was created by Team Joyguru on 26/10/23, exclusively for participation in CSV Matrix, an event by Ojass, NIT Jamshedpur, in collaboration with ACC Club. Please be informed that all content in this presentation is original and has been crafted independently. The logos and cliparts used in the presentation might have been sourced from external platforms. No content has been copied from any source other than the mentioned logos and cliparts.

Table of Contents

01

Introduction

Overview of the dataset's
significance

02

Exploration

Initial analysis and
understanding of data

03

Sales

Analysis of sales trends
and patterns

04

Categories

Top selling categories by
region

05

Customers

Insights from unique
customer IDs

06

Conclusion

Summary of key findings
and insights

Technologies/Tools Used:

We have primarily used **Python** for data analysis. The libraries used are:

- For Data Processing
 - a. Pandas
 - b. NumPy
 - c. Microsoft Excel (Software)
- For Data Visualization
 - a. Matplotlib
 - b. Seaborn



Documentations Referred:

<https://pandas.pydata.org/docs/>
<https://matplotlib.org/stable/index.html>
<https://seaborn.pydata.org/>

Link to Google Colab:

(All the Python Code used are compiled here)

<https://colab.research.google.com/drive/1XX-jrgMDJqltG3QGGE5U08htO2zVkr1X?usp=sharing>

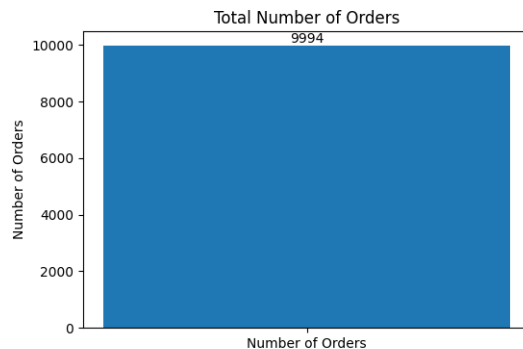
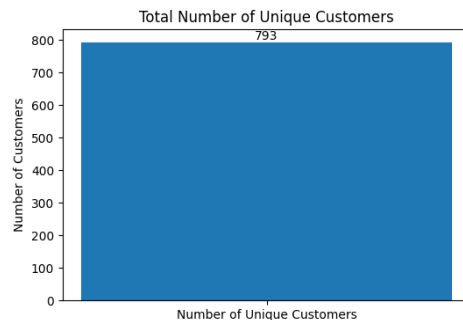
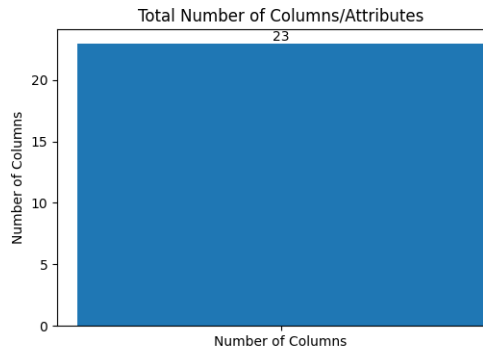
Introduction

Brief Description of the Dataset:

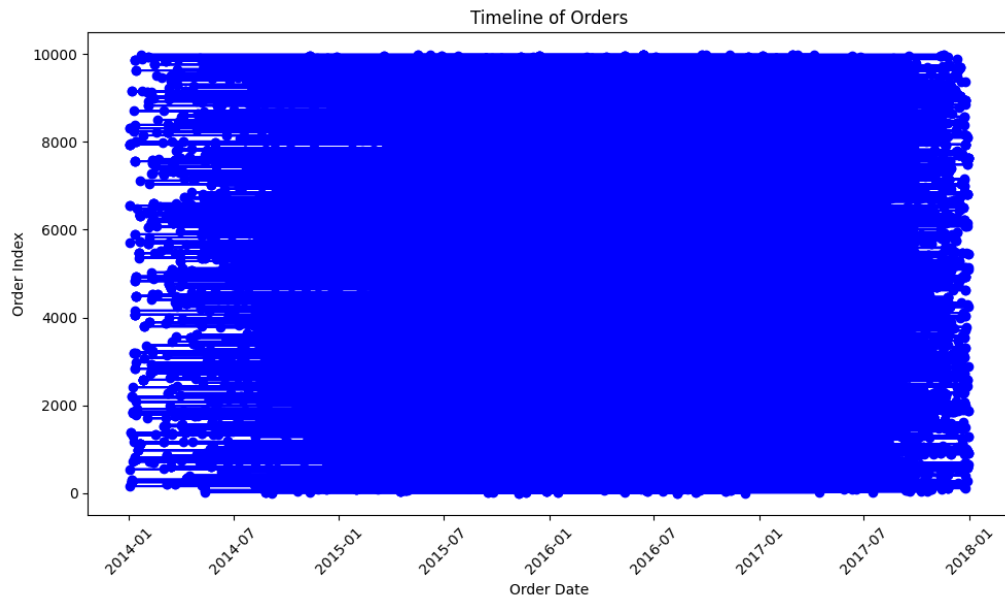
Our dataset contains various attributes representing different aspects of orders and products. For instance, 'Order ID' represents unique order identifiers, 'Sales' indicates the total sales for each order, and 'Profit' shows the profit earned from each transaction.

Dataset Size:

We have a substantial dataset comprising a total of 9994 rows. This substantial volume of data provides a rich ground for analysis and insights.



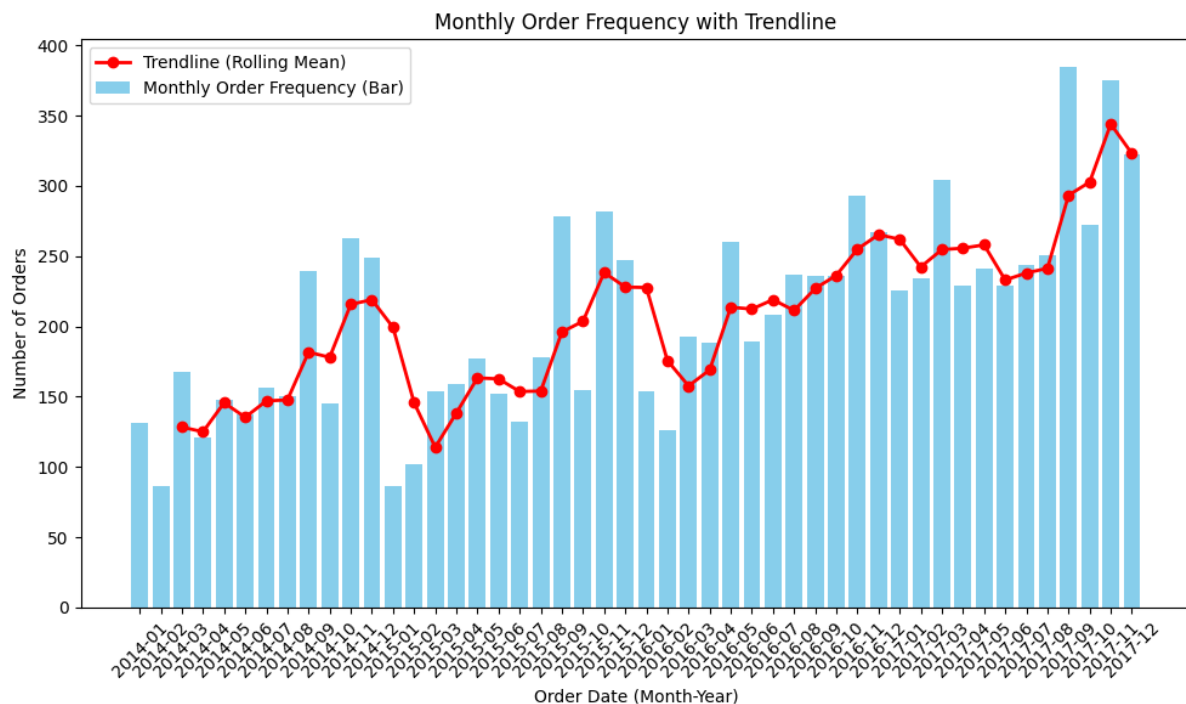
Introduction



Dataset Context:

This dataset seems to originate from an e-commerce website, representing the sales details of the e-commerce website which seems to be operating in the US. It covers a specific period (from year 2014 to 2017), allowing us to delve into the patterns and trends of the provided data.

Noticeable Trend

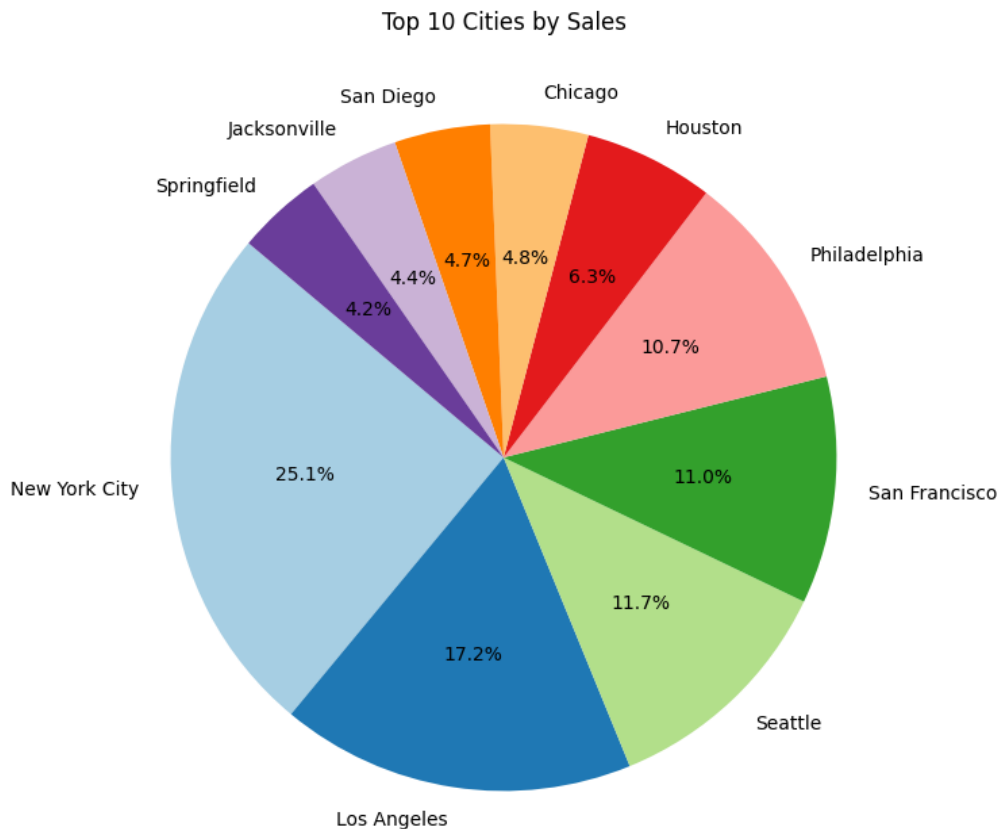


In the dataset, a distinct pattern emerges in the order frequency throughout the year. Initially, during the **early months**, there is a relatively low frequency of orders. However, as the year progresses, the order frequency steadily rises. A slight dip is observed around the month of August, followed by a significant surge in orders towards the end of the year. This cyclical trend highlights a clear seasonality in the ordering behavior, with a pronounced peak in demand during the latter part of the year."

Top 10 Cities by Sales

In our analysis, a prominent trend surfaces in the distribution of orders across cities. **New York City** emerges as the frontrunner, commanding an impressive 25.1% of all orders, making it the primary hub of activity. Following closely, **Los Angeles** captures a significant share at **17.2%**, while **Seattle** secures **11.7%** of the total orders.

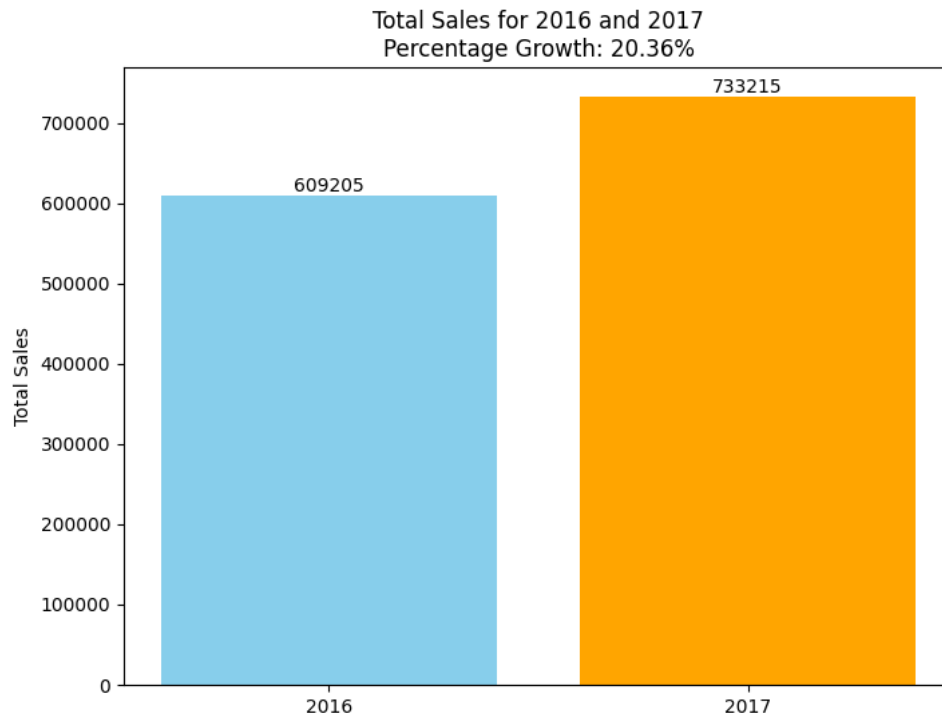
This insight was derived through strategic use of the **'groupby'** and **'sort_values()'** functions in pandas, coupled with the visualization prowess of the **'pie()'** function in matplotlib. This analytical approach allowed us to clearly visualize and comprehend the regional distribution of orders, enabling actionable insights for further strategic decisions.



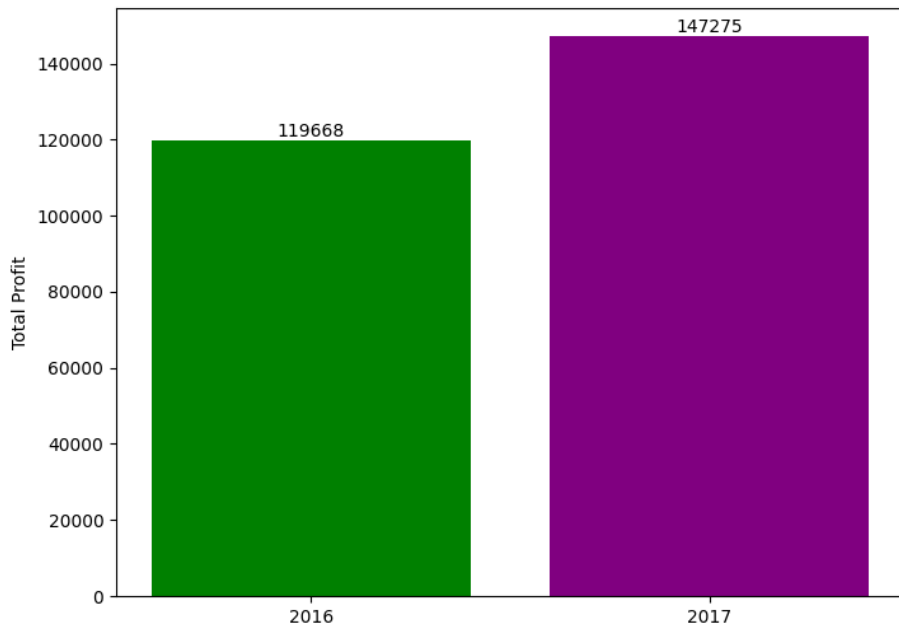
Total Sales for Financial Year 2016 and 2017 with Percentage Growth

In the year 2016, the total sales figures stood at an impressive **\$609,205.60**, reflecting a significant portion of the overall revenue. However, in the subsequent year, **2017**, a **substantial surge** was witnessed, with total sales soaring to **\$733,215.26**. This remarkable growth, denoting a substantial **20.36%** increase, ***underscores a robust and promising trajectory for the business.***

Leveraging advanced tools like Python's Pandas library, we harnessed the power of functions such as '**groupby()**' and '**sum()**' to meticulously aggregate sales data for each year. Additionally, the implementation of straightforward yet potent arithmetic calculations facilitated the determination of the growth percentage between the two pivotal years.



Total Profits for Financial Year 2016 and 2017 with Percentage Growth



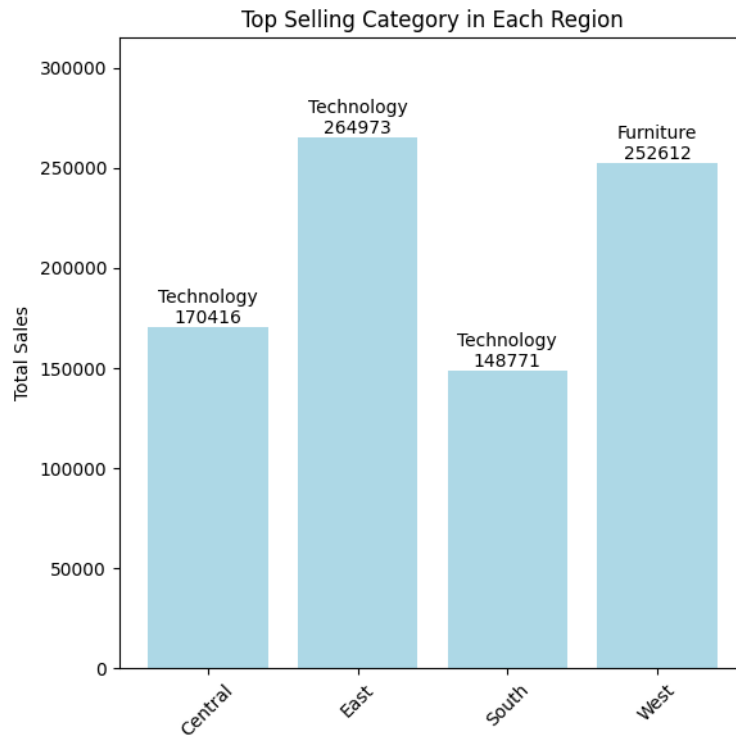
In 2016, the total profit margins stood at an impressive **\$81,795.17**, underscoring the company's financial stability. Fast forward to 2017, and the profit margins experienced a noteworthy surge, reaching **\$93,439.27**. This discernible growth marked a significant 14.24% increase in profits from the previous year, illuminating the company's financial prowess and strategic acumen.

Leveraging advanced tools like Python's Pandas library, we harnessed the power of functions such as '**groupby()**' and '**sum()**' to meticulously aggregate sales data for each year. Additionally, the implementation of straightforward yet potent arithmetic calculations facilitated the determination of the growth percentage between the two pivotal years.

Top Selling Category from Each Region

Across different regions, specific product categories emerged as consistent leaders. In the **Central** region, **'Technology'** products asserted their dominance, solidifying their position as the top-selling category. A parallel trend was observed in the **Eastern** and **Southern** regions, where 'Technology' products continued to lead the sales charge. In the **Western** region, **'Furniture'** emerged as the category of choice, showcasing the region's unique market preferences..

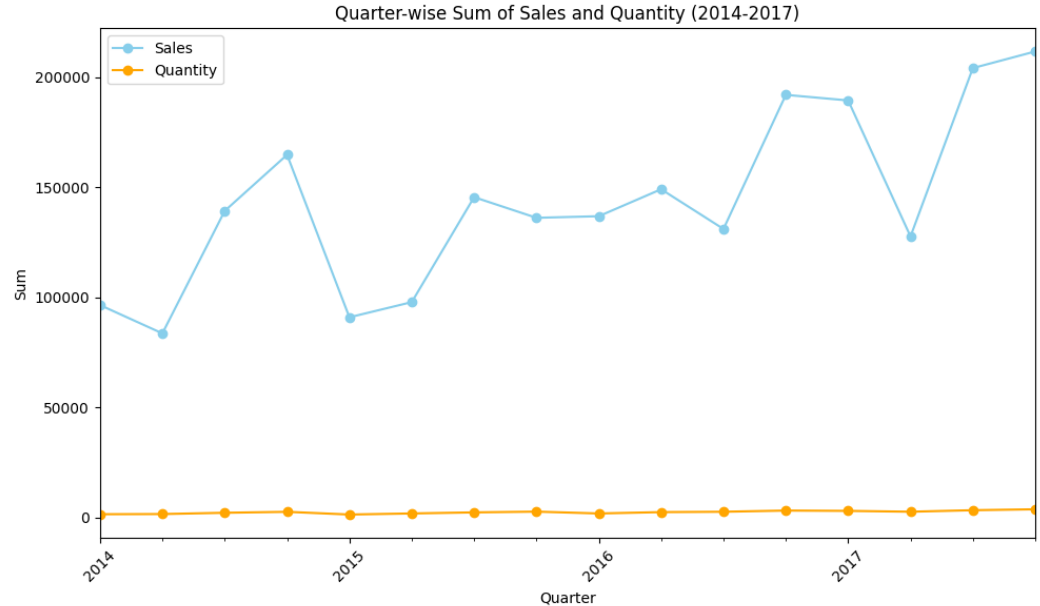
Employing Python's Pandas library, we utilized functions such as **'groupby()'** and **'max()'** to identify the top-selling categories within each region. By scrutinizing this granular data, the company gains a strategic advantage, enabling focused marketing efforts, inventory management, and product development. This analytical depth provides a roadmap for targeted business strategies, ensuring alignment with regional customer preferences and maximizing revenue potential.



Quarter-wise Sum of Sales and Quantity (2014-2017)

Delving into the quarterly data spanning from 2014 to 2017, our analysis unveiled intricate patterns in sales and quantity metrics, painting a detailed picture of the company's performance over the years. Across this period, each year was characterized by distinct sales and quantity dynamics.

In the initial quarters of **2014**, sales figures stood at **\$96,498.72** and steadily increased, culminating in an impressive **\$164,805.88** by the end of the year. Concurrently, quantity metrics showcased consistent growth, indicating a burgeoning customer base and market demand.



Quarter-wise Sum of Sales and Quantity (2014-2017)

The subsequent years witnessed fluctuating trends. In **2015**, a dip in sales during Q1 was followed by a substantial surge, reaching **\$145,554.23** in Q3. Quantity metrics mirrored this trend, reflecting market responsiveness to product offerings.

The year **2016** marked a period of renewed vigor, with Q4 sales touching a pinnacle at **\$192,059.89**, indicating heightened market penetration. Quantity metrics displayed a parallel trajectory, emphasizing increased product adoption.

2017 heralded remarkable achievements, with **Q3** emerging as a standout quarter, boasting sales of **\$204,221.99**. Quantity metrics reached an all-time high in Q4, highlighting robust customer engagement and market traction.

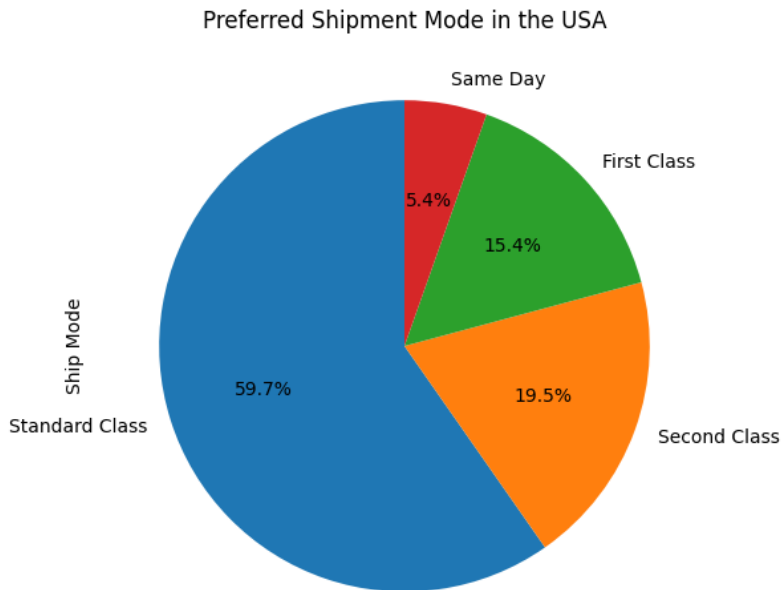


Preferred Shipment Mode in the USA

Among the various options, '**Ship Mode**' emerged as the unequivocal choice, commanding a substantial **59.7%** of the market share. This overwhelming preference underscores the significance of efficient, cost-effective, and reliable shipping services in customer satisfaction and retention.

Additionally, '**Second Class**' shipments garnered considerable attention, accounting for **19.5%** of the market share. The allure of timely delivery at reasonable costs solidified its position as a favored choice among customers.

The '**First Class**' option, valued for its speed and reliability, captured a notable **15.4%** of the market, appealing to customers who prioritize swift deliveries without compromising on quality service. Meanwhile, '**Same Day**' shipments, despite a relatively smaller percentage at **5.4%**, catered to the urgent needs of specific customer segments.



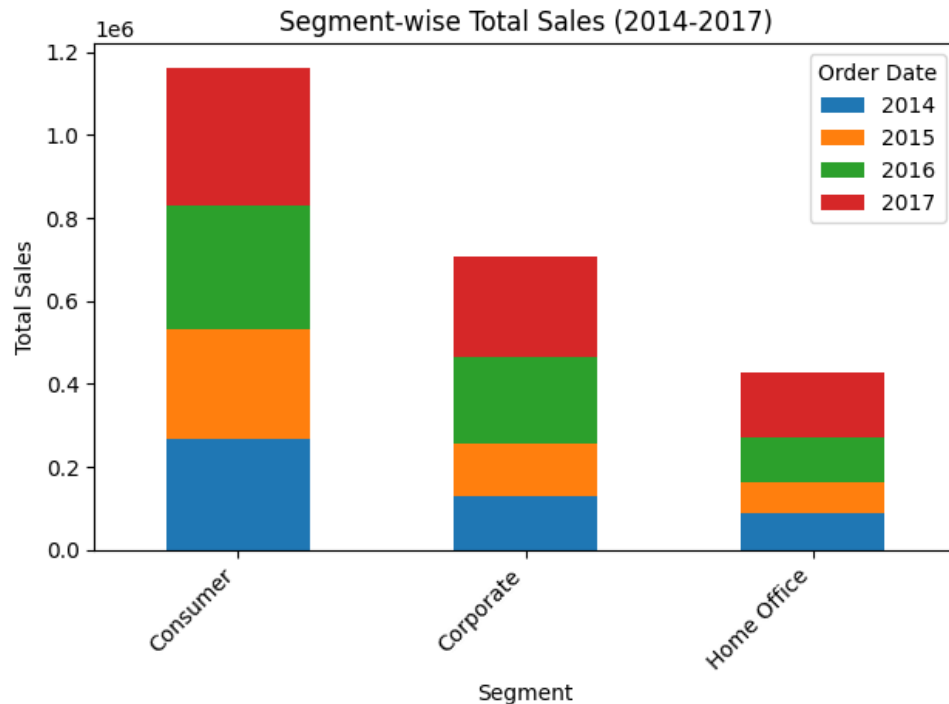
Segment-wise Total Sales (2014-2017)

Consumer Segment:

The Consumer segment showcased consistent growth, with sales steadily increasing from **\$266,096.81** in **2014** to a robust **\$331,904.70** in **2017**. This trajectory underscored the segment's significant contribution to the company's revenue stream, emphasizing the enduring appeal of products among individual buyers.

Corporate Segment:

Corporate sales, while substantial, displayed a more fluctuating pattern. Starting at \$128,434.87 in 2014, sales experienced notable peaks and troughs, ultimately reaching \$241,847.82 in 2017. Despite fluctuations, the Corporate segment's substantial market presence reaffirmed its vital role in the company's financial landscape.



Segment-wise Total Sales (2014-2017)

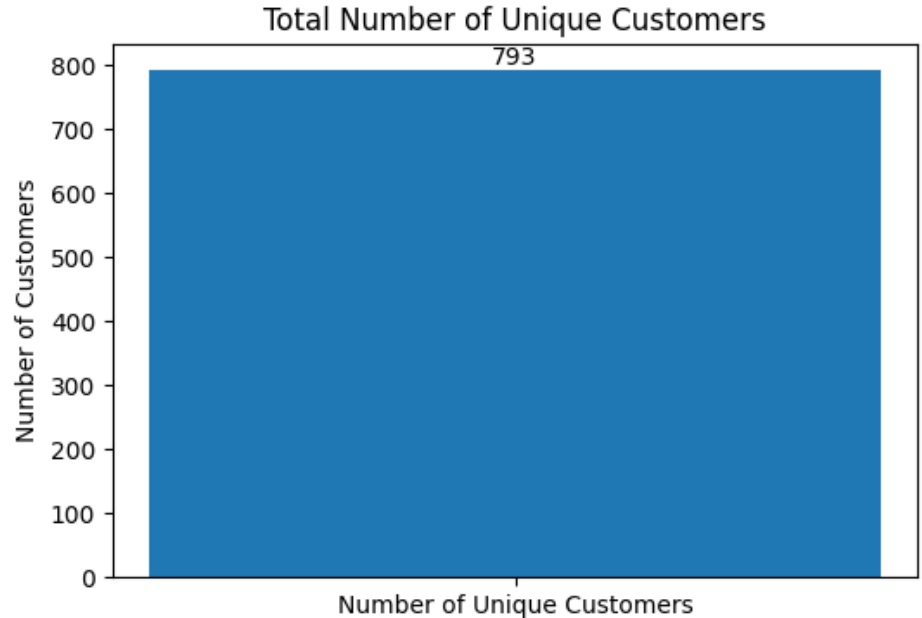
Home Office Segment:

The Home Office segment exhibited intriguing trends. Beginning at \$89,715.81 in 2014, sales faced a dip in 2015 before experiencing a significant upswing in subsequent years, reaching \$159,462.73 in 2017. This resurgence highlighted the segment's adaptability and market responsiveness, positioning it as a pivotal revenue driver.

These insights empower the company to optimize its offerings, ensuring alignment with diverse customer segments and fortifying its position in the market.

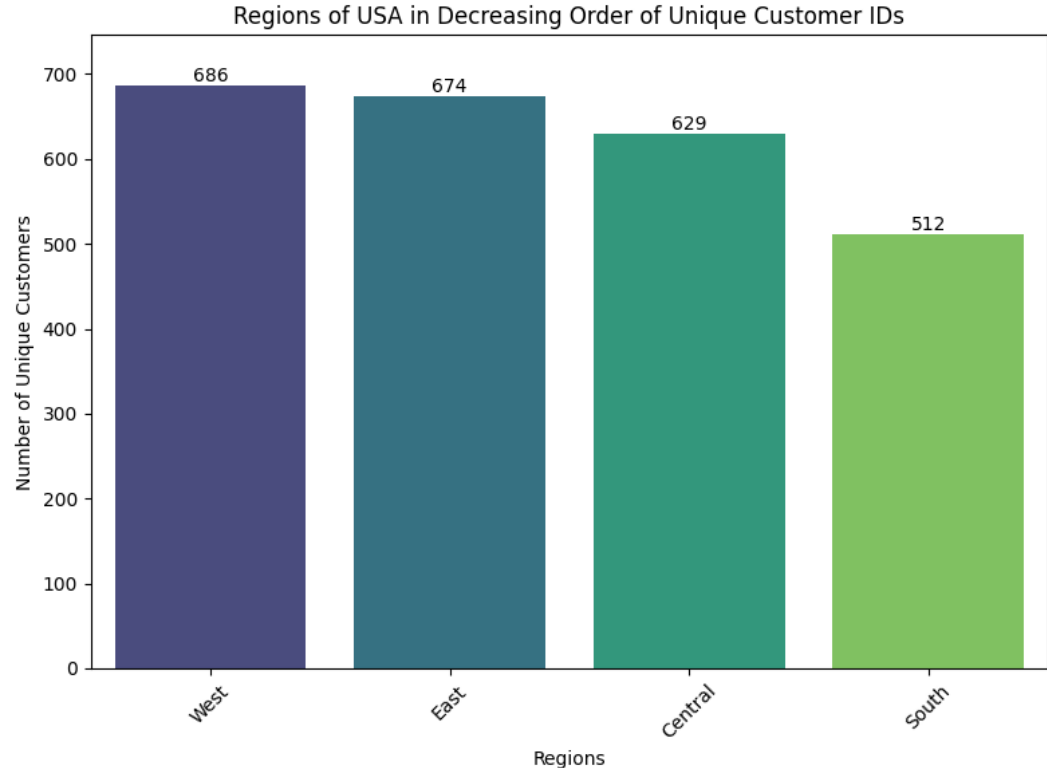
Total Number of Unique Customer IDs

In our exploration of the dataset, a striking revelation surfaced - the company has fostered connections with an impressive total of 793 unique customers. This robust customer base signifies not just the breadth but also the depth of the company's market reach and brand appeal.



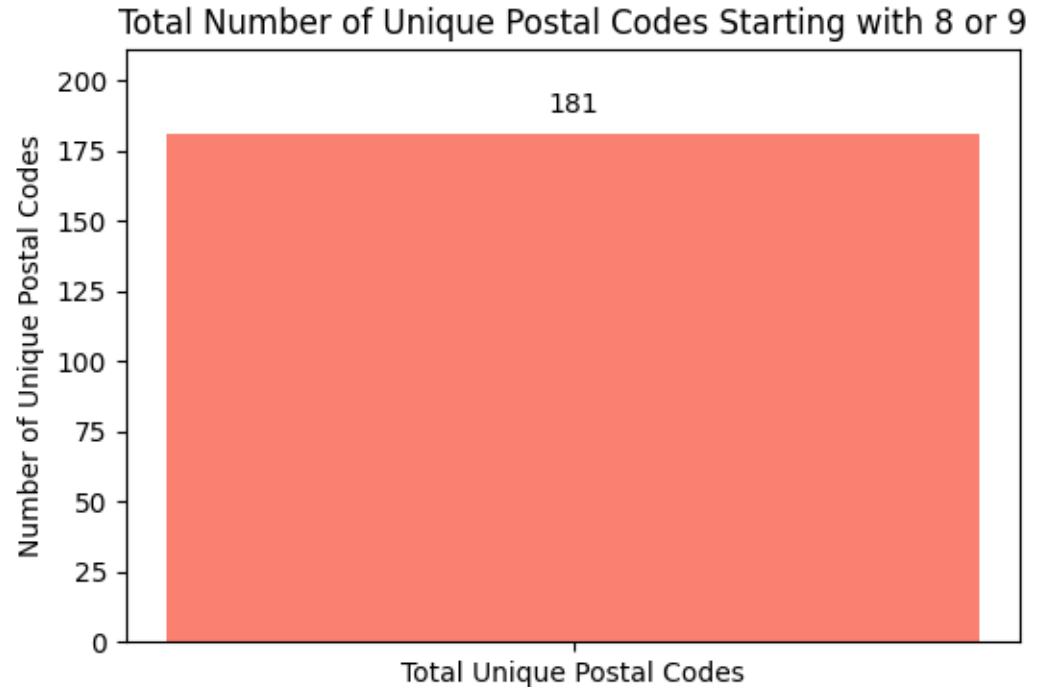
Regions of the USA in Decreasing Order of Unique Customer IDs

A detailed analysis of customer interactions across various regions in the USA yielded compelling findings. The West region emerged as a vibrant hub of customer engagement, boasting a robust network of 686 unique customer IDs. Following closely, the East region established a significant market presence with 674 unique customer IDs. In the Central region, 629 unique customer IDs underscored the company's strong foothold, while the South region, with 512 unique customer IDs, showcased a steady and growing customer base.



Total Number of Unique Postal Codes Starting with 8 or 9

Our data analysis revealed a significant presence of 181 unique postal codes commencing with the digits 8 or 9. This specific subset of postal codes delineates a distinct market segment, encompassing areas with unique geographical characteristics and potentially diverse customer needs.



Conclusions!

Conclusion 1:

Quarter 3 is very important for the company and maximum orders were reported in this time. Hence company could arrange for efficient and cost effective logistic mechanism to increase its margins by reducing operational costs. This came into notice when we found that the profit percentage increased only by 14.24% whereas its revenue increased by 20.36%, which clearly shows that the margin of the company reduced significantly.

Conclusion 2:

Technological products should be available in the stock or warehouses as they are in high demand in almost all the regions. But in Western region, furniture products should be given more preference in warehouse management. This can significantly increase the company revenue as they can deliver products much efficiently and quickly reducing delivery time and expenses and increasing customer satisfaction.

Conclusions!

Conclusion 3:

Company should focus on increase its consumer base from 793 unique customers. The company has done great in creating a robust customer base signifying not just the breadth but also the depth of the company's market reach and brand appeal. The customers are quite satisfied as life time value of the customer is \$ 18298 and the average order per customer is 6.31.

Conclusion 4:

The company should increase its holding in customer segment as it is quite resilient and also contributing to the majority share of company's revenue. In spite of that maximum orders per customer is definitely from the corporate segment. Hence company should try to crack as much deals from the corporates as much possible. Company should increase its penetration in Southern region as the consumer base is quite low.

Conclusions!

Conclusion 5:

The overwhelming preference to “Standard Class” as the Ship Mode underscores the significance of efficient, cost-effective, and reliable shipping services in customer satisfaction and retention.

Thanks!

Do you have any questions?