



Sales & Customer Analysis

Project Phase - 4

Group Name: Data Vizards

Members:

Muskan Garg

Vatsal Gohel

Engaging Summary (Elevator Pitch)

In the dynamic realm of e-commerce, where online shopping has become the norm, unlocking success hinges on deciphering customer behavior. This dataset, encompassing 286,392 entries across 36 columns, delves deep into customer orders on our e-commerce platform. This exploration has unearthed intriguing revelations that illuminate consumer tendencies and trends.

Central to our analysis is a keen examination of discount dynamics, a pivotal factor in discerning the optimal timing for launching our inaugural brick-and-mortar outlet. By dissecting the interplay between discount ratios and sales, we pinpoint the opportune month for a successful soft launch. This insight empowers us to refine pricing strategies, calibrate discount tiers, and harness the psychological allure of discounts to amplify sales.

Even amidst the digital payment surge, the tried-and-true Cash on Delivery (COD) method endures as a favorite. Its resilience challenges the assumption that all customers exclusively favor digital transactions. This paves a path towards over ideas of getting into in-store outlet chain. By embracing the rationale behind this preference, we can seamlessly meld the old-school charm of COD with modern shopping conveniences, bestowing customers with a seamless and trustworthy purchasing journey.

In essence, this dataset unveils geographically nuanced customer behaviors, the sway of discount percentages on sales, and the enduring prominence of Cash on Delivery. Armed with these insights, businesses can craft laser-focused marketing endeavors, finesse pricing strategies, fine-tune discount deployment, and diversify payment avenues. Seize this invaluable wisdom to propel growth, heighten customer contentment, and secure an enviable competitive edge.

Description of the Dataset

This dataset contains information about orders placed by customers on an e-commerce website. It includes details about the orders such as the order ID, order date, order status, item ID, quantity ordered, price, value, discount amount, total, customer ID, year, reference number, age, zip code, and discount percentage. The dataset has a total of 286,392 entries with 36 columns. There are several categorical and numerical variables such as quantity ordered, price, value, discount amount, total are numerical data, while the order date, and order status and so on are categorical data.

Data Cleaning

In order to visualize the dataset better and make the cleaning process easier, we loaded the data in a Python file. We performed some descriptive analysis on the data. We imported the necessary libraries to dive deep into the dataset to check for missing values, outliers, and duplications in the dataset. Outliers and missing values did not need to be rectified, so we did not touch the dataset. As the dataset contained sensitive information such as SSN, we removed it. There is an error with the value column as quantity ordered is decreased by a value in all rows. In order to correct the data, we created two calculated fields, "actual value" and "actual total". Likewise, we dropped those columns that were irrelevant and redundant such as value, total, Name Prefix, Middle Initial, User name, Full name, SSN. To analyze customer behavior patterns based on age group, we created a calculated field named age bin using that column.

Intended Audience

- **Sales and Marketing Teams:** The analysis can be valuable for sales and marketing teams within the organization. They can benefit from insights into sales patterns, popular products, customer preferences, and pricing strategies to optimize marketing campaigns, improve sales forecasting, and enhance customer targeting.
- **Operations and Inventory Management:** The analysis can provide insights to operations and inventory management teams. They can leverage the findings to optimize inventory levels, identify fast-selling items, manage stock-outs or overstocking situations, and improve supply chain management.
- **Finance and Accounting Departments:** The analysis, particularly with a focus on financial metrics such as gross sales, net sales, discounts, and revenue trends, can be useful for the finance and accounting departments. They can use the results to assess financial performance, evaluate pricing strategies, calculate profit margins, and track the impact of discounts on revenue.
- **Business Executives and Decision-makers:** The presented results can be relevant for business executives and decision-makers who require an overview of sales performance, customer insights, and financial analysis. These individuals can utilize the findings to make strategic decisions related to resource allocation, product pricing, marketing budgets, and overall business planning.

There can be more targeted audience as well like Online Retailers, Market Researchers, Sentiment Analysts, Marketing Professionals and Business Owners.

Limitations

The analysis of the provided dataset is subject to several important limitations that must be acknowledged for a comprehensive understanding of the results. First, the dataset represents a sample and may not encompass the entirety of the data landscape, potentially leading to incomplete insights. Moreover, the extraordinary circumstances of the global COVID-19 pandemic during the years 2020-21 introduce a significant bias, as consumer behavior and market dynamics were drastically altered. This period witnessed divergent trends in spending habits, with some individuals prioritizing essential purchases while others indulged in home-based leisure activities. Additionally, the absence of data on marketing strategies, customer preferences, and external factors hinders a holistic understanding of the sales patterns. The dataset's lack of financial metrics like profit margins limits a comprehensive evaluation of business performance. Furthermore, sensitive personal information, like Social Security Numbers, demands stringent data privacy measures. Lastly, the dataset's deficiency in contextual information about the industry and business prevents a nuanced interpretation of findings, emphasizing the importance of supplementing analysis with industry knowledge for accurate insights.

Ambiguity in the Data

The Sales dataset has few ambiguous or unclear aspects that may affect the analysis. Here are some examples:

- **Status Terminology:** The "status" column contains values such as "received," "complete," "order_refunded," and "canceled." The specific definitions and criteria for these statuses are not explicitly stated. It is unclear what triggers an order to be considered "complete" or "canceled," and what circumstances lead to an "order_refunded". Clarifying the meaning of these status values would provide more context for analysis.

- **Bi_st Column:** The "bi_st" column contains values such as "Valid", "Net", and "Gross". While we previously named it as billing status through potential interpretations, however, without further information or a clear definition provided, the exact meaning of these values remains uncertain. Understanding the intended definition or context of "bi_st" would be necessary for accurate analysis.
- **Data Privacy:** The dataset contains personal information, such as email addresses, Social Security Numbers (SSN), and phone numbers. Handling sensitive personal data requires appropriate measures to ensure compliance with data protection regulations and protect individuals' privacy.

Research questions

1. How does the geographical region impact the sales performance within a specific category? Elaborating on the variations in our performance across different regions of the country?
2. How sales fluctuate in distinct states, categorized by category on a monthly basis? How does regional disparity extend to variations in revenue generation, and are there noticeable patterns in the adoption of payment methods as we shift from one state to another?
3. What intricate insights can be gleaned from analyzing the prevalence of payment methods, considering the nuances of geographical regions, specific product categories, and even gender demographics?
4. How do gender differences impact sales and revenue across various product categories, and which products stand out as the most popular in different states based on total revenues?
5. What is the relationship between the implementation of discounts and the overarching revenue of a company? Unveiling the dynamic trends in total sales over a period and elucidate how the temporal distribution of discount offers influences the sales trajectory?

Graphs and Insights (Findings)

Interactive Dashboard

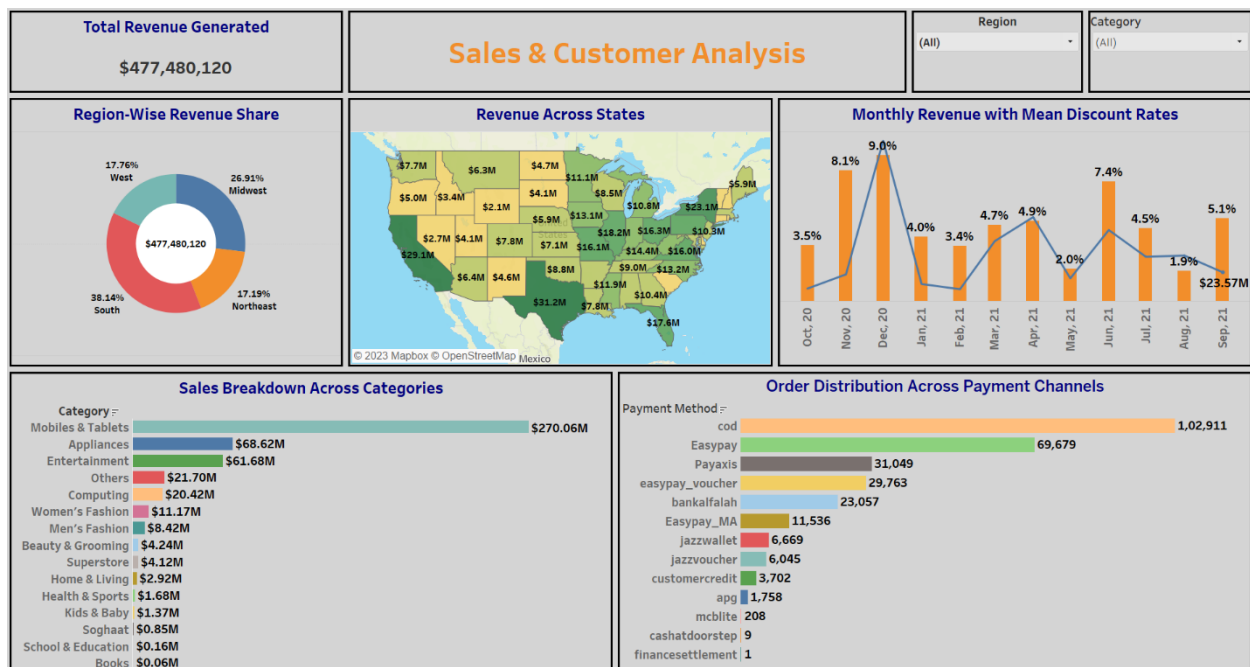
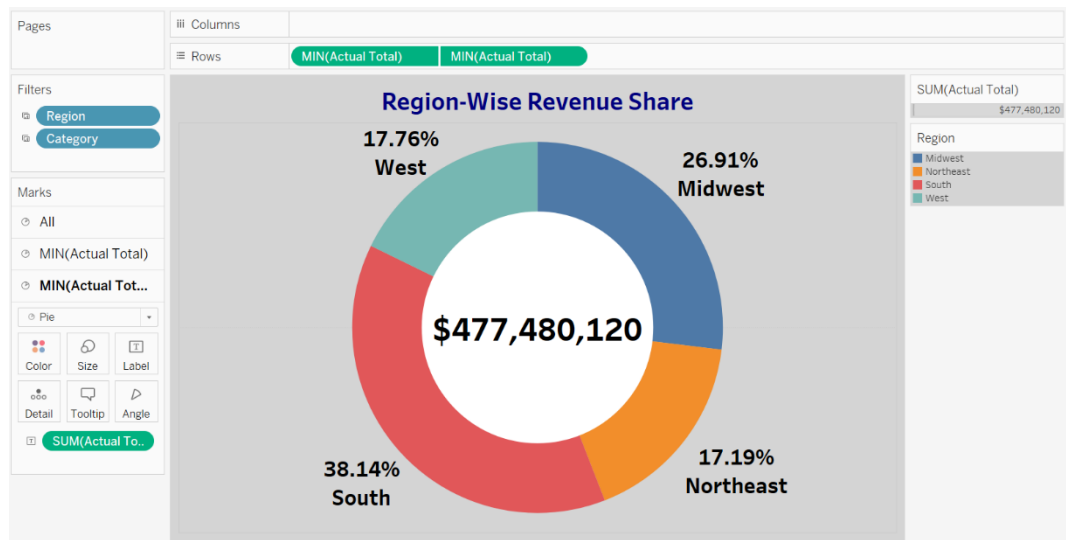
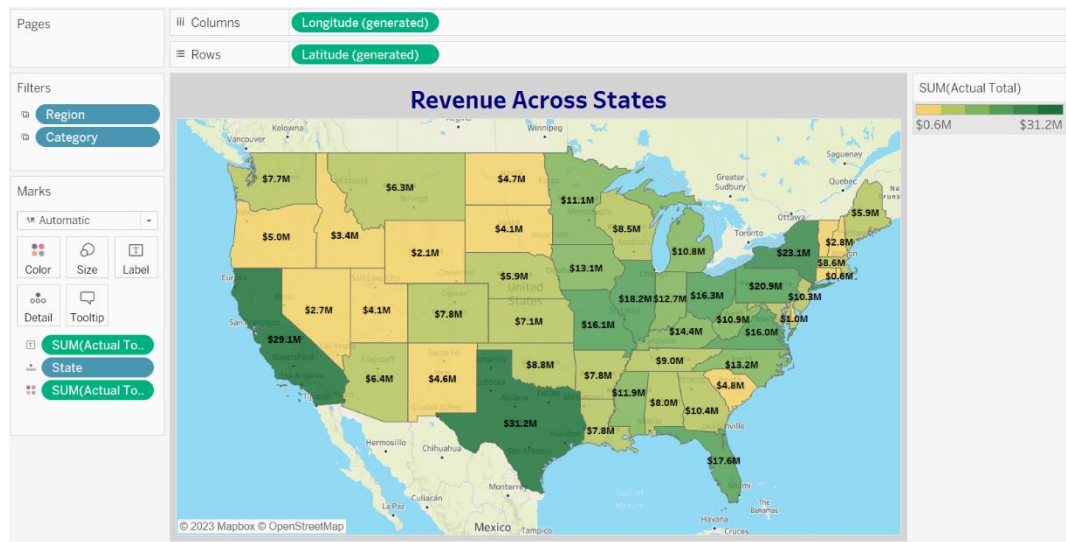


Chart 1: - Region-Wise Revenue Share



Illustrating the geographic breakdown of sales, the provided chart unveils a valuable perspective on revenue distribution. The cumulative sales revenue, amounting to \$477 million, underscores the significance of regional considerations for potential business expansion. Particularly noteworthy is the southern region's contribution, constituting an impressive 38.14% of the total revenue share. This insight illuminates a strategic pathway for deliberating on the optimal location for inaugurating our inaugural physical store. Delving deeper into the analysis by applying filters, a compelling revelation emerges: the state of Texas commands a substantial portion of revenues within the southern region. This revelation prompts us to embark on a more intricate exploration, employing methods like surveys and external campaigns to glean profound insights that can fuel our expansion strategy.

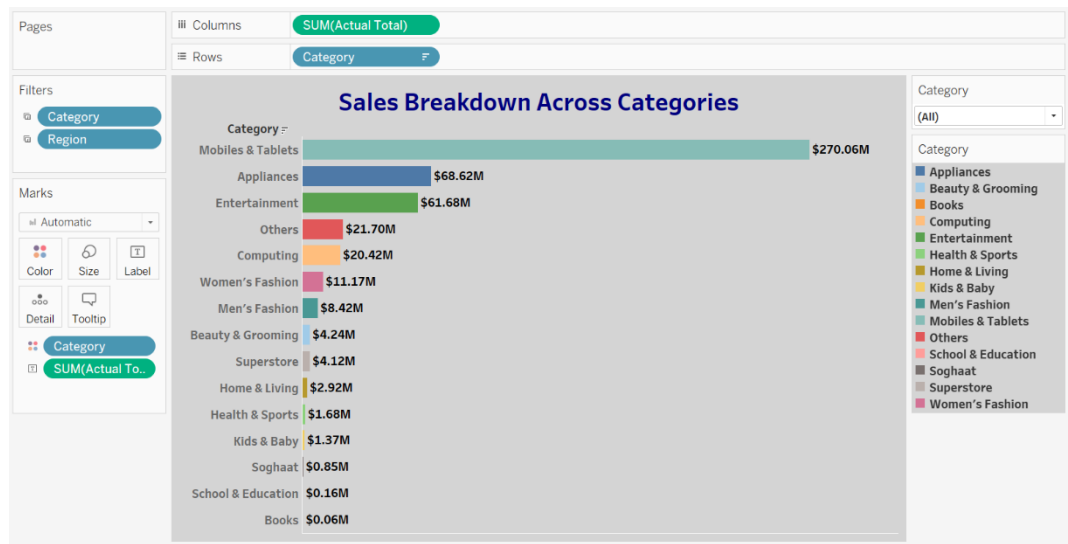
Chart 2: - Revenue Across States



The state-wise breakdown of sales offers a comprehensive view of our e-commerce website's business impact across various regions. Notably, Texas emerges as the frontrunner in terms of sales, presenting a compelling opportunity to extend our business presence beyond the digital realm. This promising scenario warrants a dual-platform approach, where we can strategically expand by establishing physical in-store outlets. To capitalize on this potential, a targeted strategy involving multiple campaigns and

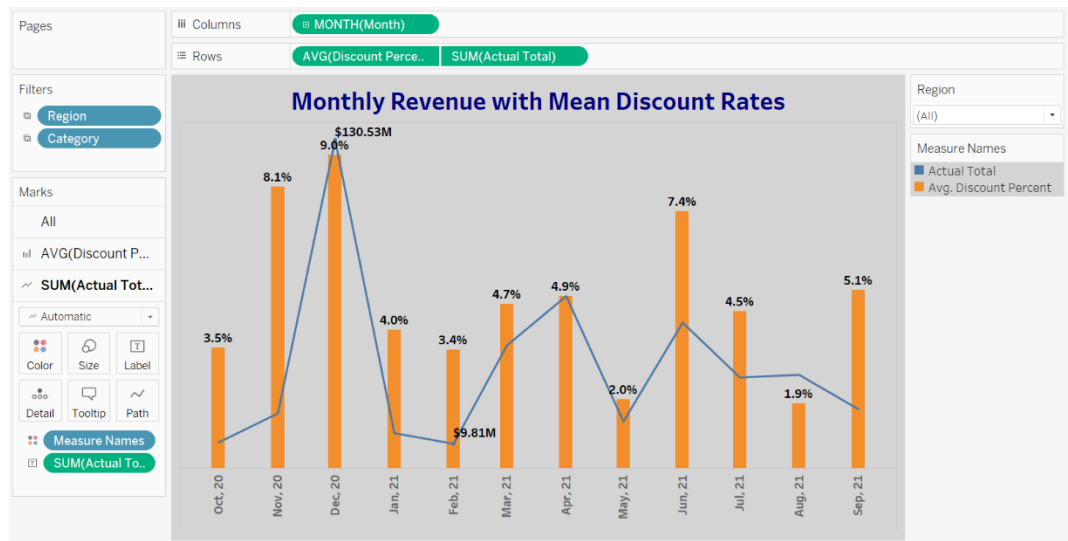
surveys can be orchestrated at a granular level. By adopting this approach, we can meticulously identify the most favorable locations to initiate a series of in-store outlets, maximizing our reach and impact.

Chart 3: - Sales Breakdown Across Categories



Delving into each category has unearthed intriguing insights. The standout revelation is the immense popularity of the "Mobiles and Tablets" category, consistently outperforming others in terms of sales. Conversely, the "Superstore" category exhibits a relatively subdued performance in the e-commerce landscape. This juxtaposition offers a strategic direction for our physical store expansion. By concentrating on the "Superstore" category, we can unlock opportunities to diversify our offerings and significantly enrich our business portfolio. This strategic move has the potential to not only broaden our market reach but also usher in substantial profits across all categories, fostering a synergistic growth trajectory.

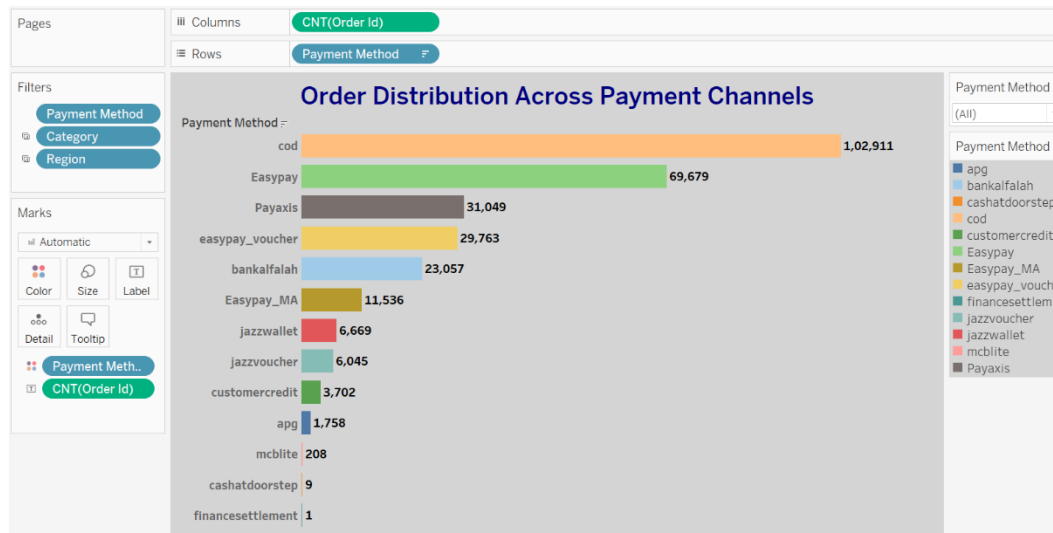
Chart 4: - Monthly Revenue with Mean Discount Rates



Presented in the chart is a depiction of the monthly average discounts alongside the corresponding total sales figures. Notably, December stands out with the highest average discount, aligning with the holiday season when sales reach their peak. Conversely, the months of July and August exhibit the lowest average discount percentages, yet still yield commendable sales performance. This juncture emerges as

a strategically opportune period to venture into the market and inaugurate a chain of physical outlets. Embarking on a soft opening during this period would offer valuable preparatory time for inventory management and pricing strategies. This foresight would position us optimally to capitalize on the upcoming sales surge, ensuring a robust and profitable market presence when the opportune moment arrives.

Chart 5: - Order Distribution Across Payment Channels



Amidst the digital transformation of payment methods, the conventional Cash on Delivery (COD) approach continues to retain its popularity. This phenomenon underscores a noteworthy reality: not all customers exclusively favor digital payments. By delving into the underlying factors driving this inclination toward COD, businesses can strategically enhance their offerings to deliver a seamless and reliable purchasing experience. This approach gracefully accommodates the preferences of individuals who lean towards more traditional payment methods, all while seamlessly integrating modern-day conveniences. Such insights instill a sense of optimism, reinforcing our potential to establish a successful chain of physical outlets. This expansion strategy presents a promising avenue for business growth, capitalizing on the harmonious coexistence of diverse payment preferences to create a thriving and inclusive marketplace.

Recommendations

We can evaluate pricing strategies after analyzing sales. Inventory management, product pricing, marketing campaigns, and customer service can all be improved by understanding market trends and sales. Analyzing the trends of which categories are more likely to be sold or recommended in more attractive ways, so that they can be stacked or recommended in a way that makes them more readily accessible to consumers.

E-commerce sales patterns can change over time, so it's crucial to regularly update our analysis and forecasts as new data becomes available. Continuously monitor the accuracy and reliability of our predictions and adjusting our strategies accordingly. Periodically reassessing our models and assumptions to ensure they remain valid and up-to-date. We can consider incorporating external data sources or market research to complement our analysis and ensuring our forecasts and recommendations reflect the current business environment.

Conclusion

Our dataset provides insight into gender-related customer behavior, discount percentages' impact on sales, and the predominance of Cash on Delivery over other payment methods. These insights help businesses develop targeted marketing strategies, optimize pricing, discounting, and payment options to cater to diverse customer preferences. Take advantage of this valuable knowledge to drive growth, enhance customer satisfaction, and secure a competitive edge. This analysis would help us recognize to get into in-store chain of outlets.

This analysis helps us understand customer behavior and a selling pattern for an e-commerce website. It gives rise to an opportunity to expand the business in realms of offline old school Markets too. More external features like sold products, external campaigns could have helped this analysis in a better way to know in-out of an e-commerce website to make high margin business.