**Semester Group Project Report: Data Visualization and Storytelling using Tableau**

***Cover page***

* **Report title: Superstore Sales(2011-2022)**
* **Group number: Group - 3**
* **Student's name whose Tableau Online account holds the visualization outcomes: Swarupa Golla , Aquil Craig Gonsalves .**
* **Group work peer evaluation**

|  |  |  |  |
| --- | --- | --- | --- |
| Considering the fairness among the group members, the participation of each member in the group collaboration is assessed in an entire group agreement base. Please keep in mind that the equal contribution is important for this team working assignment.  Complete the contribution table below by filling in each student’s name and the corresponding details. | | | |
| **First Name/ email** | **Percentage of Participation in Discussion/Development\*** | **Percentage of Contribution to the report/Tableau\*** | **Signature of Student\*\*** |
| ***swarupa.golla001@umb.edu*** | ***equal*** | ***equal*** | ***swarupa*** |
| ***a.gonsalves002@umb.edu*** | ***equal*** | ***equal*** | ***aquil*** |
| ***v.charynerella001@umb.edu*** | ***equal*** | ***equal*** | ***venu*** |
| ***m.peddi001@umb.edu*** | ***equal*** | ***equal*** | ***mihir*** |

*\*If all students contributed equally, write "equal" in each cell. If contributions vary, specify a percentage for each student (e.g., 30%). The total must sum to 100%.*

*\*\*Each student must provide their signature in the designated cell to confirm the agreement with the content of the form. Write your name for the signature.*

***Report***

1. **Superstore Sales**

This study focuses on the analysis of sales and operational of a superstore with the agenda of identifying the factors that affect profit margins, sales outcomes and increasing overall efficiency of shipping across different states in the United States. This project uses the Superstore sales dataset from 2011 to 2022 to extract meaningful insights using data visualization techniques. This research is essential as it deals with primary business issues such as sales and profit, customer satisfaction and market reach. Analyzing trends across regions will show opportunities for enhancing financial outcomes through optimizing pricing and discount strategies. In the same way, analyzing shipping duration will underline some logistical challenges that affect customer satisfaction and profitability. It will also create the potential to evaluate poor performing states regarding sales and develop focused marketing and operation plans. This visualization project aims to offer data driven insights which can put into a set of actions to boost the overall efficiency of the business focusing on sales and profit while maintaining a good customer experience to ultimately gain an edge over the competition in the retail industry.

1. **Dataset Description**
   1. **Source and content of dataset.**

The Superstore Sales dataset which is being used for this project is made by merging three datasets into a single set. The first dataset called ‘superstore dataset.csv’ includes data from 2019 to 2022 and was obtained from GitHub. The second set ‘Sample - Superstore.csv’ contained data from 2015 to 2018, and the third set ‘Superstore copy.csv’ contains records from 2011 to 2014 and both of these were acquired from Kaggle. After combining, these datasets offer a complete overview of sales, shipping details, customer information, and products for the superstore over a span of 11 years from 2011 to 2022.

* 1. **Characteristics of dataset.**

The dataset contains the following columns: Order Date, Order ID, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, Country, City, State, Postal Code, Region, Product ID, Category, Sub-Category, Product, Name, Sales, Quantity, Discount, Profit.

**Strengths:**

* The dataset contains 20 columns and data from 2011 to 2022, i.e 11 years. This large data will help to gain some meaningful insights.
* The dataset also contains columns like, daily sales and profit figures, region and state from where the purchase was made and shipped to and also information about the product and its categorization. These attributes provides many opportunities for analysis.

**Limitations:**

* Multiple datasources have been used here, which can lead to overlapping data or even inconsistencies like missing data between periods,
* The dataset does not include external factors like the socio or economic state of the region/states and also how competition can influence sales for the superstore.
  1. **Data cleaning and Integration with other datasets.**

The data was cleaned and organized using a tool called Tableau Prep Builder. This process involved taking out unnecessary columns and ensuring that all the fields had the same format. It also involved fixing any mistakes or differences between the datasets. The final combined dataset had consistent names and values for different columns. This included details like when orders were placed, how they were shipped, customer information, types of products, and financial metrics like sales and profit. This preparation made sure the dataset was ready for analysis and creating visual insights.

1. **Data Visualization and Storytelling**
2. **What factors are driving changes in profit margins across different regions?**
   * 1. **Descriptive:** What are the profit trends across regions over the years?

The trend analysis of profits over the years across different regions brings to light the fact that the performance is extremely varied. As illustrated through the heatmaps and geographical maps, regions like the West and East generate consistently higher profits compared to regions such as the South and Central. Also, state-wise analysis shows that California generates the highest profits among all states. When comparing the profitability across the states relative to California, it shows that states with high discounts relative to California have to report negative profits. Thus, states with much higher discounts, such as Texas and Illinois, demonstrate lower profitability. On the other side, states such as New York and Pennsylvania, by maintaining lower or competitive discounts, generate profitability closer to California. The above-mentioned findings therefore confirm the relevance of striking the right balance between discount and profit, offering considerable information on the regional/state-wise trends.

* + 1. **Predictive:** Can regional sales and discounts predict future profit margins?

Visualization of the data showed that regional sales and discount rates were strong factors of future profit margins. The analysis also shows a clear relationship in regions granting higher discounts are those with negative profit margins, as presented in the "Discount on Sales by Region" bar chart. For example, high average discounts on Furniture and Office Supplies in the South and Central regions are associated with low profitability. Further verbatim, this trend from a state-level analysis where, as found, excessive discounts offered in Texas and Illinois strongly predict reduced profitability and the states with balanced strategies on discounting, California and New York, fall with healthier profit margins in respect of the prediction. Thus, regions with a better spread in Technology sales, including the West, are thus observed and do come out with strong predicted margins due to the stronger intrinsic values such sale offers. These patterns point out that sales volume and discount optimization are the main levers that help predict and increase profit margins across regions.

* + 1. **Prescriptive:** What strategies should be implemented to improve profits in low-performing regions?

The insights allow the understand few prescriptive strategies for addressing the challenges in low-performing regions:

**Optimizing Discounts:** Regions like the South and Central should have more restrictive discount policies, especially for product categories such as Furniture and Office Supplies. Limiting discounts to below 15% can remarkably improve profit margins.

**State-Specific Adjustments:** For states like Texas and Illinois, reducing excessive discounts and aligning them with high-performing states like California and New York can drive profitability.

**Customer Segmentation and Targeting:** Enhancing marketing efforts for high-value customer segments, such as corporate clients in the South and Central regions, can boost sales and profits.

**Localized Strategies:** Implementing region-specific pricing and promotional campaigns to align with customer preferences and market conditions can further improve financial performance.

1. **How do profit and shipping times vary across different states, and how can we improve them?**
   * 1. **Descriptive:** What are the average shipping times for each shipping mode in different states?

Shipping time in various states, based on different shipping modes, indicates several trends. As illustrated, Standard Class is the longest in terms of average shipping time with approximately 4.5 days, followed by Second Class shipping, which takes approximately 3 days and First Class that takes about 2 days. The shortest, same day, takes less than a day. Some states, such as Wyoming, Montana, and North Dakota, always have longer average shipping times at the state level compared to other states like California and New York, which have relatively lower durations. States experiencing longer shipping times are usually plagued with poor customer satisfaction and, subsequently, lower profit margins. Also, the volume of sales is seen to be maximum with the Standard Class shipping mode, suggesting that this mode is most in demand by customers. However, as further analysis reveals, this does not necessarily translate to higher profitability.

* + 1. **Predictive:** What patterns and correlations can be identified between shipping modes, shipping times, and sales?

Predictive analytics show that there is a strong relationship between the mode of shipment, time of shipment, and how these affect sales and profitability. A scatter plot of sales vs. profit by shipping mode indicates that the faster shipping modes, such as First Class, have higher profitability despite the higher costs associated with them. Standard Class contributes more to the overall sales volume but yields lower profits because of its longer shipping times and cost constraints. Interestingly, it is observed from the scatter plot analysis that the superstore achieves the highest profit correlation with First Class shipping, evidenced by a higher R-value compared to Standard Class. This would mean, while Standard Class is leading the majority of sales volume, First Class provides a bigger margin. The states in which the Standard Class delivery time is very long have dissatisfied customers due to its long duration and, therefore, has lower repeat sales. Even more important is that shipments taking more than the threshold of 5 days also yield diminishing returns to sales, making shipment speed an essential consideration. These patterns show that there is a need to balance speed and cost-effectiveness in shipping strategies.

* + 1. **Prescriptive:** What logistics improvements can be made to optimize shipping times and increase profits?

Based on the insights derived from the descriptive and predictive analysis, the following logistics improvements are recommended to optimize shipping times and enhance profitability:

**Invest in Faster Shipping Modes:** Faster modes of shipment, such as Same Day and First Class should be used for high-value customers and products first in states with high sales potential, like California and Texas. First Class shipping is justified for select segments, where increased profitability can be derived from its implementation.

**Optimize Regional Warehousing:** Strategic placement of warehouses closer to demand-heavy regions such as New York and California can reduce shipping durations for Standard and Second-Class modes.

**Route Optimization:** Advanced routing algorithms decrease transit times of Standard Class shipments in geographical states, such as Wyoming and Montana, which is possible with implementation.

**Segmented Shipping Strategies:** High-profit states can have a customer-segmented approach, where corporate customers have faster options of shipment, while Consumer segments continue to have cost-efficient modes

**Monitoring and Analytics:** Continuous monitoring of shipping durations by state and mode, coupled with real-time adjustments to logistics processes, can further streamline operations and enhance customer satisfaction.

1. **How can we improve sales performance in states with less sales, considering factors like shipping duration and customer segments?**

#### Descriptive: What states are showing less sales?

#### For the descriptive analysis as visualized in the map of sales by region, highlights a number of states that have very low sales, including Wyoming, North Dakota, South Dakota, Maine, West Virginia and the District of Columbia. These states are consistently below the average in sales compared with the high-performing states, such as California or New York. Segment-wise, the contribution of Sales by segments in these worst-performing states is from the Consumer segment, little from either Corporate or Home Office segments. That means poor utilization of high-value customer groups might be one more reason for the reduced overall sales performance in those regions. Moreover, these states also show prolonged shipping durations compared with others, which may affect customer satisfaction and retention.

#### Predictive: Can improving shipping durations boost future sales performance?

#### The analysis of the scatter plot showed that sales in these low performing states are not closely related to shipping duration. Therefore, this factor can be ruled out as the main reason for poor sales in these low performing states. Instead, further investigation into the customer segments showed that in the high-performing states like California, sales are distributed relatively evenly across Consumer, Corporate, and Home Office segments. That distribution of segments is balanced; whereas in low-sales states, it is dominated by the Consumer segment. This would suggest that attention to customer segmentation strategies will more likely improve sales in those regions shipping speeds.

#### Prescriptive: What strategies can boost sales in these states?

#### Recommendations based on the challenges in low-sale states:

#### Focus on customer segmentation: Improve marketing and engagement with Corporate and Home Office segments in Wyoming, Maine, and North Dakota. This would better position their segment distribution closer to highly performing states, such as California, and improve overall sales.

#### Smarter Campaigns: Create campaigns to target these segments-for example, bulk discounts or special solution development of the product. Also, advertise a clear value proposition to Corporate and Home Office users through selective advertising, email campaigns, and local events.

#### Localized Strategies: Partnerships with local businesses for special corporate discounts, among others, in targeted campaigns that may raise the interest of businesses and organizations. Price policies and strategies should be adjusted to make the offerings appealing to under-representations.

#### Product Focus: The strong product focus on high margin product categories such as Technology and Office Supplies within Corporate and Home Office Segments will generate sales along with profitability.

#### Customer Loyalty and Retention: Establish the loyalty programs, offers, and events of Corporate and Home Office to retain repetitive customers for long-term relationships.

1. **Lessons learned and Implications**

The implications are practical and relevant to develop improved business strategies, operational efficiencies, and customer satisfaction. It highlights the adverse effects excessive discounts have in areas like the South and Central in particular, for categories such as Furniture. This may suggest that the company might want to fine-tune its pricing and discount strategies so as to improve profitability. Moreover, the strong positive relation of high margins with fast modes of shipping suggests that investment in logistic infrastructure in high-potential regions will pay off. The focus for low-performing states should be targeting underutilized segments-like Corporate and Home Office-with localized marketing campaigns. In this way, by emulating the trends that occur within high-performing states like California, continuous growth can be achieved with a company having better decision-making.

Visualization is the process wherein very important features and actionable insights have been shown. Enabling tools such as Tableau transformed big, cumbersome data into intuitively crafted visuals that actually helped them find out the pattern and trends quite effectively. Heatmaps and scatter plots depict the relationship between discounts with the profit margin to recommend optimized price strategies. The maps and bar charts identified low-selling states with their respective problems, such as underrepresentation of Corporate and Home Office segments, on which targeted interventions were to be carried out. Visualization thus brought out the requirement for disaggregation at a segment level; namely, the balanced customer distribution in California became a very critical factor for its top performance. Iterative refinement-for instance, eliminating shipping duration as a key driver of sales through scatter plot analysis-in turn ensured that conclusions so drawn would be evidence-based.

This experience drove us to realize that storytelling is an important virtue in data-driven decision-making as it enables us clearly and effectively to present the findings to the stakeholders concerned. It enhanced our analytics and visualization skills, preparing us with confidence to address future challenges in the best possible ways.

**Appendixes:**

**1. Changes in profit margins across different regions**

**A map of the united states

Description automatically generated**

**2. Does shipping times variation across different states affects profits, and how can we improve them?**

**A screen shot of a graph

Description automatically generated**

**3. Sales performance in states with less sales, considering factors like shipping duration and customer segments:**A screenshot of a computer

Description automatically generated