Strategic Predictions in E-Commerce: Leveraging Data to Drive Sales and Enhance Customer Engagement
Group 13

1. Research Question

a) Novel Question(s):

- How does the sales for products change with varying price levels, indicating price elasticity in the e-commerce sector?
- What is the influence of delivery times on customer satisfaction scores?
- Analysing sales distribution and product popularity by geographical regions.
- Comparative performance analysis of top sellers within the marketplace.
- Identifying loyal customer segments to tailor promotional strategies effectively.

b) Potential Audience:

The target audience for this analysis comprises e-commerce strategists, pricing analysts, operations managers, and marketing executives who are looking to optimize sales strategies, customer satisfaction, and operational efficiency.

c) Interest Justification:

- Sales managers will be interested in price elasticity to optimise pricing for revenue maximisation.
- Logistics managers will find the correlation between delivery times and review scores crucial for operational efficiency.
- Marketing teams can utilise geolocation sales data to focus campaigns in high-demand areas.
- Competitive insights from top seller performance can drive strategic decisions to enhance market positioning.

2. Data

a) Data Source: Brazilian E-Commerce Public Dataset by Olist (Link).

b) Data Period: Period of the dataset is from 2016 to 2018.

c) Level of Observation: The datasets primarily support a customer-order level of observation, with the ability to analyse daily trends (e.g., customer-order-day). Additionally, product and geographic dimensions enable analyses such as product-month and city-quarter.

3. Model

a) Outcome of Interest (Y variable): The main outcomes are sales_volume to determine price elasticity, review_score for customer satisfaction, and customer_retention as an indicator of loyalty.

b) Covariates/Predictors (X variables): Covariates/Predictors (X variables): The analysis will factor in `product_price`, `delivery_time`, and `customer demographics` along with `purchase history`, `product_category`, and `time of year` to account for seasonality impacts on sales.

c) Statistical Model(s): Employing multiple regression analysis for assessing the impact of price on sales volume, and time-to-delivery on customer satisfaction scores. For customer segmentation, machine learning clustering techniques such as K-means could be used to group customers by purchasing behavior and demographics to identify loyal segments.

4. Results/Conclusions

a) Results and Conclusions: The findings are expected to reveal the degree of price elasticity across product categories, the impact of delivery times on customer satisfaction, regional sales trends, and the performance metrics of top sellers. These insights will help in understanding the factors influencing customer loyalty.

b) Expected Conclusions: It is anticipated that shorter delivery times could enhance customer satisfaction as evidenced by higher review scores, although additional variables may also play a role. A reduction in prices is expected to correlate with an uptick in sales volumes. Sellers with a more diverse product portfolio are likely to experience greater sales. Those located in densely populated areas might see increased transactions due to greater market accessibility. Furthermore, product categories offering a broader selection are presumed to enjoy greater popularity, as reflected in sales data.