

Objective

The objective of this study is to perform customer segmentation and identify the top three products based on transactional and behavioral data.

Dataset Description

The dataset used for this study consists of two distinct sets of data: transaction data and behavior data. The transaction data contains over 260,000 rows and 8 columns, capturing essential details related to customer purchases, such as transaction ID, product name, quantity, price, and timestamps. On the other hand, the behavior data comprises more than 70,000 rows and 3 columns, providing insights into customer interactions, preferences, and engagement metrics.

To prepare the data for analysis, preprocessing steps were applied, including converting integer date formats into standard date formats. Additionally, feature engineering was performed by introducing a new column called 'package weight' and extracting brand names from the product names. These enhancements improved the dataset's usability and facilitated a more in-depth analysis of customer segmentation and product popularity.

Key Findings

During the analysis, several key findings were observed:

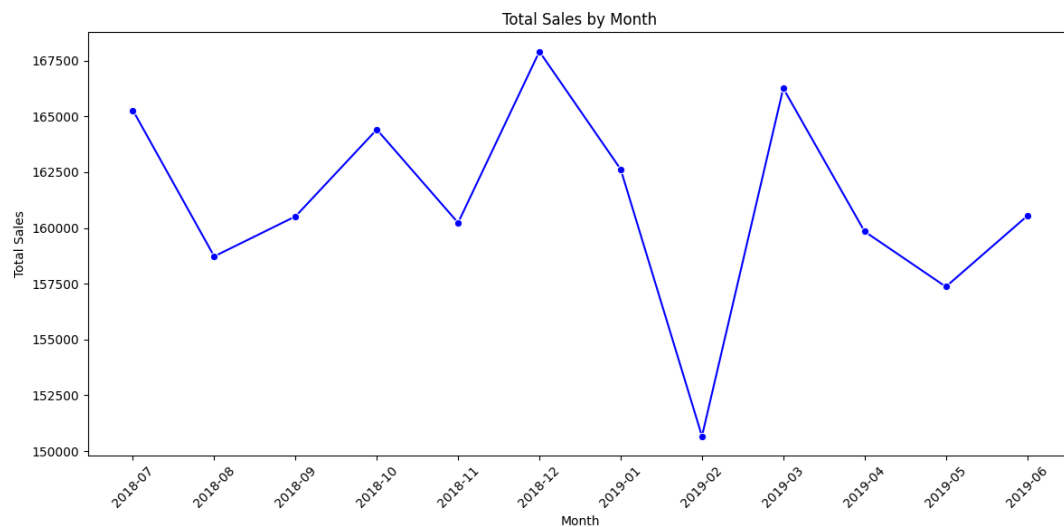
1. One date was missing from the dataset, specifically December 25, likely due to the Christmas holiday.

```
1 missing_dates = full_dates.difference(transaction_data['DATE'])
2
3 missing_dates
✓ 0.0s

DatetimeIndex(['2018-12-25'], dtype='datetime64[ns]', freq='D')
```

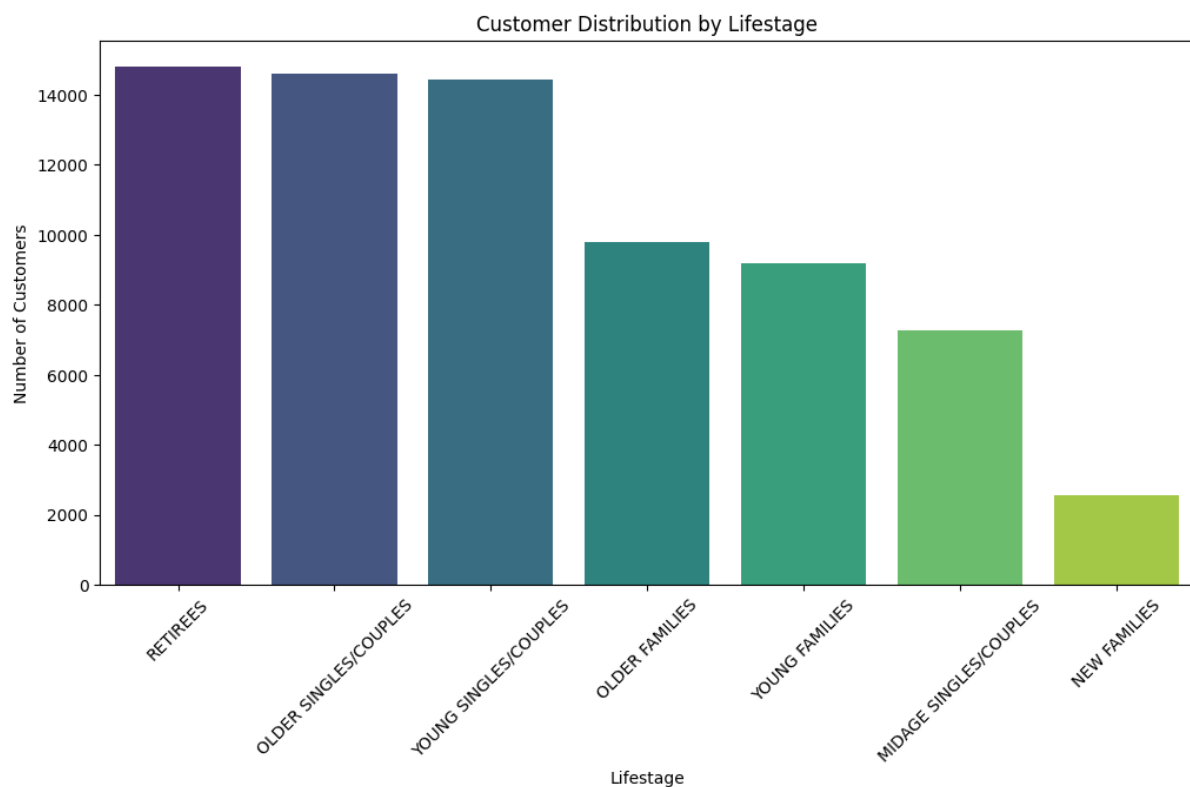
2. An outlier was identified in total sales where one transaction recorded a total sales value of 650, whereas the second highest sales value was only 29.5. Additionally, the same loyalty card number, store number, and product number appeared in this transaction, indicating an anomaly. As a result, these two items were dropped from the dataset.

3. Exploratory Data Analysis (EDA) revealed that sales tend to increase significantly during the Christmas season and drop after the Christmas and New Year sales period.

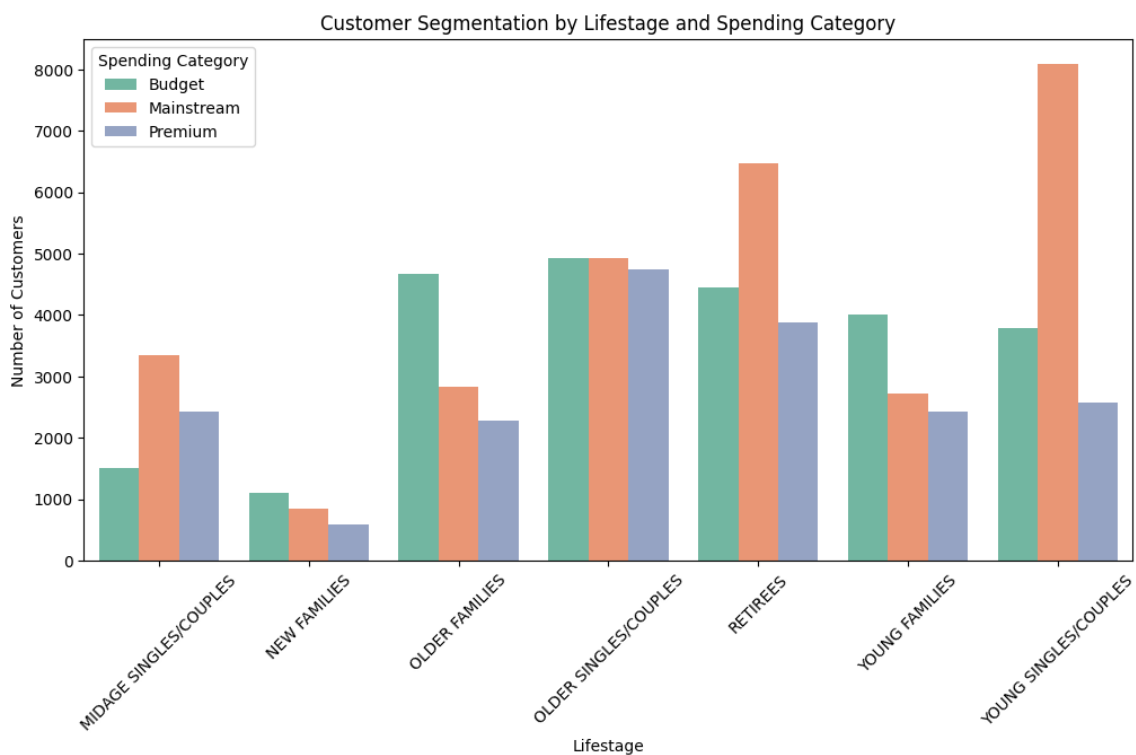
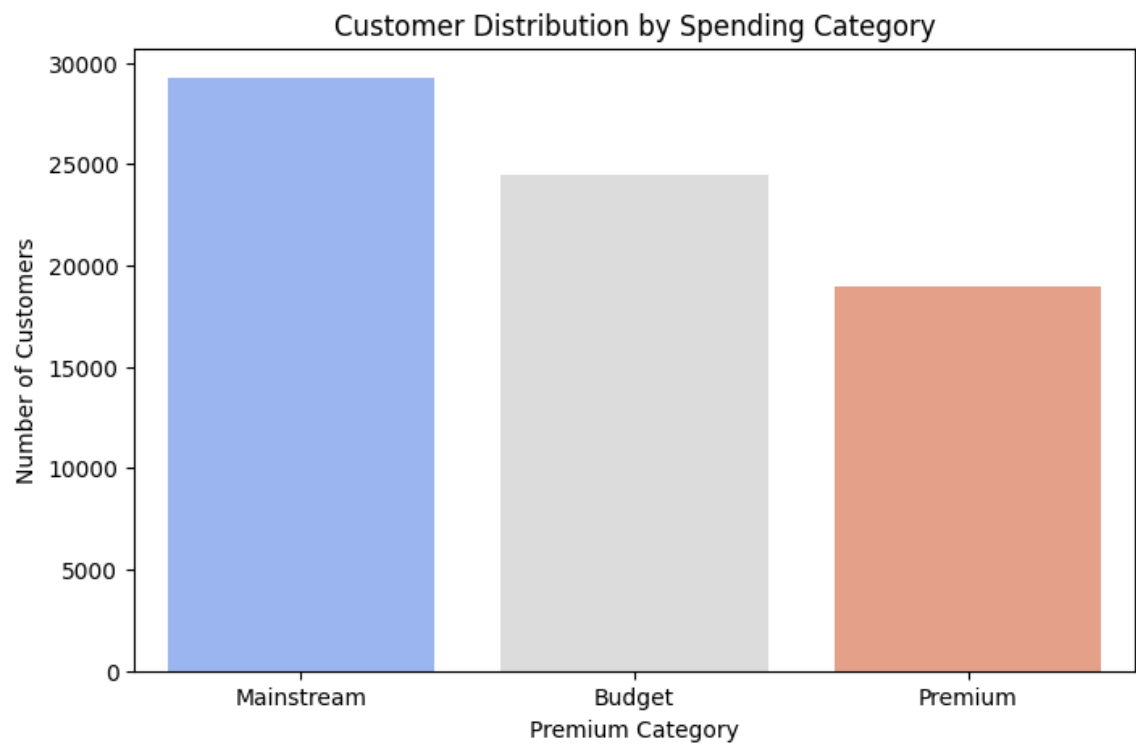


Customer Behavior Analysis

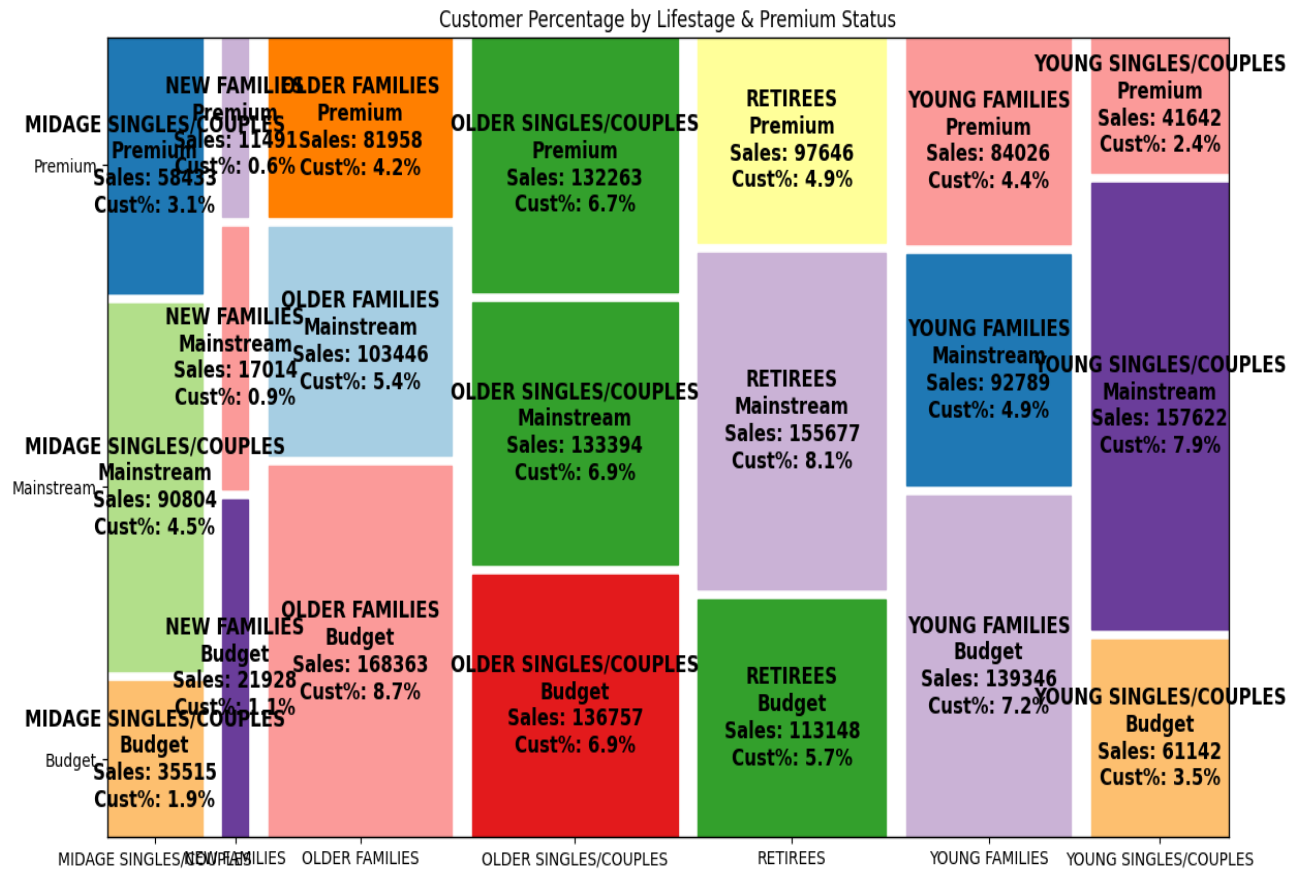
Exploratory Data Analysis was conducted to analyze customer behavior based on life stages. It was observed that retirees form the largest customer group, followed by older singles/couples and younger singles/couples. New family members represent the smallest segment, possibly due to financial constraints and other expenses.



Another analysis focused on customer distribution by spending category, revealing that majority customers are from mainstream, followed by budget and premium customers.



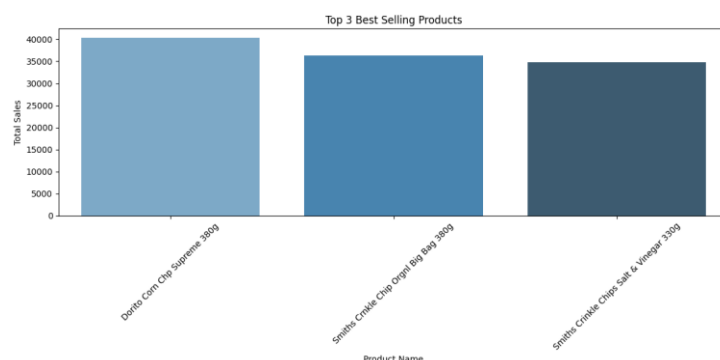
Mosaic plot is done to understand the classification of customer percentage by lifestage and premium status. It is clear that older families prefer budget items, where older single/couple prefer premium items. Also all the lifestages are buying mainstream items as well. plot is given below



Top Products and Brands

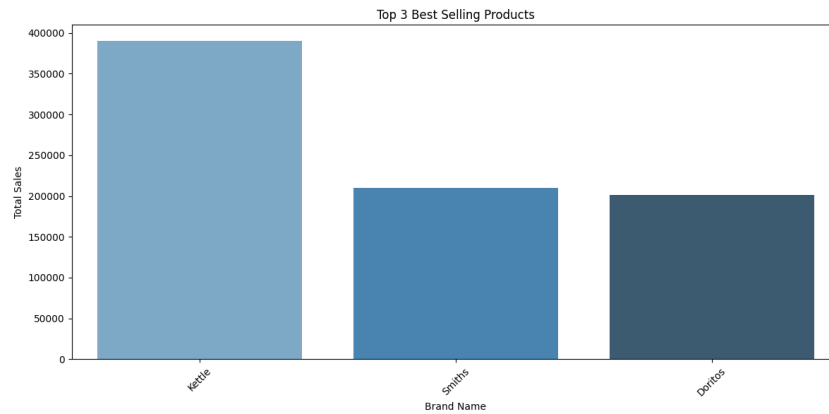
The top three products with the highest total sales are:

1. Dorito Corn Chip Supreme 380g - 40,348.75
2. Smiths Crinkle Chip Original Big Bag 380g - 36,367.60
3. Smiths Crinkle Chips Salt & Vinegar 330g - 34,804.20

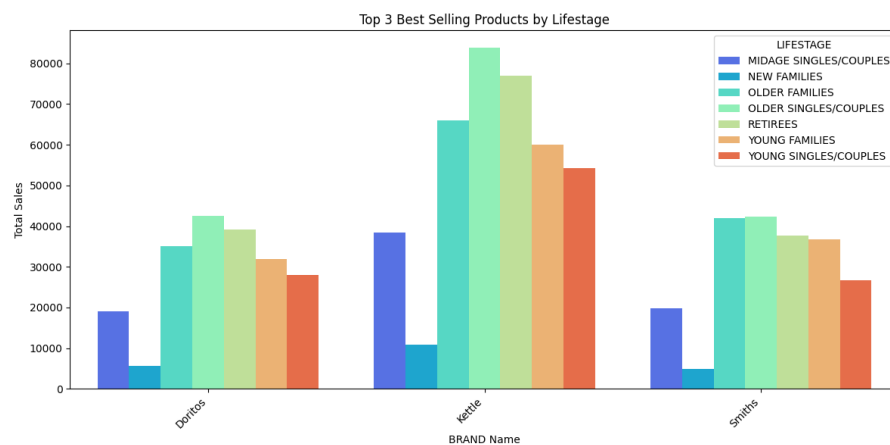
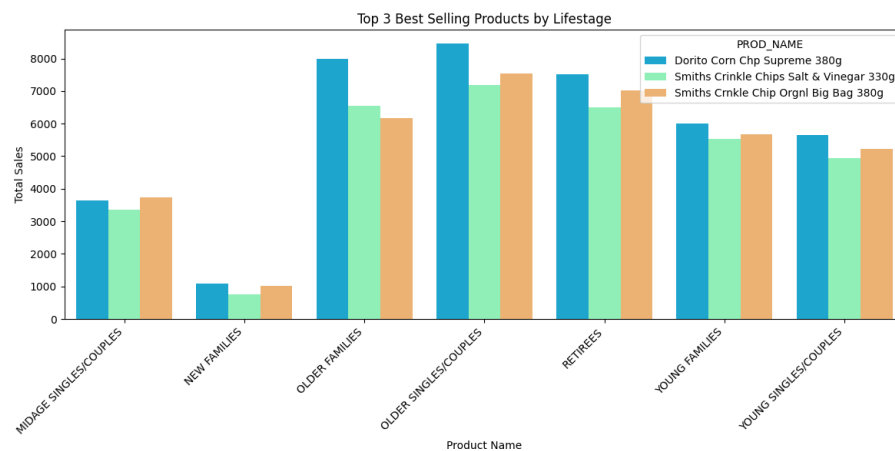


Additionally, the top three brands with the highest total sales are:

1. Kettle - 390,230.6
2. Smiths - 210,076.8
3. Doritos - 201,538.9



Apart from this, I have also find top 3 selling products by life stage and top 3 selling brands by life stage.



RFM Segmentation

To further understand customer behavior, RFM (Recency, Frequency, Monetary) analysis was conducted. This method helps in categorizing customers into different segments based on their purchase behavior. The segmentation results are as follows:

- **Moderate Customers:** 42,958
- **Best Customers:** 9,566
- **Lost Customers:** 8,537
- **Loyal Customers:** 5,520
- **Promising Customers:** 4,317
- **New Customers:** 1,739

These segments provide valuable insights into customer engagement levels, allowing for targeted marketing strategies to enhance customer retention and satisfaction.

Conclusion

The findings of this study provide critical insights into customer behavior, product preferences, and purchasing patterns. The segmentation based on RFM analysis allows businesses to target different customer groups effectively, ensuring personalized marketing strategies. The dominance of older families and retirees in premium spending highlights the importance of targeting these groups with high-value products. Additionally, mainstream and budget segments hold significant market share, necessitating tailored promotions and pricing strategies. The seasonal influence on sales further reinforces the need for strategic marketing during peak periods. Overall, the analysis underscores the importance of data-driven decision-making in enhancing customer experience and business profitability. However, due to the absence of parameters beyond total sales, profit calculations were not possible in this study, which limits a deeper financial performance evaluation.