Supermarket Sales & Customer Insights

Branch Performance, Shopper Behavior, and Operational Recommendations

by James Hamer
May 7, 2025

Project Overview

Objective: Analyze supermarket sales data from three branches to uncover customer behavior trends and opportunities for operational improvement.

Dataset Summary: Approximately 1000 transactions from 3 branches of an anonymous supermarket over a 3 month period, sourced from kaggle.com.

Python was used to load the dataset and clean the data by replacing null values with the mean or mode as appropriate.

Executive Summary

Overall sales volume is relatively consistent across the three branches, though Branch B generates the highest revenue per transaction.

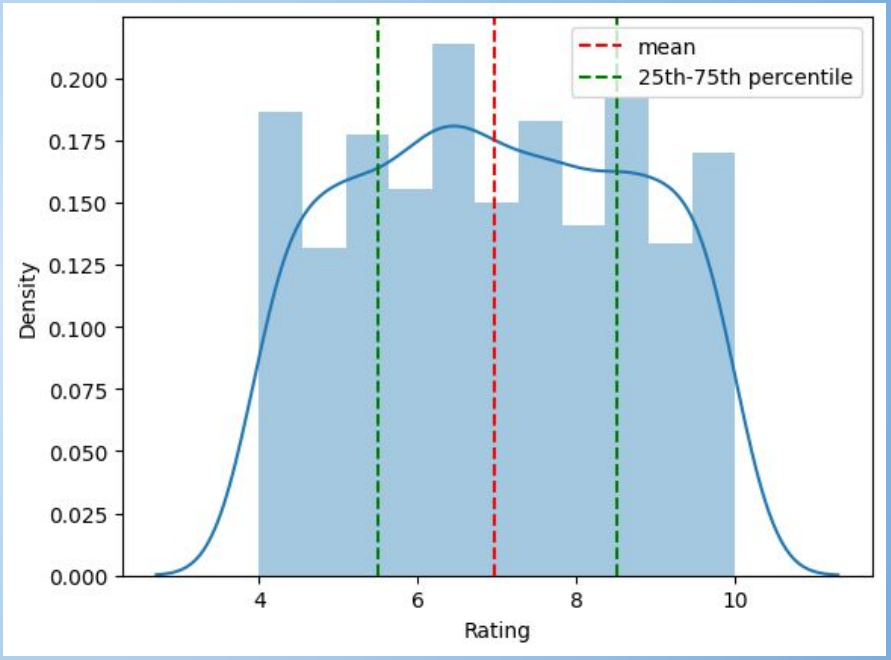
While customer satisfaction ratings are similar across branches, the lower median rating for Branch B warrants further investigation into the underlying causes.

A significant spike in sales at 7PM suggests more staffing could be needed to better assist with the increase in sales. Holidays, such as Valentine's Day, may also require more staffing, but further analysis is needed.

Members and non-members exhibit similar overall spending. Targeted promotions for members, potentially leveraging their preference for credit card or e-wallet payments, presents an opportunity to enhance sales and loyalty.

Customer Satisfaction Ratings

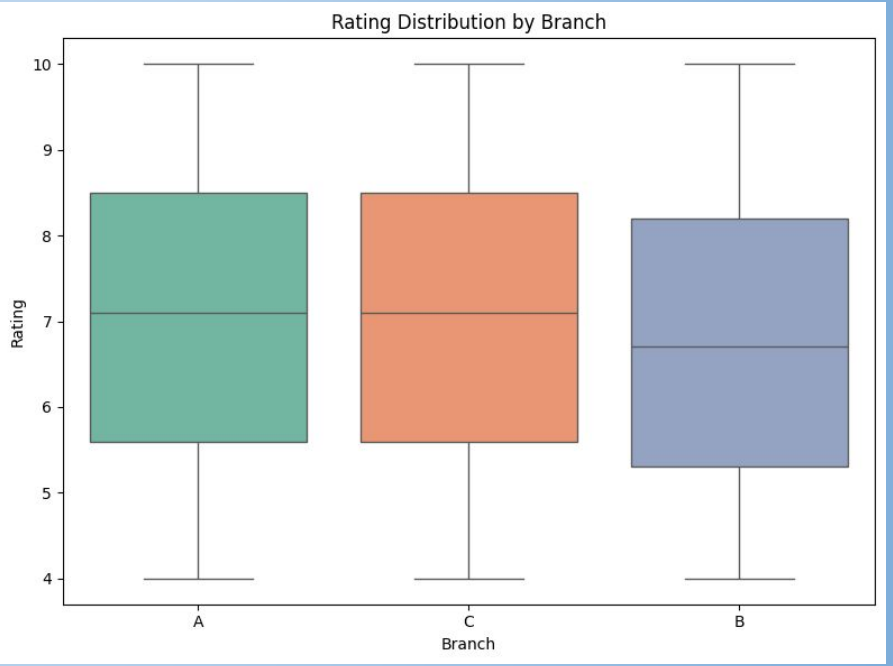
Seemingly no skew in the numbers, customers are just as likely to provide a low satisfaction rating as they are a high rating.



Customer Satisfaction by Branch

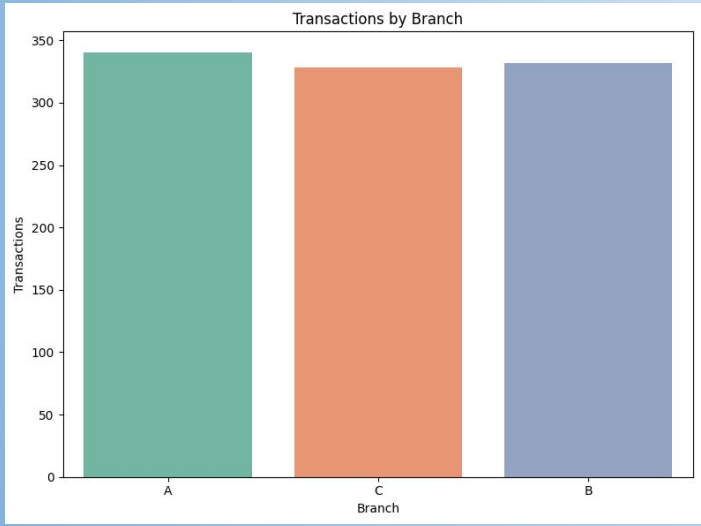
While the three branches share similar ratings, there is a noticeable difference with Branch B.

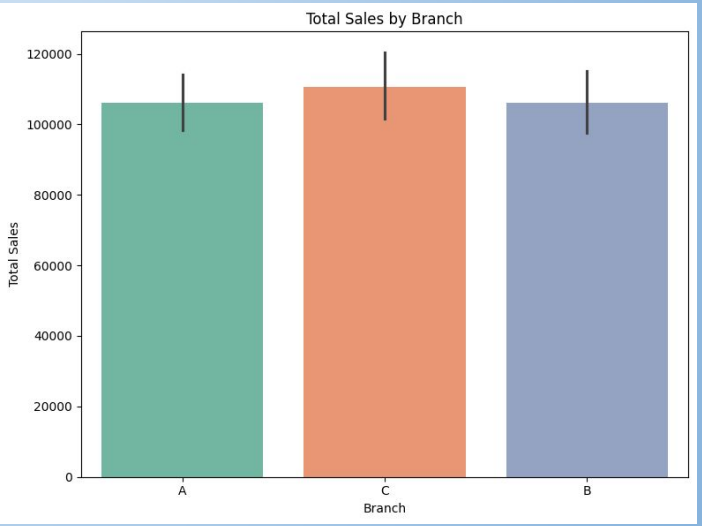
Branch B has a lower median rating than the other two branches. A deeper look into the ratings could help improve customer satisfaction.



Individual Branch Sales

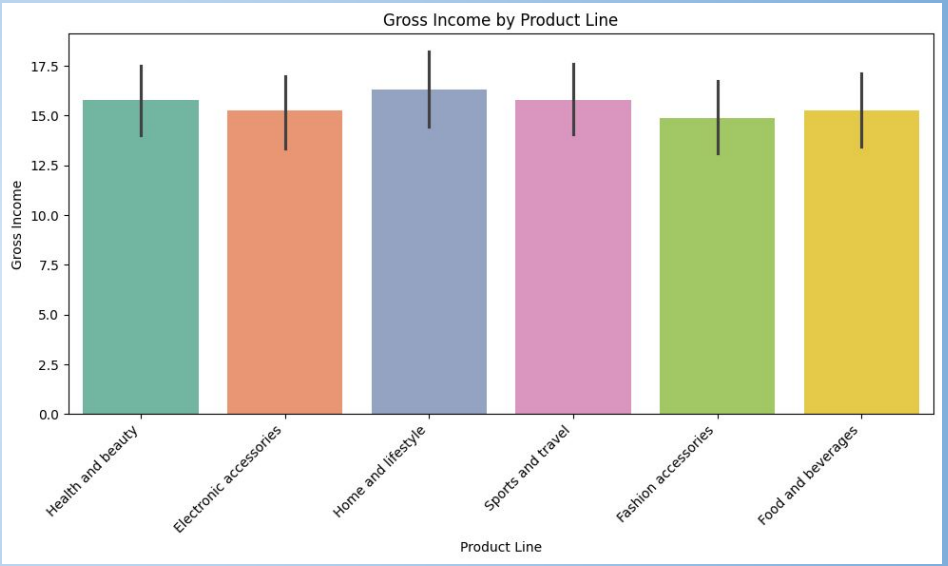
Both transactions by branch and sales by branch seem relatively even. All branches are performing about the same when it comes to sales.





Product Line Sales

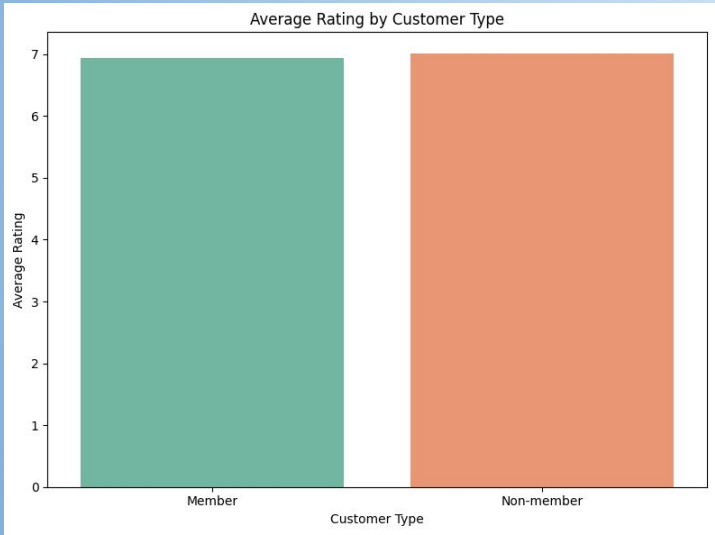
There is relative consistency in the gross income average of each of the six product lines. However the consistency of the gross income varies across product lines.

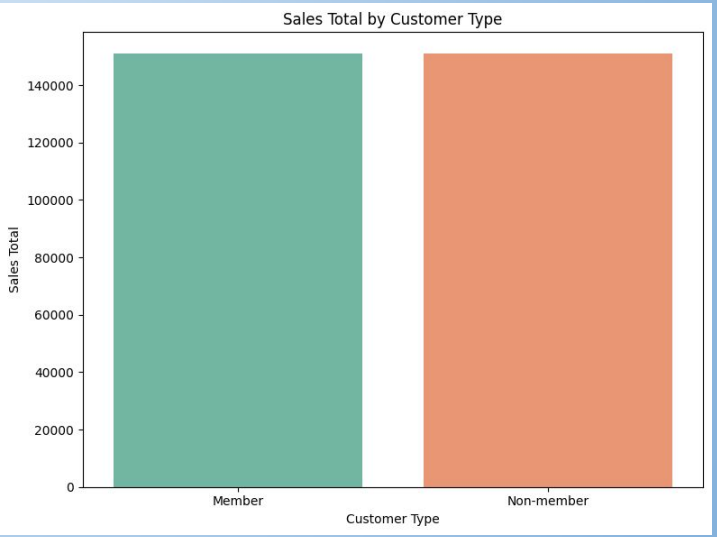


Home & Lifestyle shows the widest range of gross income. Investigation into the drivers of high and low performance could help optimize offerings.

Member vs Non-Member Behavior

Members and non-members provide similar ratings. Member status does not appear to significantly affect total sales.

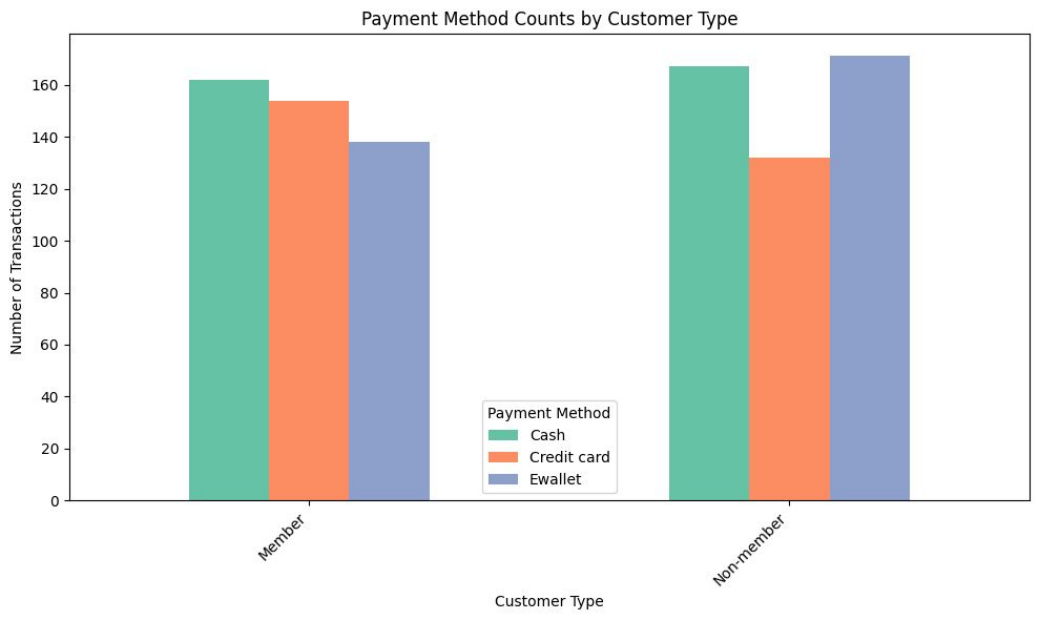




Member vs Non-Member Behavior

We can see both members and non-members prefer cash payments.

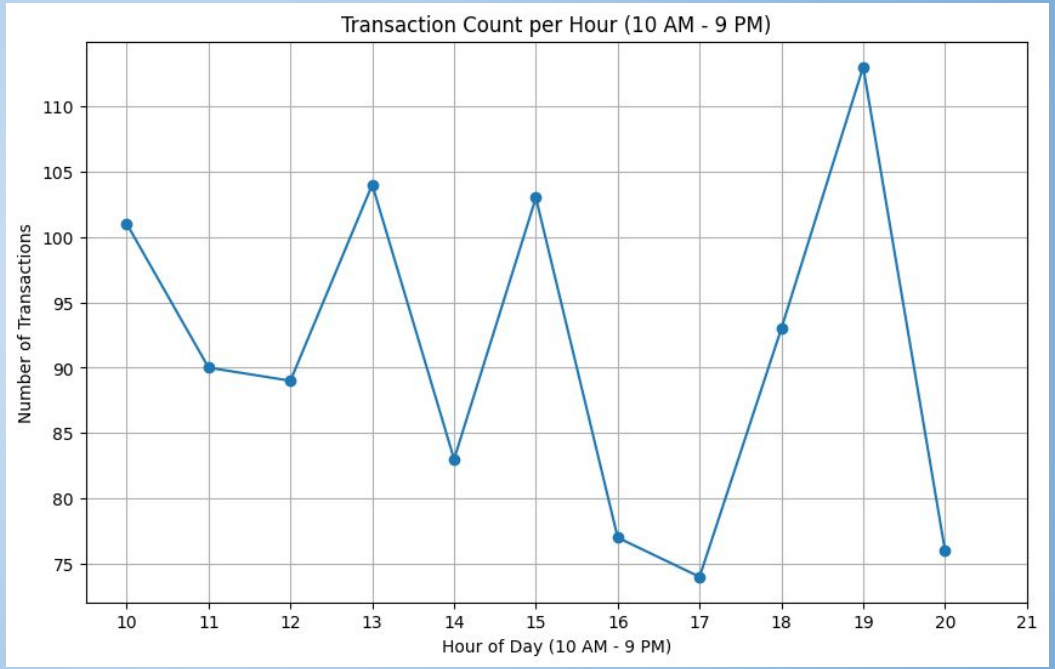
Non-members are more likely to pay be E-Wallet than members, while members are more likely to pay by credit than non-members.



Hourly Sales Trend

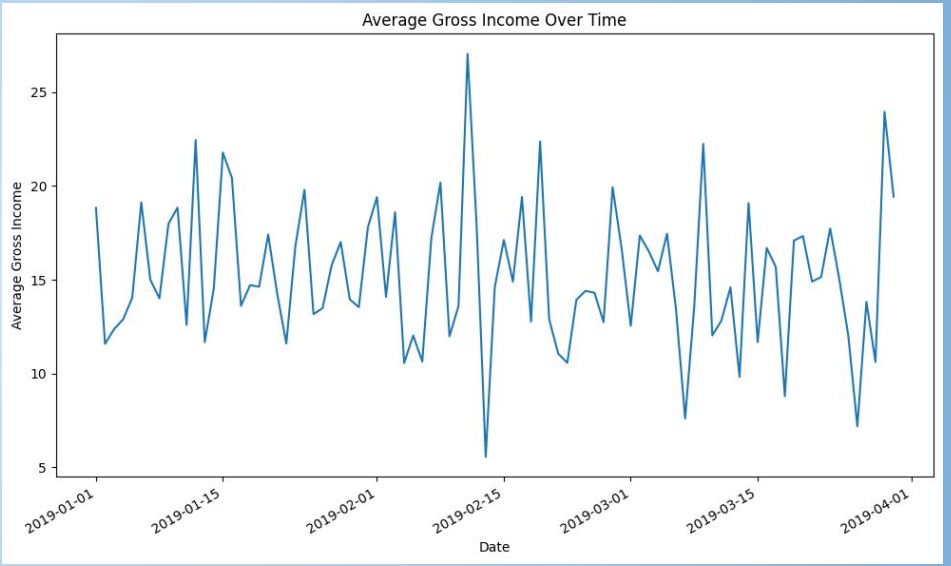
There is a noticeable spike in hourly sales at 7PM. This is likely caused by shoppers coming after leaving their jobs.

It would be beneficial to ensure the supermarkets are adequately staffed for the increase in foot traffic at that time.



Gross Income Over Time

Though sales spiked around the middle of February, the range of sales is relatively consistent over the three month period in the sample.



Could the February spike be from Valentine's Day sales? More data will be needed to see if this or other holidays can create a predictable spike of sales.

Conclusions

Branch B generates the most revenue per transaction.

Members spend about the same as non-members but prefer e-wallet payments. Providing incentives to members, such as special promotions tailored to individual members, could help increase sales.

There's a significant increase in sales around 7PM. Stores should be staffed accordingly to handle the extra sales.

Data supports both short-term staffing changes and long-term promotional strategy.

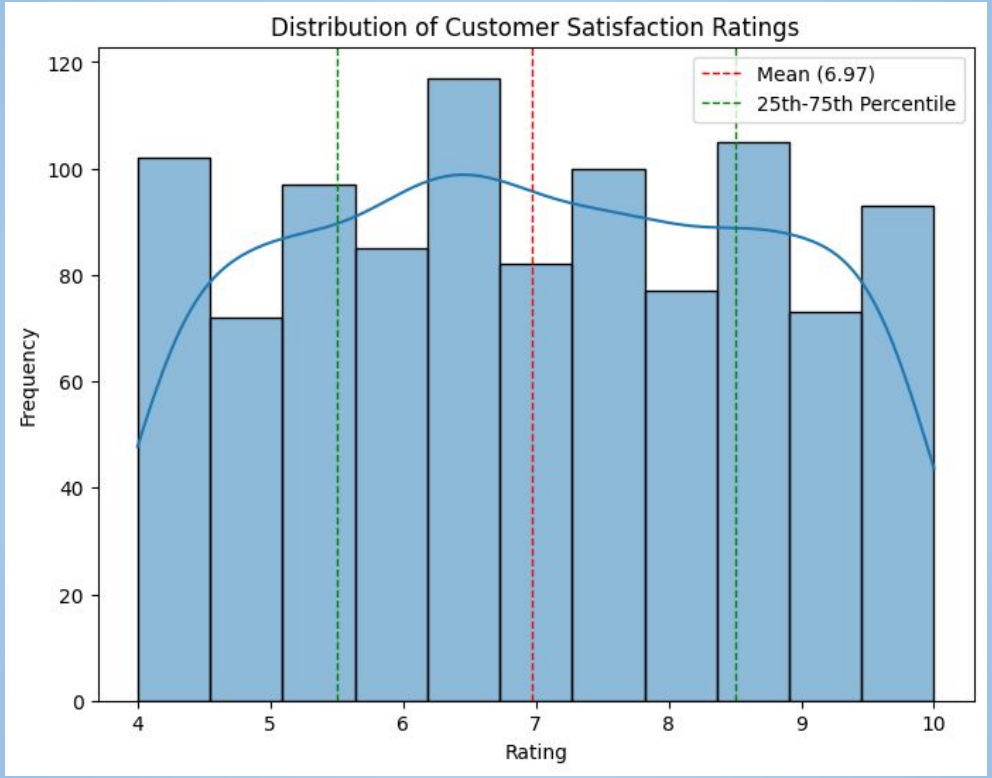
Appendix

This appendix provides basic information on the original data source and supplementary data visualizations to support the analysis in the main report.

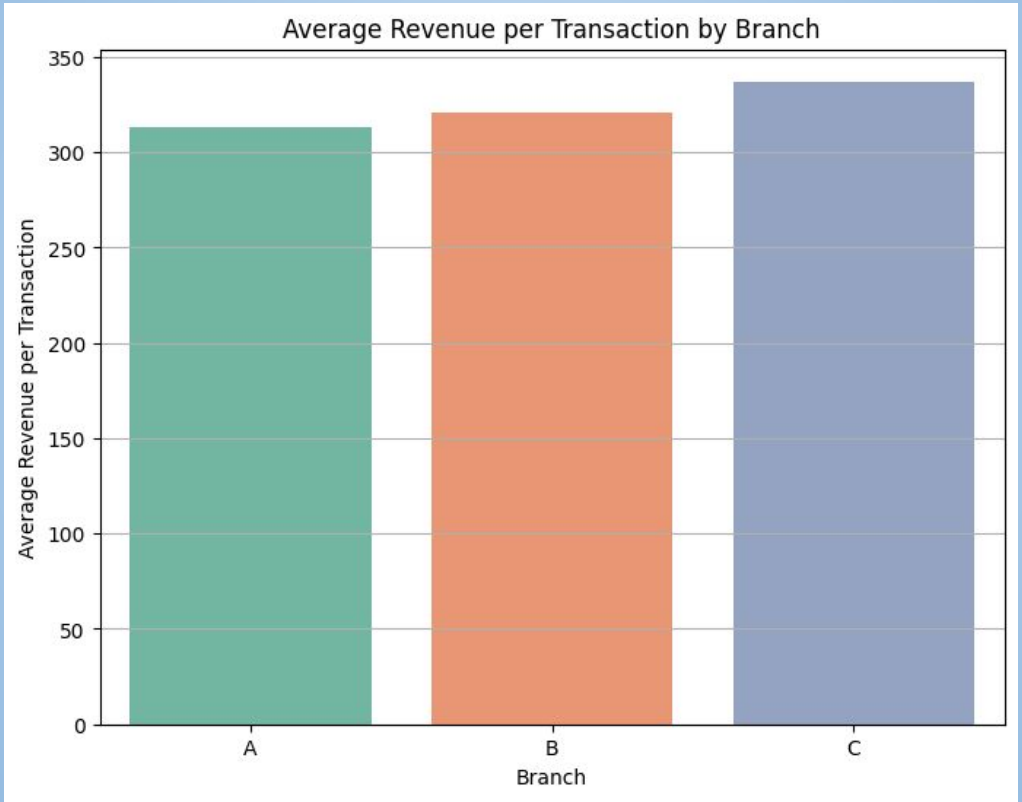
Data source: https://www.kaggle.com/aungpyaeap/supermarket-sales

Basic suggested analysis from Coursera project titled “EDA With Python and Pandas”

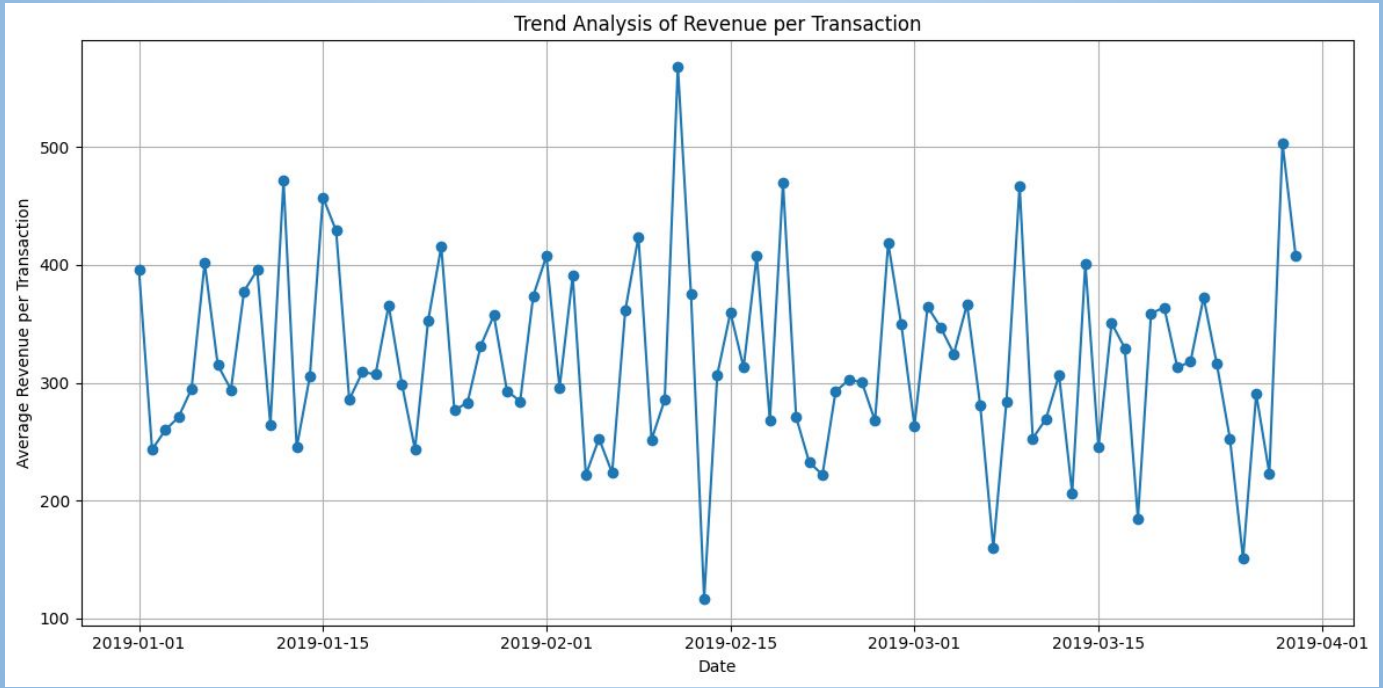
Appendix


Distribution of Customer Satisfaction Ratings
Mean (6.97)
25th-75th Percentile
Frequency
Rating
0
20
40
60
80
100
120
4
5
6
7
8
9
10

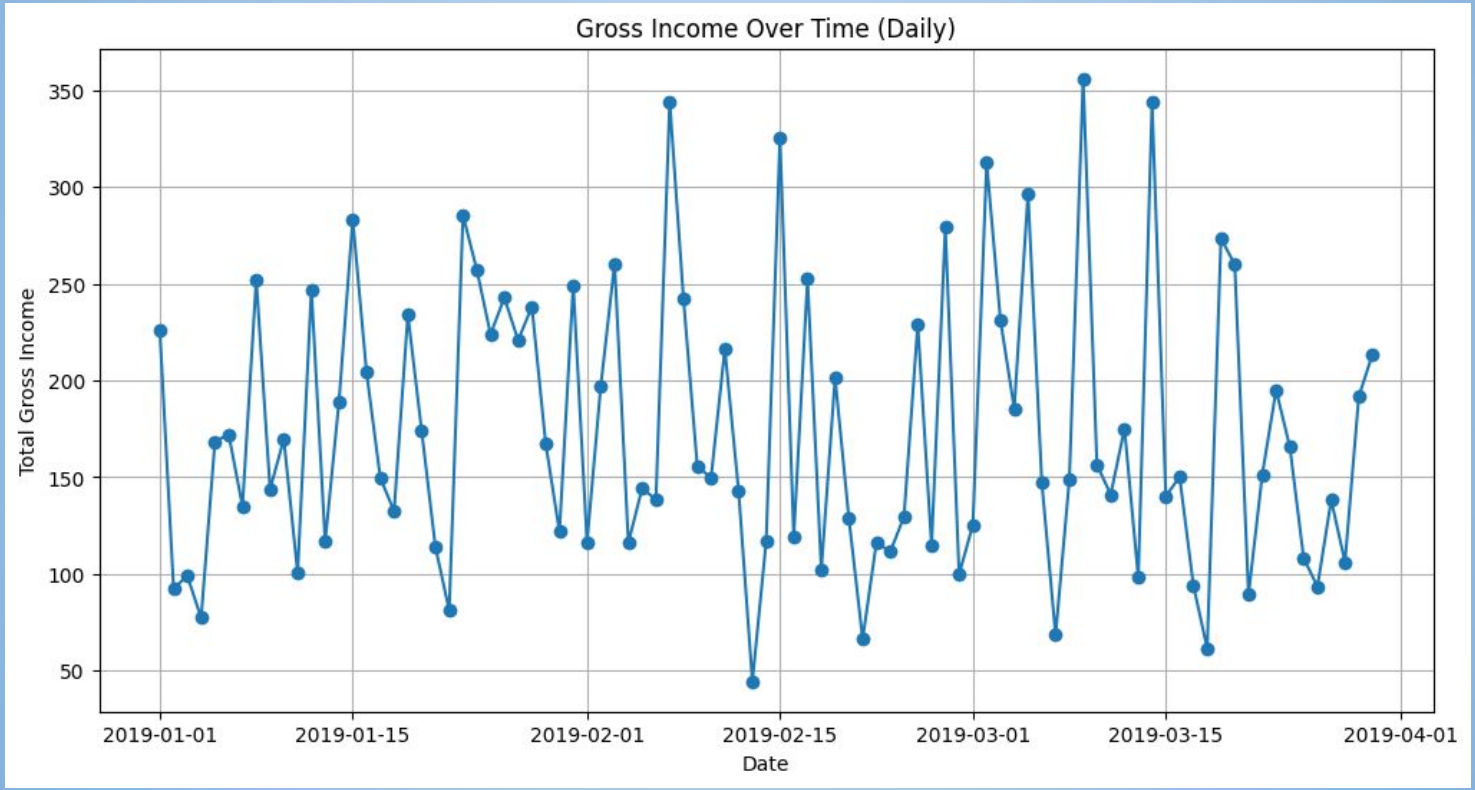
Appendix


Average Revenue per Transaction by Branch
Average Revenue per Transaction
350
300
250
200
150
100
50
0
A
B
C
Branch

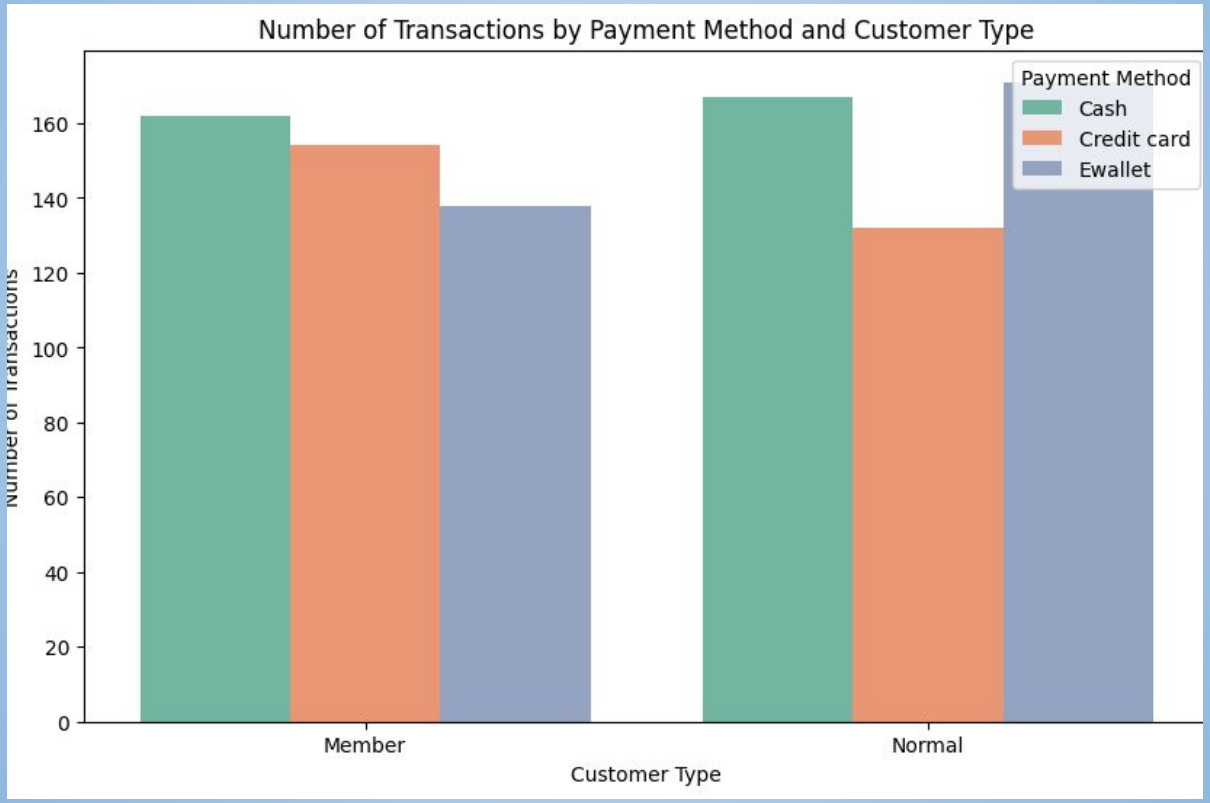
Appendix


Trend Analysis of Revenue per Transaction
Average Revenue per Transaction
100
200
300
400
500
2019-01-01
2019-01-15
2019-02-01
2019-02-15
2019-03-01
2019-03-15
2019-04-01
Date

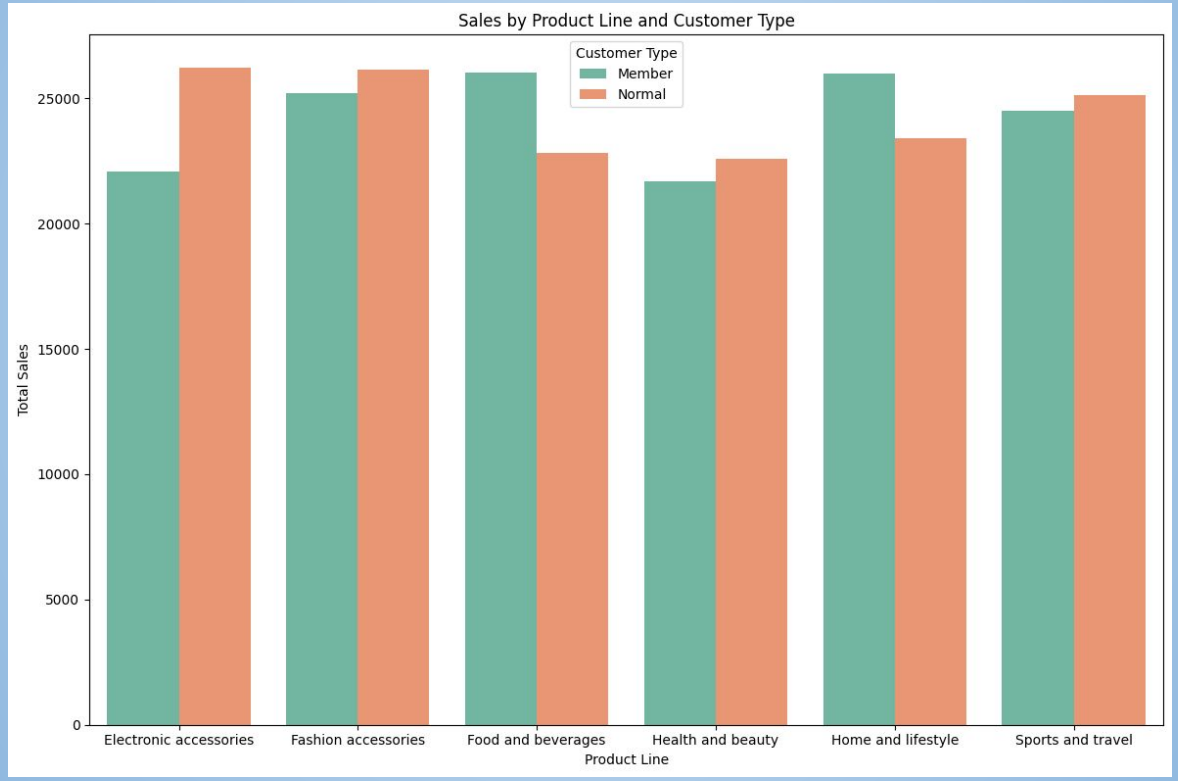
Appendix


Gross Income Over Time (Daily)
Total Gross Income
350
300
250
200
150
100
50
2019-01-01
2019-01-15
2019-02-01
2019-02-15
2019-03-01
2019-03-15
2019-04-01
Date

Appendix


Number of Transactions by Payment Method and Customer Type
Payment Method
Cash
Credit card
Ewallet
Number of Transactions
0
20
40
60
80
100
120
140
160
Member
Normal
Customer Type

Appendix


Sales by Product Line and Customer Type
Customer Type
Member
Normal
Total Sales
0
5000
10000
15000
20000
25000
Electronic accessories
Fashion accessories
Food and beverages
Health and beauty
Home and lifestyle
Sports and travel
Product Line