

## **Assignment 2**

Course: IST 707 Applied Machine Learning

Name: Prachi Sadarangani

Date: 09/14/2023

SUID: 354068727

## **Part I:**

### **Supermarket Sales – Data Analysis Report**

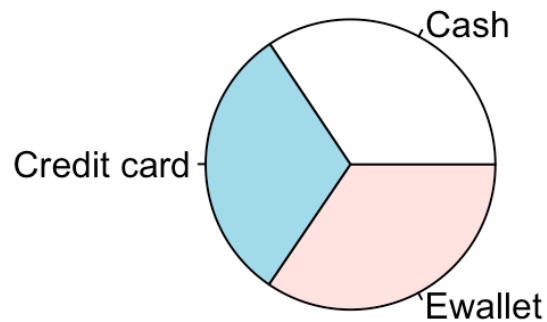
#### **Main idea:**

This report provides insights into supermarket sales data observed in three branches (A, B, C) from January 2019 to March 2019. It accounts for various factors such as the gross income, the peak hour, the gender distribution, the product lines, payment method, customer type, etc.

#### **Evidence:**

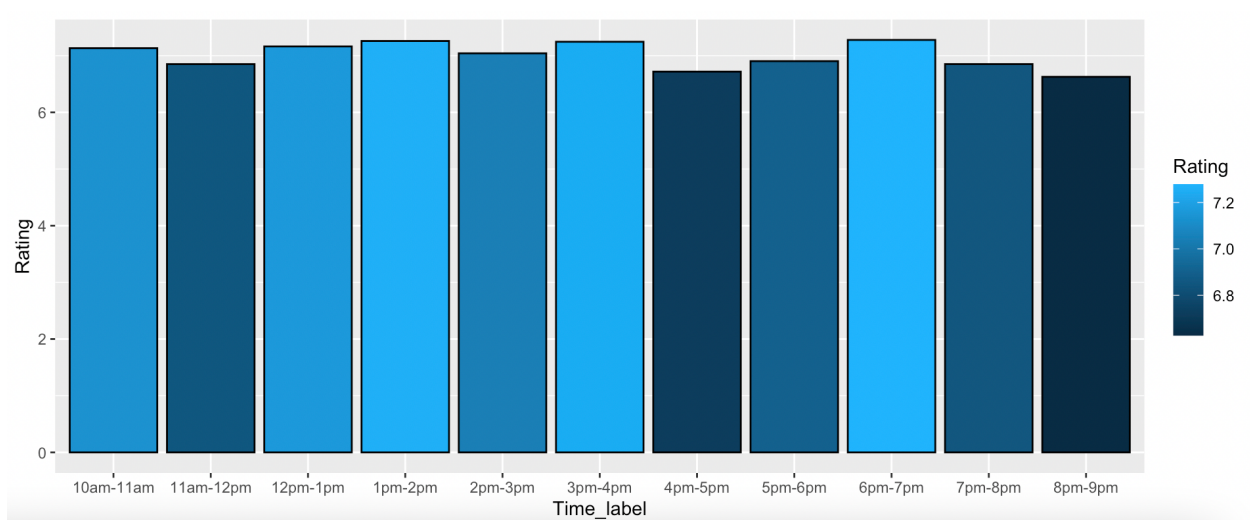
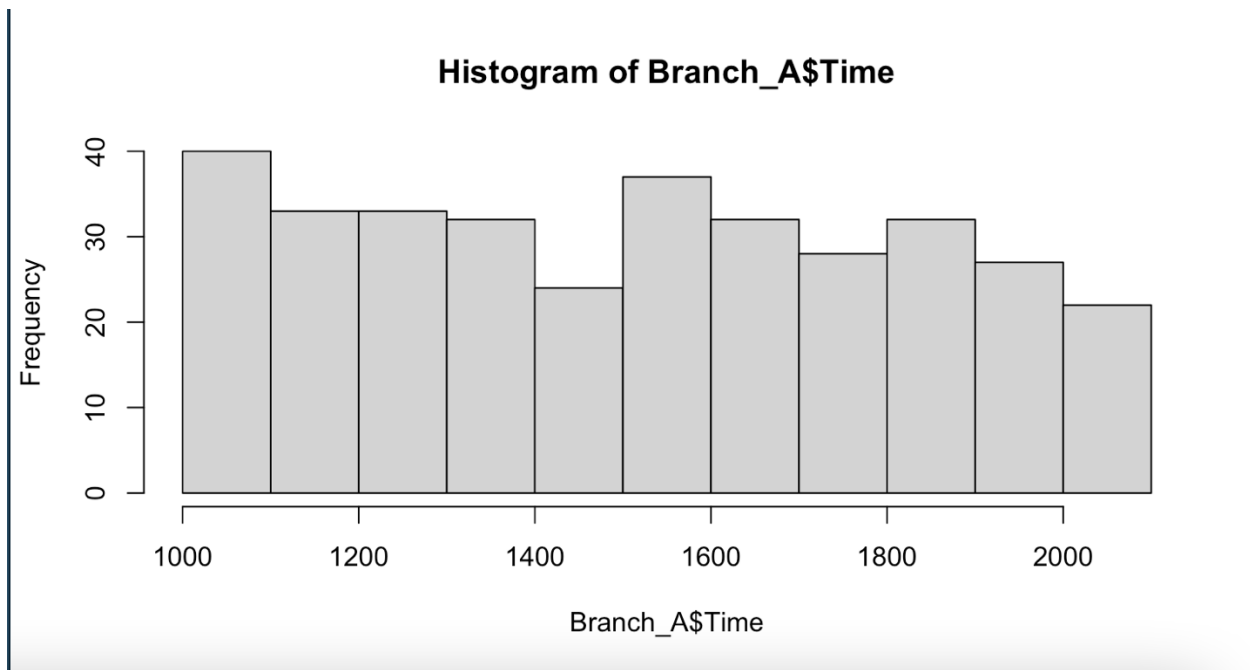
Superstore as a whole:

The gender distribution of the superstore is such that there are 499 males and 501 females. On average, male customers spend \$311 whereas female customers spend \$335. Male customers spend an average of \$348 on Health and Beauty products. Female customers spend \$380 on Home and Lifestyle products. The total income of the superstore is \$15379.37. Electronic accessories are sold the most in the superstore, there are 971 Electronic accessories products sold. As per the mode of payment, it is visible from the pie graph below that Ewallet is used the most by the customers. Additionally, when the total cost is greater than the 75th quartile then 34.8% pay using cash, 33.6% use Ewallet, and 31.6% use credit card. Out of 1000 customers, 501 are members whereas 499 are normal customers. Moreover, on average, a member spends \$327.8, and a normal customer spends \$318.12.

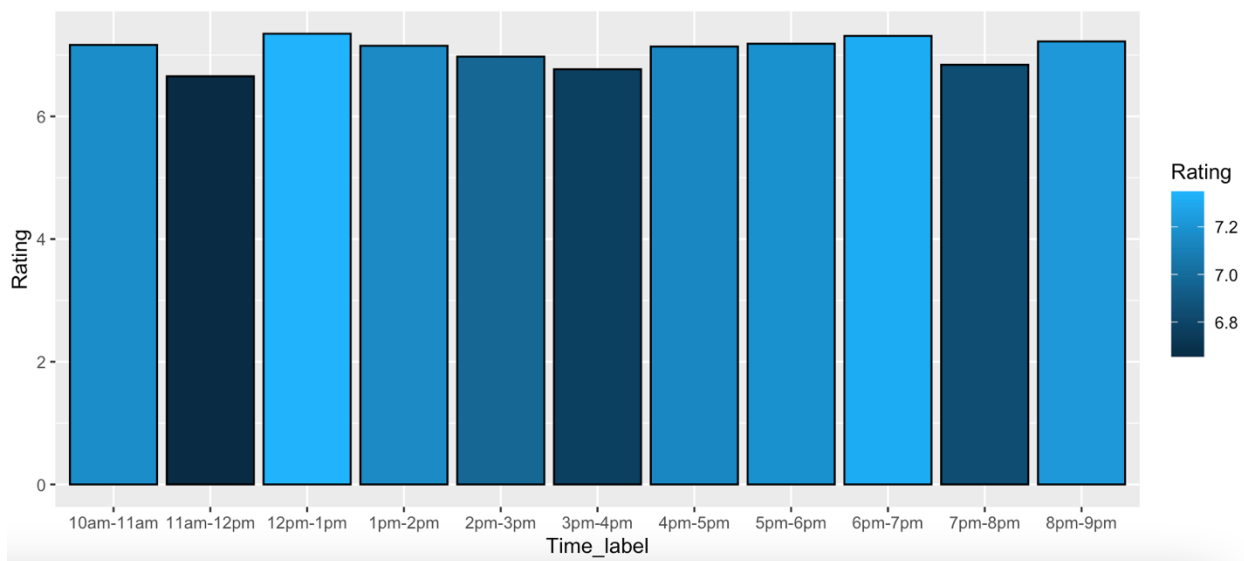
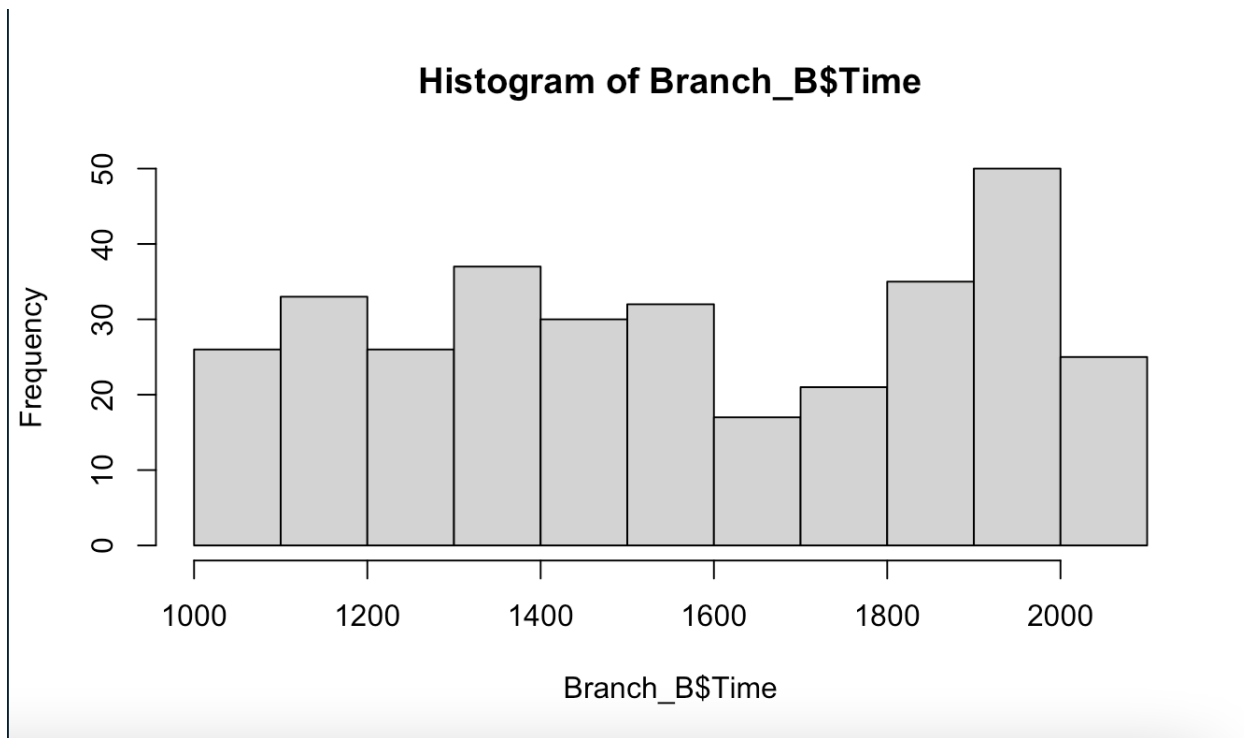


#### Branch Wise:

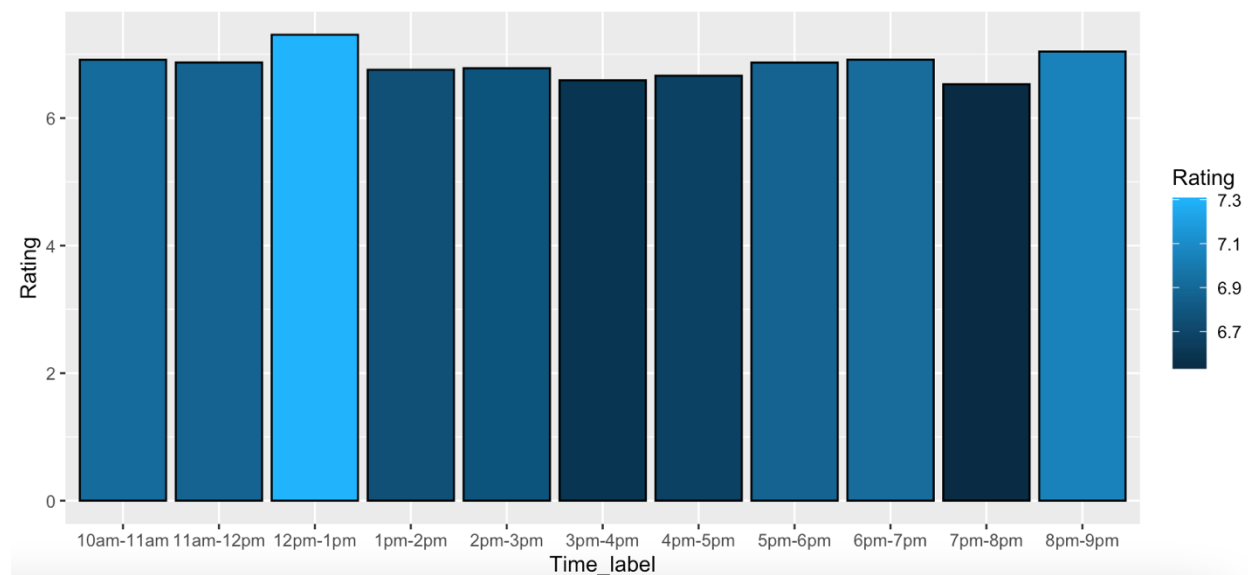
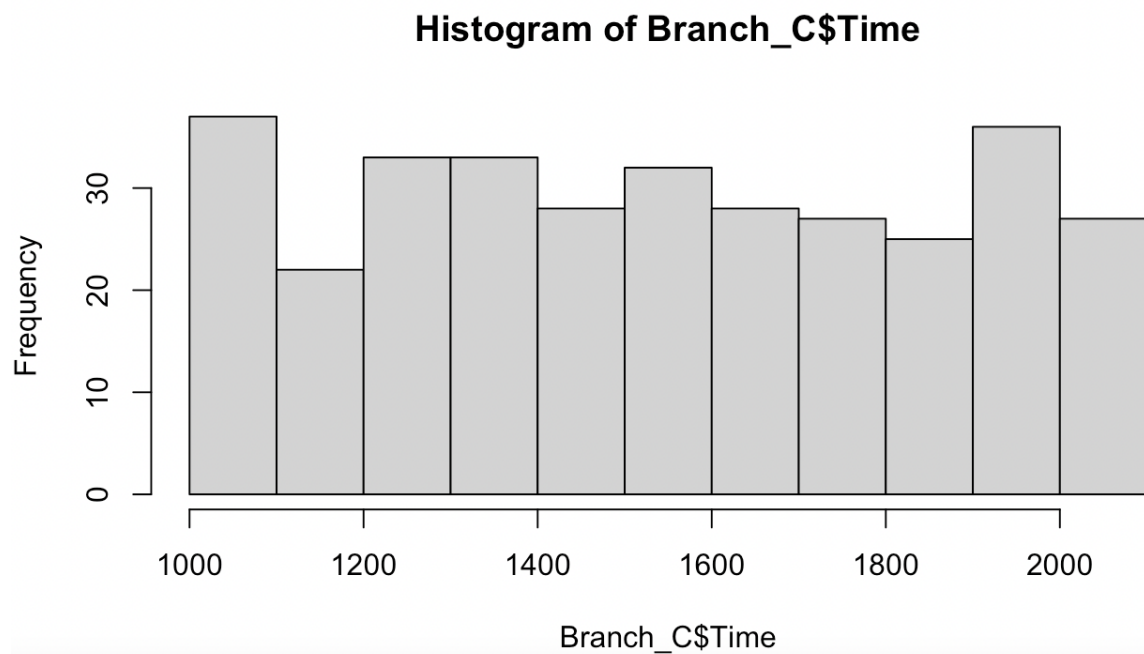
*Branch A:* Branch A has the maximum number of customers among all branches which is 340. Out of these, there are 161 female customers and 179 male customers. Additionally, the member and normal customer count is 167 and 173 respectively. The total gross income of Branch A is \$5057.16. Based on the histogram below, the peak hour for Branch A is between 10 a.m. to 11 a.m. Whereas its least busy hour is 8 p.m. to 9 p.m. Moreover, the mean rating by customers for Branch A is 7.03. Based on the 2nd histogram below, we can see that, the average customer rating throughout different times is almost similar, it lies between 6.5 and 7.5.



**Branch B:** Branch B has 332 customers. Out of all the customers, 165 are members and 167 are normal customers. Moreover, there are 162 female customers and 170 male customers. The total gross income of Branch B is \$5057.03. The below histogram shows the number of customers in the superstore at various hours. The average customer rating for Branch B is 6.81. The sports and Travel product line brings the most income in Branch B which is \$952. The 2nd histogram below focuses on customer ratings during different times. we can see that the busiest hour for Branch B is between 7 pm to 8 pm while 4 pm to 5 pm is the least populated.



**Branch C:** Branch C has 328 customers. There are 169 members and 159 normal customers. Out of the 318 customers, there are 178 female and 150 male customers. Branch C has a total gross income of \$5265.18. The below histogram portrays the number of customers in the store throughout the day at different times. Branch C gets the most income from Food & Beverages, \$1132. The 2nd histogram below shows how many customer ratings throughout different periods of time. Additionally, the average rating for Branch C is 7.07.



### Analysis:

Superstore as a whole:

The male-to-female ratio throughout the superstore is almost 1:1. On average female customers tend to spend about \$24 than male customers do. The superstore should advertise more products that the female customers would tend to buy. As per the evidence, on average, female customers spend the most money on Home & Lifestyle products followed by Food & Beverages. Whereas the male customers tend to spend

the most money on Health & Beauty Products followed by Sports & Travels. Although, overall, all customers seem to buy electronic accessories the most. Hence, there should be more electronic accessories products added to increase sales. Moreover, the evidence also suggests that most customers pay using the Ewallet, unless, the amount is high, then most customers use cash to make the payment. Hence, it's a good idea to always have change to improve customer satisfaction. Lastly, we can see that members tend to spend more than normal customers.

#### Branch Wise:

Although Branch A has the maximum number of customers, Branch C has the highest gross income. Branch A has more male customers than female customers. They also have more normal customers than those that are members. Hence, Branch A could focus more on having products that are preferred by male customers. They can also focus on the marketing of superstore membership to convert more customers into a member. From, the two histograms above, it is visible that the average customer rating for Branch A is not affected by the peak hour. However, during the closing hour, Branch A has a low average rating. Branch B has the lowest average customer rating. The histograms also suggest that the average rating of Branch B is a bit low during the peak hour. Maybe, hiring more employees to help out could improve the customer rating during the peak hour. Branch C has the highest average customer rating.

#### Link:

Based on the analysis, each Branch can focus on its customer base and cater to the marketing of products according to their demographics. Memberships could be promoted more by the superstore, since on average members tend to contribute more to the gross income than normal customers. The products that are sold the most could be stocked up more often so that the customers are satisfied, and sales are optimized. Customer ratings can further be improved by focusing on factors that are causing the ratings to go down.

#### Conclusion:

To conclude, the superstore analysis shares insights based on each branch, gender, and customer type provided above. Using these findings, data-driven decisions could be made to increase sales and profitability. Customer satisfaction can also be improved.

## **Part II: Preamble**

#### Preamble A:

Kelvin is more likely to be a Farmer. He is shy and withdrawn, he is someone who doesn't like to interact with people much. Being a farmer is essentially being by yourself and not having to interact with a lot of people. A librarian must interact with people who come into the library to borrow or return books or the ones who have questions about the library. Moreover, Kelvin is meek and tidy. As a librarian, it is their duty to make sure that people are silent and orderly. Although Kelvin is tidy, his meek personality would make it hard for him to enforce strict silence. He also has a passion for detail and farmers need to constantly monitor their products (animals and plants) to ensure a successful harvest. Additionally, statistically speaking, there are more farmers than librarians so the probability of Kelvin being a farmer is higher. Hence, Kelvin is most likely to be a farmer.

#### Preamble B:

Lucy is more likely to be a bank teller. Lucy is outspoken and deeply concerned with discrimination and social justice issues. Since Lucy cares about social justice issues, it is not necessary that her primary focus is the feminist movement. It is possible that Lucy is more inclined towards the anti-nuclear movement instead since she also participated in a demonstration regarding it. Moreover, it states that she was committed to social issues as a student. She is 31 now and a lot of people tend to lose interest in things as they grow up and their priorities change. Even though it is possible that she is still deeply concerned with social justice, there is a probability that she is not actively partaking in the feminist movement. Hence, the probability of her being a bank teller and an active part of the feminist movement seems to be less than the probability of her just being a bank teller.

*"I certify that this assignment represents my work. I have not used any unauthorized or unacknowledged assistance or sources in completing it, including free or commercial systems or services offered on the internet."*