Analysis of Starbucks using NPT and DataViz

INTRODUCTION

Starbucks was started in the year 1971. Their original business was roasting and retailing coffee, tea and spices. Currently, millions of customers are served in 33833 locations across 80 countries. It is a chain of stores that sell coffee, beverages as well as baked goods. In this study, author aims to test hypothesis regarding Starbucks factors in order to verify the relationship between these factors. Also, author shows how different age groups of customers spend money at Starbucks and how frequently they visit.

OBJECTIVE

1. To check is there a difference in Starbucks Price rate between regular customers and Nonregular customers.
2. Plotting the relationship between spending and frequency of visitation of age group of customers on items they buy at Starbucks.

METHODOLOGY

This study is based on secondary data which has been collected from the website www.starbucks.data. This dataset contains only 122 customer's data.

In this study, relationship between Starbucks factors has been checked by using hypothesis testing. To checking whether the data is normal or non normal, I have use Shapiro Wilk test. For testing non parametric data, I make use of Wilcoxon signed-rank test to check the relationship between Starbucks factors. Data visualization is done here to visualize the data. To visualize the relationship between age groups of customers and their spending, I used grammar of graphics.

ANALYSIS AND INTERPRETATION

```
> shapiro.test(starbucks_new$Price_Rate)

        Shapiro-wilk normality test

data:  starbucks_new$Price_Rate
W = 0.91251, p-value = 8.3e-07

> shapiro.test(as.numeric(starbucks_new$Loyalty))

        Shapiro-wilk normality test

data:  as.numeric(starbucks_new$Loyalty)
W = 0.51375, p-value < 2.2e-16

> wilcox.test(starbucks_new$Price_Rate ~ starbucks_new$Loyalty)

        Wilcoxon rank sum test with continuity correction

data:  starbucks_new$Price_Rate by starbucks_new$Loyalty
W = 509.5, p-value = 8.053e-07
alternative hypothesis: true location shift is not equal to 0

> # where PriceRate is numeric and Loyalty is A binary factor
> |
```

From Shapiro Test of both Variables, Price Rate & Loyalty(Regularity) of customers; We can see that the P-value is less than 0.05.The data is not normal, hence I performed the Wilcoxon signed-rank test to further analyze the data. Considering the following hypothesis as follows;

H_0: There is no significant difference in Starbucks price rate between regular and Non-regular customers.

H_1: There is significant difference in Starbucks price rate between regular and Non-regular customers.

Decision Criteria: Wilcoxon test shows that the P-value is again 8.053e-07 which is less than 0.05. Hence, we reject the NULL Hypothesis.

Conclusion: There is significant difference in Starbucks price rate between regular and Non-regular customers.
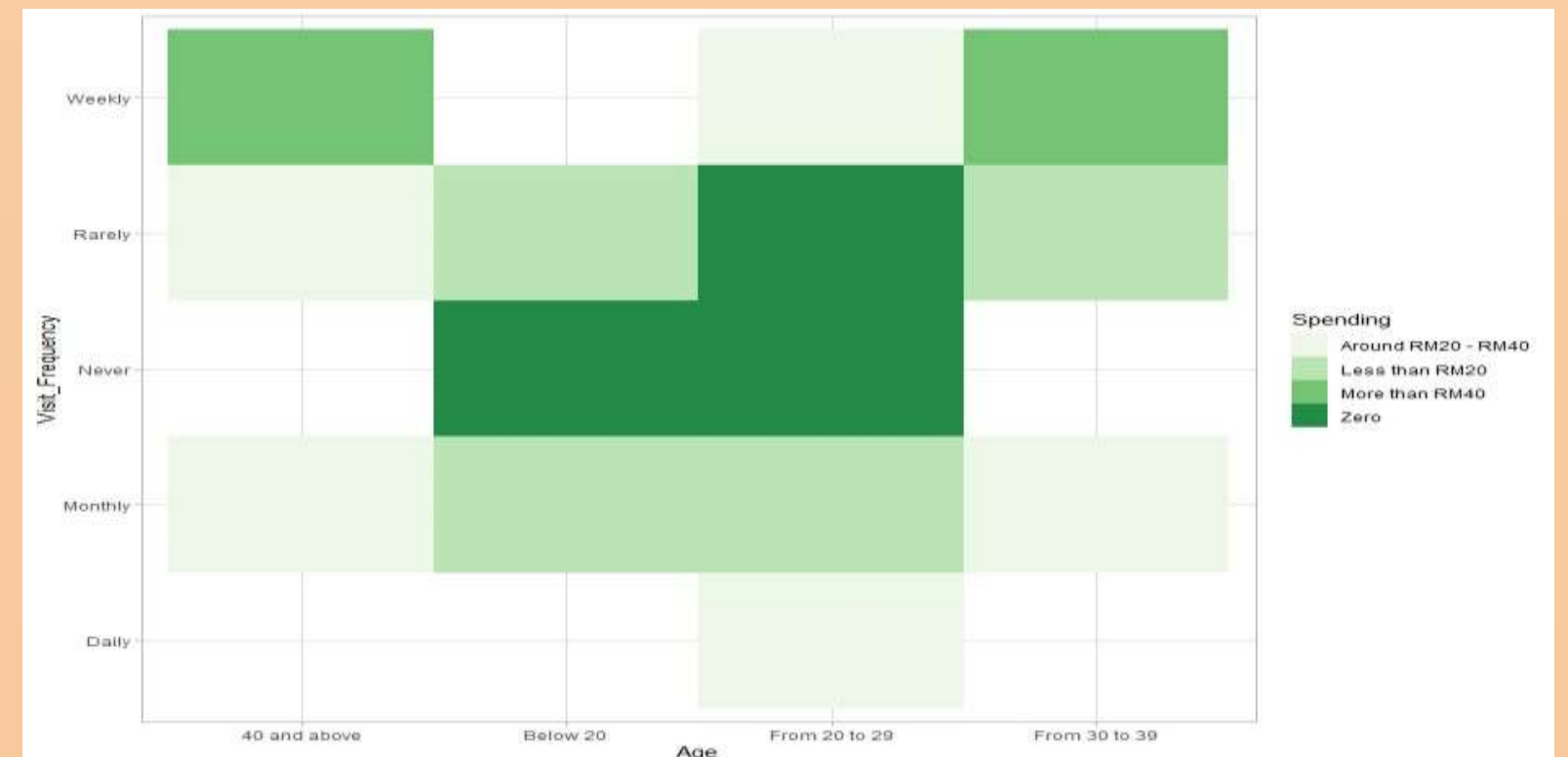
The visualization indicates visit frequency as per the age groups that visit the Starbucks store with respect to the amount, they spend in Starbucks for buying their items. We can see that ages above 30 spend more than RM-40 and they visit weekly to the Starbucks store. we can also see that age is below 30 spend less than RM20 or zero amount towards the Starbucks buying and their visit frequency is rarely or never. almost all age groups visit monthly to the Starbucks store and spend somewhere Below RM40. people below age 30 who never visit Starbucks store spend 0 amount. Age is below 30 has the most visiting frequency than ages above 30.

CONCLUSION

1. Customer's Rating the price of Starbucks items depends on whether the customer is regular or not.
2. The Starbucks management team can plan on improving their sales based on the visualization that shows visit frequency age and the spending of each customer.
3. The visualization is extremely helpful in understanding customer needs according to their visit frequency and the amount that they can spend as per their age.

LIMITATIONS

1. The dataset contains only 122 customer's data. This is very small for obtaining the accuracy of calculating results
2. Study questions do not cover the whole data from dataset.
3. We don't know how the data was collected, there is a possibility for customer to submit answers in different formats or spellings. This would make the data collection fallacious.

REFERENCES

- **https://www.starbucks.in/articles** .
- **https://en.wikipedia.org/wiki/Starbucks** .
- **https://www.geeksforgeeks.org/wilcoxon-signed-rank-test-in-r-programming/** .
- **https://ggplot2.tidyverse.org/reference/geom_tile.html** .

ACKNOWLEDGEMENTS

Author Name: Mandar Vasant Tarmale
Vidya Prasarak Mandal's B. N. Bandodkar College of Science,Thane(W)