Starbucks Customer Survey

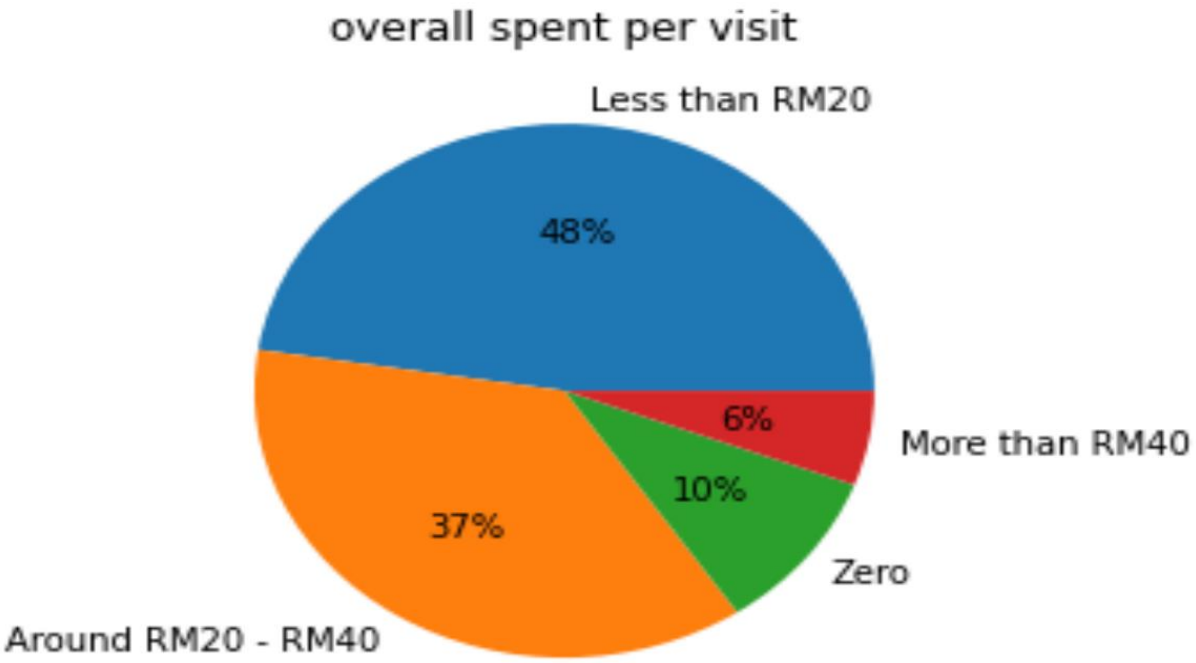
This dataset is composed of a survey question of over 100 respondents for their buying behaviour at Starbucks.

The survey dataset initially contained category columns as object types and all the columns had long names which makes it difficult for a data analyst while making calculations. I have cleaned the dataset and refactored the column names accordingly.

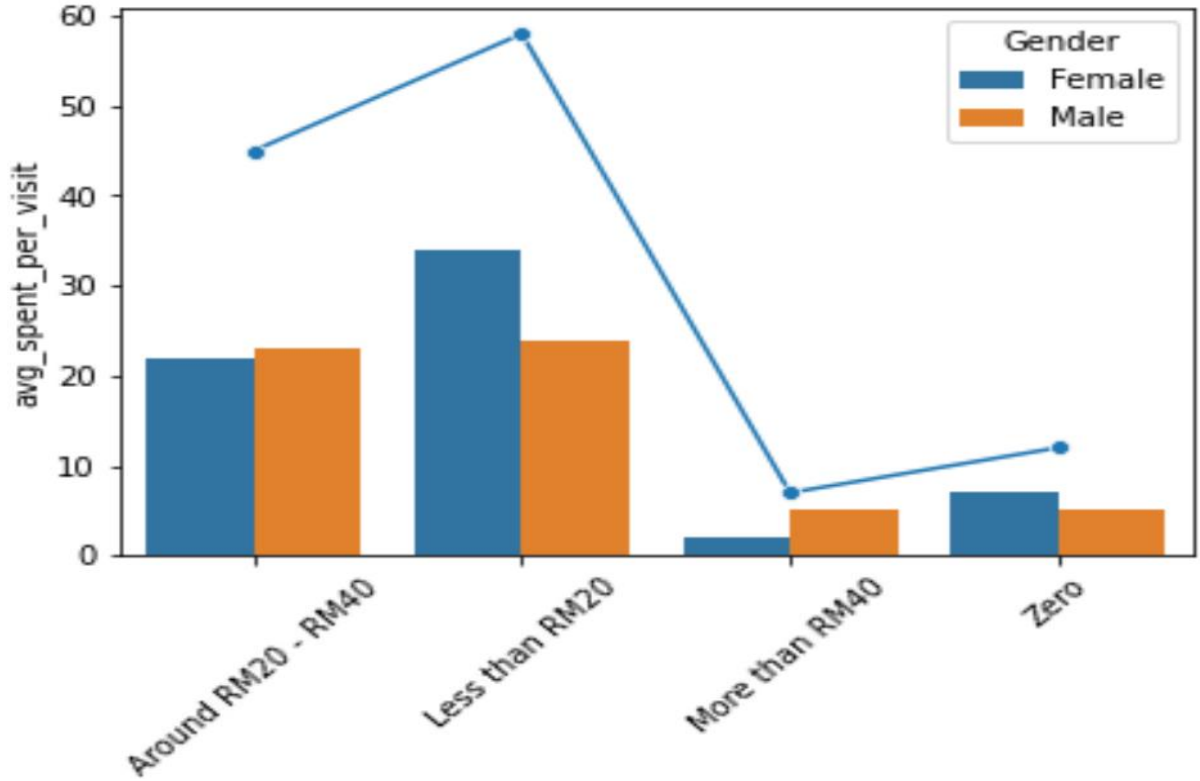
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 122 entries, 0 to 121
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Timestamp             122 non-null    object
 1   Gender                122 non-null    category
 2   Age                   122 non-null    category
 3   Employment            122 non-null    category
 4   Income                122 non-null    category
 5   visit_frequency       122 non-null    category
 6   visit_type            121 non-null    category
 7   visit_time_spent      122 non-null    category
 8   nearest_starbucks     122 non-null    category
 9   memcard_available     122 non-null    category
 10  frequent_purchase     122 non-null    category
 11  avg_spent_per_visit   122 non-null    category
 12  brand_rating          122 non-null    int64
 13  price_rating          122 non-null    int64
 14  promotion_rating      122 non-null    int64
 15  ambience_rating       122 non-null    int64
 16  wifi_rating           122 non-null    int64
 17  service_rating        122 non-null    int64
 18  choosing_stb_rating   122 non-null    int64
 19  promotion_heard_from  121 non-null    object
 20  willing_to_visit_stb  122 non-null    object
dtypes: category(11), int64(7), object(3)
memory usage: 13.6+ KB
```

Any business to be sustainable or remain afloat needs to generate subsequent sales. So, in this dataset the column which describes the sales detail is **avg_spent_per_visit**. We will be ignoring the column **frequent_purchase** as there doesn't seem to be much variability in the product names.

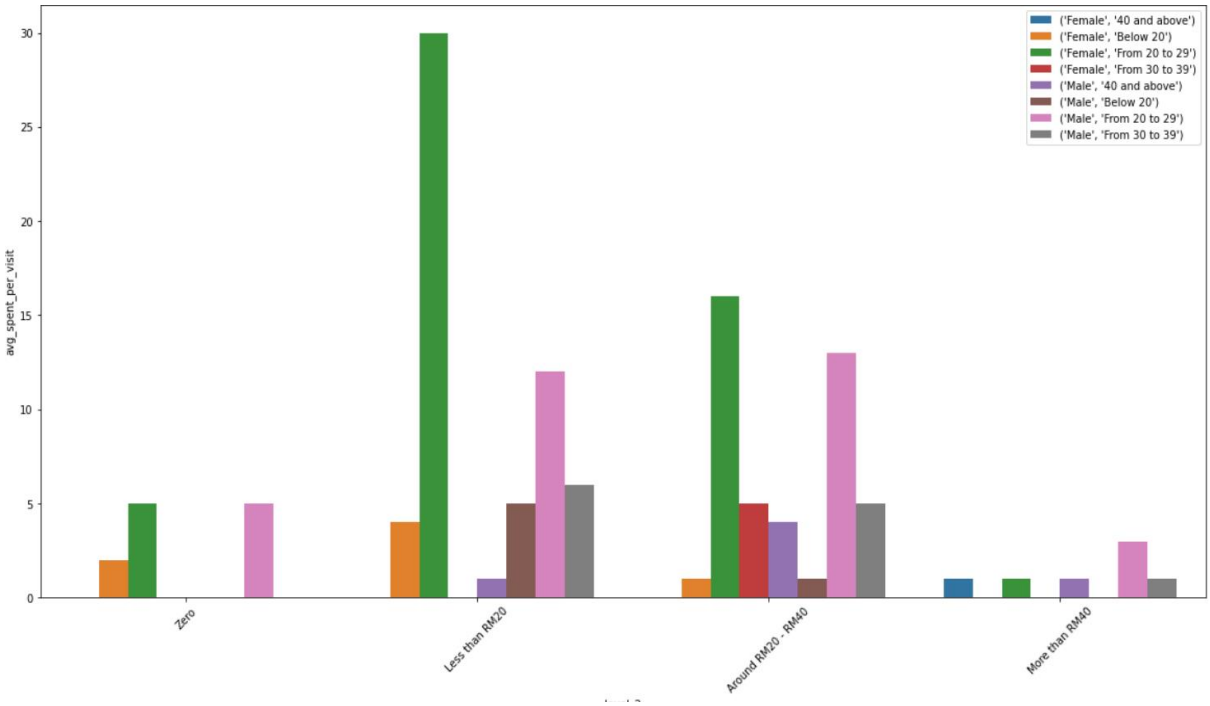
When we start analyzing the **avg_spent_per_visit**, we see that there seems to be a large sales volume in the categories **Less than RM20(5.81 CAD)** and **Around RM20 – RM40 (11.63CAD).**



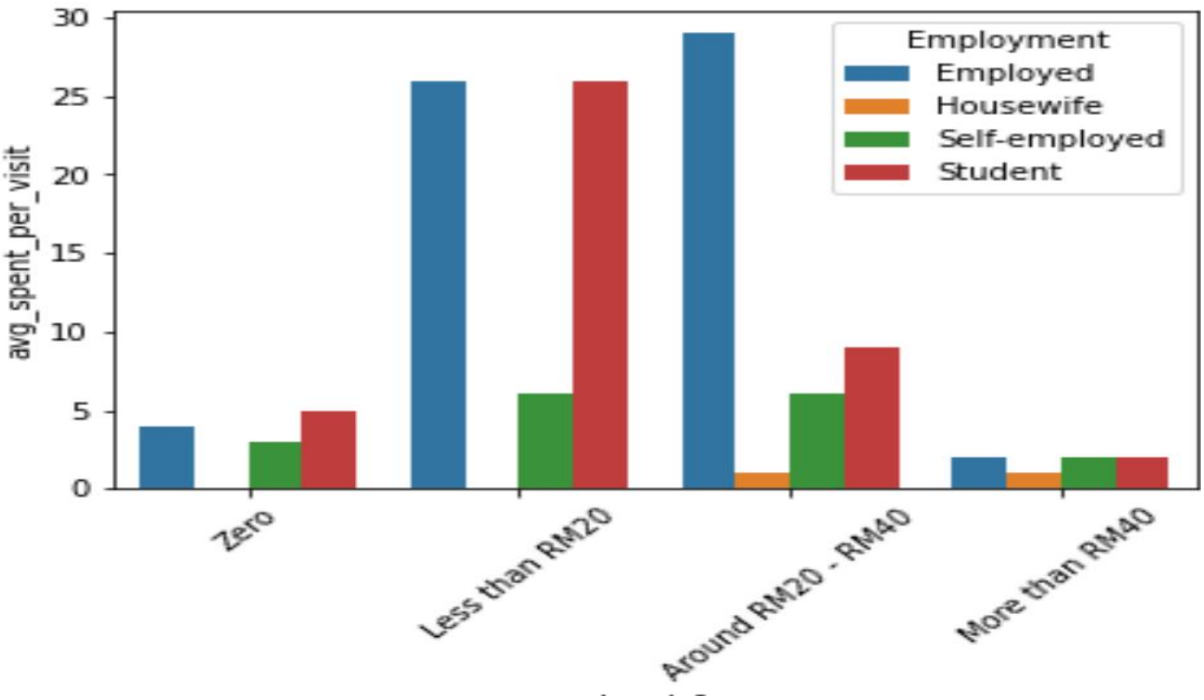
When drilling down deeper among Males and Females, we can observe that females spent more in the category Less than RM20.

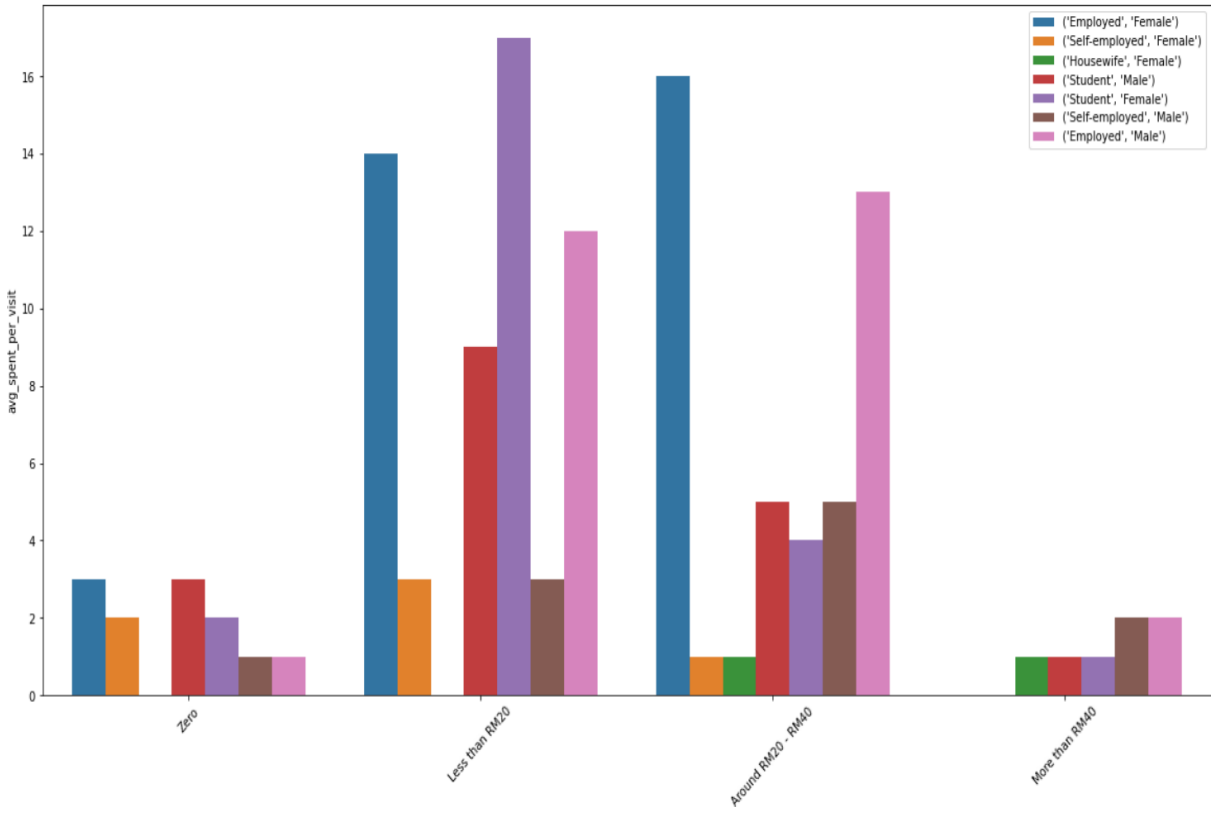


The same follows for different age categories among Males and Females.

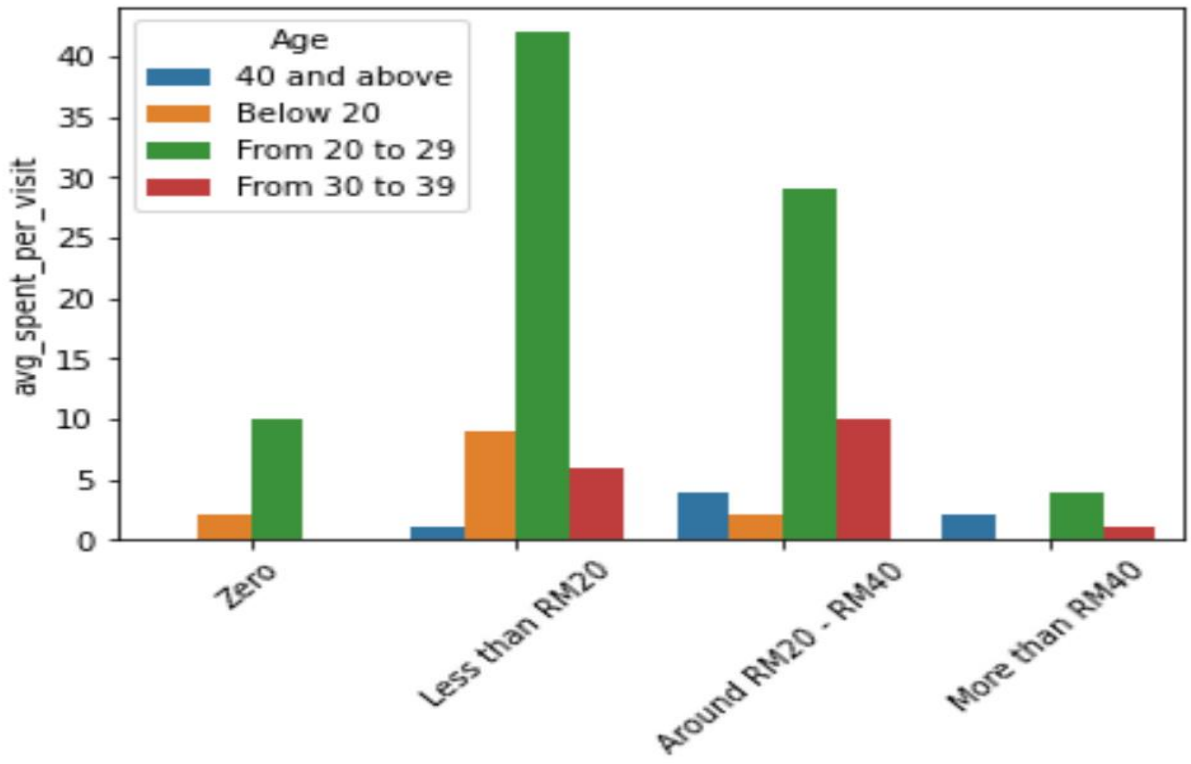


While checking if Employment type had been a confounder for less sales in the category of **More than RM40**, we can outright ignore this assumption since many employed customers too are not willing to spend in the category **More than RM40.**

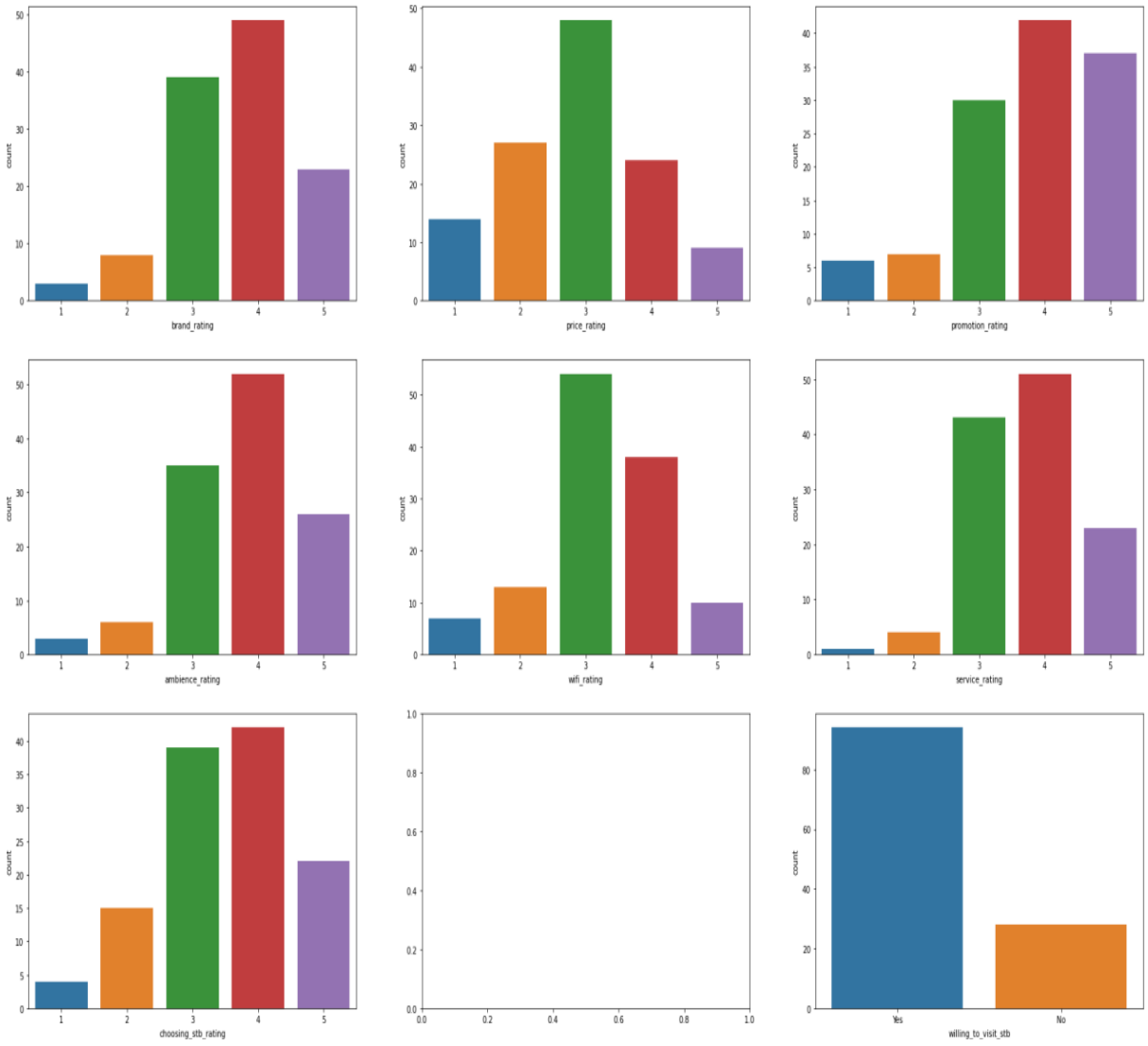




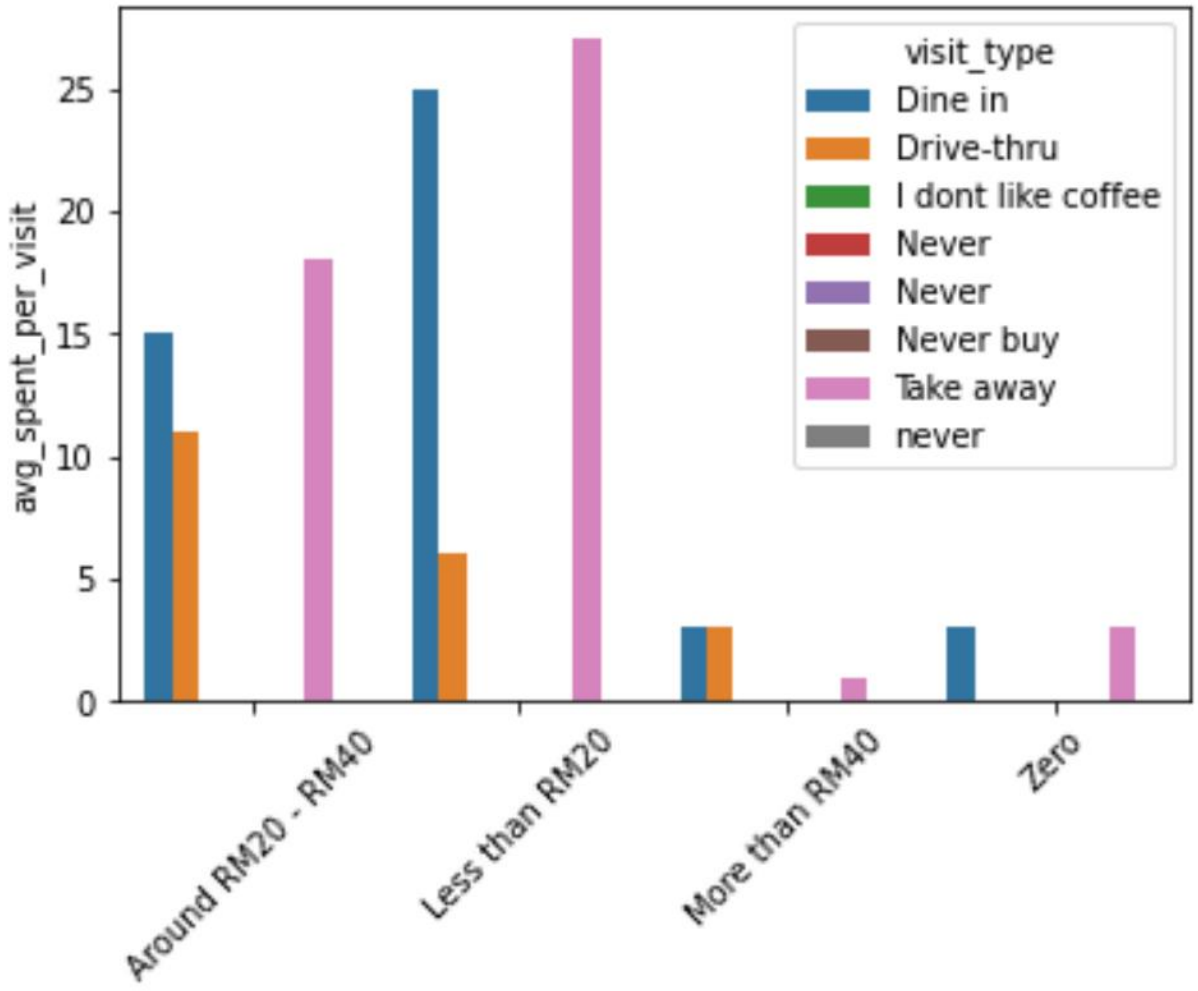
While considering Age into the picture, millennials seem to spend very less in the More than RM40 category products. And interestingly, people in the range 30 to 40+ age range when visited the store always seem to spend on something.



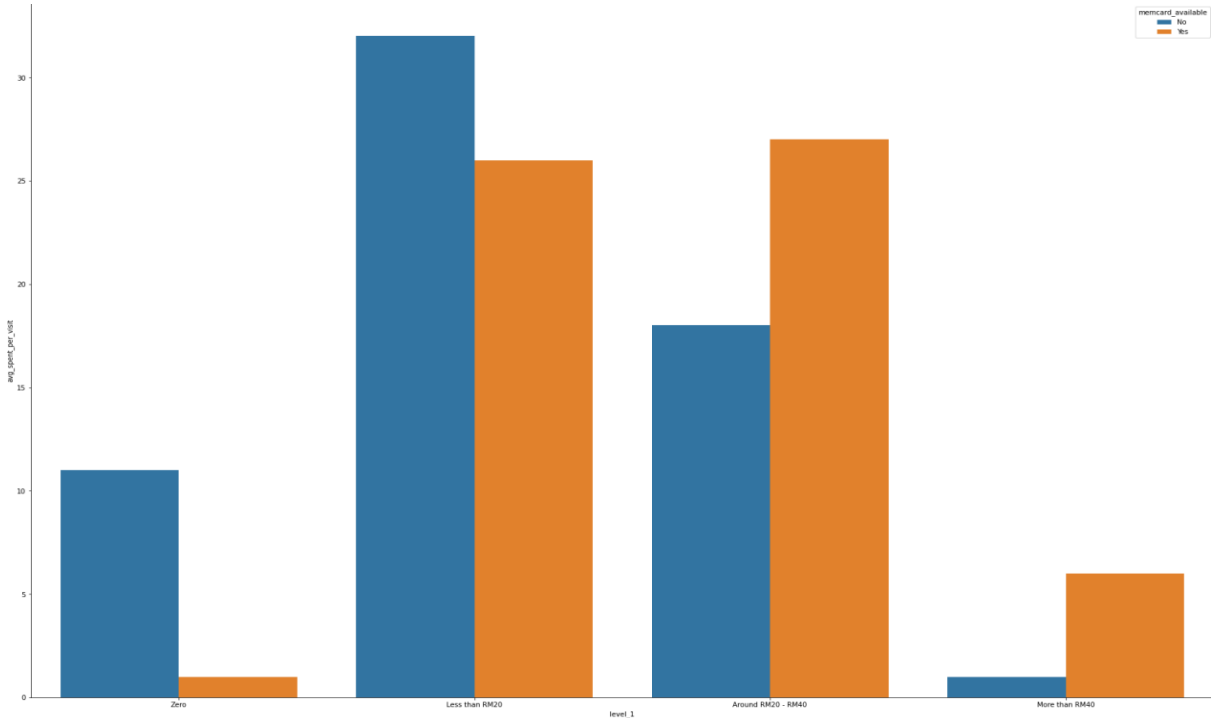
Plotting the count plot of all the ratings column, we see that most of them follow a right skewed distribution towards 5 star rating, **except price and wifi which seem to follow a perfect normal distribution. So there's a chance for improvement.**

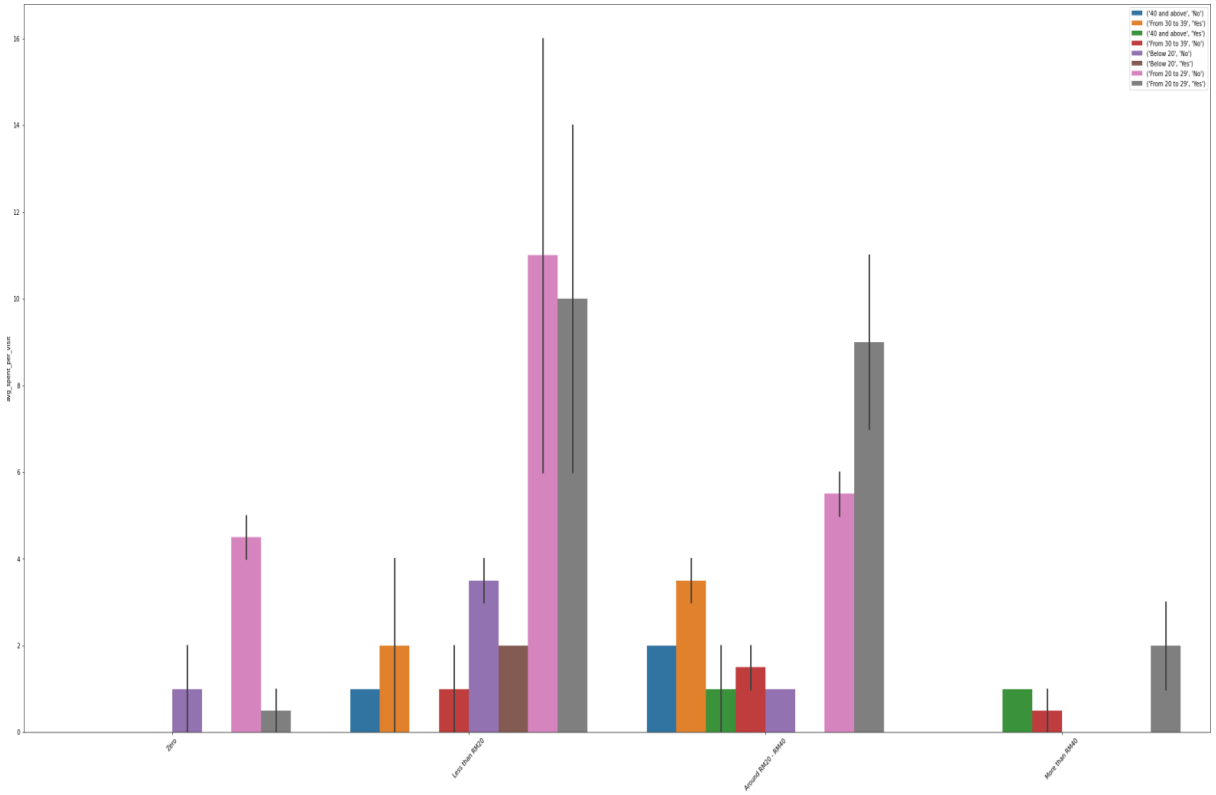


Hypothetically there's a chance for a substantial sales increase from Dine in customers by improving the wifi network. Since most people who Dine in tend to use laptops, mobile phones. As of now, there's no data to prove this hypothesis but they mostly seem to be correlated.



Customers who have **membership** card seems to have a sense of belonging, as most customers with membership card always seem to spend and sometimes even more in higher category products.





To conclude the analysis, by testing these below suggestions there a chance for subsequent increase in sales

- Adding few more products to the category ‘Less than RM20’ and Around RM20 – RM40’
- Improving the WIFI network
- Signing up more customers who are above 20 into members

Group G:

Dheshoju Kalyan Kumar

Gurdaan Walia

Manuel Paredes

Keerat Singh Sandhu