MINI PROJECT

Name: Aditya Agarwal
Roll No. : 2110110854
SNU ID: aa501

INTRODUCTION

The goal of this report is to study the data given to me in the form of an excel sheet, namely “Serve.xlsx”.

And to provide suggestions to the company after analyzing it successfully.

Few of the things we are going to analyze are as follows:

- Tipping difference when the gender of waiter changes
- Tipping difference when the time of meal changes
- Bill amount on different days
- Bill amount for different meals
- Average bill amount for different party sizes
- Average bill amount per person for different party sizes
- Average party size for different days of the week
- Average party size for different meals
- Average bill amount for smokers and non smokers

DATA CLEANSING

Library used for reading data from the excel sheet was openpyxl.

Outlier Handling

Interquartile range method:

The range, which is the minimum subtracted from the maximum, is one indicator of how spread out the data is in a set (note: the range is highly sensitive to outliers—if an outlier is also a minimum or maximum, the range will not be an accurate representation of the breadth of a data set). Range would be difficult to extrapolate otherwise. Similar to the range but less sensitive to outliers is the interquartile range. The interquartile range is calculated in much the same way as the range. All you do to find it is subtract the first quartile from the third quartile.

Removing further ambiguity

To handle further ambiguity present in the data, we are using the mean method almost everywhere. And in places where using the mean method doesn't make sense, we have used Scatter plots and Line of best fit.

Invalid Entry Handling

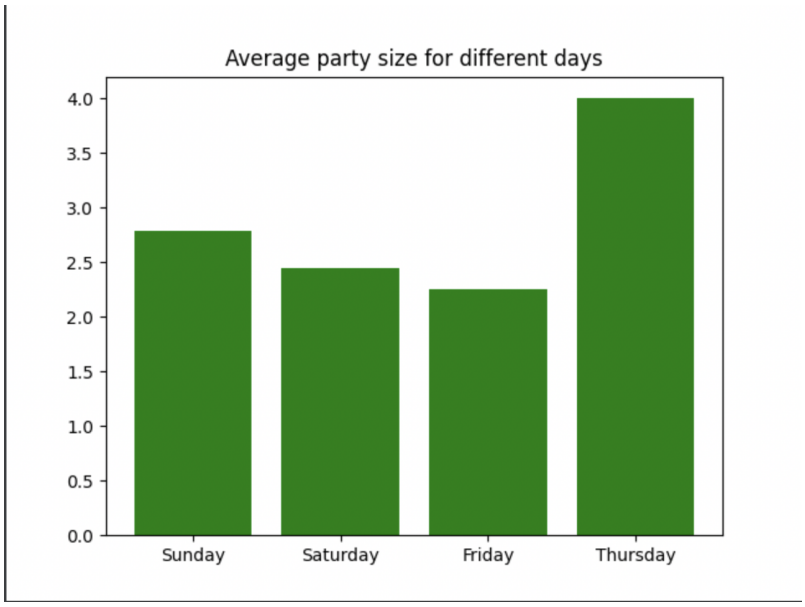
Upon going through the excel sheet it was noticed that a lot of entries were inavlid (eg. the party size for one entry was left blank, amount for one entry was given in numbers). This was handled through 2 means.

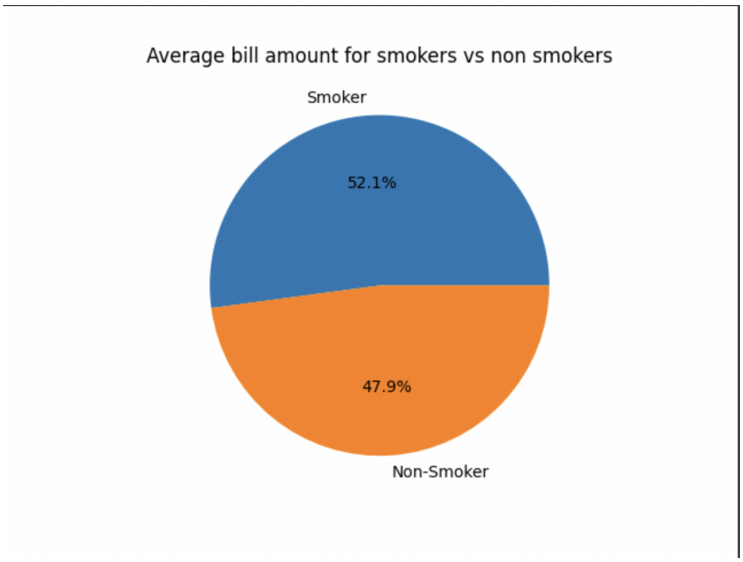
- First one being the isfloat function. This was defined in the beginning of the python file and checks whether the value passed through it is a legal floating value or not.
- The second method being, the if statement used at the beginning of every for loop to check if a cell is empty, i.e contains "None" value or not.

DATA VISUALIZATION

Python libraries used for data visualization were Numpy and Pandas.

Here are some sample Graphs generated using the above libraries.





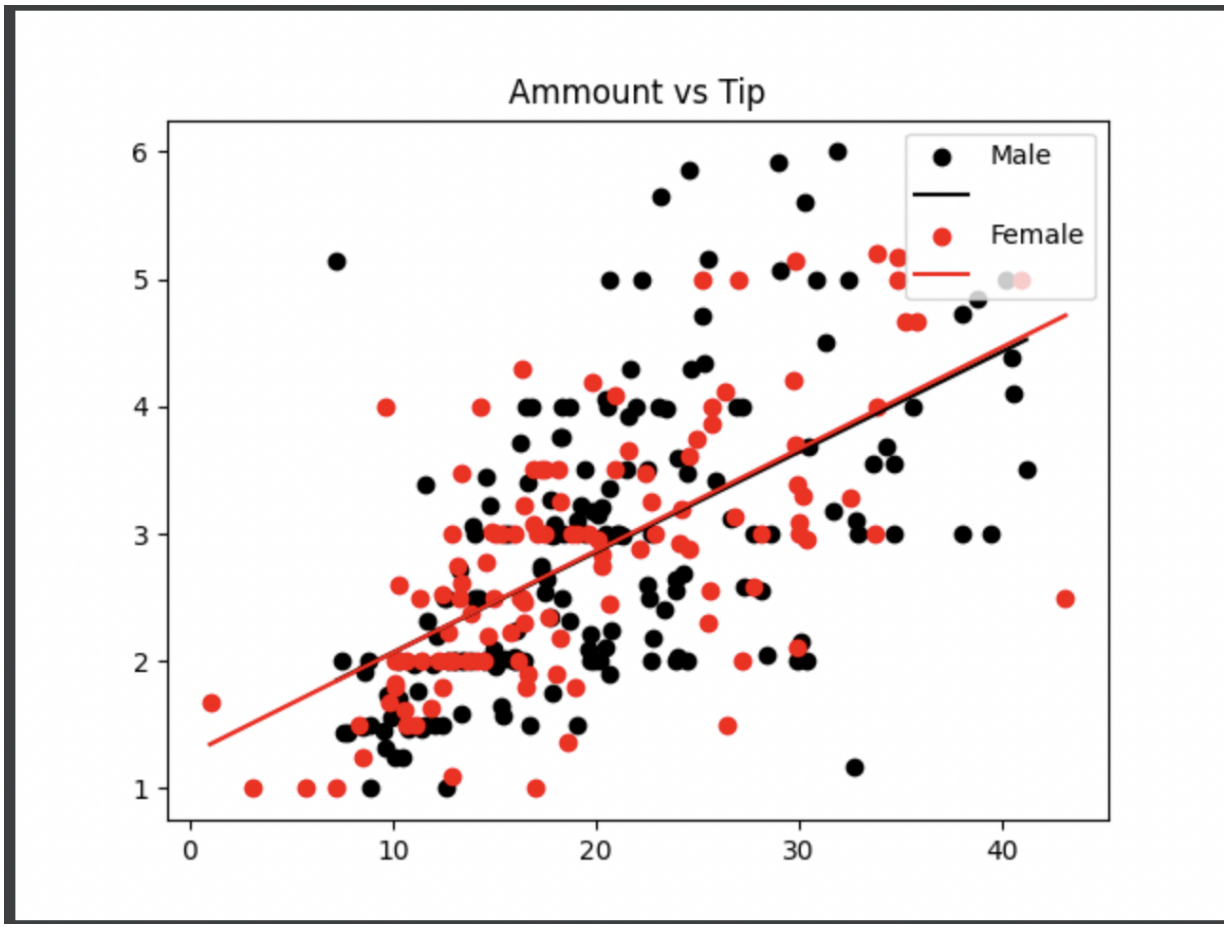
FINDINGS AND SUGGESTIONS

Tipping difference when the gender of waiter changes

After cleansing the data through the above mentioned methods, the amount was taken on the X axis and tip given to the waiter/waitress was taken on the Y axis.

This was then plotted on to a graph with the red dots representing the data points representing cases when the waiter was female, and black representing the cases when the waiter was male.

The red line represents the Line of best fit of the female waitresses and the black line represents the Line of best fit of the male waiters.



Upon analyzing the above graph we can see that there is no major tipping difference between male and female waiters.

But we also notice that female waiters are tipped when the bill amount is considerably low.

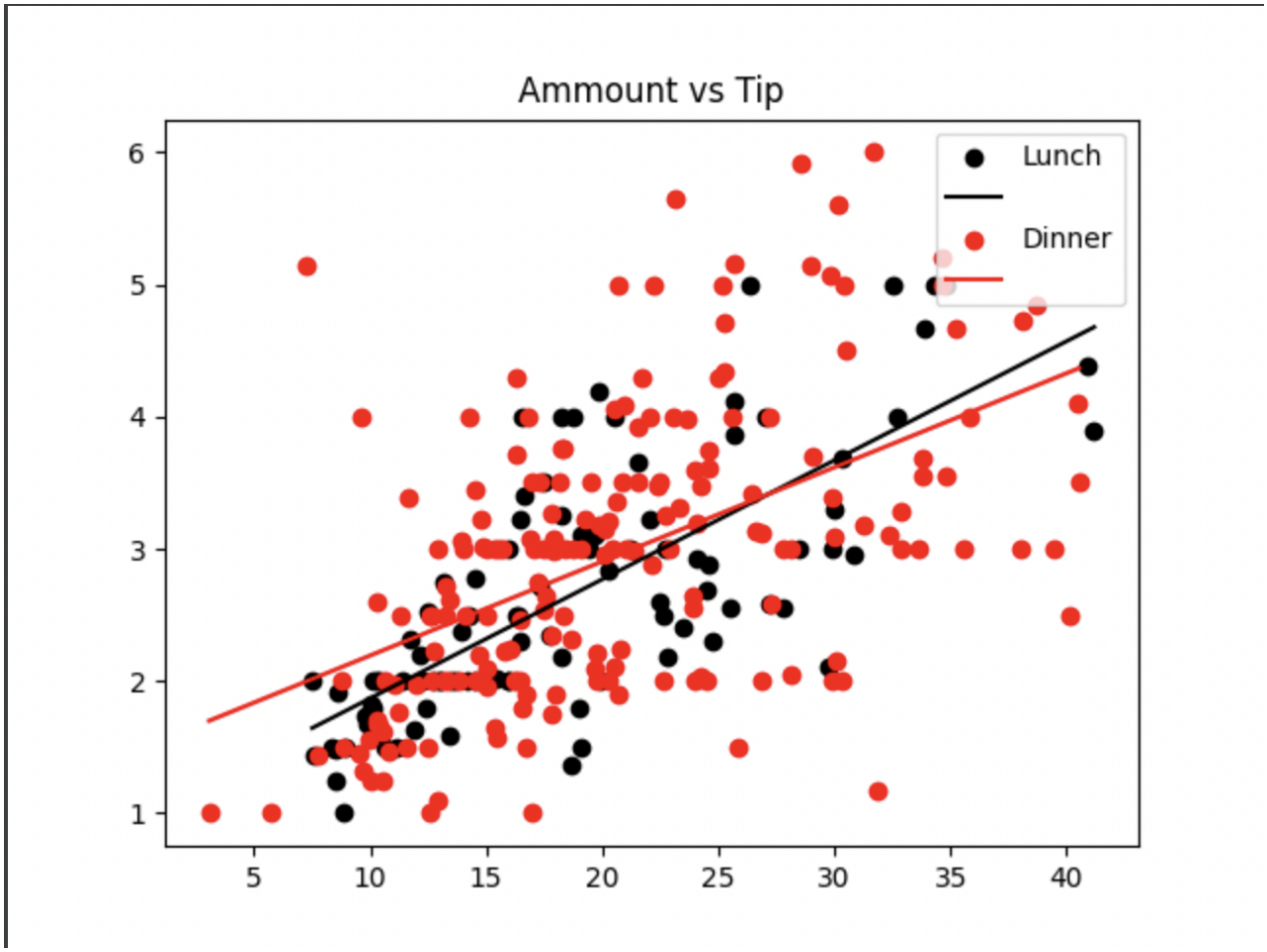
Therefore I would suggest that we send female waitresses to the table from which we expect small bill amounts.

This would increase the chances of them being tipped.

Tipping difference when the time of meal changes

Similar to the previous graph, we have plotted amount on the X axis and the tip on the Y axis.

And this time black represents data points where time of the meal is Lunch and red represents data points where time of the meal is dinner.

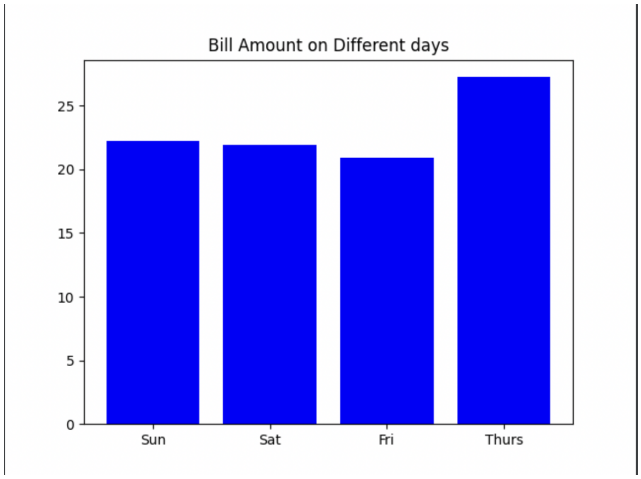


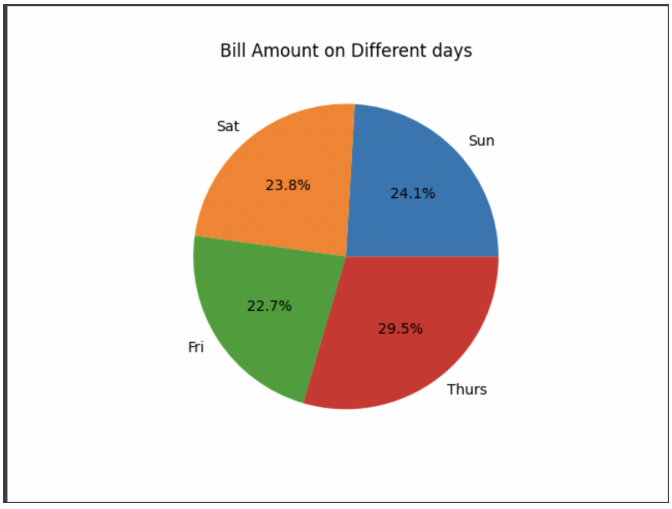
Upon analyzing we see that customers do not tip well during dinner when the bill amount is high.

Thus I would suggest that we improve the service quality during dinner hours, to improve the amount customers tip.

Bill amount on different days

Average bill amount for different days was calculated and is shown in the form



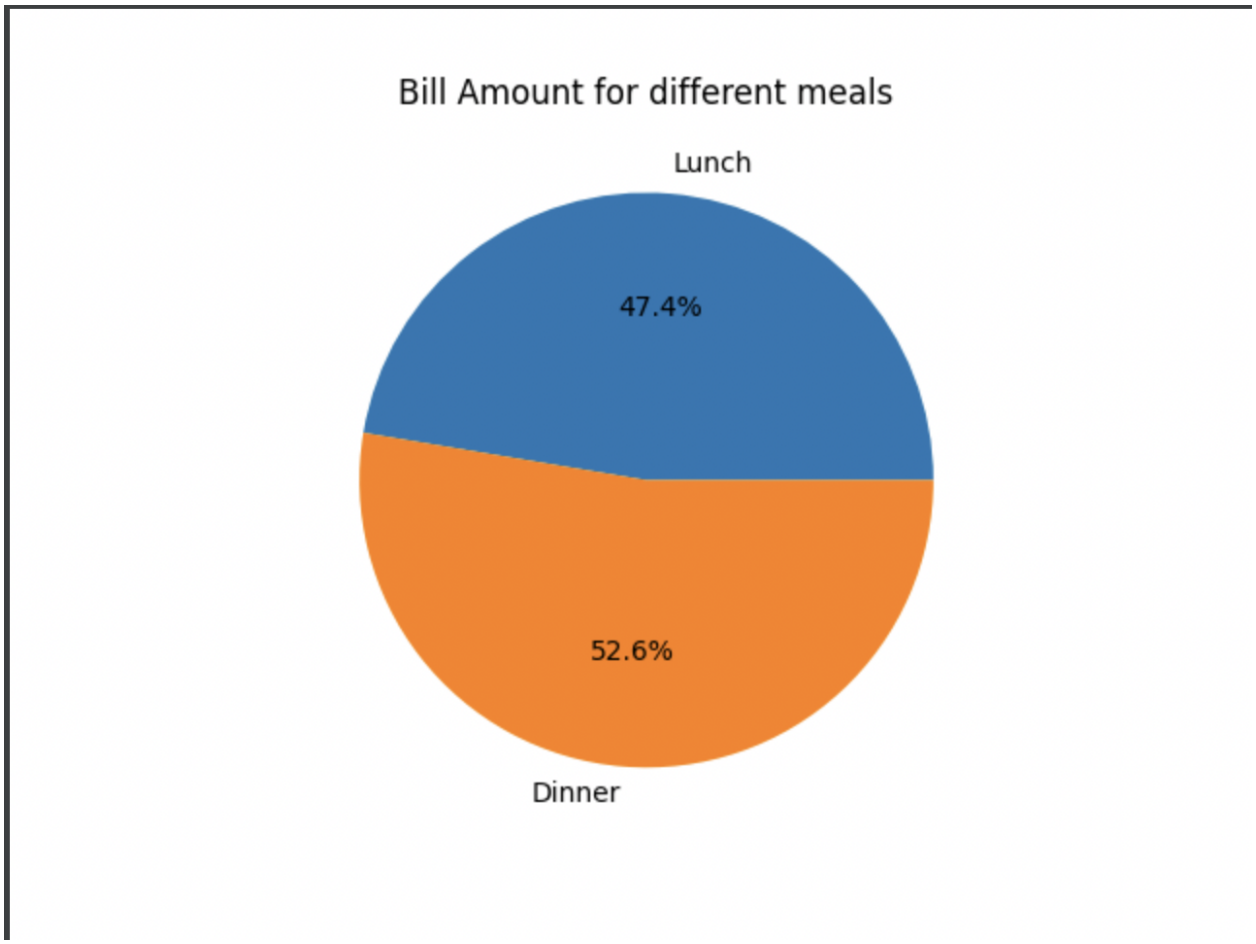


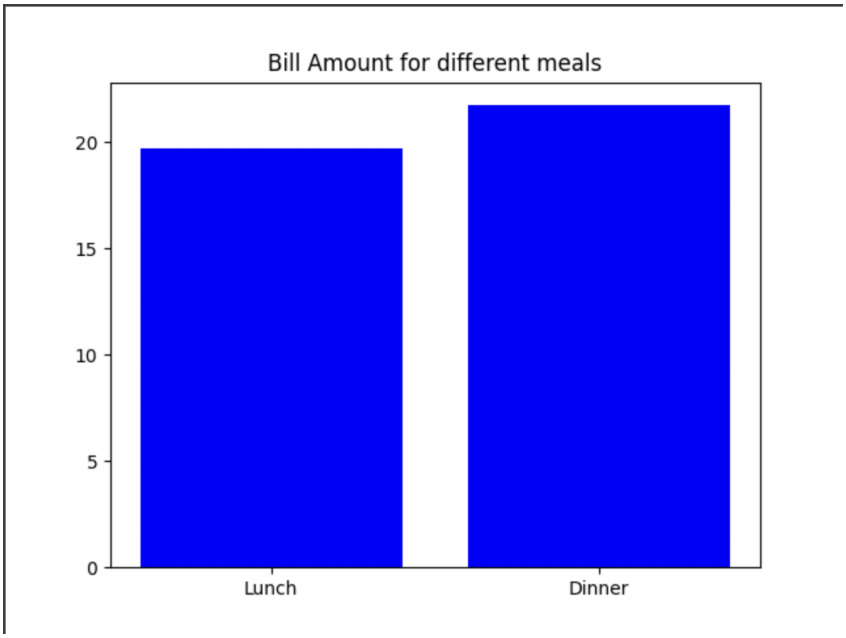
Upon analyzing the above graph and chart we can conclude that, the average highest amount of bills is generated on Thursdays.

Thus I would suggest that we are prepared on Thursdays to handle a large number of orders by stocking up on different raw materials and by having enough waiters ready to reduce the customer waiting time.

Bill amount for different meals

The following graphs are similar to what was shown up, except this time we have taken different time of meals into consideration instead of different days.





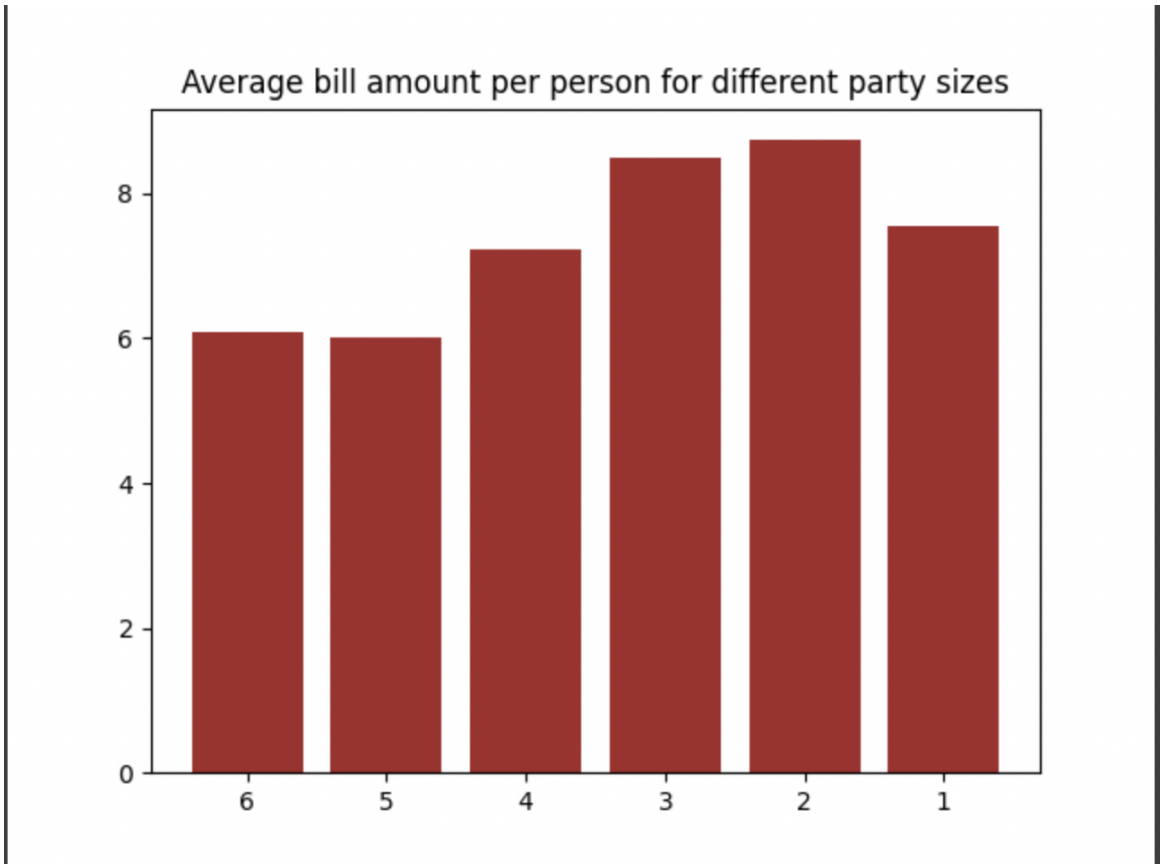
The above table doesn’t show much disparity between the bill amount when customers come during lunch and during dinner, hence I have no suggestions in this case.

Average bill amount for different party sizes

The below graph shows the average bill amount per person when they come in different party sizes.

That is, if a party of 2 has a bill of 20, then the per person amount would be 20/2 = 10.

Average amount per person was taken on the Y axis and party size they arrived in is taken on the X axis.

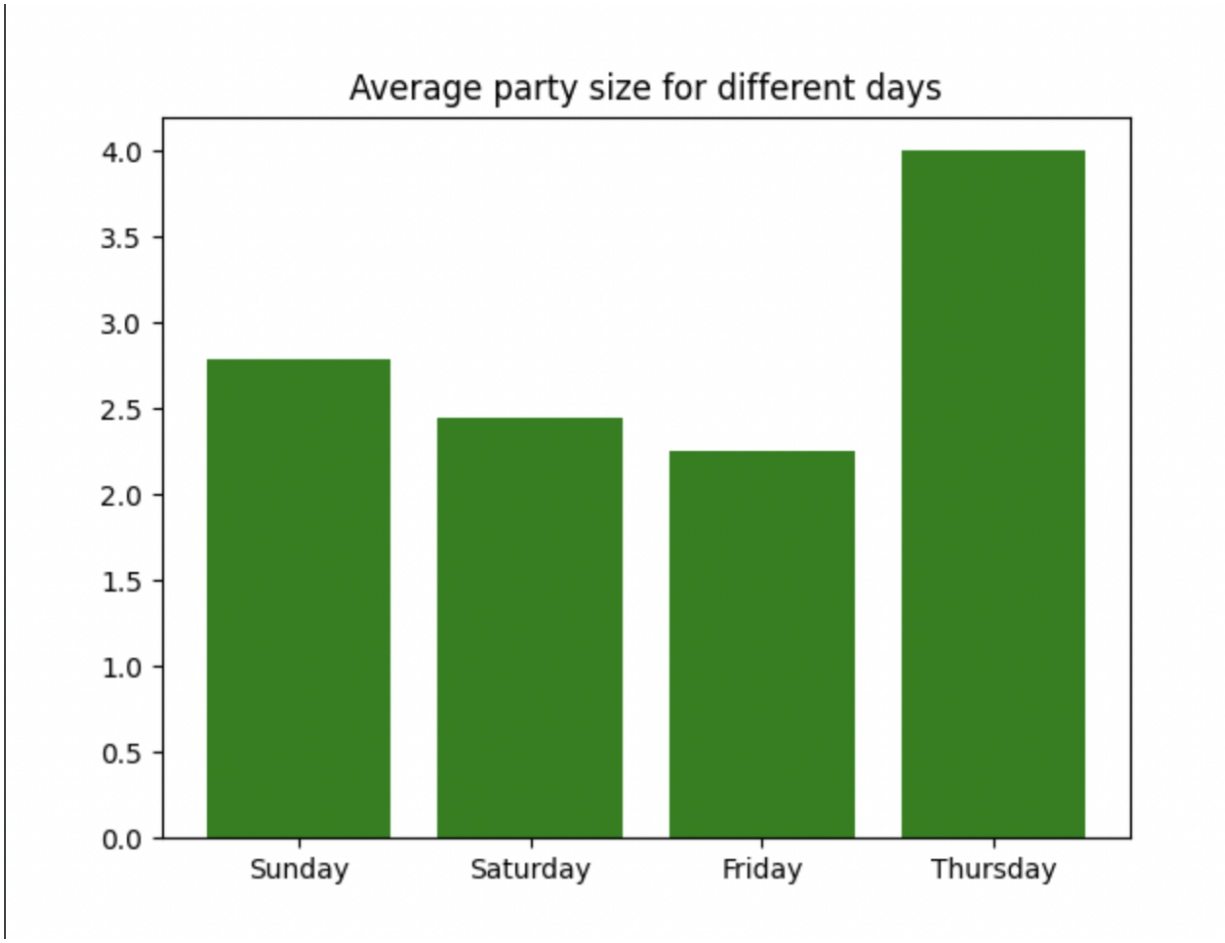


We can see that, highest average bill per person is, when people come in party size of 2.

Thus I would suggest that we make the restaurant more couple friendly, to attract more people in groups of two, to maximize our bill amount and ultimately increase our profits.

Average party size for different days of the week

Average party size for each day was calculated as shown in the bar graph present below.



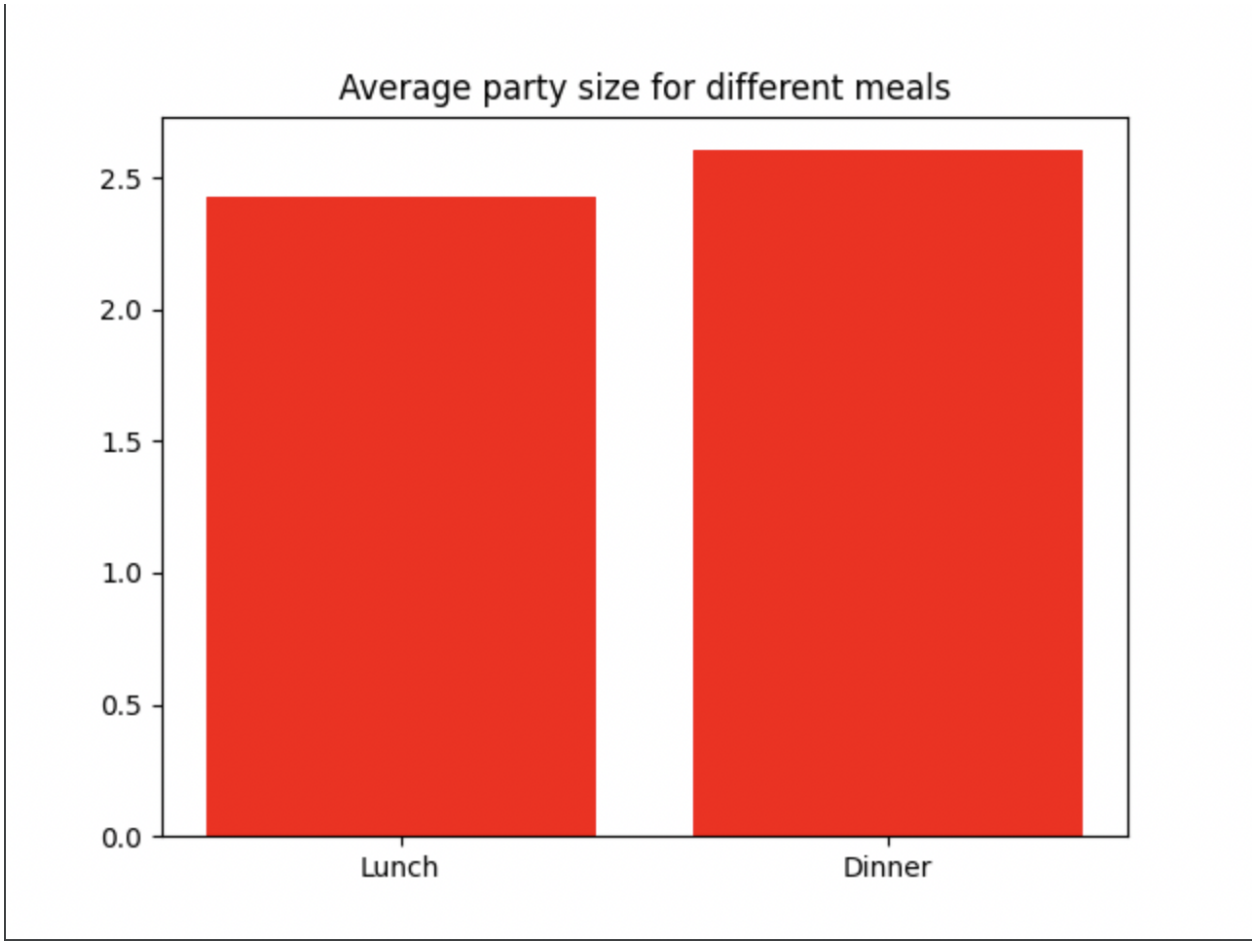
In concurrence to our earlier finding, of Thursday having the highest bill amount, it is no surprise that Thursday also has the highest party size.

Thus I would once again suggest what I had suggested earlier.

I would suggest that we are prepared on Thursdays to handle a large number of orders by stocking up on different raw materials and by having enough waiters ready to reduce the customer waiting time.

Average party size for different meals

Similar to the above graph, average party size for different times of meals was calculated and is displayed in the form of a bar graph.

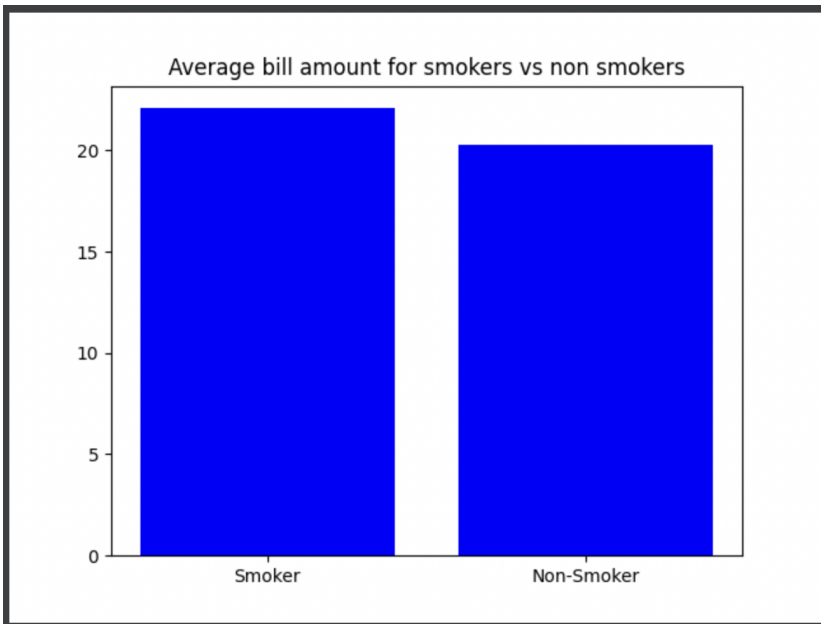


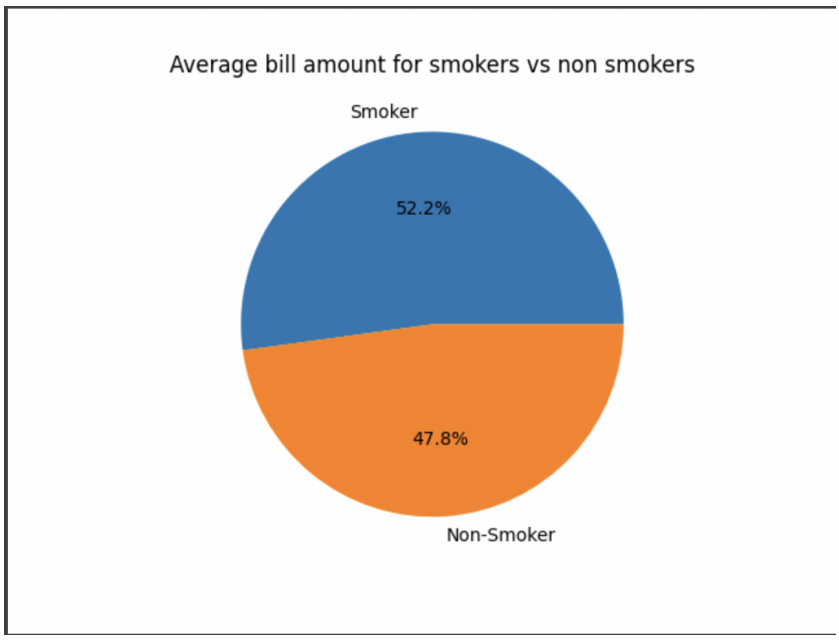
It is evident from the bar graph that there is no noticeable difference between the average party size for lunch and average party size for dinner. Thus I would not make any suggestions in this case.

Average bill amount for smokers and non smokers

The bar graph below shows the average bill amount of a smoker and a non smoker.

The pie chart below shows the percentage difference between the bill amount of a smoker and a non smoker.





Upon studying the above graphs we can see a small difference between the bill amount of a smoker and a non smoker.

In this case I would like to make 2 suggestions to the business:

- **First one being, that we start selling cigarettes at our restaurants to attract more smokers, that would eventually increase the bill amount because smokers tend to have a higher bill amount (inferred from the graph)**
- **Second one being that we consider building a smoking room for smokers, which would again attract more smokers and increase our bill amount and eventually our profits at the end of the day.**

REFERENCE LINKS

- INTERQUARTILE METHOD
- IS FLOAT FUNCTION
- CONTINUE
- NUMPY
- PANDAS
- OPENPYXL