Project: Zomato API II

This project is submitted by Gurpreet Singh, email-singh.gurpreet513@gmail.com, phone number-8837841113.

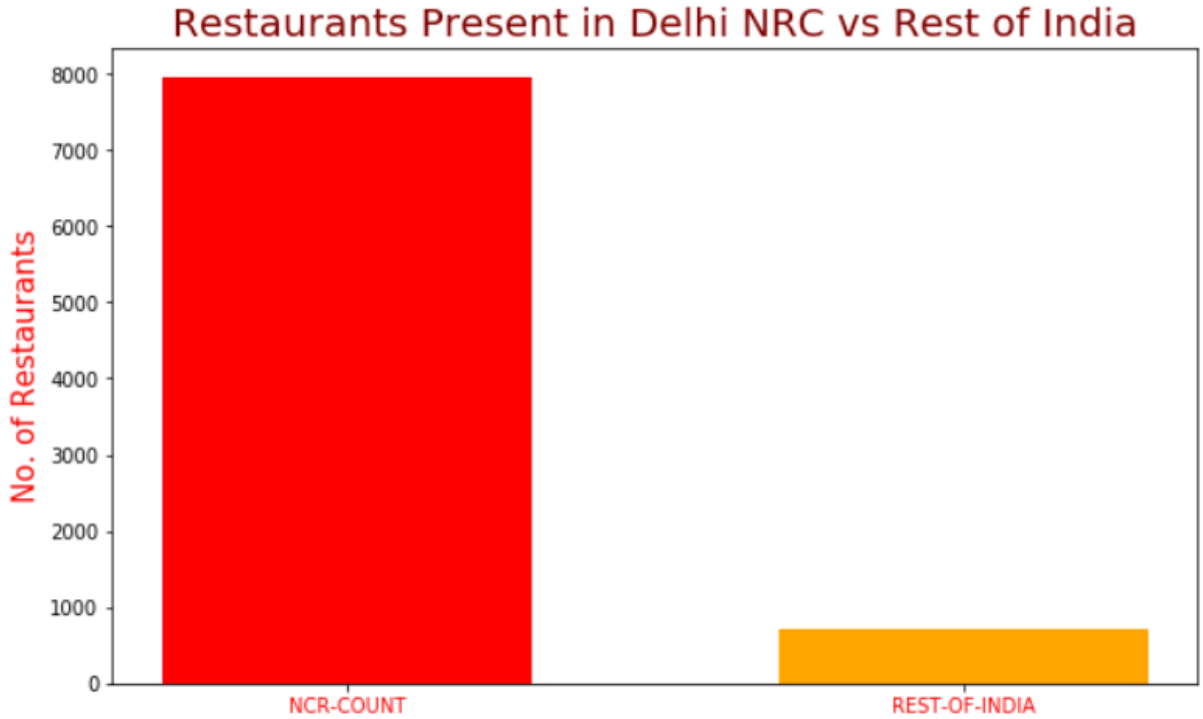
Question 1. The dataset is highly skewed toward the cities included in Delhi-NCR. So, we will summarise all the other cities in Rest of India while those in New Delhi, Ghaziabad, Noida, Gurgaon, Faridabad to Delhi-NCR. Doing this would make our analysis turn toward Delhi-NCR v Rest of India.

1.1 Plot the bar graph of number of restaurants present in Delhi NCR vs Rest of India.

Answer 1.1

```
NRC Count: 7947
Rest Count(India-NRC): 705
```

Plot:



Justification: First we will read our csv file using the read_csv()function of our pandas library. After that we will simply remove all the NaN values. After that I used a variable data_india where I only chose The data that was of my use. Now we have to count total restaurants in Delhi NCR vs rest of India. For that I used a function in which I Passed my dataframe and simply counted umber of restaurants for each Place and added them for Delhi NRC. For rest of india, I subtracted the Delhi NCR restaurant from the total restaurants in India.
Now we can simply plot the required bar graph using the Matplotlib.pyplot function i.e. the bar()function for our required bar graph.

1.2 Find the cuisines which are not present in restaurant of Delhi NCR but prese nt in rest of India.Check using Zomato API whether this cuisines are actually not served in restaurants of Delhi-NCR or just it due to incomplete dataset.

Answer 1.2

```
Cuisines that are NOT in NCR but are in rest of india are as follows :-
German
Malwani
German
BBQ
Cajun

Cuisines which are NOT in Delhi NCR but are in rest in india by dataset
through ZOMATO API are:-
Cajun
German

Justification: First we will create a list with all their cuisines in
the target cities for Delhi NCR. Then we will use a loop and append all
the cuisine names into an empty list(which for now will have many
repeating cusines). Then we will convert that list into a set for
unique values only which will be converted back to a list. Now in a
same way we will pick the cuisines for the cities which for the rest of
India. A new empty list will be created to append all the values of
cuisines into the empty list which are not in our Delhi NCR list using
a loop. Now we can simply print the final list of all the cuisines
which are served in the rest of India and not in Delhi NCR.
Now to confirm this using our Zomato API, we will send a request to
The api link 'https://api.zomato.com/v2.1/cities' and pass the param
'q':'Delhi' and for the headers we will pass the Accept key and its
Value according to the documentation with the user key that we can
Generate from the Zomato api. Now using our response, we can get the
City Id for our city Delhi. Another api get request will be made
To the url 'https://api.zomato.com/v2.1/cuisines' to fetch the
Cuisines for delhi by passing its city id in the params with the
Previously used headers. Using a loop we can store all the cuisine
Names into a list. l=set(REST_CUISINES_NOT_NCR).difference(li) will
Help us to get the cusinies which are not served in delhi ncr using api
, and l can then be printed to get the cuisines.
```

1.3 Find the top 10 cuisines served by maximum number of restaurants in Delhi NCR and rest of India.

Answer 1.3

```
------------------------------------------------------------------------
Top 10 cuisines present in DELHI NCR are as follows
North Indian 3597
Chinese 2448
Fast Food 1866
Mughlai 933
Bakery 697
South Indian 569
Continental 547
Desserts 542
```

```
Street Food 538
Italian 535
--------------------------------------------------------------------------
--------------------------------------------------------------------------
Top 10 cuisines present in REST IN INDIA are as follows
North Indian 349
Chinese 242
Continental 177
Italian 147
Cafe 136
Fast Food 97
South Indian 62
Mughlai 59
Desserts 55
Mexican 50
```

Justification: First lets create a list by merging rows of city and Cuisines from our Indian data usin zip and list function. Then we can Create an empty dictionary and in that using a loop we can see which Cuisines is being served for how many restaurants ii our target cities . Then we will sort the dictionary and using a loop print the first 10 cusines and using the same loop we will append the values of the Cuisine name and the no. of occurences in 2 lists which will help us Plot the graph. Similarly we can find cuisines and their no. of occuren ces for rest of India, sort it and print 10 cuisines. Now using the lis le we created for the cuisines and no. of occurences, we can simply plo t the bar graphs for delhi ncr and outside delhi ncr.

1.4 Write a short detailed analysis of how cuisine served is different from Delhi NCR to Rest of India. Plot the suitable graph to explain your inference.

Answer 1.4

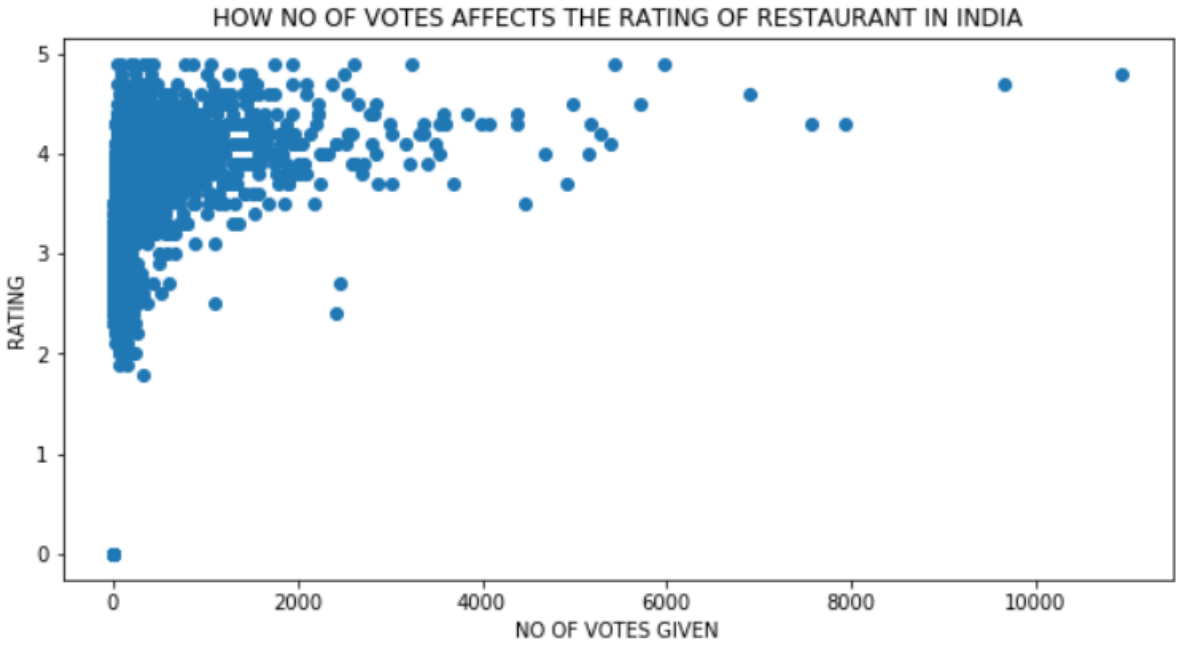
I suppose Question 1.3 is quite the justification for question 1.4 where I have presented the necessary graphs and the top 10 cuisines served in delhi ncr and outer delhi ncr. But taking a common for both the results we can conclude that North Indian, Chinese, Fast Food, Italian, Mughlai, South Indian, Desserts are the common top 10 cuisines. The result could be better for analysis if we had the complete data not only skewed towards Delhi NCR.

Question 2. User Rating of a restaurant plays a crucial role in selecting a restaurant or ordering the food from the restaurant.

2.1 Write a short detail analysis of how the rating is affected by restaurant due following features: Plot a suitable graph to explain your inference.

2.1.1 Number of Votes given Restaurant

Answer 2.1.1

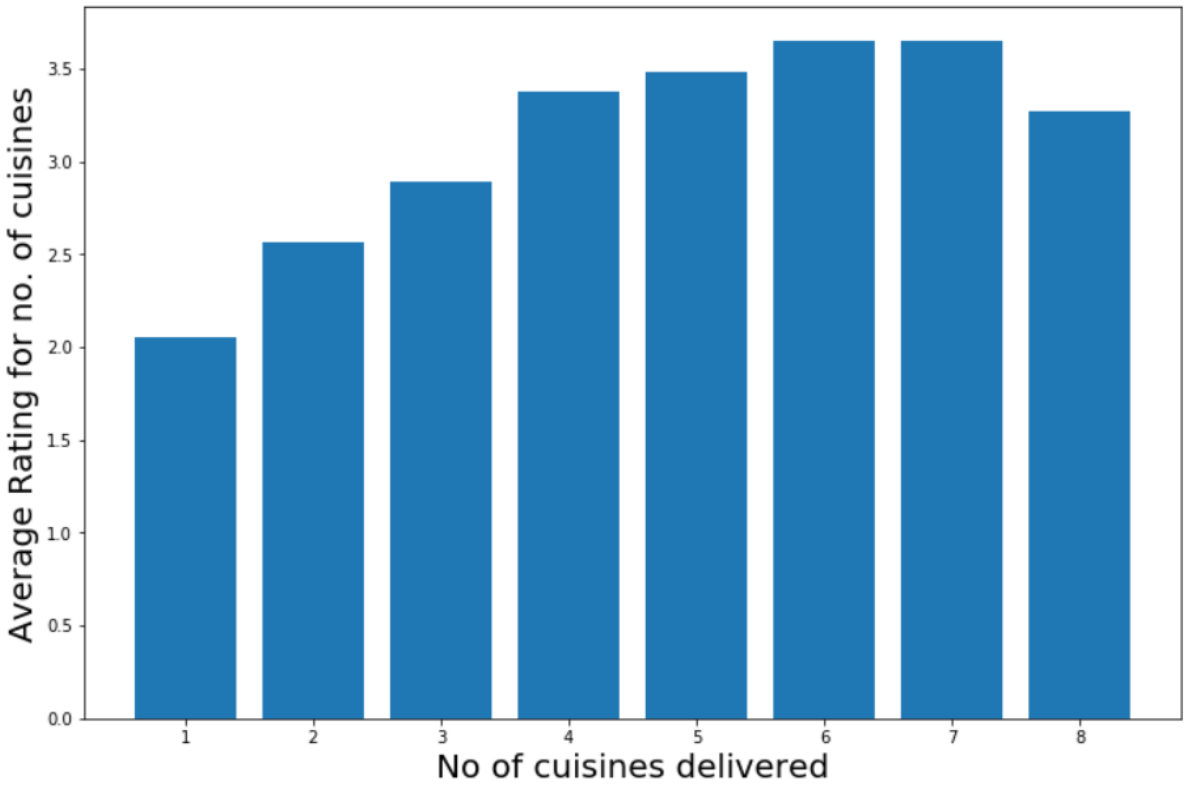


Justification: Checking how the number of votes can actually affect the rating of a restaurant is simple. We can simply pick the rating and the no. of votes to form the graph. Now this graph will actually help us to sort restaurants according to the ratings. The higher the rating, better is the restaurant is supposed to be corresponding to the no. of votes for that restaurant.

2.1.2 Restaurant serving more number of cuisines.

Answer 2.1.2

We can conclude that the restaurants serving 6 or 7 cuisines are more likely to get more rating. While the restaurant serving only 1 cuisine has very low rating.

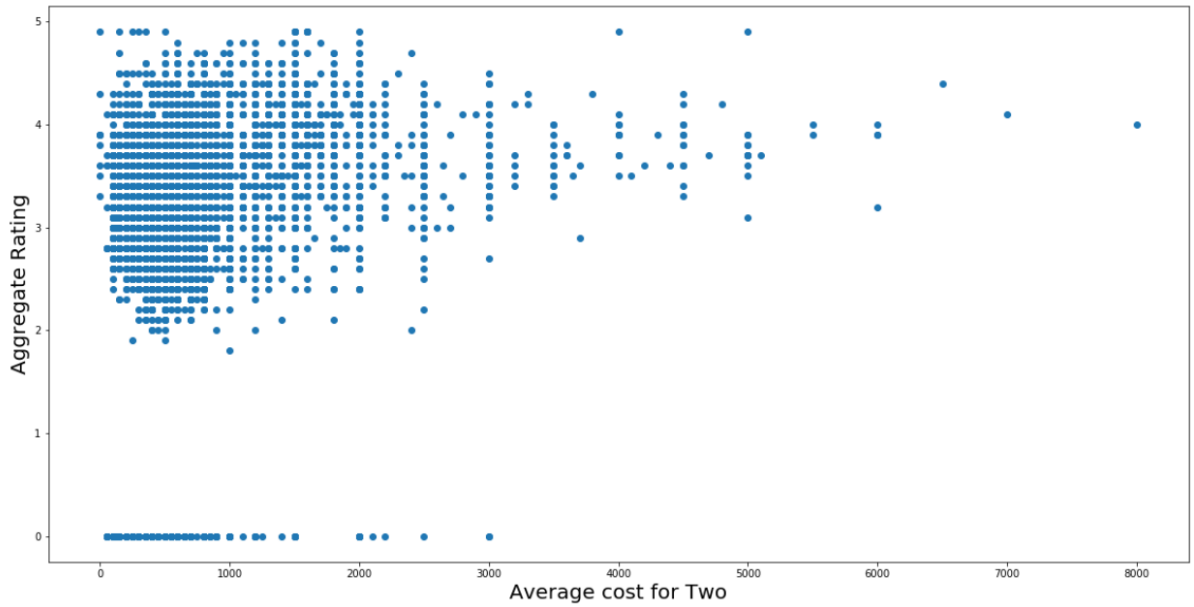


Justification: We will create a new column in our dataframe using the apply method. With the apply method, the Cuisines Column will be passed. Each row in the new column i.e. the count column will have the count of no. of cuisines corresponding to the restaurants using our function f. Now we will create another list combining 2 cols i.e the aggregate rating column and count column. We will create an empty dictionary d, 2 empty lists with names x and y. Using a loop in our dictionary d, we will set keys as the no. of cuisines in restaurants and the values will be lists containing the ratings for each restaurant corresponding to the no. of cuisines. Then using another loop, we will set the values for the keys in our dictionaries to the average of the ratings present in the values-list. Then using another loop we will append all the keys in list x and all the values of dictionary d in y. Now we can simply plot the bar graph, where x axis will represent the no. of cuisines and the y axis will represent the average rating corresponding to those no. of cuisines.

2.1.3 Average Cost of Restaurant

Answer 2.1.3

We can conclude from the graph that lower the Average cost of food, the higher is the rating of the restaurant supposed to be.

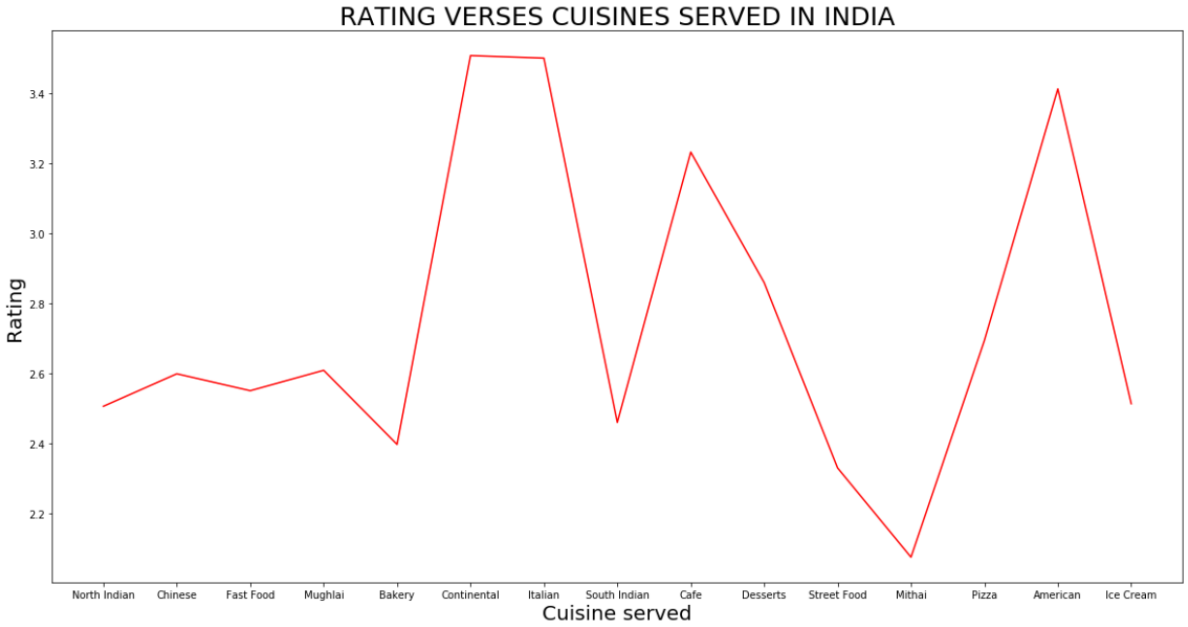


Justification: we can create a list using 2 rows for each element of list in form of a tuple. Then we will plot a graph of Average cost for two on x axis against the Aggregate Rating.

2.1.4 Restaurant serving some specific cuisines

Answer 2.1.4

We can see that Continental Cuisine has the maximum rating while Mithai has the lowest.



Justification: As done earlier also, we will first find the top 15 cuisines served in most restaurants of India and print each one of them. Then for each cuisine we will take the average of ratings on similar lines as done before. And in the end, print cusines and the average rating corresponding to them.

2.2 Find the weighted restaurant rating of each locality and find out the top 10 localities with more weighted restaurant rating?

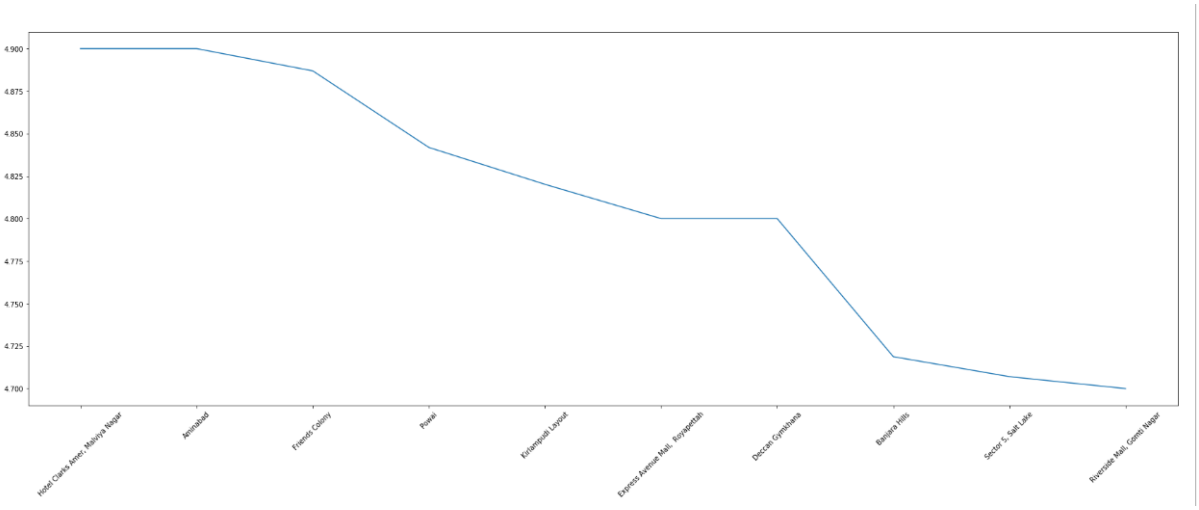
2.2.1 Weighted Restaurant Rating=Σ (number of votes * rating) / Σ (number of votes)

Answer 2.2.1

```
--------------------------------------------------------------------------
                          TOP 10 LOCALITIES HAVING BEST RATING

Hotel Clarks Amer, Malviya Nagar        |            4.90
Aminabad                                |            4.90
Friends Colony                          |            4.89
Powai                                   |            4.84
Kirlampudi Layout                       |            4.82
Express Avenue Mall,  Royapettah        |            4.80
Deccan Gymkhana                         |            4.80
Banjara Hills                           |            4.72
Sector 5, Salt Lake                     |            4.71
Riverside Mall, Gomti Nagar             |            4.70
```

We can see that Hotel Clarks Amer, Malviya Nagar locality has the Maximum rating whereas Riverside Mall, Gomti Nagar is lowest in the List.

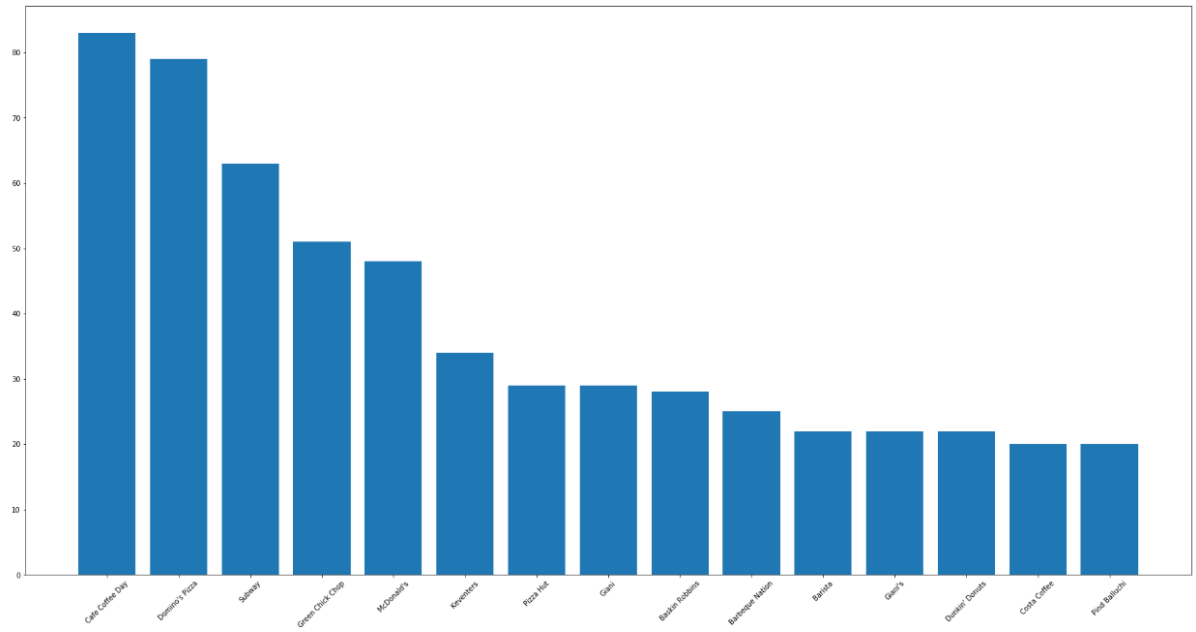


Justification: First we will create a column by the name weighted_average with values as votes Multiplied by aggregate rating for each row. Now we will create another list comprising two cols i.e. locality and weighted_average. Now we will find the sum of the column Votes. We will create a list ind and it will have the names of all the localities. Now using a loop in an empty dictionary we will set the key s as the names of localitites and the values will be the weighted avera ge according to the formula given. Then we will sort our dictionary and Print the top 10 localities having best rating and then plot the graph.

Question 3. Visualization

3.1 Plot the bar graph top 15 restaurants have a maximum number of outlets.

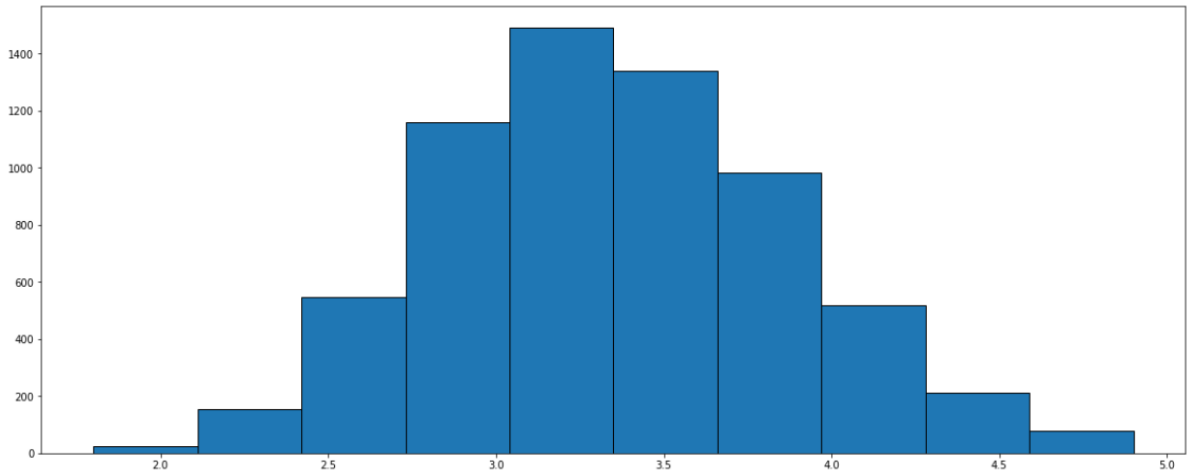
Answer:



Justification: We will take the Restaurant Name col and check the occurences for each restaurant using our value_counts() function and then the plotting has to be done only for top 15 ones.

3.2 Plot the histogram of aggregate rating of restaurant(drop the unrated restaurant).

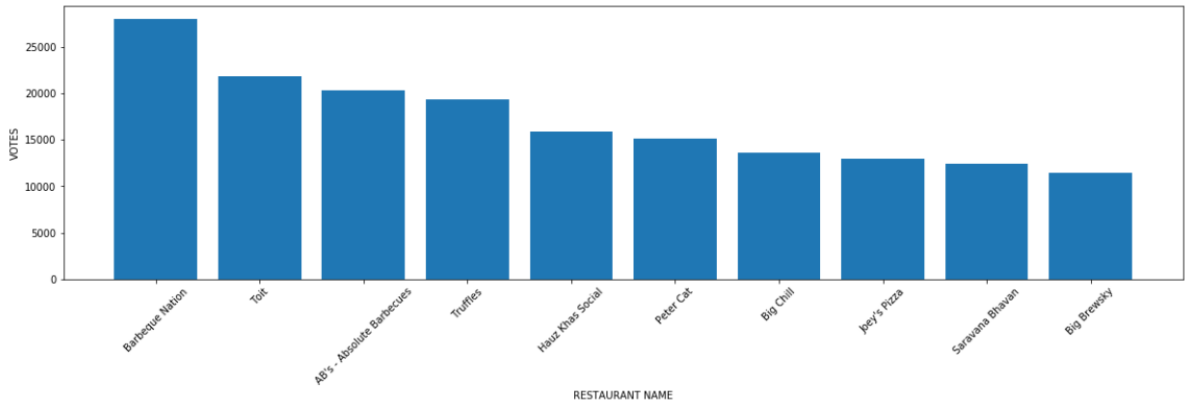
Answer 3.2



Justification: After not selecting the Cols with Aggregated restaurants which are unrated, we can create a histogram for the rating and its no.

3.3 Plot the bar graph top 10 restaurants in the data with the highest number of votes.

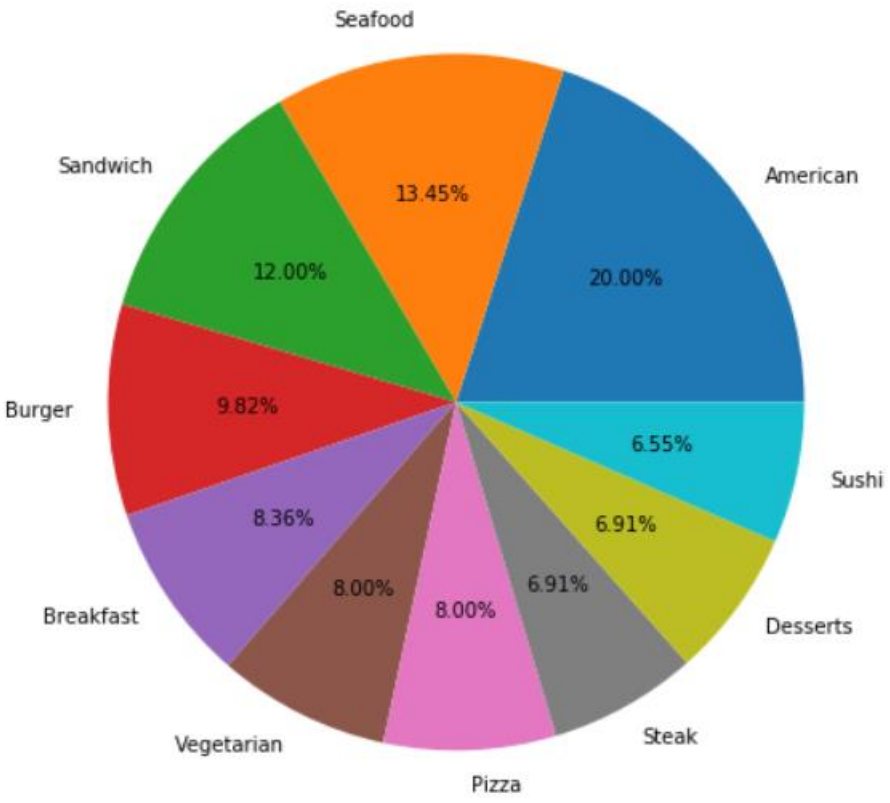
Answer 3.3:



Justification: We will create a dictionary which will have the total number of votes for each restaurant including all its outlets. Then after sorting we will plot a bar graph for the no. of votes.

3.4 Plot the pie graph of top 10 cuisines present in restaurants in the USA.

Answer 3.4:



Justification: First we will select the data for USA using its country code. Then using loops on the same lines as done before we will find the occurences o each cuisine and then plot a pie graph according to the percentage ratio of each cuisine.

3.5 Plot the bubble graph of a number of Restaurants present in the city of India and keeping the weighted restaurant rating of the city in a bubble.

Answer 3.5: