PROJECT

ZOMATO API

(PART – II)

Plots and Explanation

By :

PAWAN KR. MISHRA

Question 1:

The dataset is highly skewed toward the cities included in Delhi-NCR. So, we will summarize 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.

a) **Plot the bar graph of number of restaurants present in Delhi NCR vs Rest of India.**

Explanation:

Given information:

It is provided that the dataset is highly skewed toward the cities included in Delhi-NCR. So, we will summarize 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.

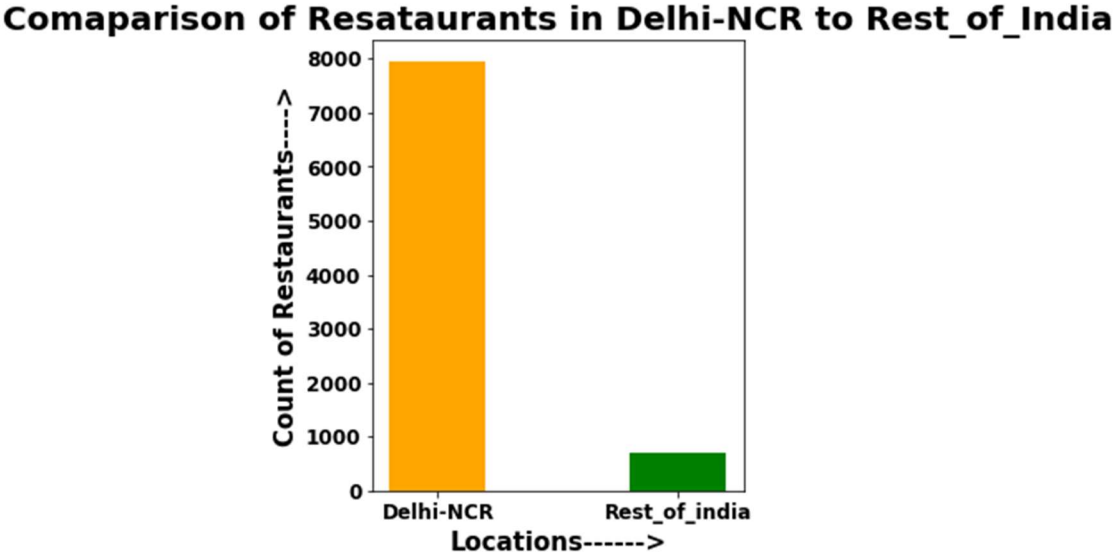
Steps performed to get the count of Restaurants at both location:

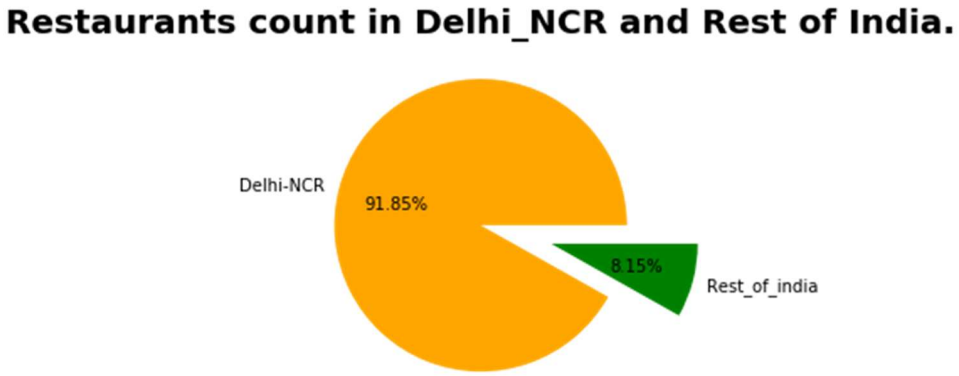
- For getting the counts of restaurants in both the locations I have created the city array for the cities of India.
- I have iterated over the city array and marked the cities New Delhi, Ghaziabad, Noida, Gurgaon, Faridabad to Delhi-NCR, And the remaining cities under Rest of India.
- Since each row gives details of unique restaurant so while iterating just increasing the count by 1.
- After the iteration is over we got the locations with number of restaurants present there using the information to plot graph.
- I have plotted the bar graph and pie chart based upon the information gathered.

Tables and Graphs:

The location with count of restaurants are:

Location	Count of Restaurants.
Delhi-NCR	7947
Rest_of_India	705





From the above representation it is clear that Delhi-NCR location has more number of restaurants as compared to rest of India.

b) Find the cuisines which are not present in restaurant of Delhi NCR but present 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.

Explanation:

Given Information:

We need to find out the Cuisines which are present in rest of India but not in Delhi-NCR location and by using Zomato API we need to find out whether they are not served or it is due to incomplete dataset.

Steps performed in finding the Cuisines present in rest of India but not in Delhi-NCR:

- I have created the city and cuisines array to get the cuisines present in restaurants of each city.
- I have iterated over the city array and marked the cities New Delhi, Ghaziabad, Noida, Gurgaon, Faridabad to Delhi-NCR, And the remaining cities under Rest of India.
- I have created the set to store the Cuisines served in Delhi-NCR and for the rest of India respectively.
- While iterating over the cuisines array there are cases where multiple cuisines are present for particular restaurant, then in that condition I have split the names.
- After getting the set of cuisines I will subtract the cuisines set of Delhi-NCR with the cuisines set of rest of India to get the set of cuisines that are not present in Delhi-NCR.
- From the API we have collected all the cuisines that are served in Delhi-NCR in a set, then I have subtracted this set from the set of cuisines collected using the data frame after which I have found that two cuisines we not present in the data frame.

Tables and Graphs:

Cuisines not in Delhi are
German
BBQ
Malwani
Cajun

Missing Cuisines are
Malwani
BBQ

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

Explanation:

Given Information:

We need to collect the top 10 cuisines that are served in Delhi-NCR and rest of India.

Steps performed to get top 10 cuisines are:

- I have created the city and cuisines array to get the cuisines present in restaurants of each city.
- I have iterated over the city array and marked the cities New Delhi, Ghaziabad, Noida, Gurgaon, Faridabad to Delhi-NCR, And the remaining cities under Rest of India.
- I have created 2 different dictionaries to store the cuisines and the count of restaurants serving those cuisines as key value pair for both in Delhi-NCR and rest of India.
- While iterating over the cuisines array there are cases where multiple cuisines are present for particular restaurant, then in that condition I have split the names.
- After completing the iteration, we got the dictionary of cuisines for both Delhi-NCR and Rest of India, now I have sorted the dictionary based upon values in descending order to get the top 10 cuisines.

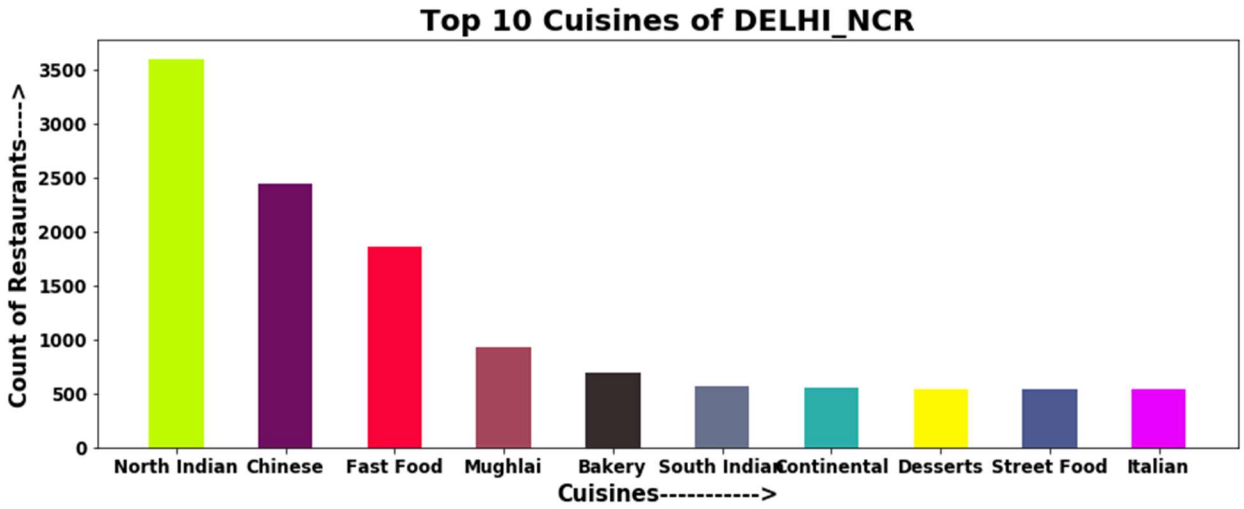
Tables and Graphs:

Top 10 Cuisines in Delhi-NCR are:

Cuisines	Count of Restaurants.
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 in Rest-of-India are:

Cuisines	Count of Restaurants.
North Indian	349
Chinese	242
Continental	177
Italian	147
Cafe	136
Fast Food	97
South Indian	62
Mughlai	59
Desserts	55
Mexican	50



1.) The first bar graph presents the top 10 cuisines that are served by restaurants of Delhi-NCR.

2.) The second bar graph presents the top 10 cuisines that are served by restaurants in India other than excluding Delhi-NCR region.

3.) From the graph it is clear that North Indian cuisines and Chinese are the cuisines that are served by maximum restaurants of region including and excluding Delhi-NCR (North Indian preferred more over Chinese).

4.) There are some cuisines that are common in both the graphs (Mughlai, Continental, Fast Food, South Indian, Desserts, Italian) but they differ in their counts of restaurants serving them.

5.) There are some cuisines that are new in both graphs (Bakery, Mexican).

d) 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.

Explanation:

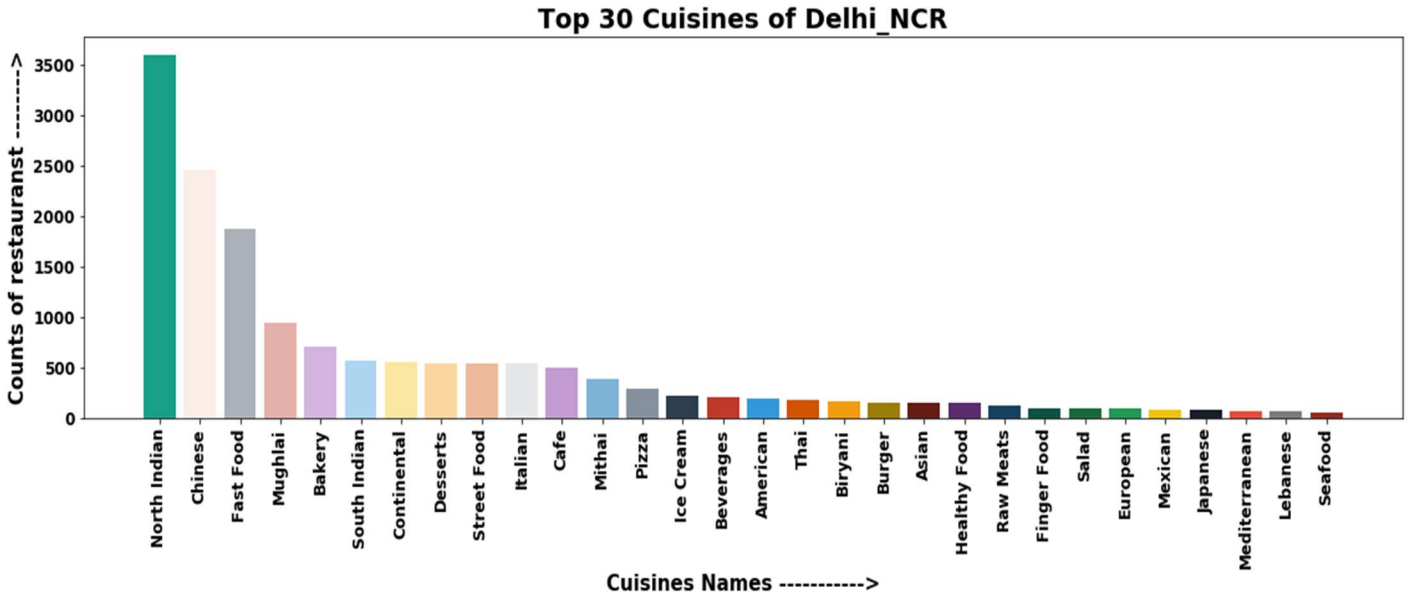
Given Information:

We need to find the difference between how cuisines served in Delhi-NCR to cuisines served in Rest of India.

Steps performed to get the difference of cuisines served:

- I have created the city and cuisines array to get the cuisines present in restaurants of each city.
- I have iterated over the city array and marked the cities New Delhi, Ghaziabad, Noida, Gurgaon, Faridabad to Delhi-NCR, And the remaining cities under Rest of India.
- I have created 2 different dictionaries to store the cuisines and the count of restaurants serving those cuisines as key value pair for both in Delhi-NCR and rest of India.
- While iterating over the cuisines array there are cases where multiple cuisines are present for particular restaurant, then in that condition I have split the names.
- After completing the iteration, we got the dictionaries of cuisines for both Delhi-NCR and Rest of India, now I have sorted the dictionaries based upon values in descending order to get the top 30 cuisines.
- On the basis of data collected I have provided the analysis and drawn graphs to support my inferences.

Tables and Graphs:



Analysis:

Here I have plotted the graphs of top 30 widely served cuisines that are served in Delhi-NCR and Rest of India, from graph it is clear that North Indian cuisine is the cuisine that is served by maximum number of restaurants of India (In Delhi-NCR=3597, Rest of India=349). After that Chinese cuisine is widely served (In Delhi-NCR=2448, Rest of India=242). After these two we have fast food as widely served (In Delhi-NCR=1866, Rest of India=97). After this we have Mughlai cuisine (In Delhi-NCR=933, Rest of India=59), and there are many more cuisines as show in the graph.

Question 2:

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

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.

a. Number of Votes given Restaurant .

Explanation:

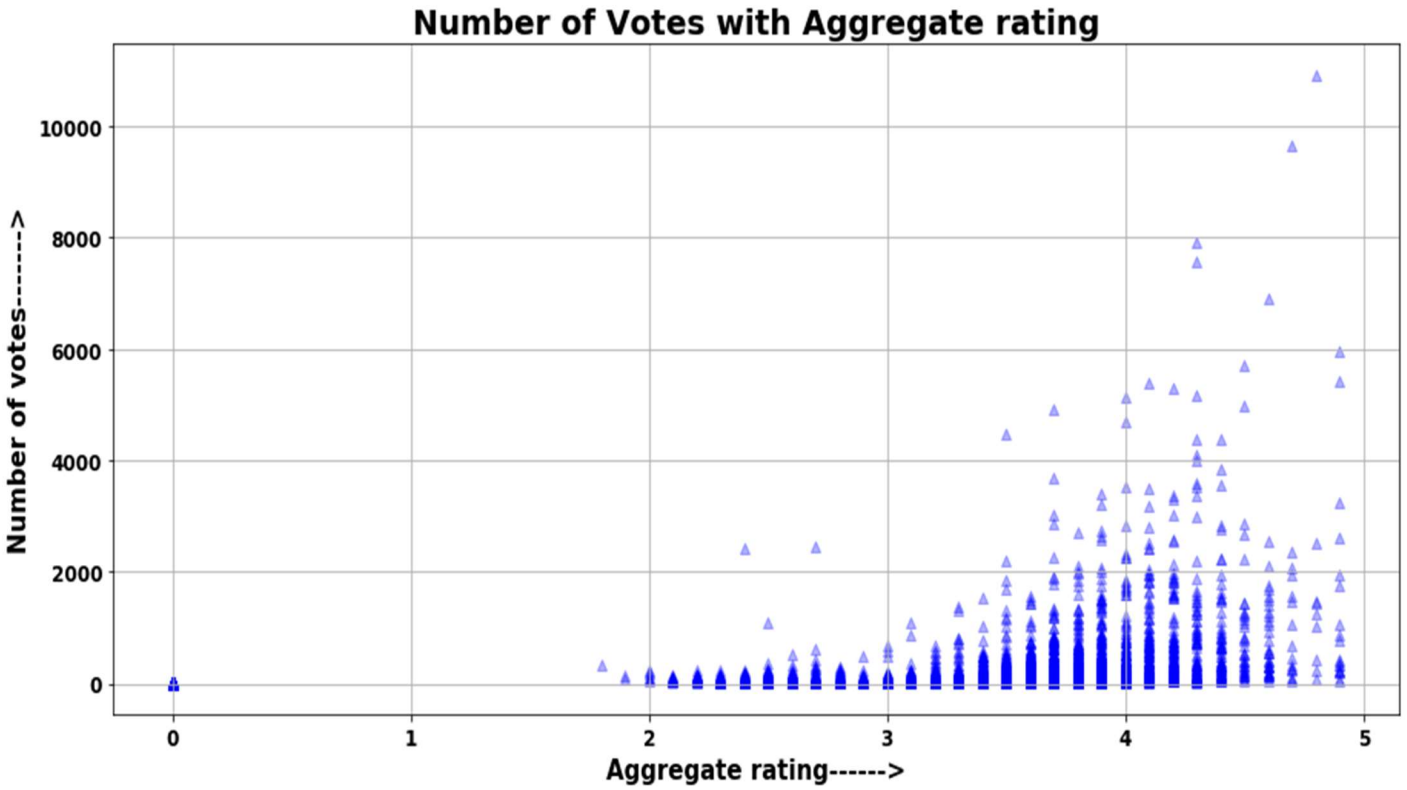
Given Information:

We have to analyze how rating gets affected to due to the number of votes given to a restaurant.

Steps performed to Analyze the relation between rating and votes of restaurants:

- I have analyzed the data for the restaurants of India.
- I have created the rating array and votes array from the Aggregate Rating and Votes column of data frame.
- I have plotted the scatter graph to Analyze the relation between the rating and votes to a restaurant.
- The analysis I have drawn from graph is mentioned below.

Tables and Graphs:



Analysis:

Clearly as the number of votes increases the chances of getting higher ratings also increases. It is clear from the graph that any restaurants with more than 5000 votes get ratings between 4 and 5. when the votes are less than 2000, there is a diversity of ratings, which varies from 2 to 5. So for a restaurant to have a probability of getting higher aggregate rating, it should have more than 5000 votes. Also there are many restaurants having votes between 1000 and 2000 with very good ratings, that is between 4 and 5.

b. Restaurant serving more number of cuisines.

Explanation:

Given Information:

We have to analyze how rating gets affected to due to the number of cuisines served by the restaurants.

Steps performed to Analyze the relation between rating and votes of restaurants:

- The analysis is done only for the restaurants present in India.
- To get the count of number of cuisines served by the restaurants we need to know the number of cuisines served by each restaurant, for which I have added the cuisine count column in the selected data frame and have created an array for the same.
- I have created dictionary to store the count of restaurants of restaurants particular number of cuisines.
- I have created an array to store the rating and plotted the graph between number of cuisines served and what is there rating.
- Below mentioned is the analysis I have drawn from the graph.

Tables and Graphs:

Number of restaurants serving the provided number of cuisines.:

Number of Cuisines served	No. of restaurants serving given number of cuisines
2	3144
1	3097
3	1596
4	552
5	155
6	70
7	26
8	12

Analysis:

From the graph itself, it is clear that when the number of cuisines provided increases from 3 to 8, generally the rating seems to converge between 3 and 4. restaurants providing a greater number of cuisines are not much likely to get higher ratings, especially when the number of cuisines provided exceeds 6. while if the restaurants provide a smaller number of cuisines, for example 1 or 2 cuisines, they are more likely to get higher ratings. In fact, many restaurants in India are providing 2 cuisines. there is a very a smaller number of restaurants providing 8 cuisines. it seems like when a restaurant provides too many cuisines, its focus on the quality of food offered diverges. while restaurants providing less cuisines focus on the quality of food to get good aggregate ratings.

c. Average Cost of Restaurant.

Explanation:

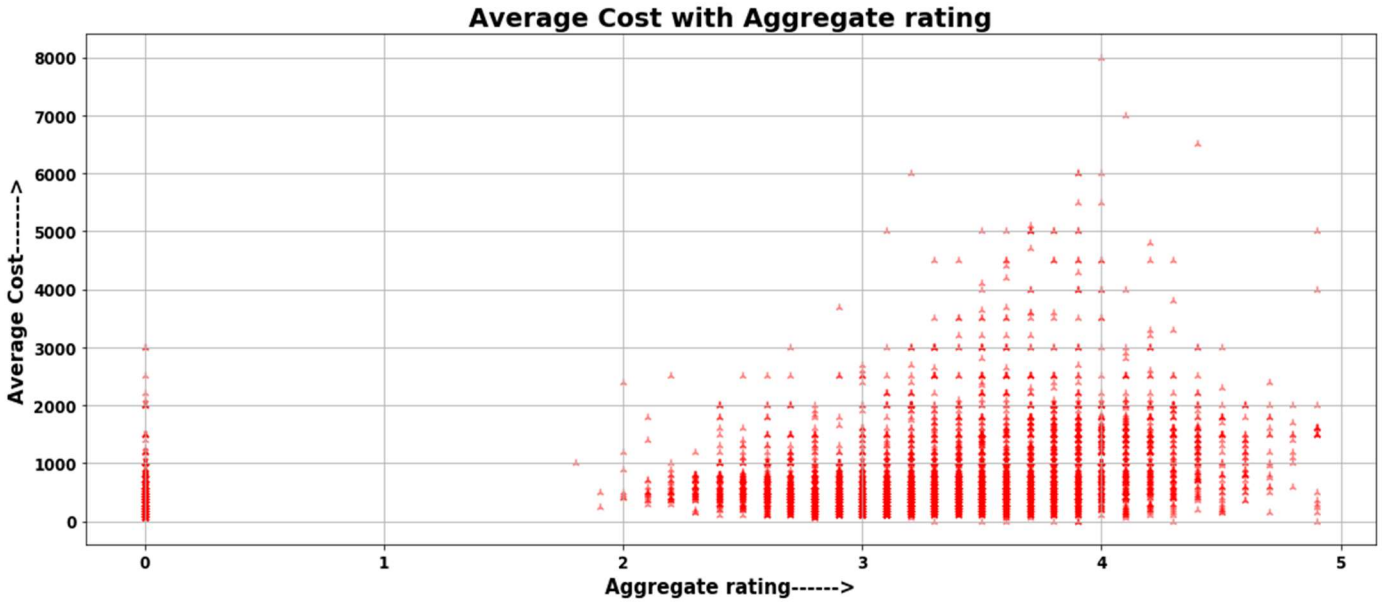
Given Information:

We have to analyze how rating gets affected to due to the Average Cost for two people in a restaurant.

Steps performed to Analyze the relation between rating and Average Cost for two:

- This analysis is for the restaurants of India.
- To get the graph between the Average Cost and rating I have created array for both of them.
- I have plotted the graph based upon the data provided.
- Below mentioned is the analysis I have extracted from the graph.

Tables and Graphs:



Analysis:

As the average cost goes in between 2000, to 6000, it seems like the food is not worth the money because in that case the average rating is generally between 3 and 4. whereas when the average cost is between 1000 to 2000, the rating varies from 3 to 4.5 mostly. But for the higher price ranges the rating is between 3 and 4

d. Restaurant serving some specific cuisines.

Explanation:

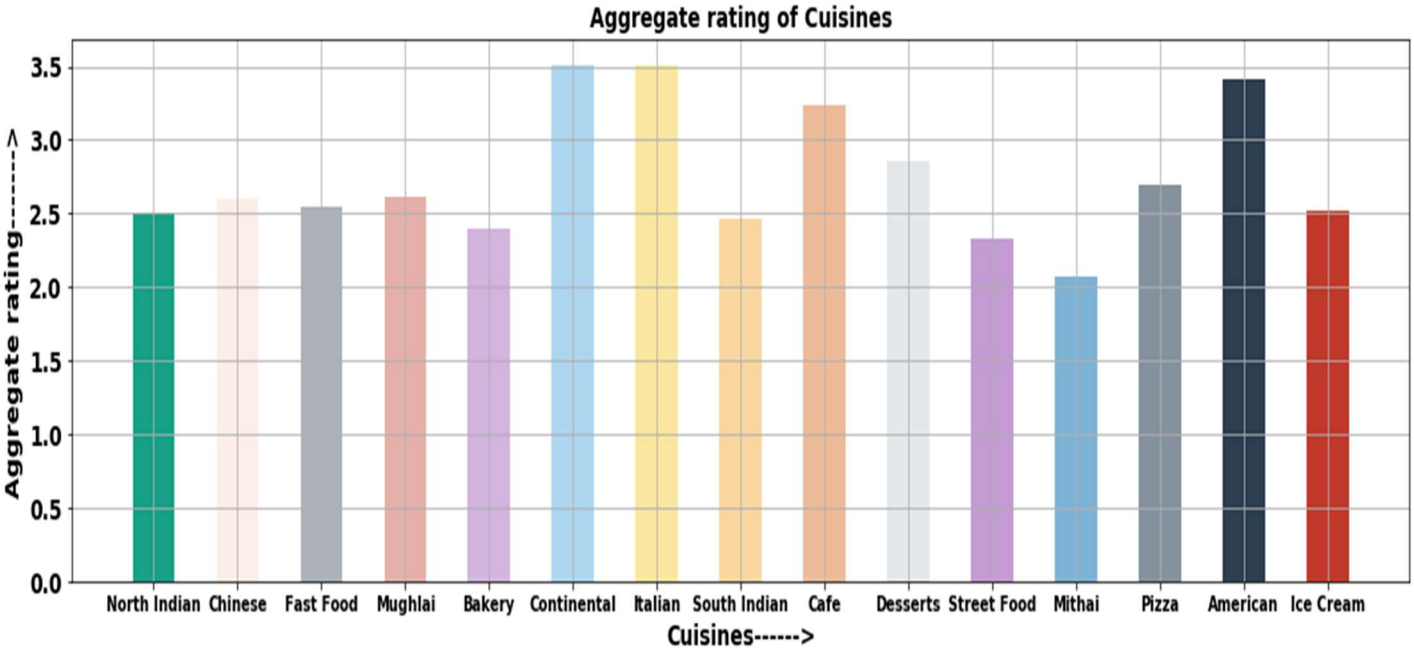
Given Information:

We have to analyze how rating gets affected to due to some specific cuisines served by the restaurants.

Steps performed to Analyze the relation between rating and specific number of cuisines served by the restaurants:

- The analysis is done for the country India.
- Here I have collected the top 15 cuisines for their analysis.
- I have created the average rating array to store the average rating for each of the cuisine by adding rating of all restaurants serving it and then dividing the sum by count of restaurants serving it.
- I have plotted the bar graph between cuisines and average rating.
- Below mentioned is the analysis I have drawn from it.

Tables and Graphs:



Analysis:

These are top 15 cuisines that served maximum number of times all the restaurant. Calculated average rating for top cuisines and most of the cuisines having good average rating and Italian cuisines is at top. From above graph we can see restaurant serving Cuisine 'Italian','Continental','American' have more average rating compared to other cuisine

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

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

Explanation:

Given Information:

We have to find the weighted restaurant rating for each locality and find the top 10 localities with more weighted restaurant rating.

we have to calculate the average rating by using the below mentioned formula i.e.

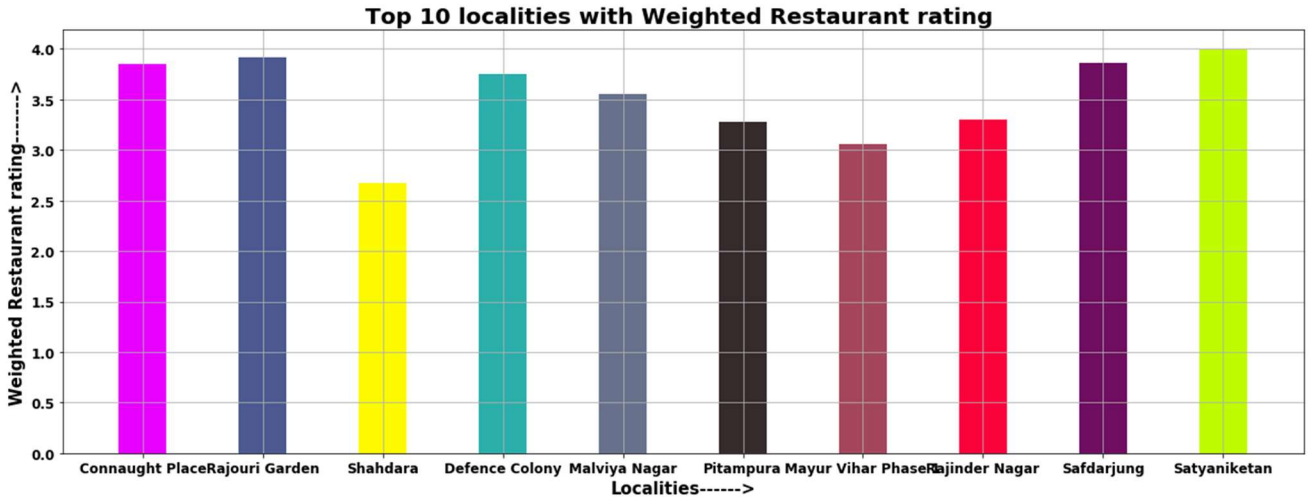
Weighted Restaurant Rating=Σ (number of votes * rating) / Σ (number of votes)

Steps performed to calculate the weighted restaurant rating for each locality:

- I have created a list of tuples where each tuple contains the locality, votes and aggregate rating of each row as its element.
- After that I have collected the top 50 localities which are having maximum number of restaurants and stored it in array named locality.
- After that I have iterate over the locality array and calculated the weighted restaurant rating with the help of list created in step 1.
- After that I have plotted the graph for top 10 localities with maximum aggregate rating.

Tables and Graphs:

Locations
Connaught Place
Rajouri Garden
Shahdara
Defence Colony
Malviya Nagar
Pitampura
Mayur Vihar Phase 1
Rajinder Nagar
Safdarjung
Satyaniketan



Question 3:

Visualization

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

Explanation:

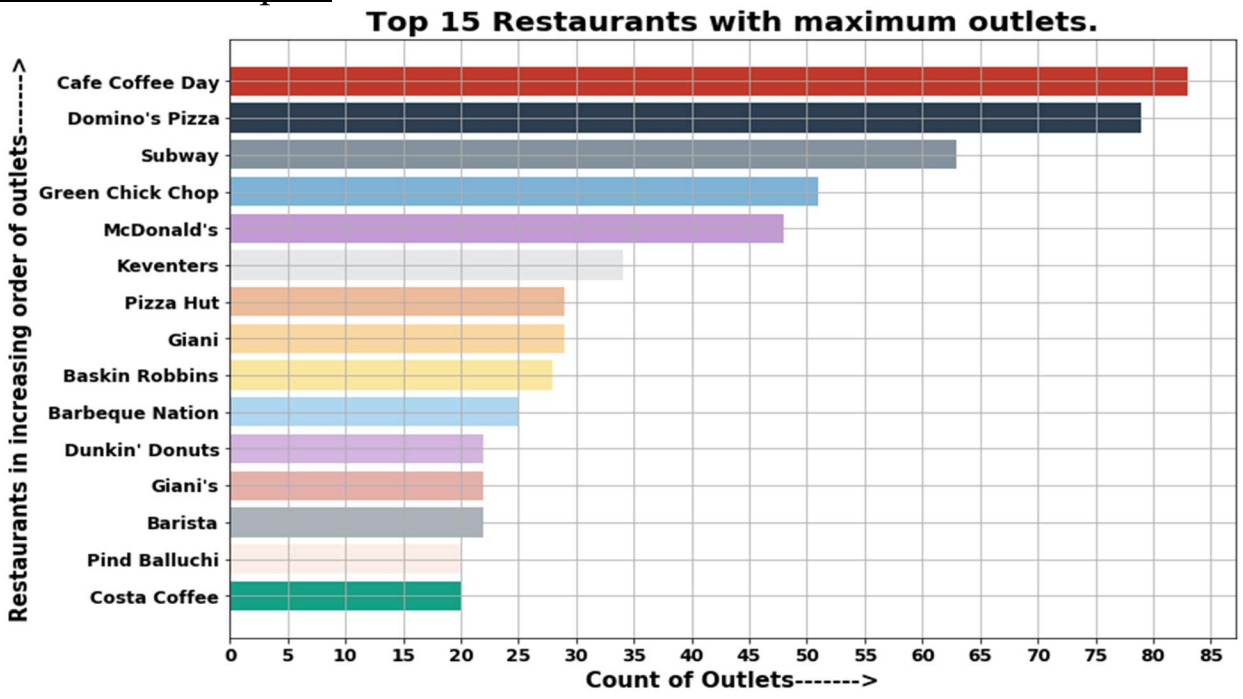
Given Information:

Here we are asked to plot the bar graph of top 15 restaurants which have maximum number of outlets.

Steps performed in extracting the list of top 15 restaurants which have maximum number of outlets:

- With the help of Restaurant name column, I have created the restaurant array with the purpose to iterate over it and count the number of outlets for each restaurant.
- I have created dictionary to store the particular restaurant and its outlets as key value pair.
- I have sorted the dictionary based upon value so that we can easily get the name of our required 15 restaurants.
- I have plotted the graph of top 15 restaurants and its outlets count.

Tables and Graphs:



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

Explanation:

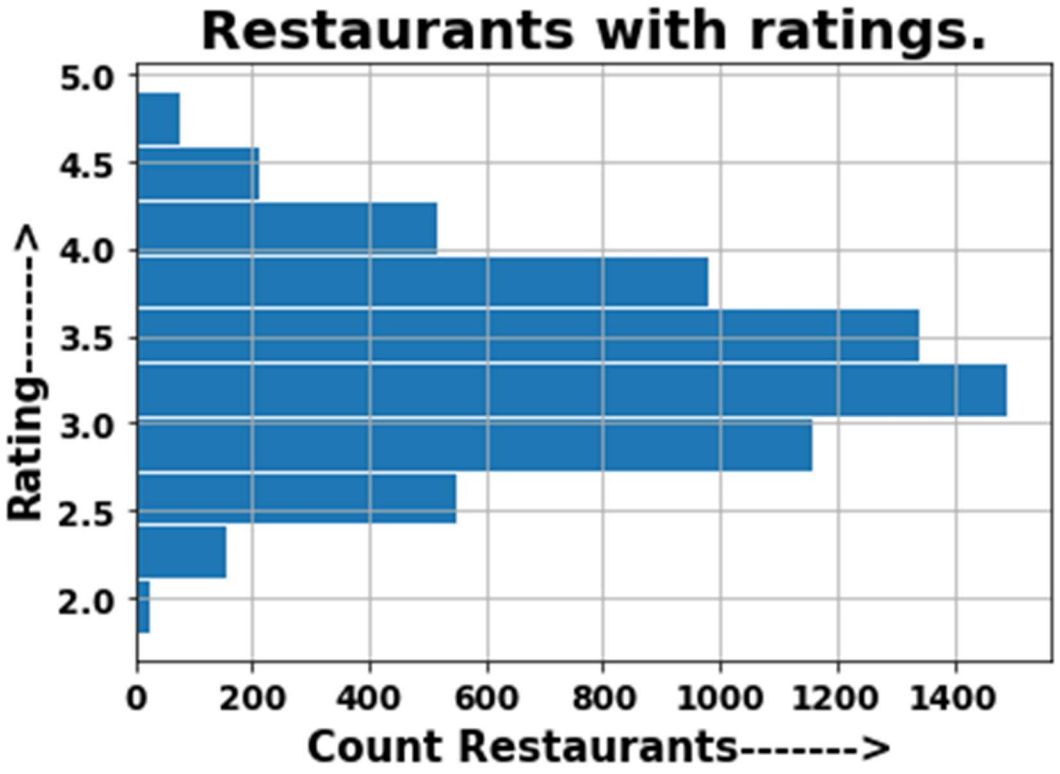
Given Information:

Here we are asked to plot the histogram for the aggregate rating of restaurants by removing the unrated restaurants.

Steps performed to plot the histogram of aggregate rating of restaurants:

- I have created the rating array using the Aggregate rating column of the data frame by dropping the value of restaurants that are unrated.
- Plotting the histogram using the rating array.

Tables and Graphs:



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

Explanation:

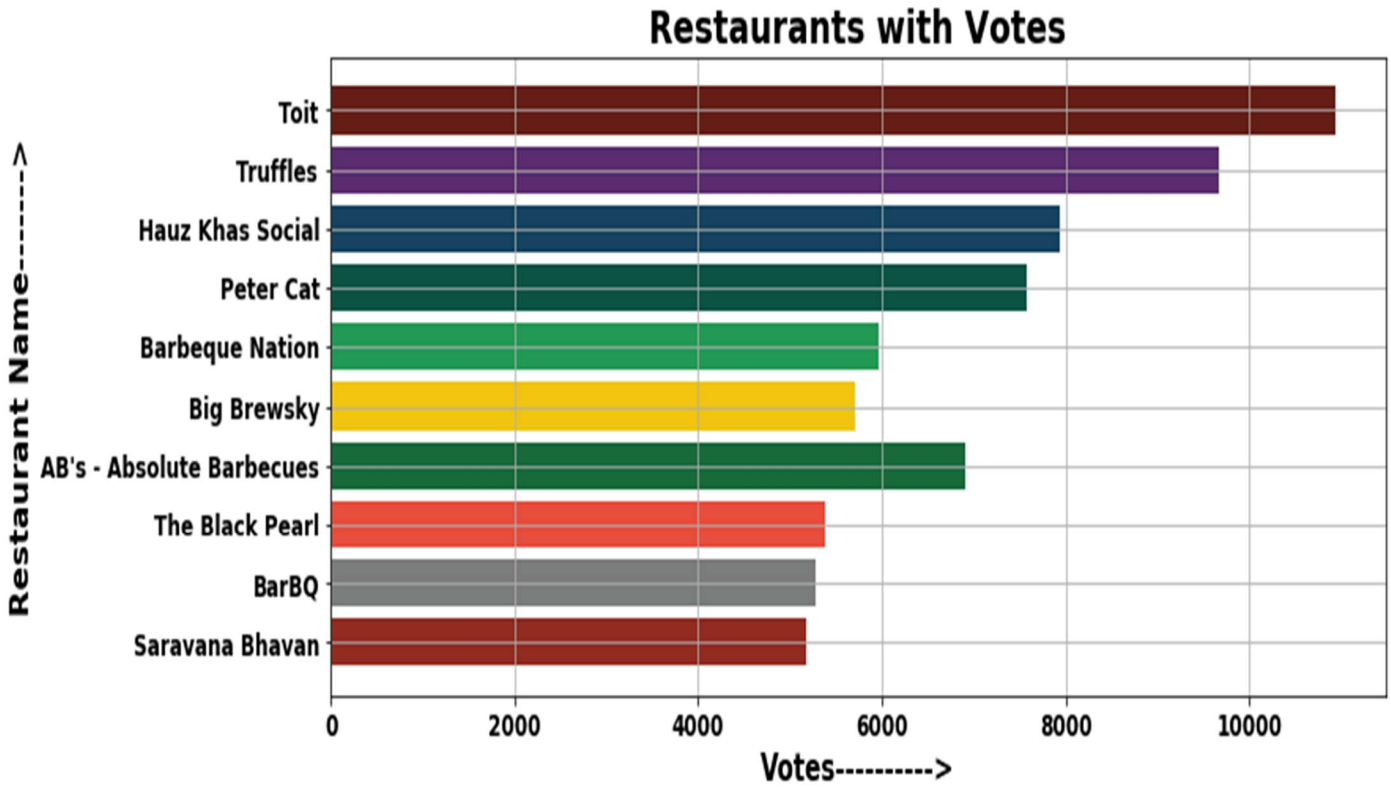
Given Information:

Here we are asked to plot the bar graph of top 10 restaurants with maximum number of votes.

Steps performed to extract the top 10 restaurants with maximum number of votes:

- I have sorted the data frame based upon the values of votes column in descending order.
- Now iterating over the values of data frame and storing the name of restaurant array and its corresponding vote in vote array.

Tables and Graphs:



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

Explanation:

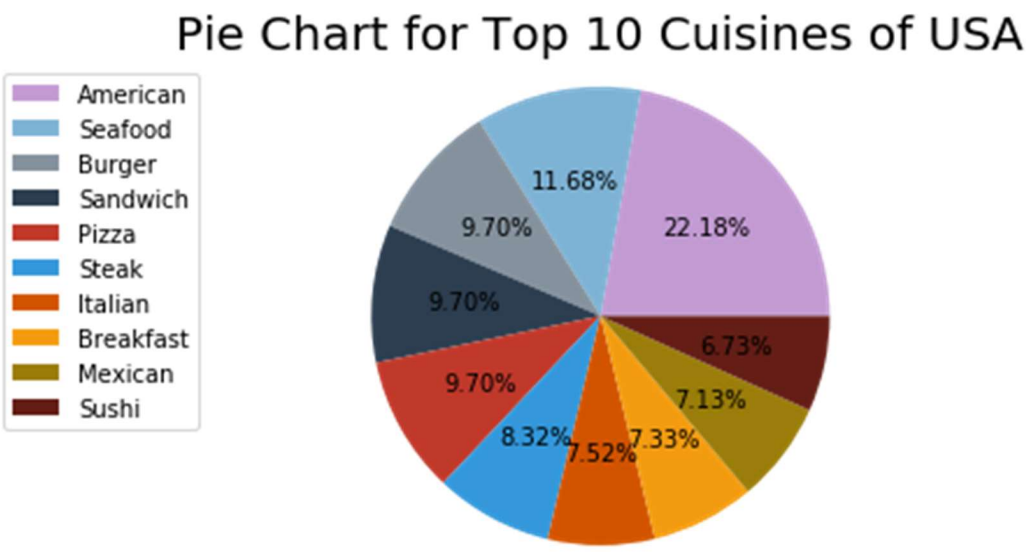
Given Information:

Here we are asked to plot the bar graph of top 10 restaurants with maximum number of votes.

Steps performed to extract the top 10 Cuisines present in restaurants of USA.

- By using the country code of USA, I have created the copy of data frame.
- I have dropped the null entries present in cuisines table.
- I have created the dictionary to store the cuisines and their counts as key value pair.
- After sorting the dictionary, we, extract the top 10 cuisines from it, using which I have plotted the pie chart.

Tables and Graphs:



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**.

Explanation:

Given Information:

We have to create the bubble graph for the restaurants present in Indian cities with weighted restaurant rating for each city, by using below mentioned formula.

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

Steps performed to bubble graph for the restaurants present in Indian cities:

- I have created the arrays to store the city and count of restaurants in that city.
- While iterating over the city array I have created the copy of data frame for each particular city.
- In the copied data frame of each city I have added new column with name rate which stores the product of aggregate rating and vote for each row.
- Using the sum method to calculate the total rate and the total count of votes from rate and vote column.
- Appending the weighted restaurant rating in the array named weighted restaurant for each city.
- After collecting the weighted restaurant rating for each city I have plotted the bubble graph for each city.

Tables and Graphs: