

Zomato API Case Study

Ques1)

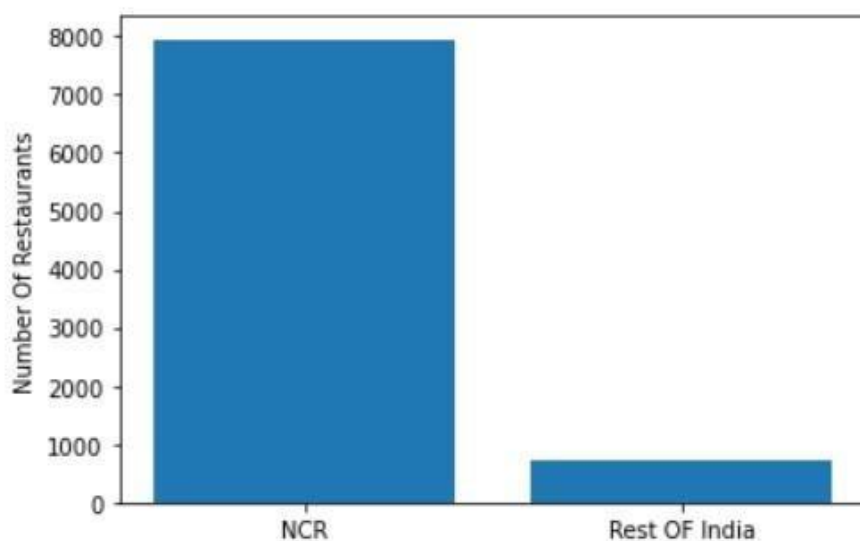
The dataset is highly skewed toward the cities included in Delhi-NCR. So, we will summarise all the other cities in Rest of India while those in New Delhi, Ghaziabad, Noida, Gurgaon, Faridabad to Delhi-NCR. Doing this would make our analysis turn toward Delhi-NCR v Rest of India.

1. Plot the bar graph of number of restaurants present in Delhi NCR vs Rest of India.
2. 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.
3. Find the top 10 cuisines served by maximum number of restaurants in Delhi NCR and rest of India.
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.

a)

Answer: **Explanation:** The code first imports all the libraires and then opens the data using pandas and filters the data for Country India with code 1 and then filters it for cities present in NCR. It then calculates the length of whole data from India and subtracts it from the data from NCR. Then it plots the bar graph for the same.

Figure:



b)

Answer) {'BBQ', 'German', 'Malwani', 'Cajun'}

Explanation: The code first imports all the necessary libraires and then it filters the data according to Country Code with India and then the cities in NCR and Rest of India. Then it applies function to each of the cuisines present in the restaurant and stores it in the two sets, one for NCR and other for Rest of India. Then the resultant set set (NCR) – set (Rest of India) is the required ans.

When I examined the Zomato database, some of the dishes are not present in this dataset as it not the updated one and has errors in it.

Using the Zomato API these results were found:

{'Malwani', 'BBQ', 'Cajun', 'German'}

c)

Answer)

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

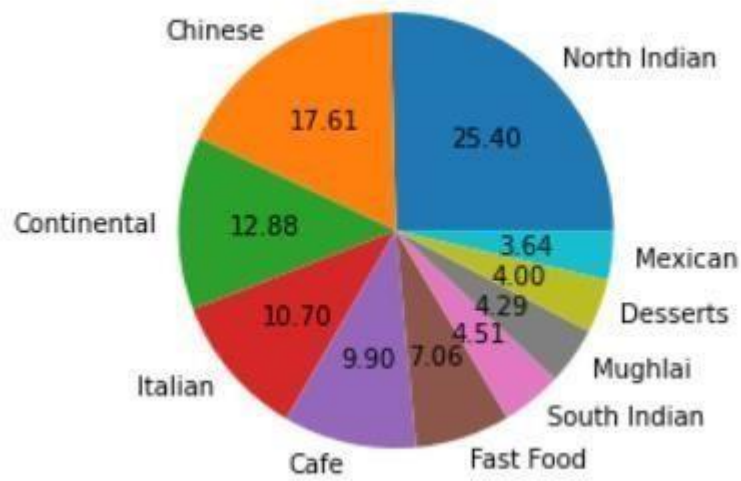
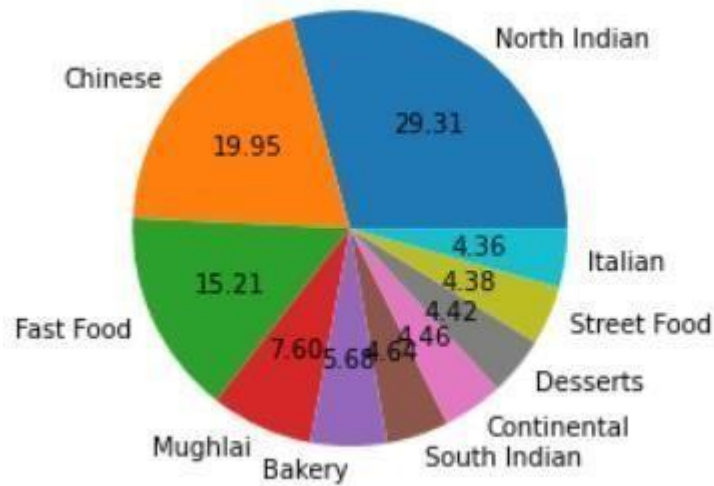
-----FOR REST CITIES-----

Name of Cuisine	Count
1. North Indian	349
2. Chinese	242
3. Continental	177
4. Italian	147
5. Cafe	136
6. Fast Food	97
7. South Indian	62
8. Mughlai	59
9. Desserts	55
10. Mexican	50

-----Delhi NCR-----

Name of Cuisine	Count
1. North Indian	3597
2. Chinese	2448
3. Fast Food	1866
4. Mughlai	933
5. Bakery	697
6. South Indian	569
7. Continental	547
8. Desserts	542
9. Street Food	538
10. Italian	535

Figure for NCR and Figure for Rest of India

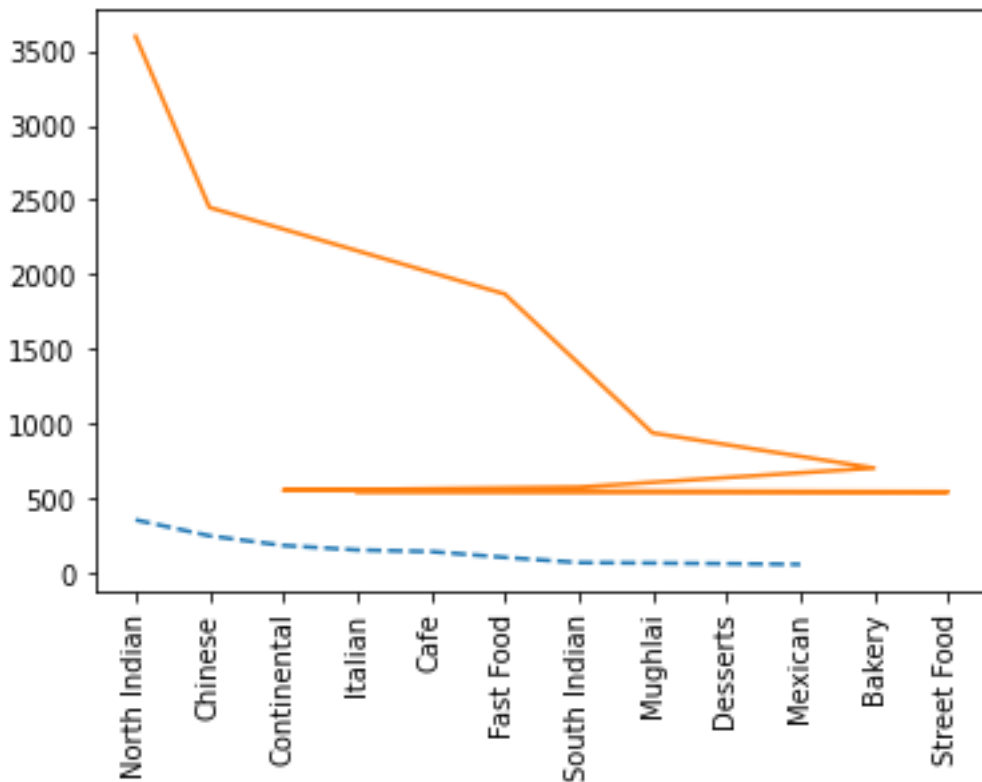


d)

d)

Answer The plot below shows how the data related to cuisines is different for both NCR and Rest of India

Explanation: The code first imports the necessary libraries and then it filters the data according to NCR and Rest of India and then applying loop to store the values in an array. Which is further plotted in the shown figure. It shows that regions in NCR serve more cuisines than the rest of India.



Question2)

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

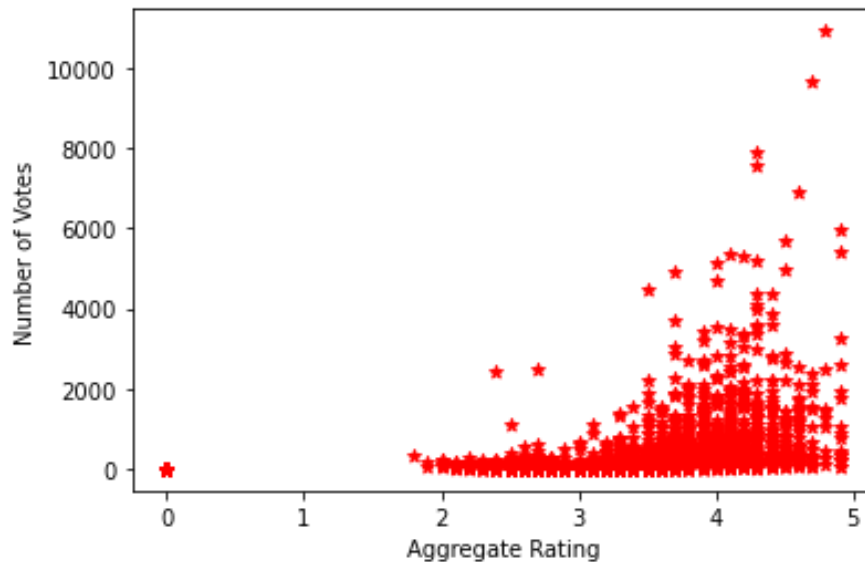
1. Write a short detail analysis of how the rating is affected by restaurant due following features: Plot a suitable graph to explain your inference.
 1. Number of Votes given Restaurant
 2. Restaurant serving more number of cuisines.
 3. Average Cost of Restaurant
 4. Restaurant serving some specific cuisines.

a)

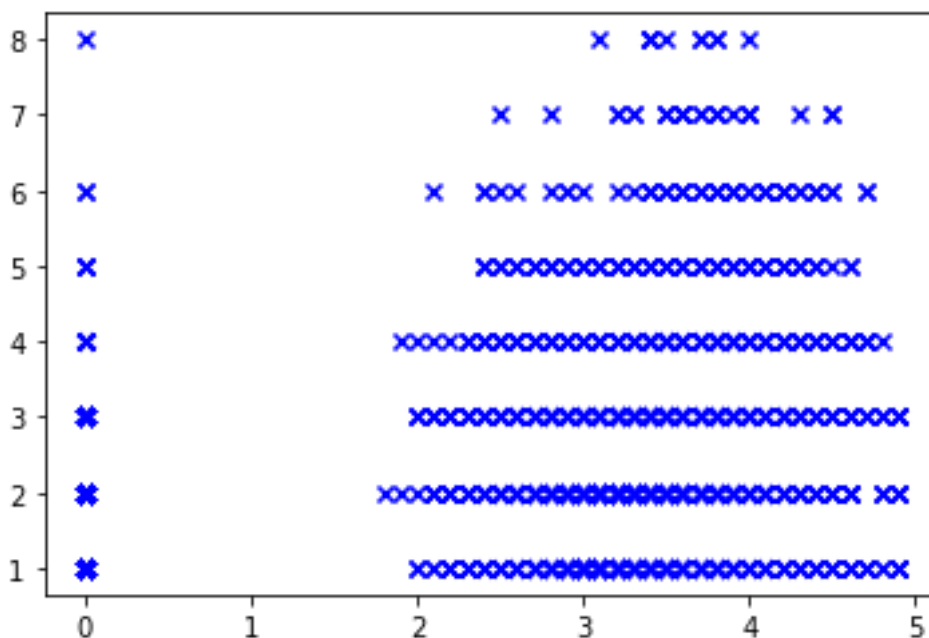
Answer: Rating v/s Number of Votes.

It can be clearly observed by looking at the graph that the relation between number of votes and Rating is positive but it can also be seen that some restaurants have very high rating around 5 but still have low number of votes.

Explanation: The code first sorts or filters the data so as the Country code is 1 and then it stores the number of votes and aggregate rating at same indexes in two arrays which is further used to plot the graph. Here a scatter graph is plotted.



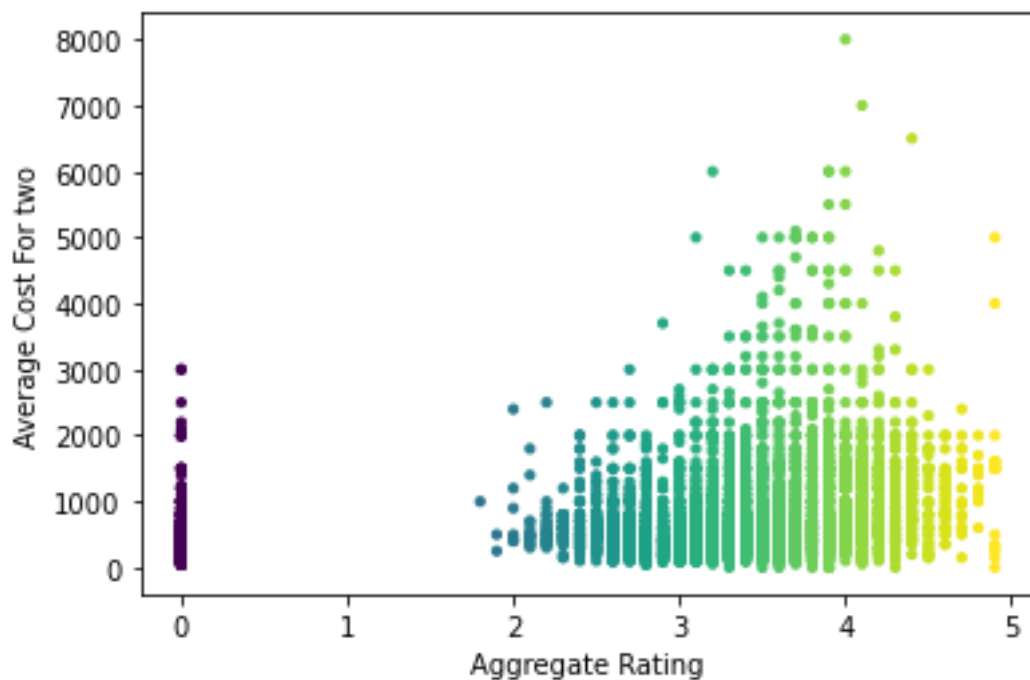
b) Number of Cuisines v/s Rating Figure:



It can be observed using this image that the relation between rating and number of cuisines is somewhat ambiguous and for both restaurants with good and bad rating there are more and less amount of cuisines.

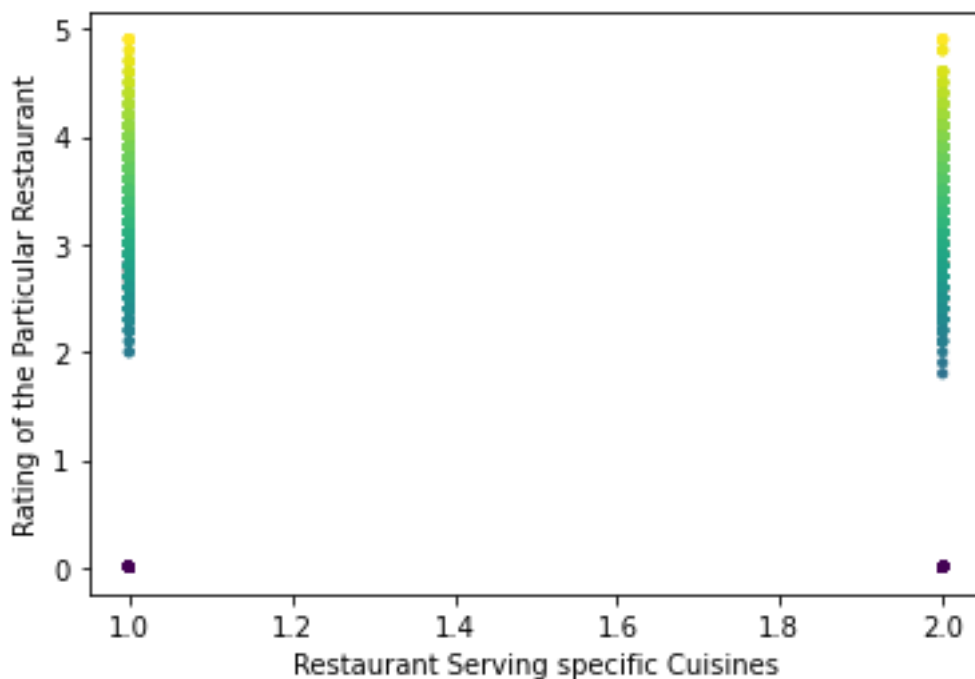
c) Average Cost for two v/s Rating

Figure:



It can be observed using the plot that the most of the restaurants that have low average-costfortwo hold both high and low rating, so there is no correlation between these parameters. It would not be a adequate choice to rate a restaurant on the basis of average cost for two as both quality and service matters more when it comes to restaurants. d)

Answer) Specific cuisines Serving Restaurants v/s Rating Figure:



It can be observed from the figure that specific cuisines serving restaurants also have both high and low rating which means that the number of cuisines served in the restaurants does not matter unless the quality and service of the food is good

2-part2)

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

1. Weighted Restaurant Rating = $\frac{\sum (\text{number of votes} * \text{rating})}{\sum (\text{number of votes})}$.

Answer) Top localities with more weighted restaurant rating:

Locality	Rating
Hotel Clarks Amer, Malviya Nagar	4.900000
Aminabad	4.900000
Friends Colony	4.886916
Powai	4.841869
Kirlampudi Layout	4.820161
Deccan Gymkhana	4.800000
Express Avenue Mall, Royapettah	4.800000
Banjara Hills	4.718762
Sector 5, Salt Lake	4.707023
Riverside Mall, Gomti Nagar	4.700000

The code first imports all the necessary libraries and then it filters the data for country code 1 and then it Initialises a rating column according to the given formula and then drops the nan values and then sorts the value on the bases of rating in descending order and prints the first 10 values.

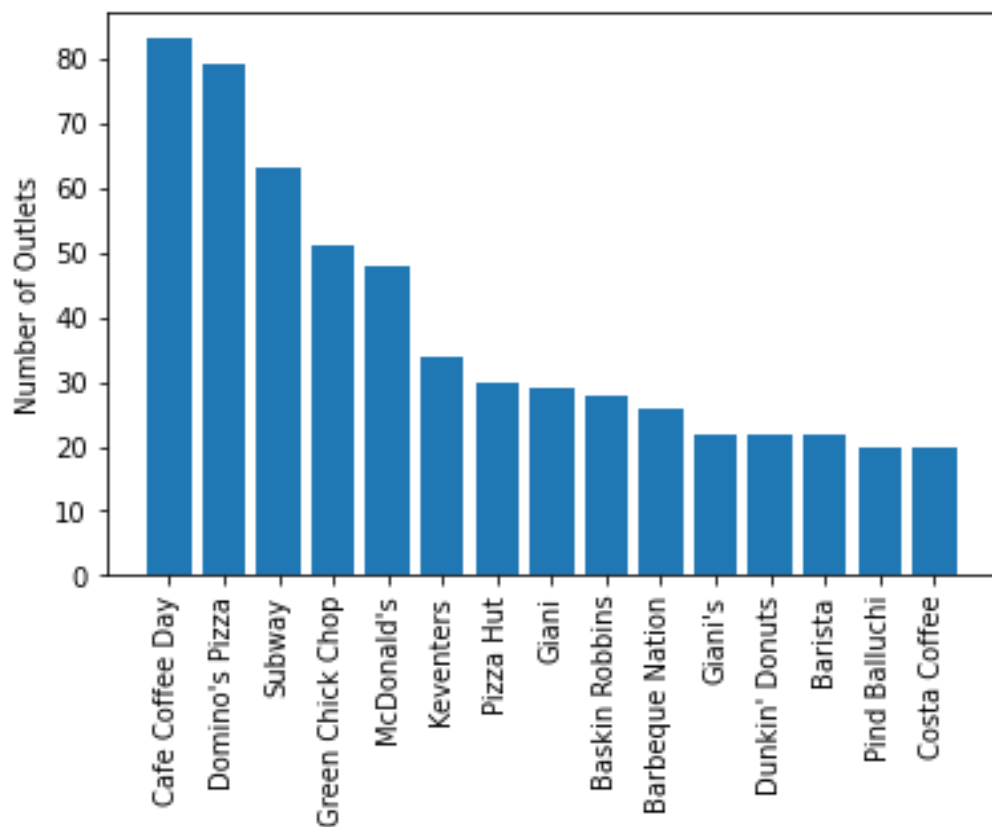
Question3)

Visualization

1. Plot the bar graph top 15 restaurants have a maximum number of outlets.
2. Plot the histogram of aggregate rating of restaurant(drop the unrated restaurant).
3. Plot the bar graph top 10 restaurants in the data with the highest number of votes.
4. Plot the pie graph of top 10 cuisines present in restaurants in the USA.
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.

a)

Answer) Figure:

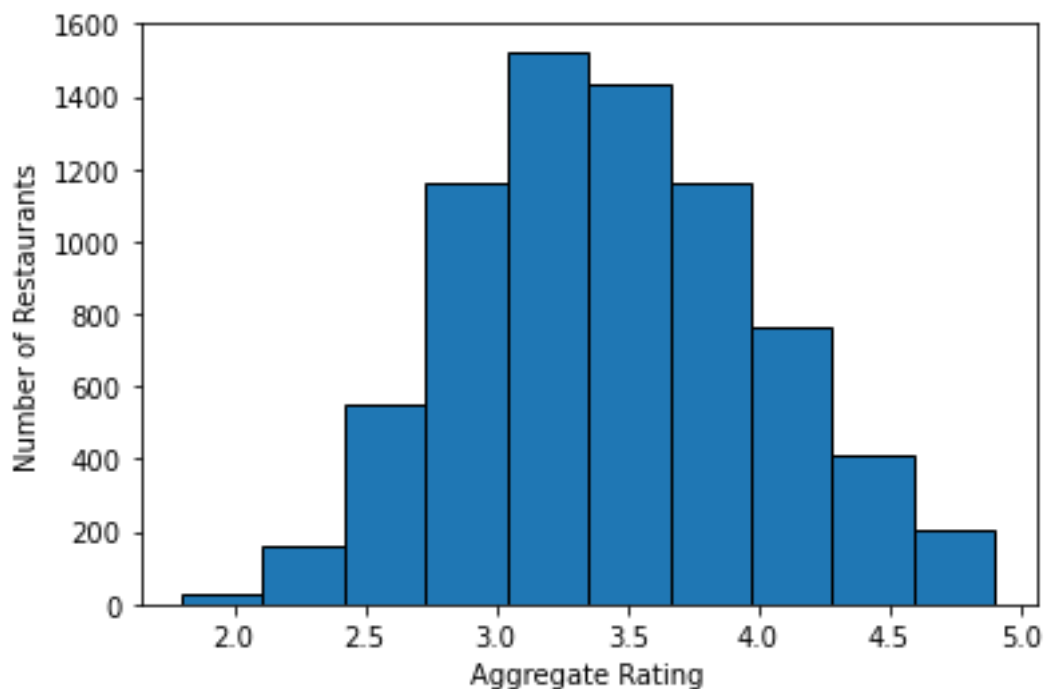


Top 15 restaurants having a maximum number of outlets.

Cafe Coffee Day	83
Domino's Pizza	79
Subway	63
Green Chick Chop	51
McDonald's	48
Keventers	34
Pizza Hut	29
Giani	29
Baskin Robbins	28
Barbeque Nation	25
Giani's	22
Dunkin' Donuts	22
Barista	22
Costa Coffee	20
Pind Balluchi	20

b)

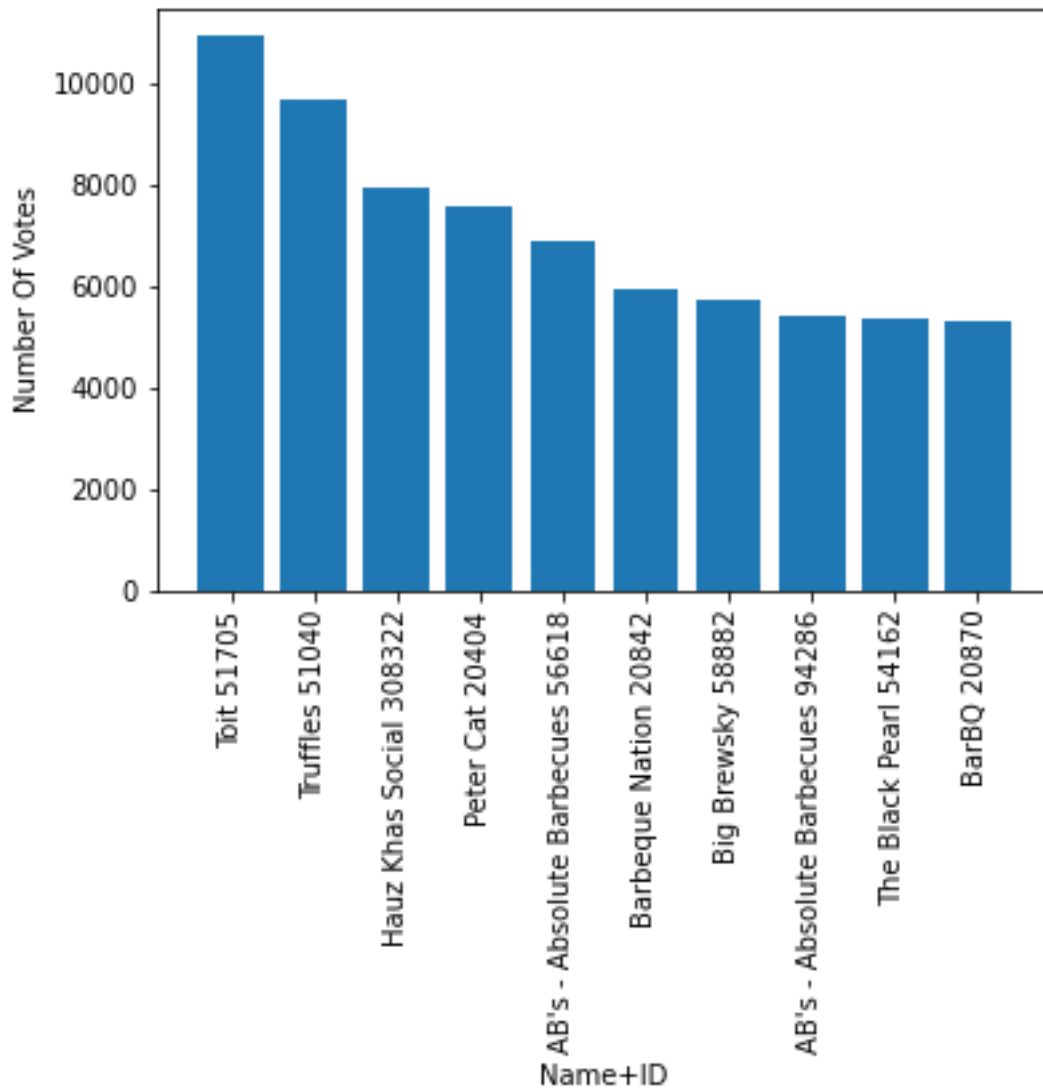
Answer) Figure:



The code first filters the data for rating! =0 and then it plots the bar graph with Number of restaurants on Y axis and Aggregate rating on x axis

c)

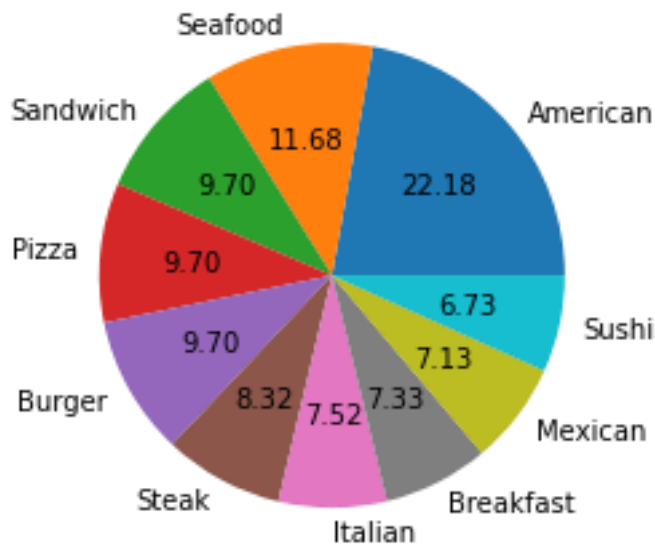
Restaurant Name (ID)	No of votes
Toit (51705)	10934
Truffles (51040)	9667
Hauz Khas Social (308322)	7931
Peter Cat (20404)	7574
AB's - Absolute Barbecues (56618)	6907
Barbeque Nation (20842)	5966
Big Brewsky (58882)	5705
AB's - Absolute Barbecues (94286)	5434
The Black Pearl (54162)	5385
BarBQ (20870)	5288



The code first filters the data based on votes and sorts in descending order. Then it stores the ID + NAME in one array (so that the restaurants can be distinguished) and Votes in another which is then plotted in a bar graph with different attributes

d)

Answer) Figure:



Top 10 cuisines present in restaurants in the USA.

Name	Count
American	112
Seafood	59
Burger	49
Pizza	49
Sandwich	49
Steak	42
Italian	38
Breakfast	37
Mexican	36
Sushi	34

The code first sorts the data based on the requirement of country code being 216 i.e. USA. Then it applies function fun1 to convert them into strings. It then drops the nan values and applies fun2 to Cuisines in order to split and store them in a dictionary which is further sorted and then name and count is stored in different arrays which is further plotted in a pie chart with their percentage depicted.

e)

Answer) Figure:

The Question first filters the data according to country Code 1 as required and then it calculates the agrating by multiplying it by votes and then it groups the data on the basis of city and calculates its Count. It also calculates the sum and then divides it by the Votes so as to Calculate the Weighted Rating and now the bubble chart is plotted with x axis as City name and y axis as the count and the size of the bubble as agrating.

