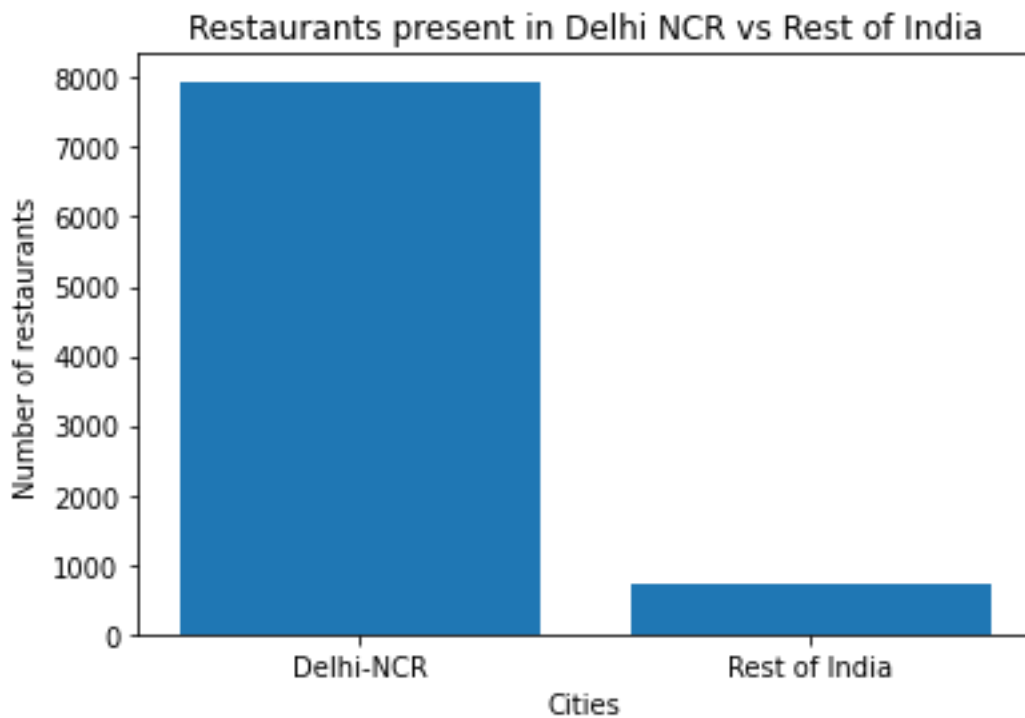# ZOMATO API-2

1.1 First I have filtered out all countries which is not having Country Code as 1.
Next, I initiated two counters one for Delhi NCR one for the rest of India.
Next, I created an array city from the dataframe. After this I traversed the whole city array and if the value of city was New Delhi or Gurgaon or Noida or Ghaziabad, or Faridabad I would increment the Delhi NCR counter by one else I would increment the rest of India counter by 1. At the end I have displayed both the counters and plotted a bar graph.

OUTPUT:

Restaurants present in Delhi-NCR: 7947
Restaurants present in Rest of India: 705

1.2 First I have filtered out all countries which is not having Country Code as 1.

Next, I have created a new dataframe for rest of India where the value of city is not among New Delhi, Gurgaon, Noida, Faridabad or Ghaziabad.

Next, I created a dataframe for Delhi NCR, to do this I have firstly replaced all the cities New Delhi, Gurgaon, Noida, Faridabad and Ghaziabad with City name: Delhi NCR. After this I created the Delhi NCR dataframe by filtering the cities.

Next, I created 2 sets 1 each for Delhi NCR and 1 for the rest of India. After this I traversed the cuisines columns of both the dataframes and added them to their respective sets. After this I took the set difference between rest of India and Delhi NCR, and displayed it.

OUTPUT

Malwani
Cajun
German
BBQ

1.3 First I have filtered out all countries which is not having Country Code as 1.

Next, I have created a new dataframe for rest of India where the value of city is not among New Delhi, Gurgaon, Noida, Faridabad or Ghaziabad.

Next, I created a dataframe for Delhi NCR, to do this I have firstly replaced all the cities New Delhi, Gurgaon, Noida, Faridabad and Ghaziabad with City name: Delhi NCR. After this I created the Delhi NCR dataframe by filtering the cities.

Next, I created 2 dictionaries to store cuisines and the number of restaurants serving those cuisines, 1 each for Delhi NCR and 1 for the rest of India. After this I traversed the cuisines columns of both the dataframes and added them to their respective dictionaries and kept incrementing the count parallely. After this I sorted the dictionaries by the count and displayed the top ten cuisines for both Delhi-NCR and for the rest of India

OUTPUT

```
Top 10 cuisines served by maximum number of restaurants in Delhi NCR:
North Indian
Chinese
Fast Food
Mughlai
Bakery
South Indian
Continental
Desserts
Street Food
Italian
-----------------------------------------------------------------------
---
Top 10 cuisines served by maximum number of restaurants in rest of Indi
a:
North Indian
Chinese
Continental
Italian
Cafe
Fast Food
South Indian
Mughlai
Desserts
Mexican
```

1.4 First I have filtered out all countries which is not having Country Code as 1.

Next, I have created a new dataframe for rest of India where the value of city is not among New Delhi, Gurgaon, Noida, Faridabad or Ghaziabad.

Next, I created a dataframe for Delhi NCR, to do this I have firstly replaced all the cities New Delhi, Gurgaon, Noida, Faridabad and Ghaziabad with City name: Delhi NCR. After this I created the Delhi NCR dataframe by filtering the cities.

Next, I created 2 sets 1 each for Delhi NCR and 1 for the rest of India. After this I traversed the cuisines columns of both the dataframes and added them to their respective sets. After this I took the set difference between rest of India and Delhi NCR and the set difference between Delhi NCR and rest of India. Next I have plotted two pie-charts to display the cuisines exclusive to Rest of India and to Delhi NCR.

OUTPUT

Cuisine served in Rest of India but not in Delhi NCR
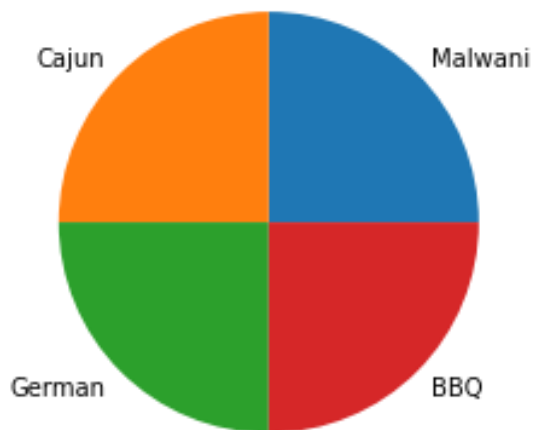
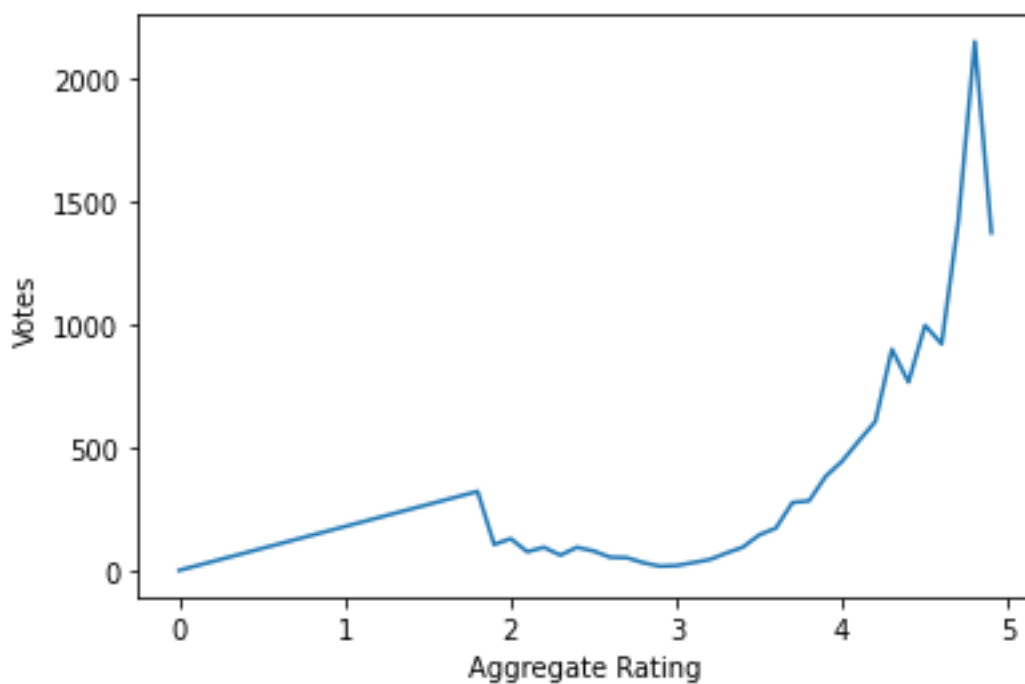Cuisine served in Delhi NCR but not in Rest of India
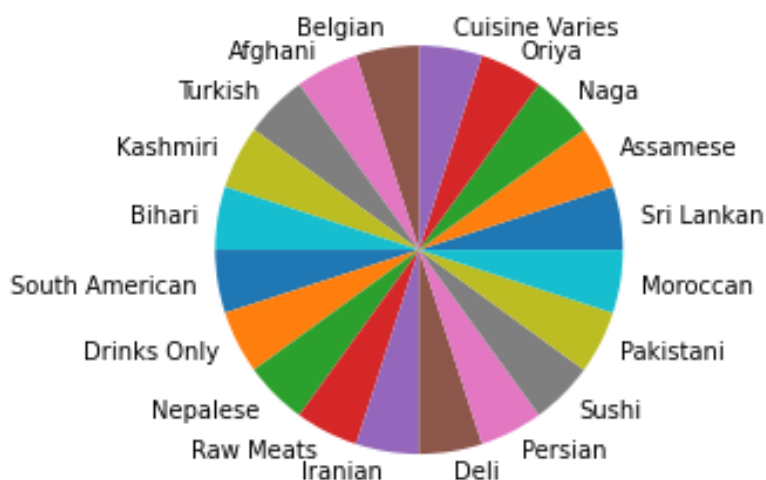
2.1.1 First I have filtered out all countries which is not having Country Code as 1.

Next, I created a dictionary to store the votes given to a restaurant corresponding to its rating. To do this I kept a track off total number of votes and no or restaurants corresponding to a single rating value. After this I divided the total votes by the total number of restaurants to get average number of votes corresponding to 1 rating value.

After this I sorted the dictionary based on the rating and plotted a line graph between Aggregate rating and votes.
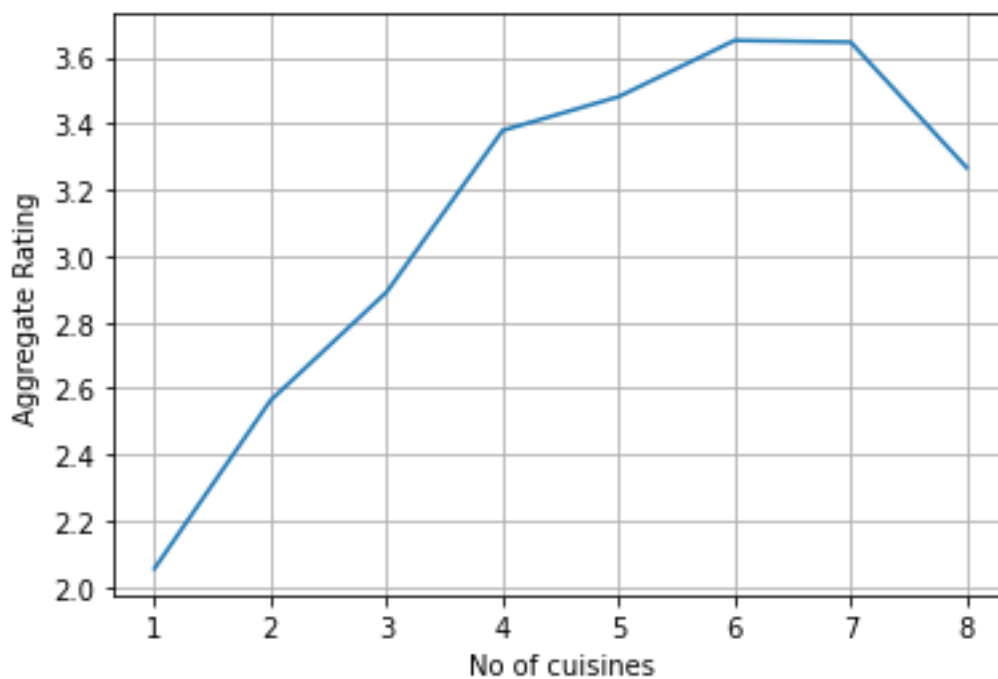
OUTPUT

2.1.2 First I have filtered out all countries which is not having Country Code as 1.

Next, I created a dictionary to store the rating of a restaurant corresponding to the number of cuisines served in a restaurant. To do this I made a new column in the dataframe to keep track of number of cuisines served in a restaurant using the cuisine column. To do this I kept a track off rating of restaurants corresponding to the number of cuisines served in that restaurant. After this I divided the total rating by the total number of restaurants to get average rating corresponding to the number of cuisines served.

After this I sorted the dictionary based on the number of cuisines and plotted a line graph between number of cuisines and Aggregate rating.
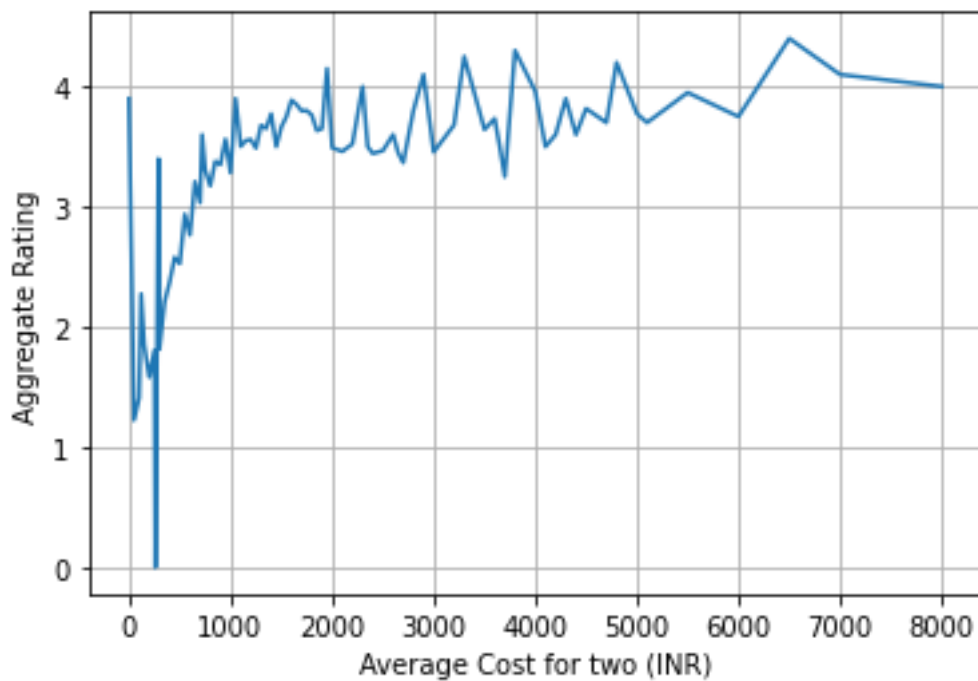
OUTPUT

2.1.3 First I have filtered out all countries which is not having Country Code as 1.

Next, I created a dictionary to store the rating of a restaurant corresponding to the Average cost in a restaurant. To do this I kept a track off rating of restaurants corresponding to the average cost eating in that restaurant. After this I divided the total rating by the total number of restaurants to get average rating corresponding to the average cost.

After this I sorted the dictionary based on the Average cost and plotted a line graph between Average cost and Aggregate rating.
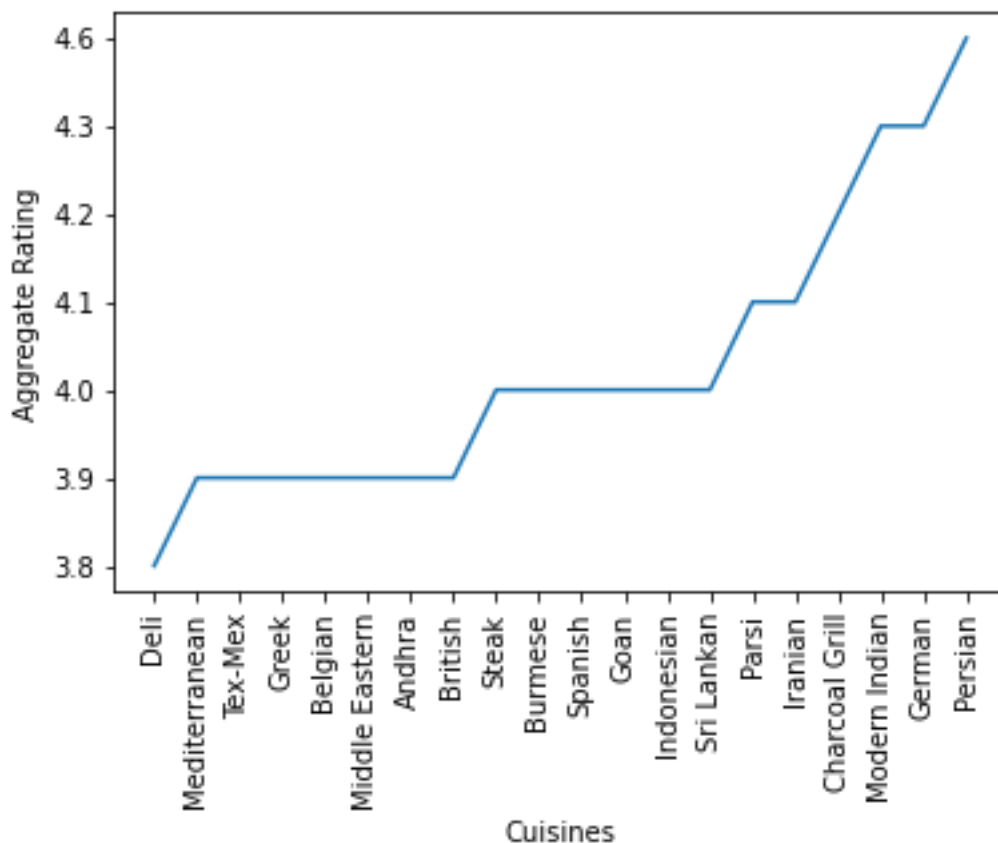
OUTPUT

2.1.4 First I have filtered out all countries which is not having Country Code as 1.

Next, I created a dictionary to store the rating corresponding to a cuisine type. To do this I made a cleaned the cuisine column in the dataframe to keep track of cuisines served in a restaurant in a list format. To do this I kept a track off rating of restaurants corresponding to the cuisine. After this I divided the total rating by the total number of restaurants serving that cuisine to get average rating corresponding to a cuisine type.

After this I sorted the dictionary based on the rating and plotted a line graph between cuisines and Aggregate rating.

OUTPUT

2.2.1 First I have filtered out all countries which is not having Country Code as 1.

Next, I created a dictionary to store the product of rating and votes and votes corresponding to restaurant for its locality. To do this I made use of locality, votes, aggregate rating columns of the dataframe. After this I calculated the weighted average rating for that locality.

Next, this I sorted the dictionary based on the rating and plotted a line graph between locality and weighted average rating.

OUTPUT
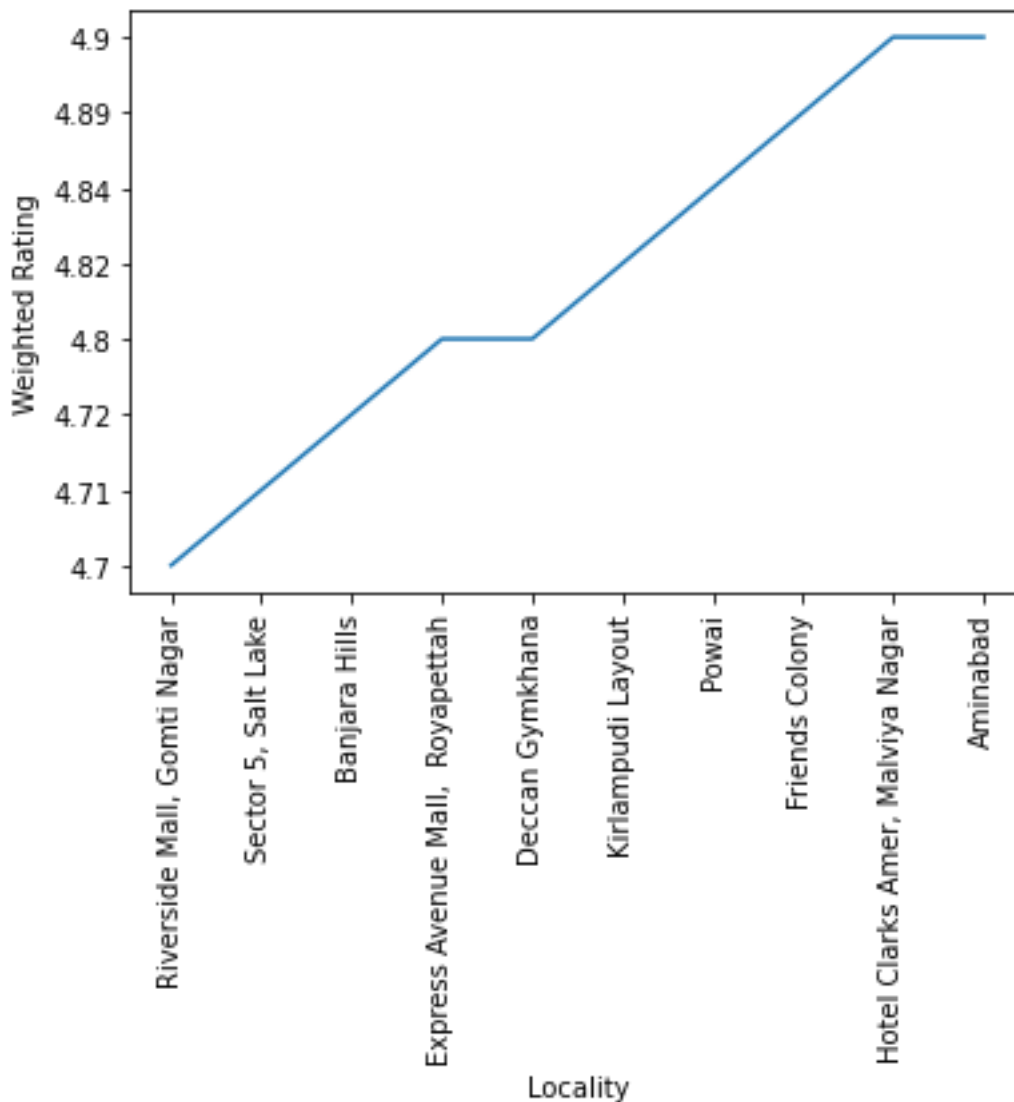
3.1 Here I have considered only Indian Restaurants as it has been mentioned on the top of the Questions. Here I have filtered out only Indian Restaurnats first by using boolean indexing. Once that's done I have made one dictionary in which I stored the outlet name and made it as key and frequency as the value. After iterating over all the names I got my dictionary fully prepared for the data visualization but as the we need to print in descending order so I have used one list of list named restaurant_names_counts which stores the name and the frequencey. After that I have sorted the list of list based on the frequency. Note: I have assumed no two rows with same outlet name and locality present in the dataset. If that would have been considered then I would have need to filtered out those if already appeared.

OUTPUT
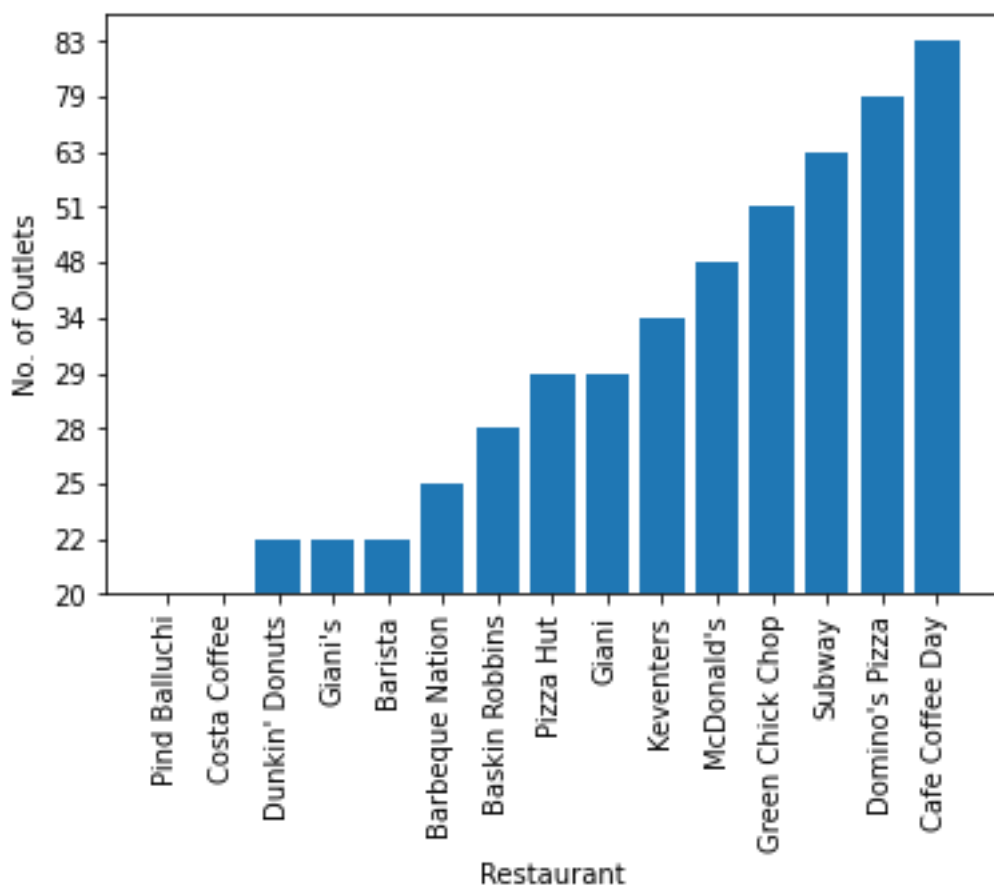
3.2 As mentioned in the question I need to drop all the unrated restaurants so I have checked the Rating text column if the text is not Not Rated then consider otherwise exclude and I am storing in a new data frame that is zomato_ratings. After that I am picking values from only Aggregate rating column and storing it to ratings. Just to make sure there is no nan values I have checked and store the values which don't have value as nan. after that I am ploting it graph and made the bin to auto so that it decide automatically and made xticks to 0-5 as our rating lies between 0-5. From the graph we can see that number of zero rating is drastically high as compare to other ratings. If we exclude 0 rating then we can notice most of the restaurants got rating in between 3-4.

OUTPUT
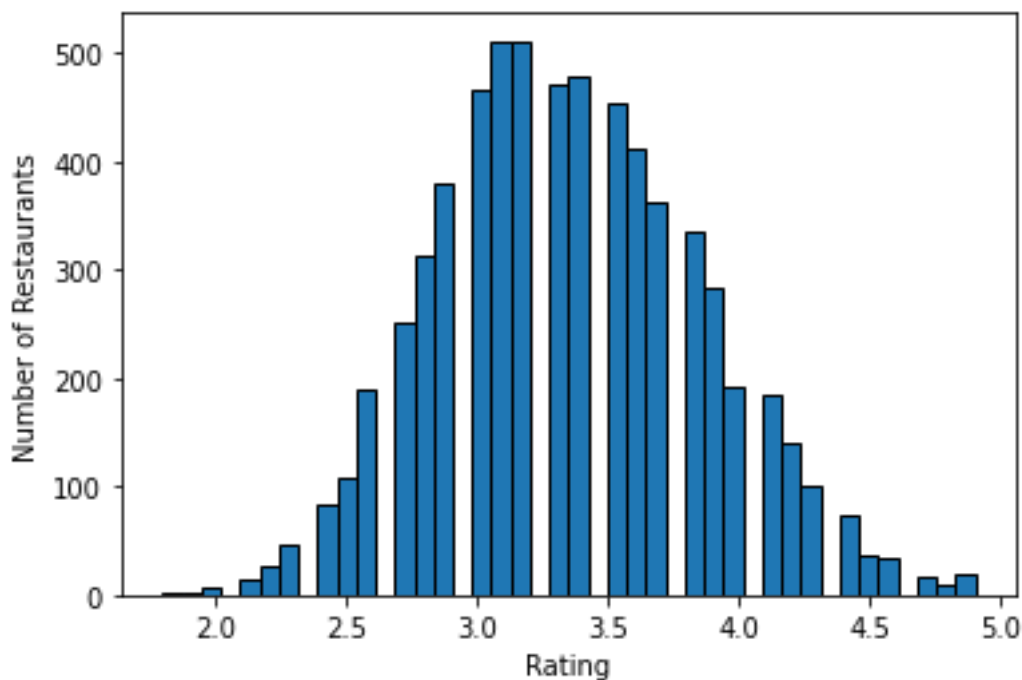
3.3 To get the top 10 restaurants I have one dictionary to get the name of restaurant and number of votes but here one case that can happen is that with same restaurant in different location get less votes and other one get ample number of votes. As it hasn't been clearly mentioned in the question statement that do I need to consider only name or not, so I have considered them as different. In the dictionary I am storing the name with its locality so it will be unique and will not repeat even after the name is same for two or more restaurants. Here name with locality works as the key and number of votes as the value. I have made one column wherein I have combined name and locality & votes with ### & $$$. Reason for using two delimeter is that ### will be used to split into two parts i.e name_locality and votes; later $$$ has been used to split into two parts where one part is Name of the restaurant and the locality. I am storing the values after calling the function get_votes. Once I stored all the values I have made one list where 0th column contains votes and 1st contains name of the restaurant. Now we have the correct data now I simply sort them in descending order based on the votes and picked top 10 restaurants.

OUTPUT
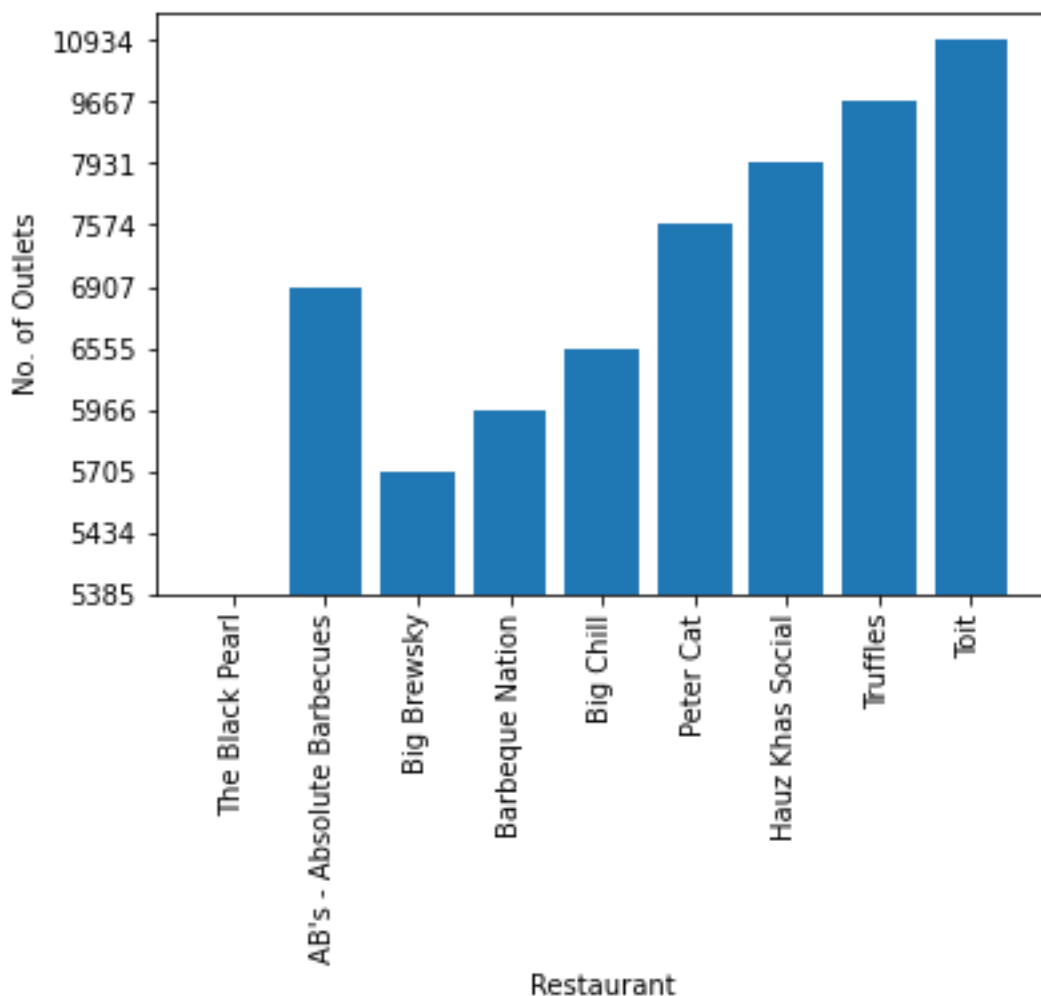
3.4 Here we need to filter out based on Country Code first and I have done and stored in zomato_usa_df. I have dropped all nan values present in the cuisines column. I have maintained on dictionary which will help me to keep track of cuisines count. One restaurants might be offering several cuisines. I have used one function inside which I am storing the count. Before that I am making sure that I am not skipping any cuisines. Cuisines column consists of all the cuisines names together so I have converted the type and then splited based on combination of space and comma; and updating the count which are in the cuisines list. After I got all the count stored in the dictionary then I have made one list of list which store values like 0th column will store the count and 1st column will store the cuisine name. After that I have sorted the list of list based on counts. Now I just took only top 10 cuisines names and its values.

OUTPUT
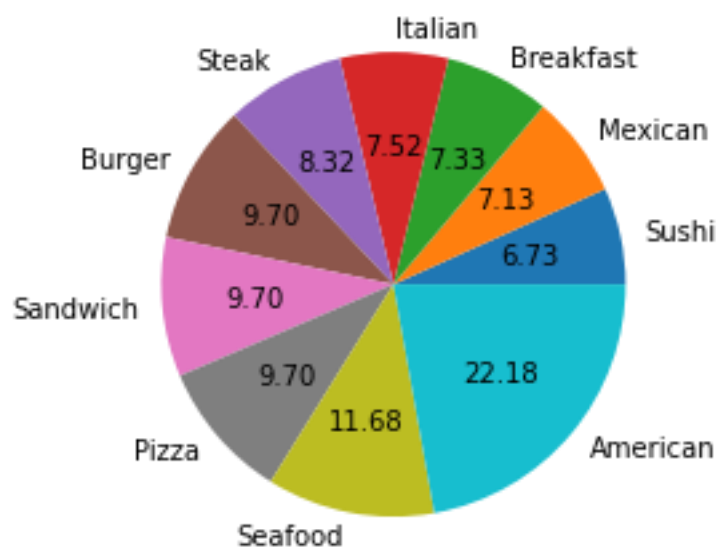
3.5 As we need to store the city weighted restaurant rating so I am making one city dictionary which will help me get the city names and the weighted ratings. I have used the same formula which have been mentioned in one of the previous question to find out the. But here I need to keep track of the count of restaurant present in the city so I am maintaining that by updating the count by one whenever I encounter the same city name. So basically, here key is city and value is the list wherein 0th storing the summation of votes*ratings, 1st column storing the summation of votes and 2nd column. storing the count of restaurants present in that city. Once I got the dictionary filled with values I have calculated the weighted rating and stored in a list named city_ratings with city name and total count. after that I have sorted based on the count of restaurants present in the city. So number of cities can be huge so I have considered top 20 cities in India. Delhi topped among the of cities. After that I have plotted them in the graph and provided the size of the bubble based on the weighted rating of the city.

I have subtracted 3 from each value if any value gets negative, I have made it positive. As they are very close to each other so I have multiplied with 100 to get the proper visualization.

OUTPUT