Coding Ninjas Project: Zomato API – II

Note – Consider only Indian restaurants in this analysis

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.

Solution 1:- Firstly we will read the zomato.csv file using the pandas and save the Dataframe in df. Now we will extract the Indian Restaurants data using the country code equals to 1 and save the dataframe in df_India. Now we will extract the Delhi_NCR Region and Rest of India and store the dataframe in df_Delhi_NCR and df_ROI respectively.

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

   Solution 1.1:- In this firstly we will find the length of df_Delhi_NCR and df_ROI which will give number of restaurants present in Delhi NCR and Rest of India and now we will plot a bar graph of this data using the matplotlib. Number of restaurants present in Delhi NCR and Rest of India are:-
   - Delhi-NCR 7947
   - Rest of India 705
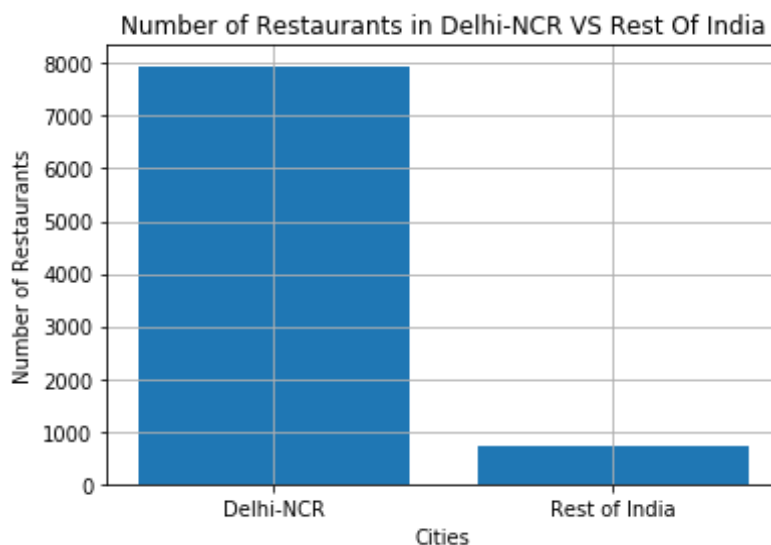
2. Find the cuisines which are not present in restaurant of Delhi NCR but present in rest of India.Check using Zomato API whether this cuisines are actually not served in restaurants of Delhi-NCR or just it due to incomplete dataset.

Solution 1.2:- In this we will extract the values of cuisines in the form of list from the df_Delhi_NCR and df_ROI respectively and apply a function named cuisines which will return the list of cuisines in that region. After that we will run a for loop and check the cuisines that are not served in Delhi-NCR. The cuisines which are not present in restaurant of Delhi NCR but present in rest of India are:-
- German
- Malwani
- BBQ
- Cajun

Now we will use the zomato-api to check whether these cuisines are actually not served in restaurants of Delhi-NCR or just it due to incomplete dataset. Firstly we will generate a zomato api key from the website and use it as a user-key in headers. Now we will find the city id for Delhi-NCR region by making a get request to https://developers.zomato.com/api/v2.1/cities and passing headers and params. After obtaining the city id we will make a get request to https://developers.zomato.com/api/v2.1/cuisines by passing the headers as user-key and params as city id which will give all the cuisines served in the Delhi-NCR region. So the cuisines which are not present in restaurant of Delhi NCR but present in rest of India are:-
- German
- Cajun

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

Solution 1.3:- In order to find the top 10 cuisines served by maximum number of restaurants in Delhi NCR and rest of India we will firstly extract the cuisines from df_India and run through a function cuisines which will return a dictionary with cuisine and count as a key-value pair. Now we will a sort the dictionary in reverse order on the basis of count and make it into the list and extract the top 10 from the list. So the top 10 cuisines served by maximum number of restaurants in Delhi NCR and rest of India are:-
- f-India
- North Indian

- Chinese
- Fast Food
- Mughlai
- Bakery
- Continental
- Italian
- South Indian
- Cafe
- Desserts

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.

Solution 1.4:-In this we will compare the two regions on the basis of cuisines and number of restaurants serving the cuisines. As done previously we will extract the top 10 cuisines in Delhi-NCR and ROI separately and plot the two line graphs accordingly

Cuisines In DELHI-NCR

Cuisines In Rest-of-India
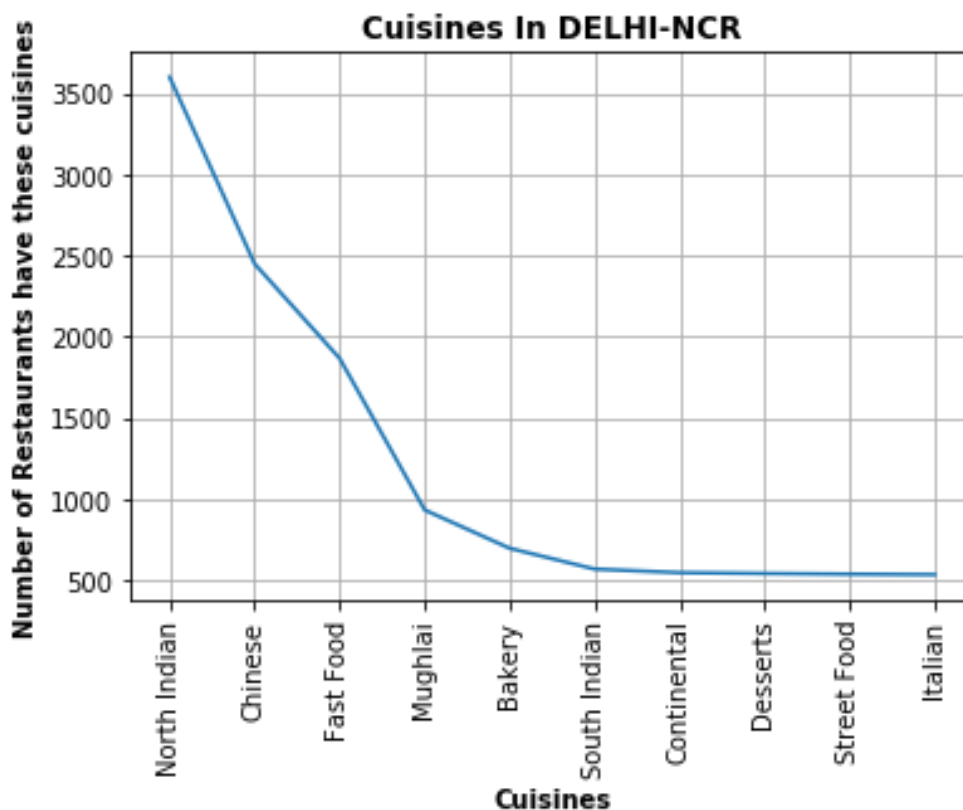
From the above two graphs we can see that in Delhi-NCR region there are more than 3500 restaurants selling North Indian cuisine while there are only 350 restaurants which sell North Indian cuisine in the Rest of India which is a lot of difference. Moving on to Chinese cuisine there are about 2500 restaurants selling Chinese cuisine in Delhi-NCR while there are about 250 restaurants which sell Chinese cuisine in the Rest of India. Moreover there are 86 cuisines that are sold in Delhi-NCR region while there are 70 cuisines sold in Rest of India.This makes a huge difference as Delhi-NCR region restaurants sell more cuisines than that of Rest of India

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

Solution 2:- Firstly we will make a copy of the df_India and store it in df_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.

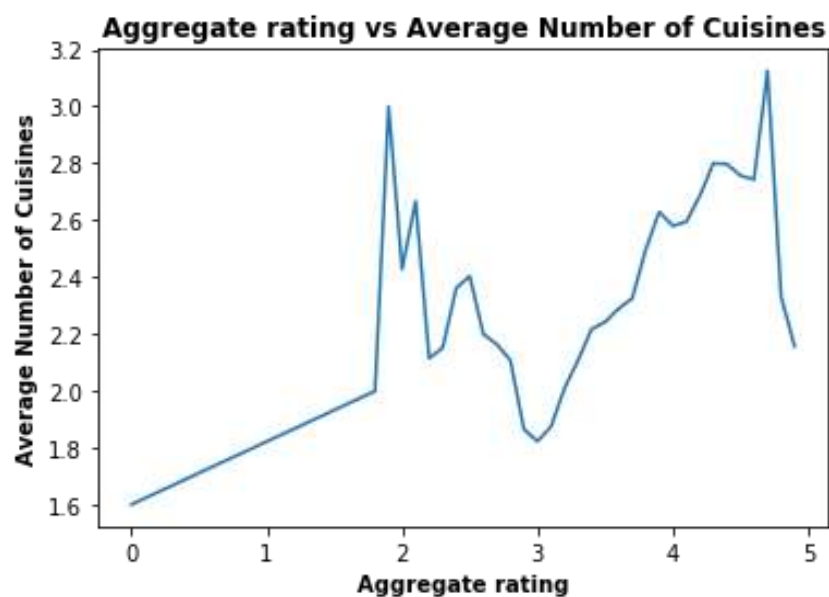   1. Number of Votes given Restaurant

    Solution 2.1.1:-We will now group the df_2 by Aggregate rating and find the mean of votes and sort them according to index in ascending order. We will extract the rating from the index and the votes from the values and plot a line graph between the rating and the votes.

Aggregate Rating Vs Votes
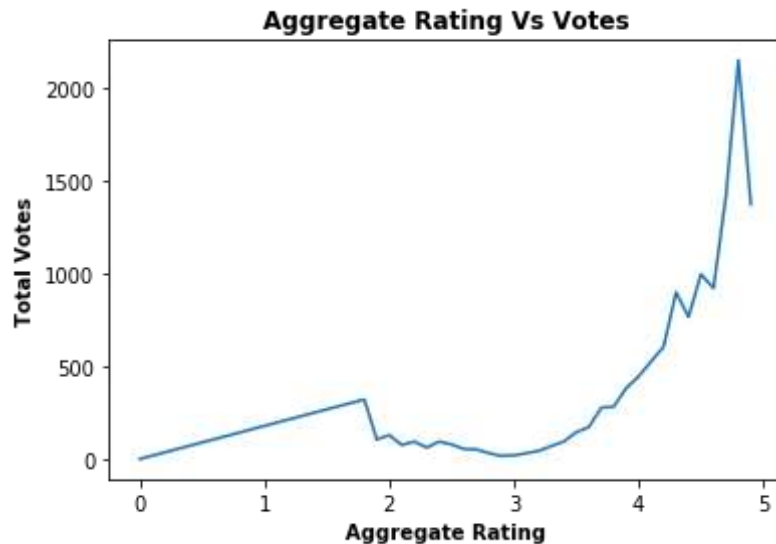
From the given graph we can see that number of votes have huge impact on the ratings.So with increase in number of votes it results in increase of rating.

2. Restaurant serving more number of cuisines.

Solution 2.1.2:- In this we will make a new column in df_2 named Cuisines_count with the help of count_cuisines function which returns the count of the cuisines a restaurant serves. Now We will group the df_2 by Aggregate rating and find the mean of Cuisines_count and sort them according to index in ascending order. We will extract the rating from the index and the Cuisines_count from the values and plot a line graph between the rating and the votes.

Aggregate rating vs Average Number of Cuisines

From the graph we can see that more the cuisines more is the rating but for some restaurants this is not applied as the rating depends on other factors also.

3. Average Cost of Restaurant

Solution 2.1.3:- We will now group the df_2 by Aggregate rating and find the mean of Average Cost for two and sort them according to index in ascending order. We will extract the rating from the index and the Average Cost for two from the values and plot a line graph between the rating and the votes.
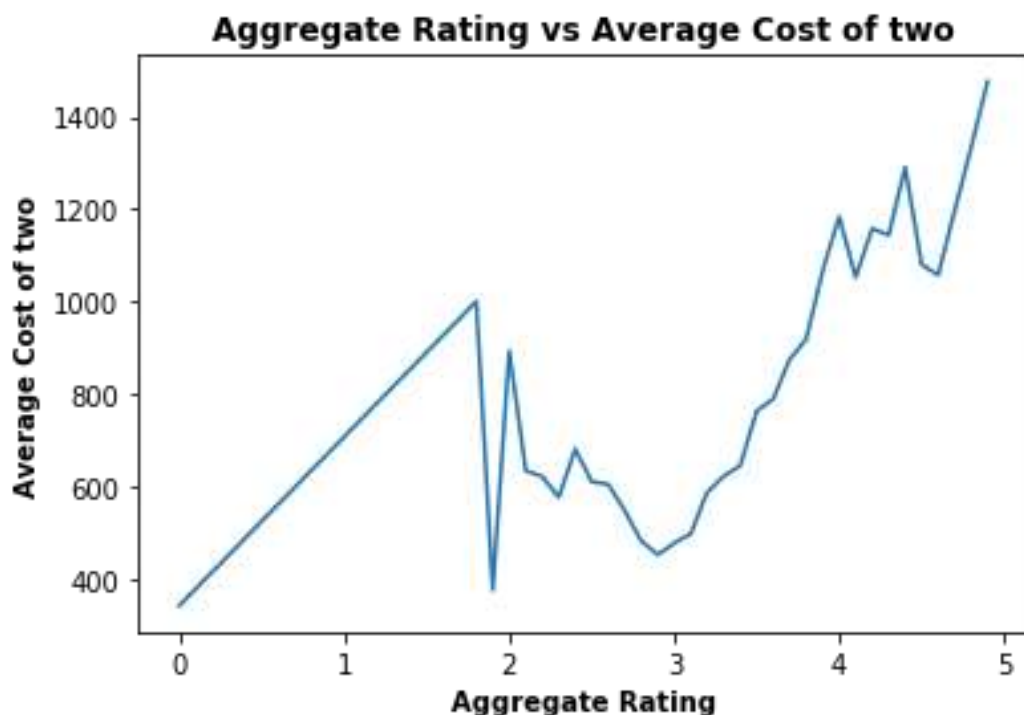
4. Restaurant serving some specific cuisines

Solution 2.1.4:- In this we will firstly extract the unique values of Rating text using value_counts and then whose rating text is Excellent and Very Good we will extract that cuisines and make a dictionary of these cuisines and their count. Then we will sort the dictionary in the reverse order on the basis of count and make the list of special cuisines. Then using specialcuisines function we will make a new column which stores the special cuisine of that restaurant. Then we will group the df_2 by Special_Cuisine and find the median of Aggregate rating and sort them according to values in descending order. We will extract the rating from the values and the cuisines from the index and plot a bar graph between the cuisines and the rating.

Special Cuisines vs Rating
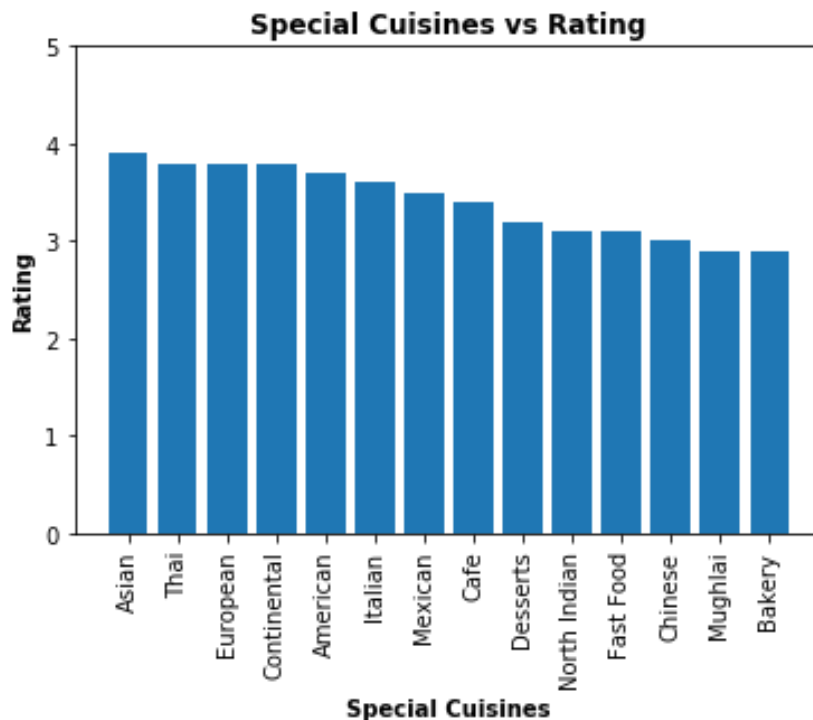
Restaurants serving some special cuisines like Asian, continental, Italian have more ratings than the other restaurants

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

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

   Solution 2.2.1:- In this we will make a new column of Weighted Restaurant Rating which will be number of votes * rating. Then we will group df_2 by Locality and find the sum of Weighted Restaurant Rating and store it in a1. After that we will group df_2 by locality and find the sum of Votes and store it in a2. Then we will divide a1 by a2 and sort them on the basis of values in descending order and extract the top 10 localities and then with the help of this data we will plot the bar graph of locality vs Weighted Restaurant Rating.
   The top 10 localities with more weighted restaurant rating are:-
   - Aminabad   4.9
   - Hotel Clarks Amer, Malviya Nagar   4.9
   - Friends Colony   4.886916367367881
   - Powai   4.841868613138686
   - Kirlampudi Layout   4.820161290322581

- Express Avenue Mall, Royapettah   4.8
- Deccan Gymkhana   4.8
- Banjara Hills   4.718761726078799
- Sector 5, Salt Lake   4.707022552098202
- Riverside Mall, Gomti Nagar   4.7

Locality vs Weighted Restaurant Rating
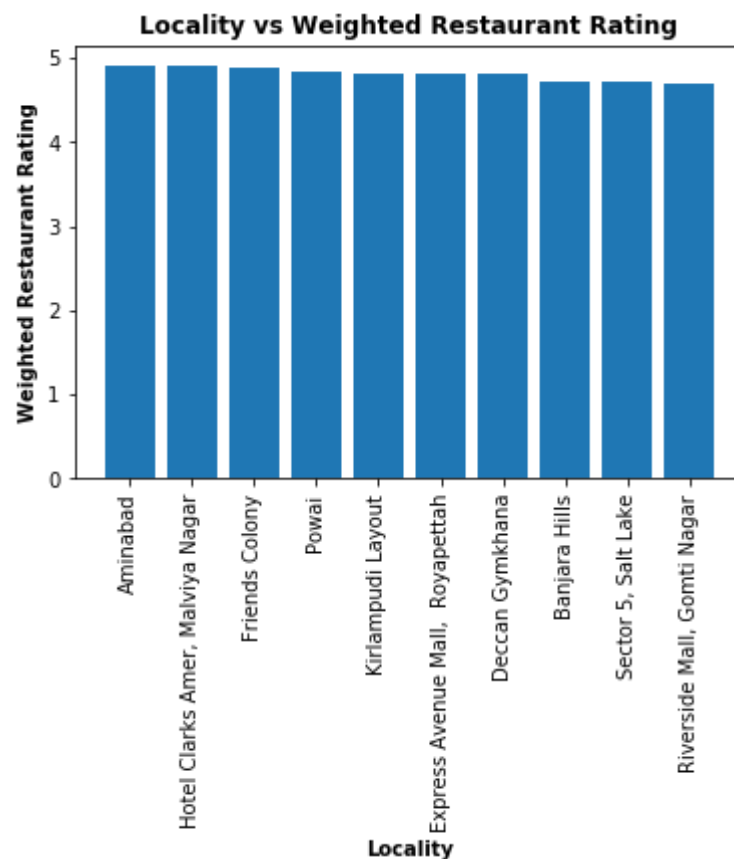
3. Visualization

Solution 3 :- Firstly we will copy the df_India and store it in df_3.

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

    Solution 3.1:- In this we will group df_3 by Restaurant Name and find the value_counts() and sort it according to descending order and extract top 15. We will extract the restaurants from the index and the count of outlets from the values. So with this data we will plot the bar graph of Restaurants vs Number of Outlets. Top 15 restaurants have a maximum number of outlets are:-
    - 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
- Dunkin' Donuts   22
- Giani's   22
- Barista   22
- Pind Balluchi   20
- Costa Coffee   20
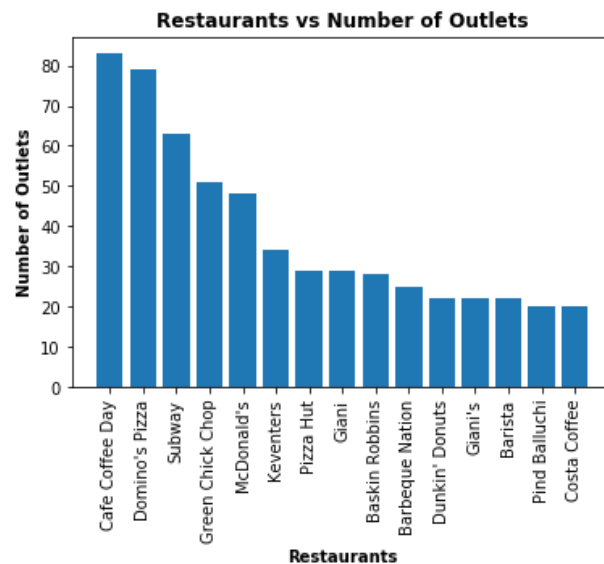
2. Plot the histogram of aggregate rating of restaurant(drop the unrated restaurant).

Solution 3.2:- In this we will make a df_hist in which Rating text is not equal to Not rated. After that we will store the values of Aggregate rating in the data and make the historgram with this data keeping the bins as auto.

Aggregate Rating vs Number of Restaurants
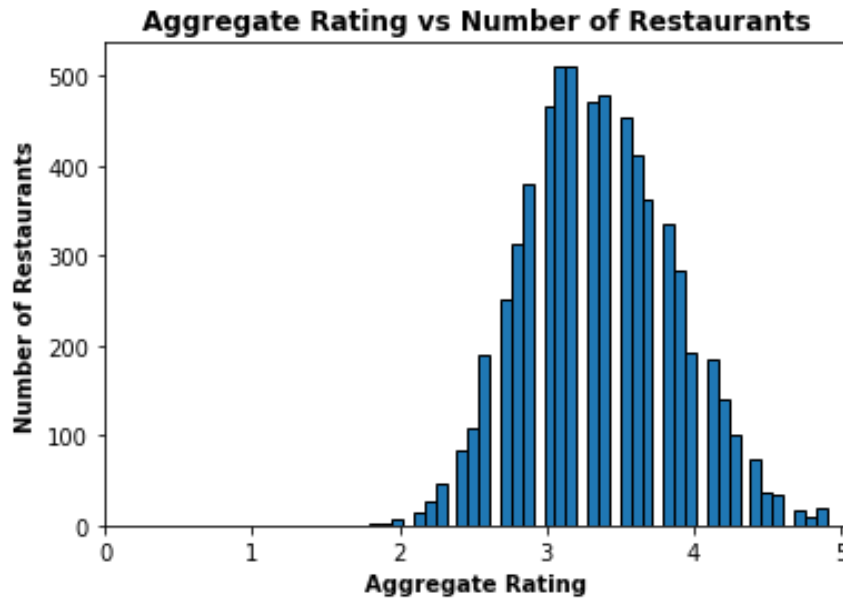
3. Plot the bar graph top 10 restaurants in the data with the highest number of votes.

Solution 3.3:-In this we will group df_3 by Restaurant Name and find the sum of the votes and sort by the values in descending order and extract the top 10 restaurants. We will extract the restaurants by the index and the votes by the values. Now we will
Make a bar graph with the help of this data. Top 10 restaurants in the data with the highest number of votes are :-

- Barbeque Nation   27835
- AB's - Absolute Barbecues   13400
- Toit   10934
- Big Chill   10853
- Truffles   9682
- Farzi Cafe   9189
- Chili's   8156
- Hauz Khas Social   7931
- Joey's Pizza   7807
- Peter Cat   7574

Restaurants VS Votes
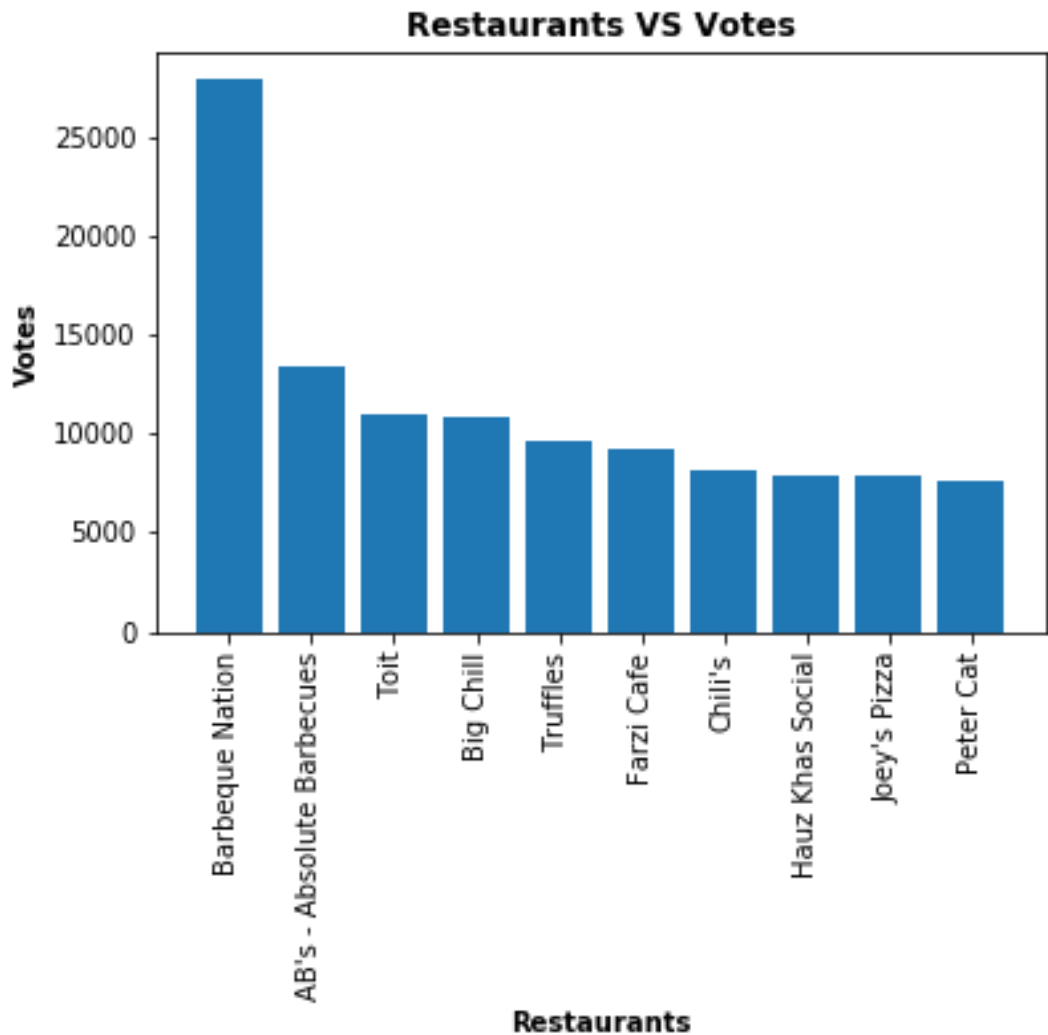
4. Plot the pie graph of top 10 cuisines present in restaurants in the USA.

Solution 3.4:- In this firstly we will extract the data of USA from the df by using Country Code as 216 and store it in df_USA and we will drop the na values from that. Now we will extract the cuisines values from the df_USA and with the help of cuisines function we will make a dictionary which stores the cuisines and its count. Now we will sort the dictionary on the basis of count and convert it into list and get the top 10 cuisines. Now with the help of this data we will plot a pie chart. The top 10 cuisines present in restaurants in the USA are:-

- American   112
- Seafood   59
- Burger   49
- Sandwich   49
- Pizza   49
- Steak   42

- Italian   38
- Breakfast   37
- Mexican   36
- Sushi   34

Top 10 cuisines present in restaurants in the USA
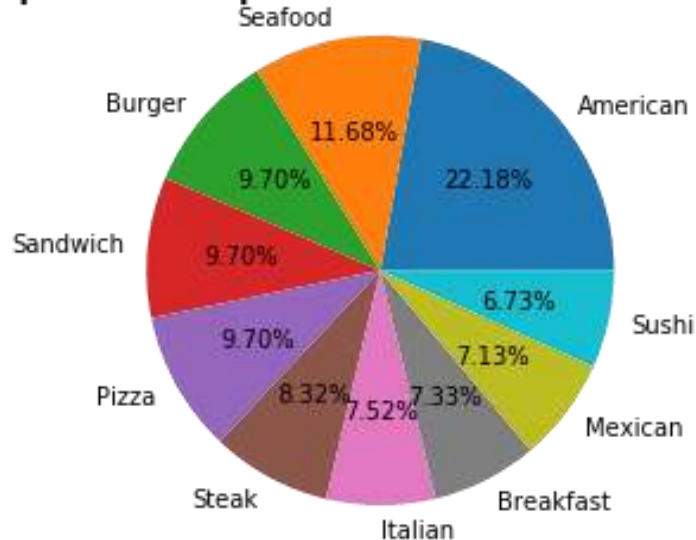
5. Plot the bubble graph of a number of Restaurants present in the city of India and keeping the weighted restaurant rating of the city in a bubble.

Solution 3.5:- In this we will do exact thing which we have done in 2.2.1 but the only difference is that we will group the data by cities. Now we get the weighted_restaurant_rating. Now we will extract the rating by values and cities by index. Then we will find the number of restaurants in the city with the help of value_counts() and sort on the index in the ascending order and store it's values in number_of_restaurants_in_city. Now we will plot a bubble graph of the cities and number_of_restaurants_in_city using s as weighted_rating*9.

City VS Number of Restaurants