

1.

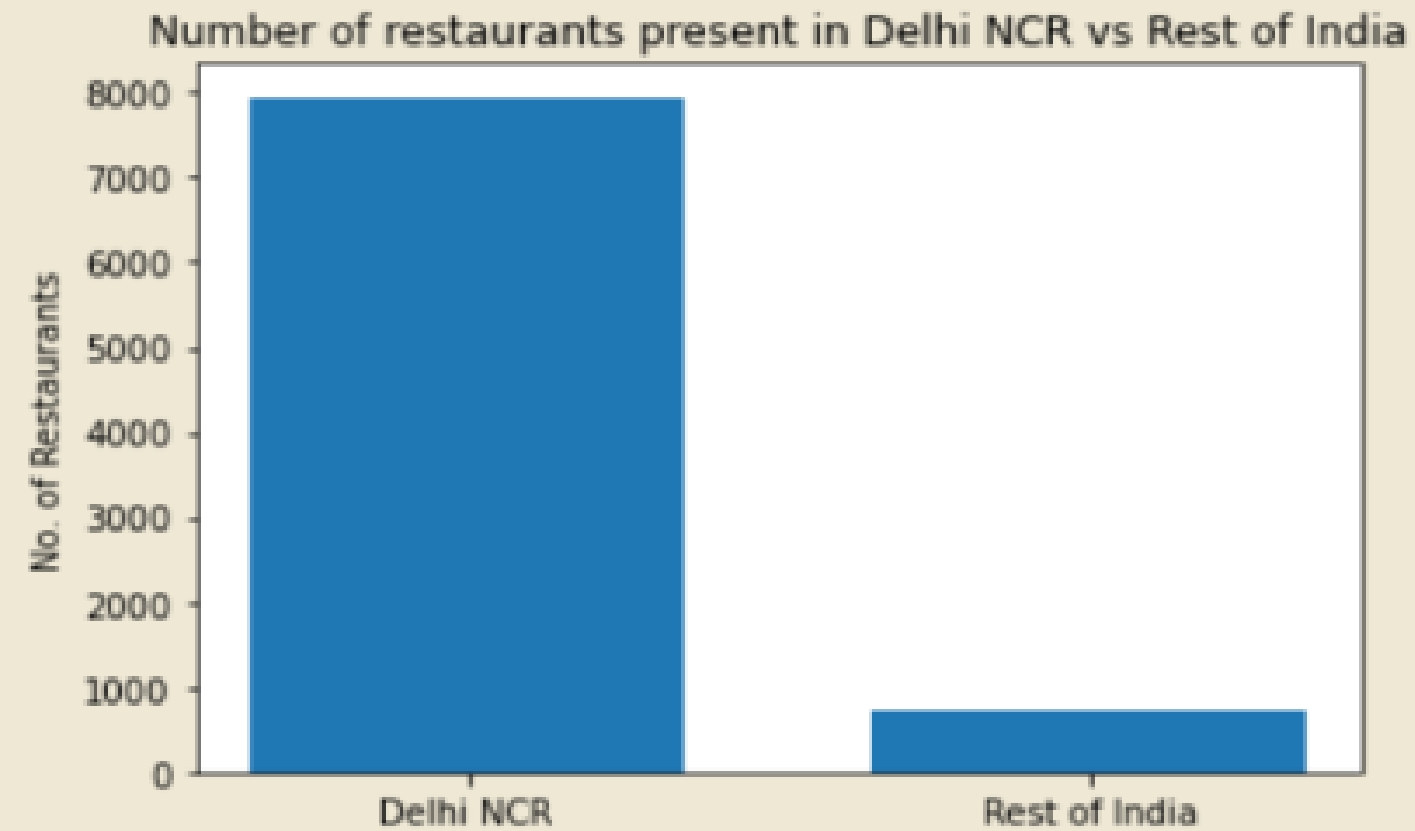
(1.) Plot the bar graph of number of restaurants present in Delhi NCR vs Rest of India.

- For this, we have to first make separate dataframes of Delhi NCR and Rest of India, by implementing the conditions of cities ("New Delhi", "Ghaziabad", "Noida", "Gurgaon", "Faridabad") to get Delhi NCR dataframe. And for Rest of India dataframe, (Country code == 1) and then the reverse condition for these 5 given cities.
- Now count the number of restaurants in both the dataframes using 'Restaurant ID' column.
- Store the counts and name of categories in different lists, will be used in plotting.
- Now using matplotlib.pyplot library, we can plot the bar graph.

No. of restaurants present in :-

Delhi NCR : 7947

Rest of India : 785



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

- First identify unique cuisines of Delhi NCR and Rest of India region, and store them in different lists. For identifying uniqueness, apply check condition.
- Now form an empty list, which will store our final answers.
- As we have to find cuisines which are present in Rest of India but not in Delhi NCR. So iterate over Rest of India's list, and check which cuisines are not in Delhi NCR list.
- Append them in our empty list while iteration. Now this list is our answer.

```
Cuisines which are not present in restaurant of Delhi NCR but present in rest of India :  
['German', 'Malwani', 'BBQ', 'Cajun']
```

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

- In previous problem, we used list to get unique cuisines. But here use dictionary, through which we will get unique cuisines as well as there countings in respective dataframes.
- Now sort the dictionary using values(), and in reverse order.
- Now get first 10 entries from the respective sorted dictionary, which will be top 10 served cuisines of the restaurants of corresponding regions.
- The answer printed by the code are in next slide.

Top 10 cuisines served by maximum number of restaurants in Delhi NCR :-

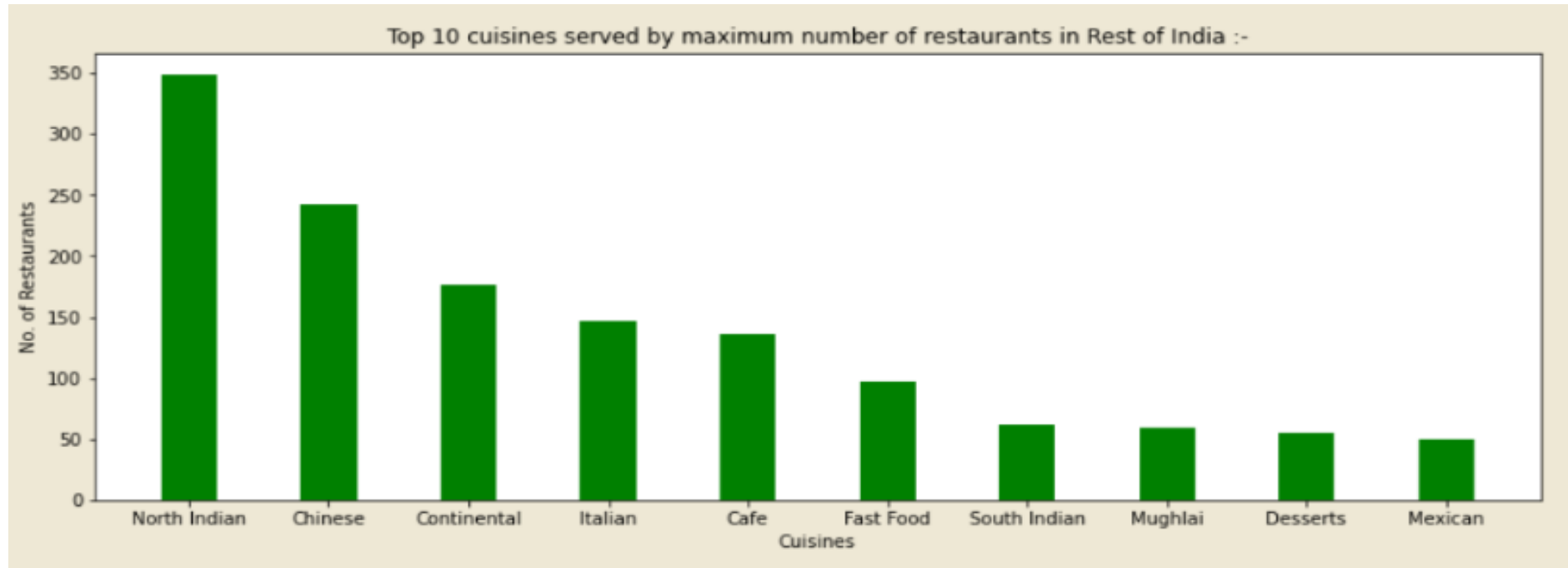
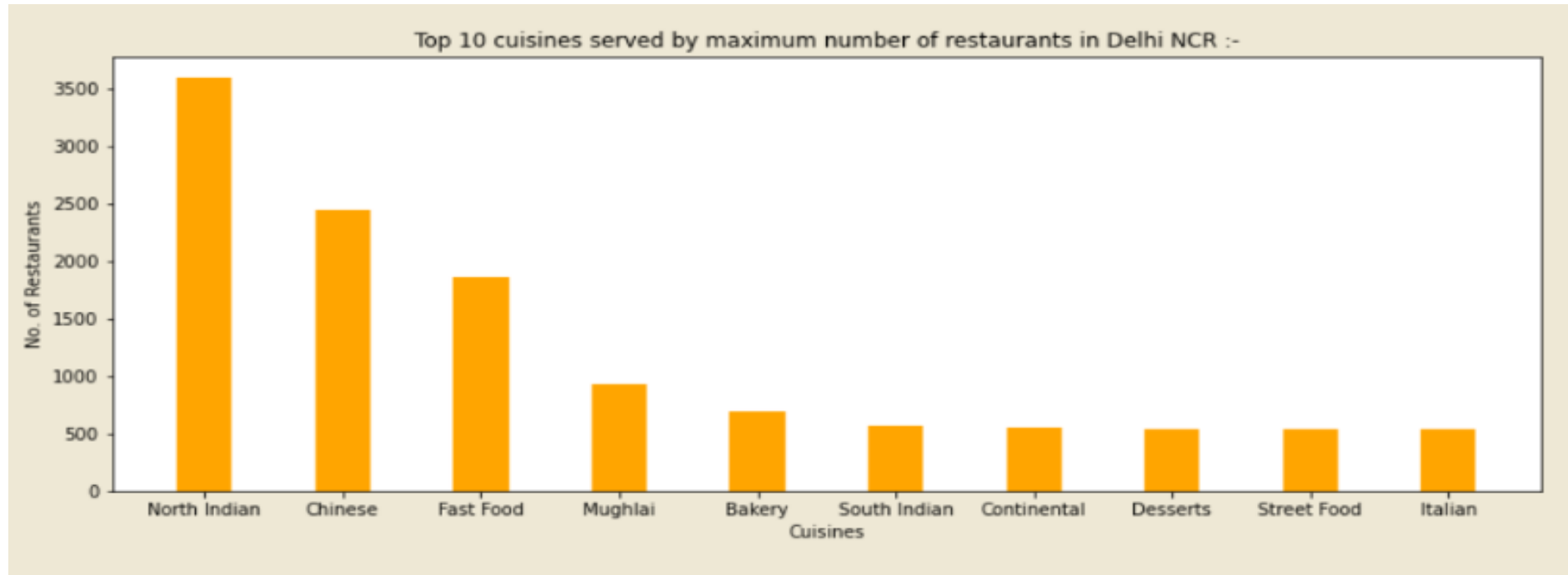
- 1 . North Indian - 3597
- 2 . Chinese - 2448
- 3 . Fast Food - 1866
- 4 . Mughlai - 933
- 5 . Bakery - 697
- 6 . South Indian - 569
- 7 . Continental - 547
- 8 . Desserts - 542
- 9 . Street Food - 538
- 10 . Italian - 535

Top 10 cuisines served by maximum number of restaurants in Rest of India :-

- 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

(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.

- The Cuisine serving distribution ratio is similar in both regional categories, just the cuisines differ.
- 'North Indian' and 'Chinese' cuisines have higher popularity in both regions.
- 'South Indian' and 'Deserts' also share similarity in both regions, i.e. almost at bottom of top 10 list.
- 'Continental' and 'Italian' have higher servings in Rest of India, but not in Delhi NCR region.
- 'Fast food' and 'Mughlai' have higher servings in Delhi NCR region, but not in Rest of India.
- 'Bakery' is served in Delhi NCR but not quite often in Rest of India.
- 'Mexican' is served in Rest of India but not quite often in Delhi NCR.

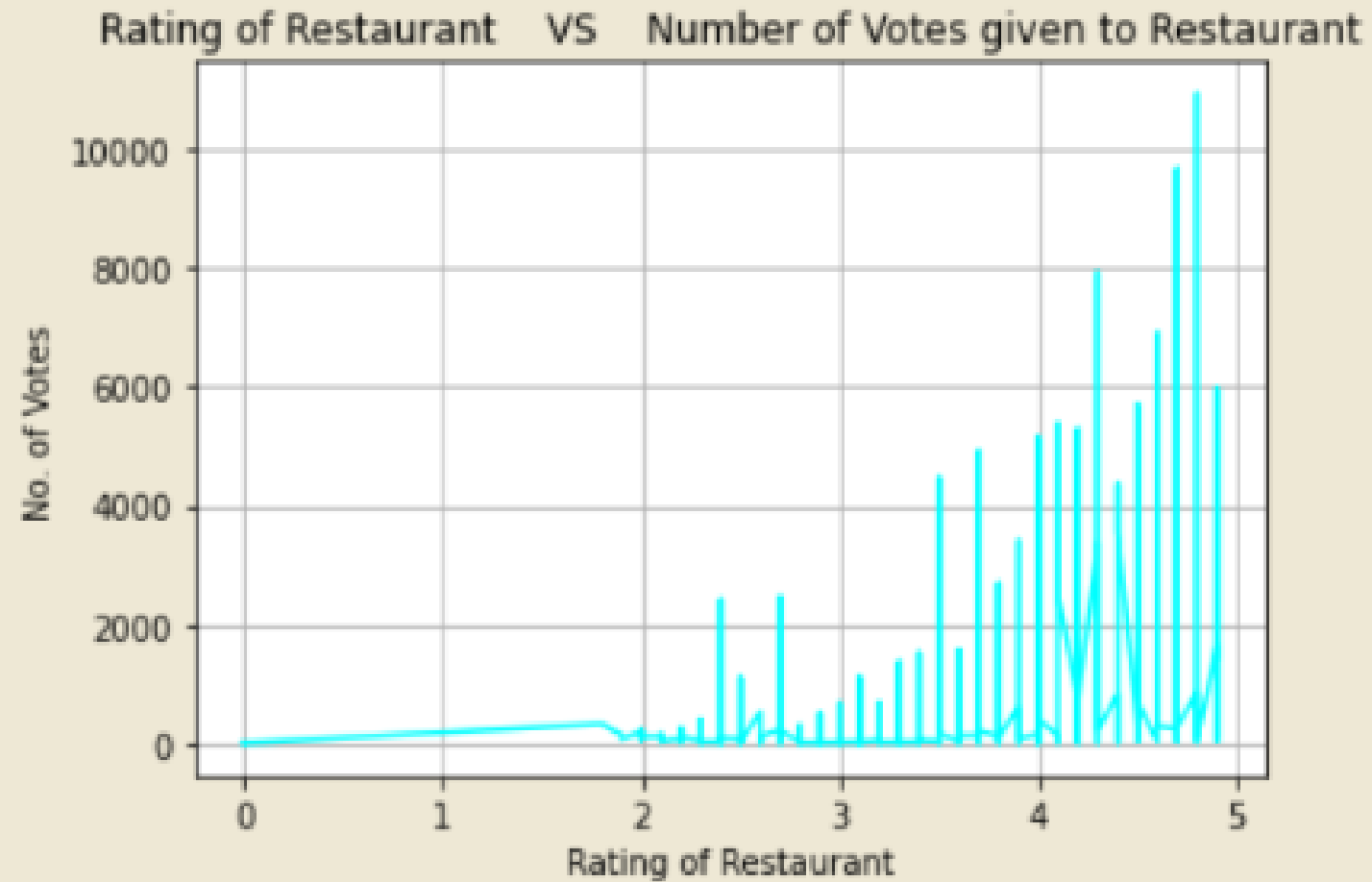


2. User Rating of a restaurant plays a crucial role in selecting a restaurant or ordering the food from the restaurant.

(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.

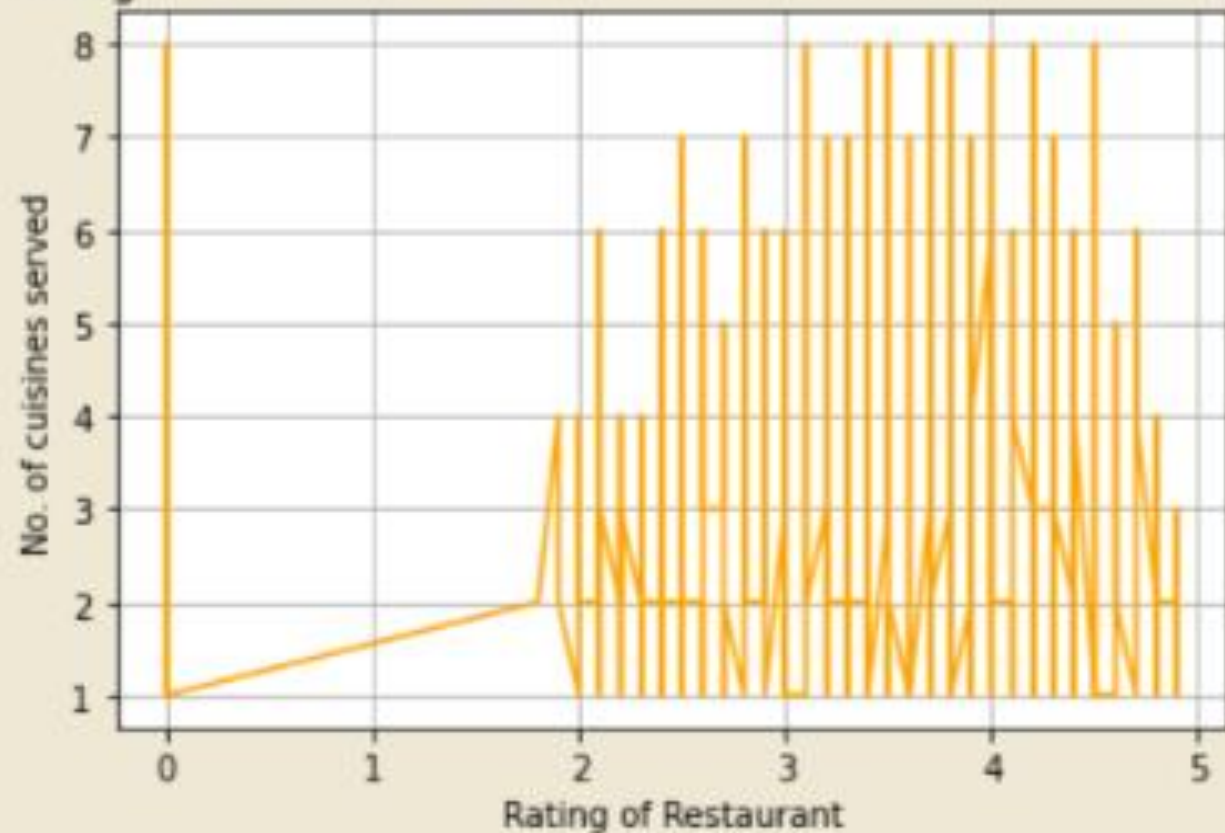
- 1. Number of Votes given Restaurant
- 2. Restaurant serving more number of cuisines.
- 3. Average Cost of Restaurant
- 4. Restaurant serving some specific cuisines.

- The plots and the short detail analysis are stated in the next slides.
- For first 3 cases, there are plots along with analysis statement, but for 4th case there is no plot as the reason stated is against the outcome that will come up from plot.
- For 4th case, the plot can be formed by categorising the ratings based on cuisine (using dictionary) and summing up the ratings of corresponding categories. And simultaneously building another dictionary which stores the counts of corresponding cuisines. After getting both dictionaries, dividing the 1st dictionary values by 2nd dictionary values will give the data too plot the graph.

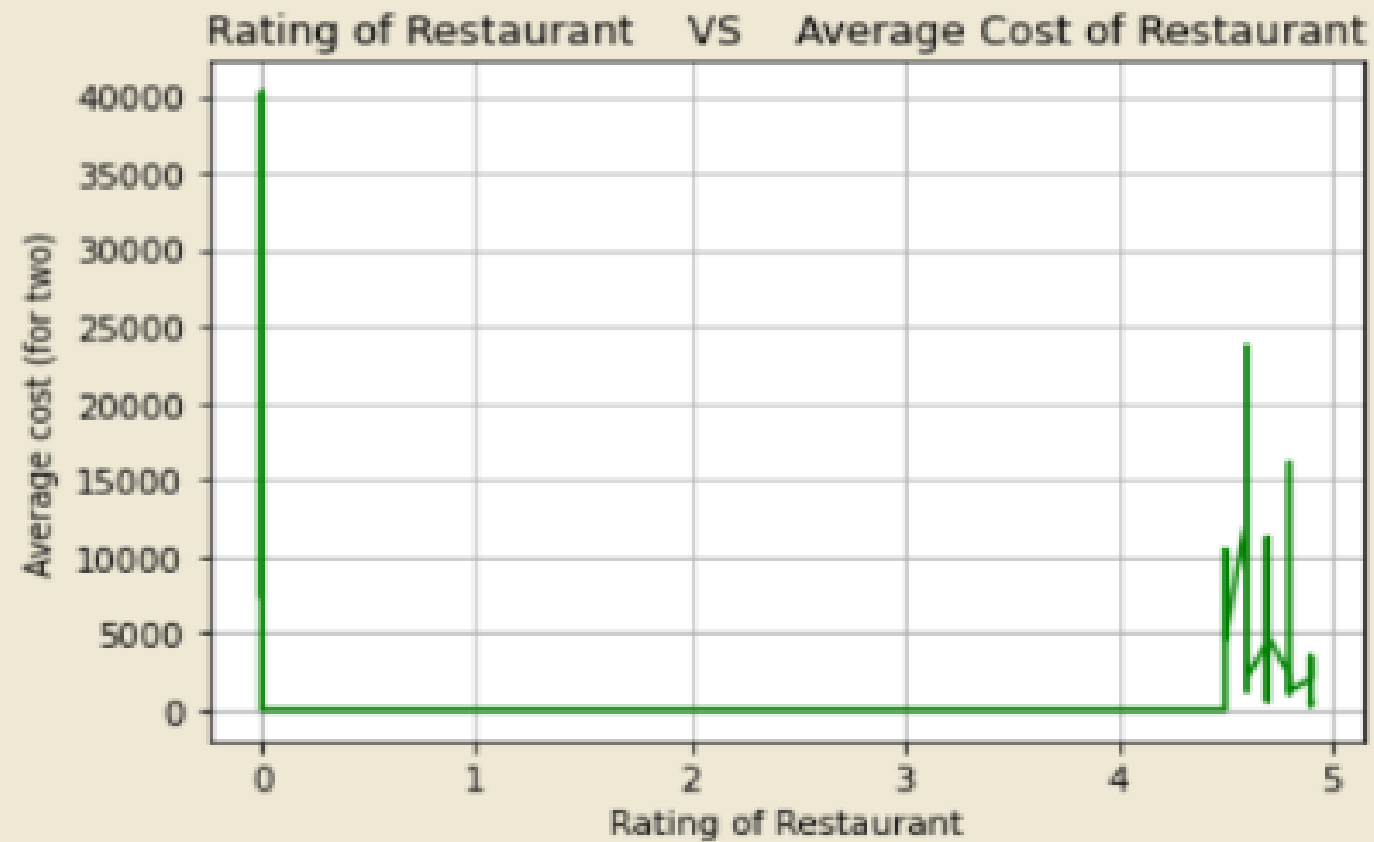


As no. of votes given to a restaurant increases, it's rating also increases, as seen in plot.

Rating of Restaurant VS No. of cuisines served in a Restaurant



When no. of cuisines served are maximum i.e. about 7 to 9, rating is average, i.e. about 3 to 4. And when rating of restaurants get at it's peak, no. of cuisines are mostly around 3 to 6.



When Average cost is neither too large nor too small, the ratings are good.
Otherwise, when cost is at extremes, the ratings of restaurants are bad.

As different cuisines are served in different regions, so people from different regions will rate a same cuisine differently. So just by knowing cuisines served by a restaurant, you can't predict its rating.

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 = $\Sigma (\text{number of votes} * \text{rating}) / \Sigma (\text{number of votes})$.

- First work on the formula without applying the filter based on localities, i.e. get Total number of votes and then,
- $((\text{No. of votes}) * (\text{rating})) / (\text{total votes})$, row wise. \rightarrow Weighted rating
- Now for getting average rating based on localities, first get sum and count of these weighted ratings.
- The divide the sums by the counts to get average. And get top 10 entries out of the sorted dictionary in reverse order.

Top 10 localities with more weighted restaurant rating :-

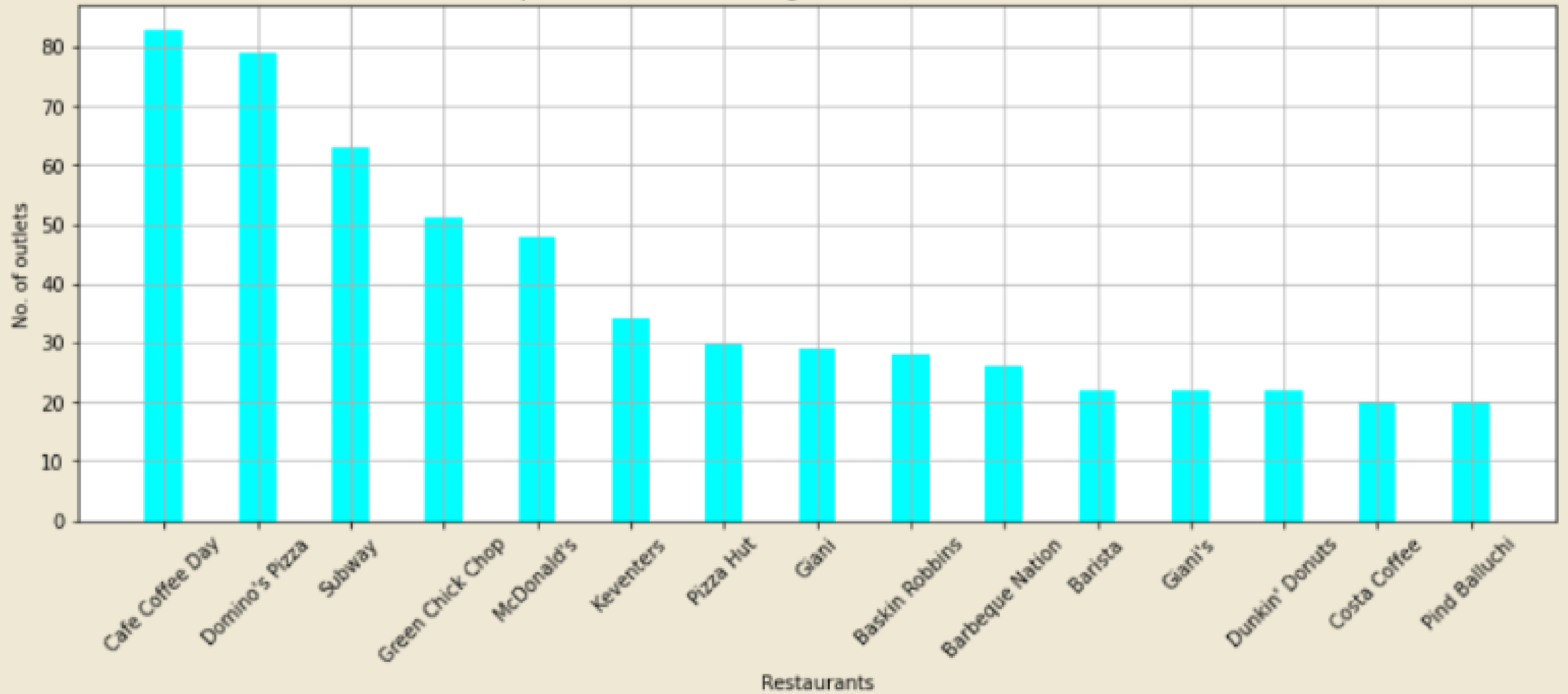
- 1 . Connaught Place - 0.3291606084162694
- 2 . Cyber Hub, DLF Cyber City - 0.09305559355284274
- 3 . Hauz Khas Village - 0.08553740211991501
- 4 . Khan Market - 0.07649016278037828
- 5 . Rajouri Garden - 0.06772137497539445
- 6 . Indiranagar - 0.060105094935758635
- 7 . Sector 29 - 0.0563167394546407
- 8 . Park Street Area - 0.05270861344748089
- 9 . Sector 18 - 0.05226474582039108
- 10 . Delhi University-GTB Nagar - 0.05139489338702628

3. Visualization

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

- For getting no. of outlets count, use 'Restaurant Name' column.
- Use dictionary to count no. of occurrence of a unique name in the dataframe column.
- Sort the dictionary in reverse order based on the values() of dictionary.
- Get the top 15 entries from the sorted dictionary to get top 15 restaurants have a maximum number of outlets.
- Print the bar graph by using matplotlib.pyplot.

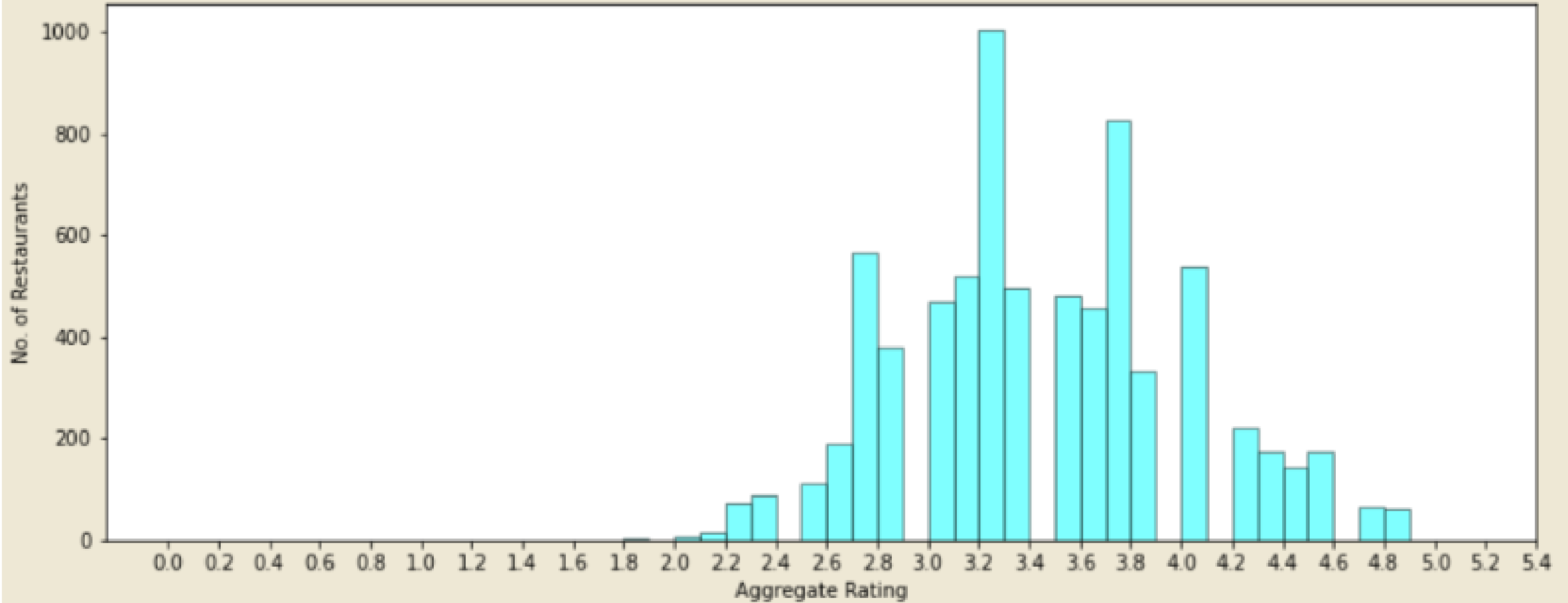
Top 15 Restaurants having maximum number of Outlets



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

- We have to drop the unrated restaurant, so 'Rating text' column will be used, and the filter those entries having entry 'Not rated' in this column.
- Now chose only those entries from the 'Aggregate rating' column, corresponding to which the condition in above point is fulfilled.
- Now .hist() method of matplotlib.pyplot will plot the histogram based on the no. of occurrences on different entries in the list of chosen entries.

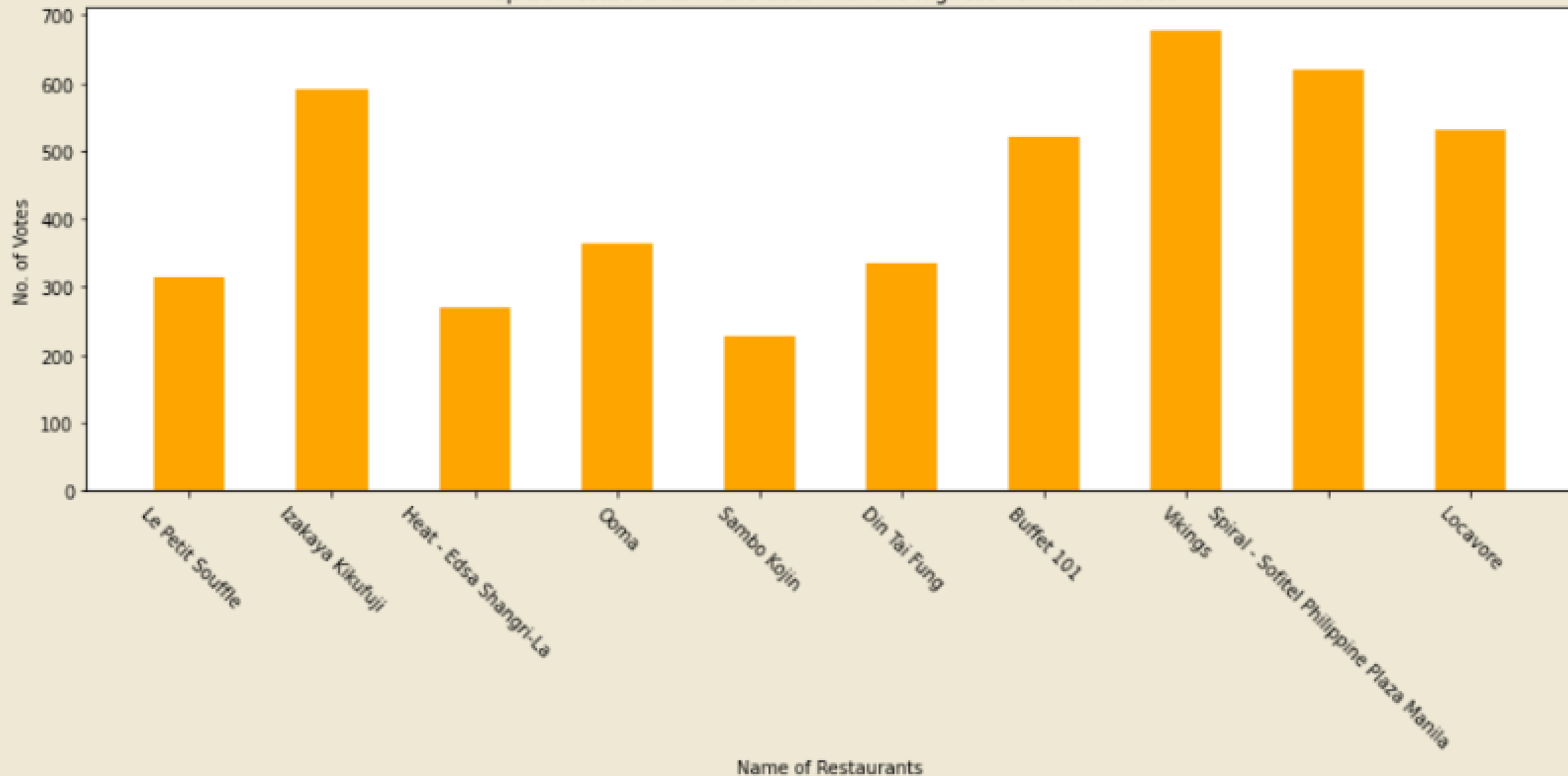
Histogram of aggregate rating of restaurant :-



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

- Sort the whole dataframe based on 'Votes' column. Now top 10 restaurants in the data with the highest number of votes are already the top 10 entries of the dataframe.
- Now get the Restaurant names & No. of votes of the top 10 entries.
- Plot the bar graph using `.bar()` functionality of the library `matplotlib.pyplot`.

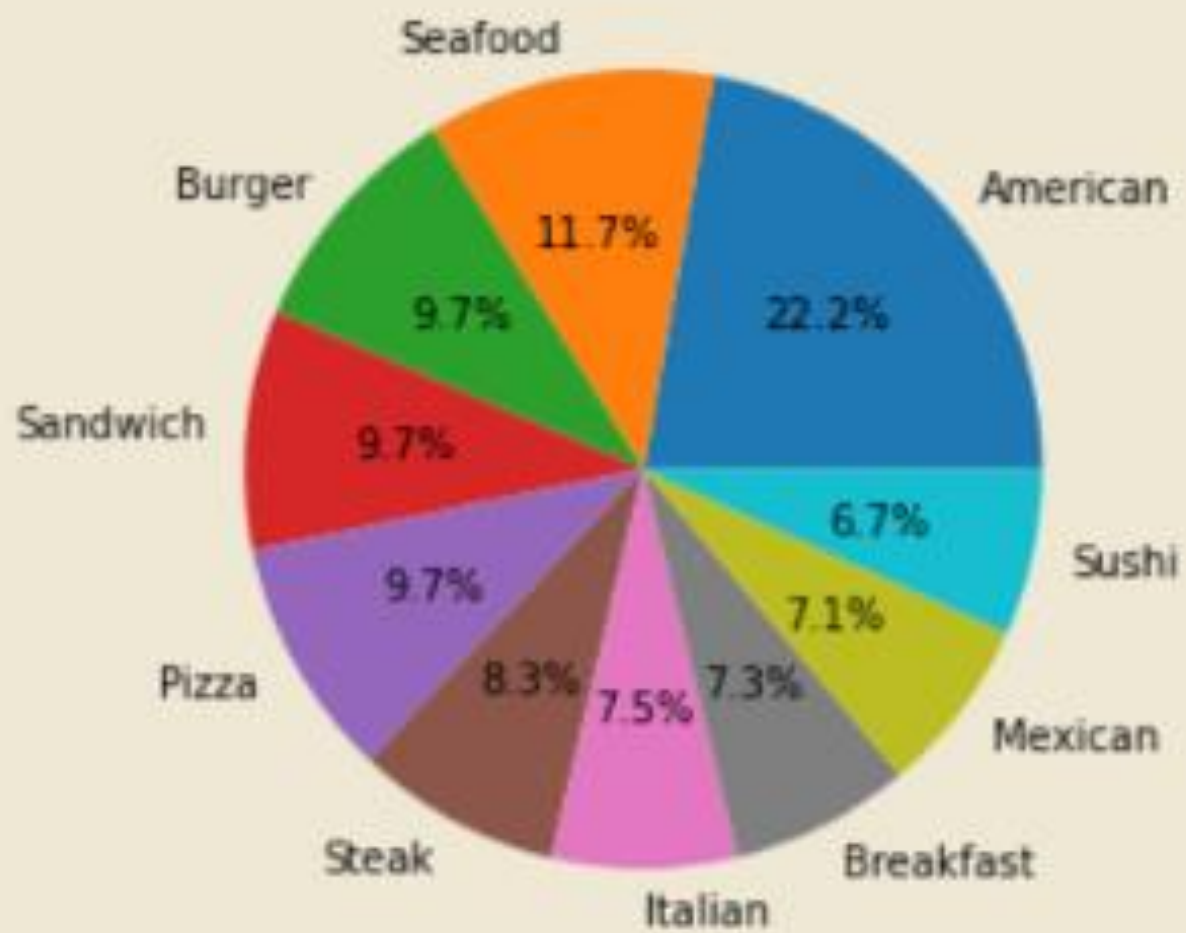
Top 10 Restaurants in the data with the highest number of Votes



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

- Like we got dataframe filtered on the basis of location in previous problem (Delhi NCR vs Rest of India). Similar process will be done in here based on 'Country Code' column, and it's entry referring to USA, i.e. 216.
- Now from the filtered dataframe, get counting of unique cuisines using python dictionaries.
- Sort the dictionary in reverse order and get the first 10 entries, i.e. cuisine names and No. of restaurants serving it.
- Plot pie graph using .pie() functionality of matplotlib.pyplot library.

Top 10 cuisines present in restaurants in 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.

- First get the weighted restaurant rating of the city, as we have got that of localities in a previous problem. Just by using 'City' column inspite of 'Locality' column and applying similar approach.
- Then using python dictionaries, get no. of restaurants in each unique cities.
- Now get the three values, i.e. City_name , Number_of_restaurants , City_weighted_res_rating in three different lists, because the function used to plot bubble graphs only operate on lists.
- Plot bubble graph using .scatter() functionality of matplotlib.pyplot library.
- As there are so many cities of India in this dataframe, so they are not visible clearly in the bubble graph axis labelling.

Bubble graph

