

Zomato API

Analysis on Data Collected By Zomato API

Problem:-

1.) The dataset is highly skewed toward the cities included in Delhi-NCR. So, we will summarise all the other cities in the 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 vs the Rest of India.

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

2.) Find the cuisines which are not present in the restaurant of Delhi NCR but present in the Rest of India.

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

4.) Write a short detailed analysis of how cuisine served is different from Delhi NCR to the Rest of India. Plot the suitable graph to explain your inference.

Solution Description:

For each sub-problem:

- Import the libraries.
- Load data in dataframe.
- Extract Indian Restaurant's data.
- Add a column storing region.
- Reset Index

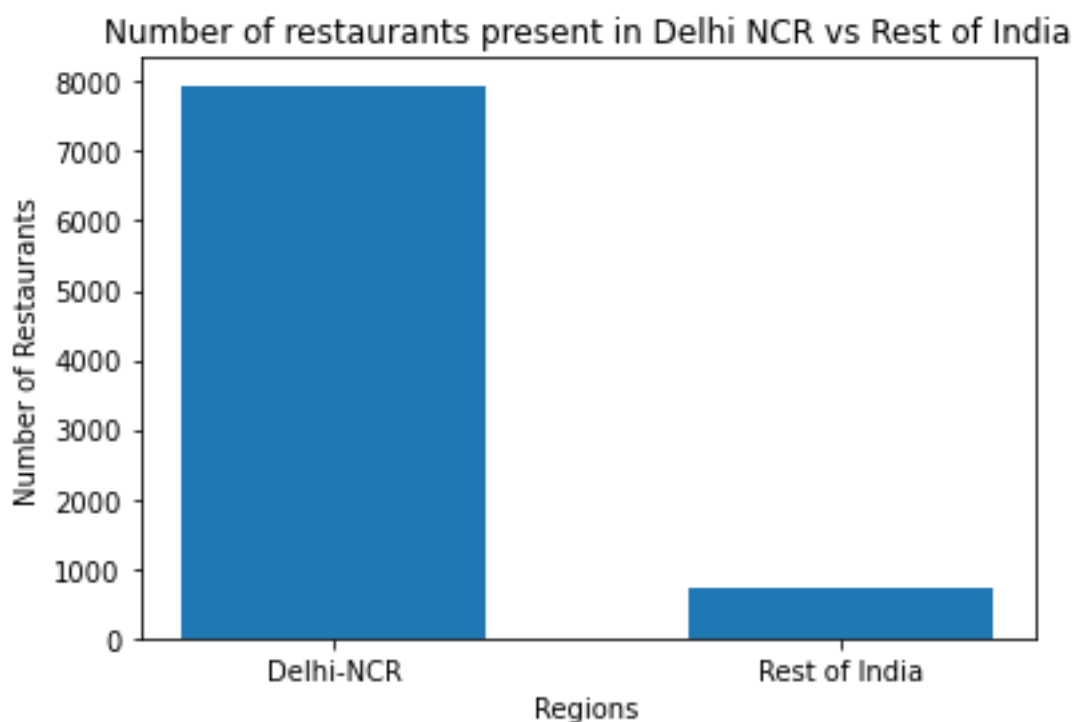
Solution steps for 1.1):-

- a) Use `value_counts()` method on Region Column and get the number of outlets in each region.
- b) Plot the bar graph.

Result:-

{Delhi-NCR: 7947, Rest of India: 705}

Bar Graph:-



Solution steps for 1.2):-

- a) Initiate 2 empty sets, for Delhi-NCR and the Rest of India.
- b) Iterate each row of dataframe and store cuisines based on the region in the respective sets.
- c) Now subtract Delhi-NCR's set from the Rest of India's set.

Result:-

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

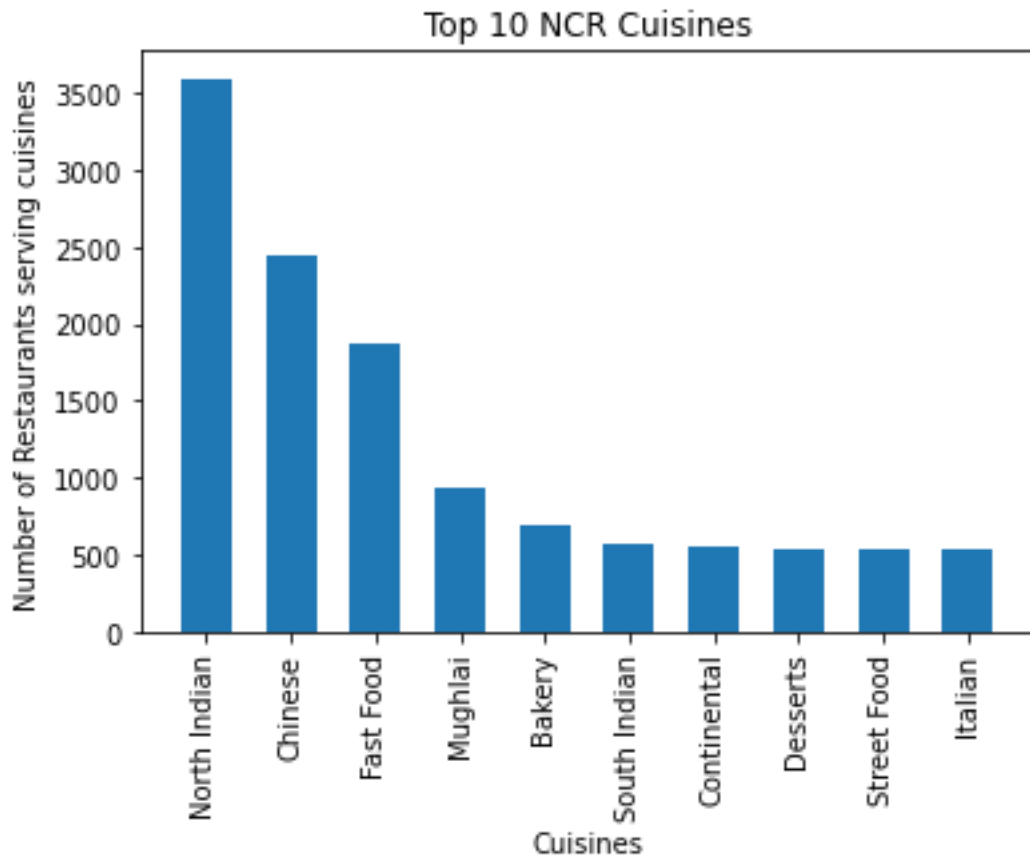
Solution steps for 1.3):-

- a) Extract Delhi-NCR region data from dataframe.
- b) Initiate a dictionary to store the count of cuisines.
- c) Iterate each row of this extracted dataframe and store cuisines and their count based.
- d) Extract the top 10 cuisines with the highest number of cuisines from the dictionary.
- e) Repeat the above steps for the Rest of India.

Result:-

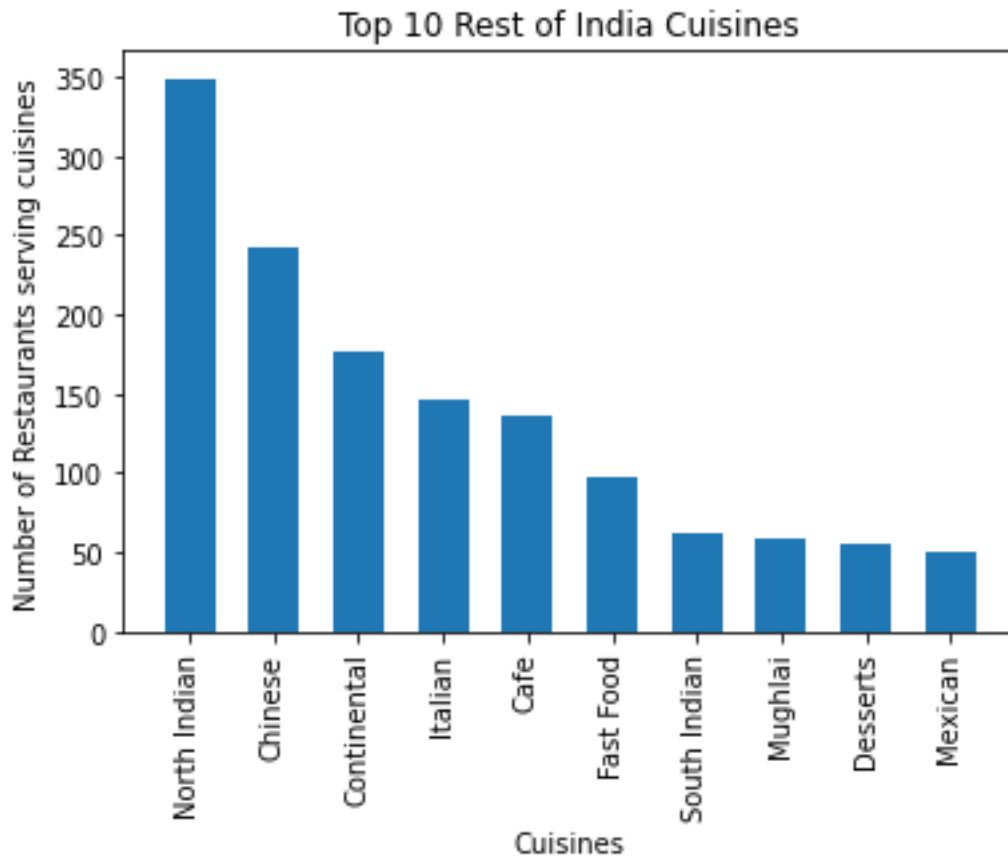
Top 10 Cuisines of Delhi NCR are:

- ***['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 of India except Delhi-NCR Region are:

- ***['North Indian' '349']***
- ***['Chinese' '242']***
- ***['Continental' '177']***
- ***['Italian' '147']***
- ***['Cafe' '136']***
- ***['Fast Food' '97']***
- ***['South Indian' '62']***
- ***['Mughlai' '59']***
- ***['Desserts' '55']***
- ***['Mexican' '50']***



Analysis description of 1.4):-

I have achieved the following results by analyzing the data-

a) Number of Cuisines in both regions:

Delhi-NCR Region: 86

Rest of India Region: 70

b) Unique Cuisines served in one region but not in another region:

After analyzing the cuisines set of both regions, the conclusion is:-

- Unique Cuisines served in Delhi NCR Region, but not in Rest of India are: ***{'Turkish', 'Assamese', 'Bihari', 'Cuisine Varies', 'Pakistani', 'Afghani', 'Kashmiri', 'Raw Meats', 'Sushi', 'Moroccan', 'Drinks Only', 'Belgian', 'Nepalese', 'South***

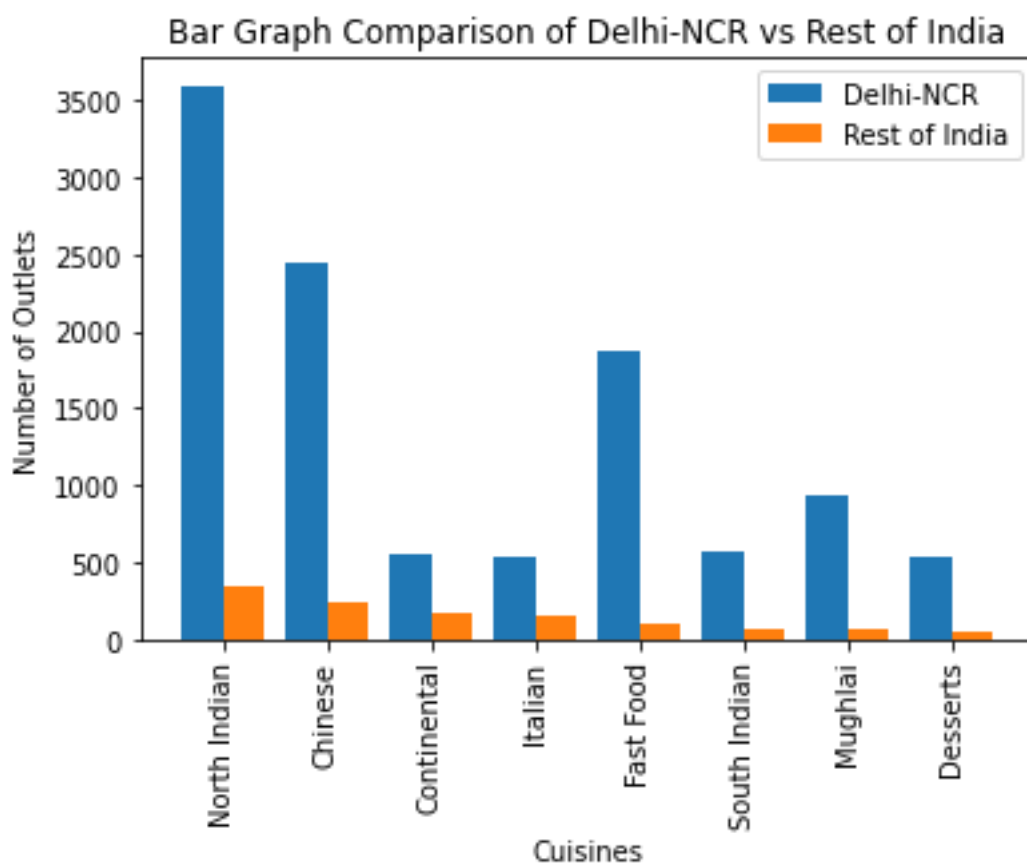
'American', 'Persian', 'Sri Lankan', 'Iranian', 'Naga', 'Deli', 'Oriya'

- Unique Cuisines served in Rest of India Region, but not in Delhi-NCR: ***{'Malwani', 'German', 'BBQ', 'Cajun'}***

c) Comparison of common cuisines in top 10 cuisines of both regions using graphs:-

After analyzing data of both regions top 10 cuisines based on the number of outlets, the conclusion is:-

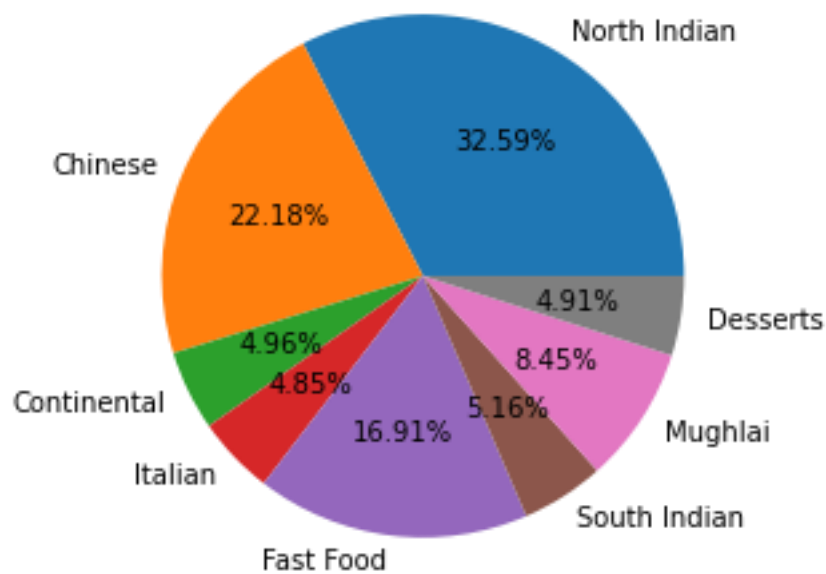
- Common Cuisines among top 10 cuisines of each region are: ***{North Indian, Chinese, Continental, Italian, Fast Food, South Indian, Mughlai, Desserts}***



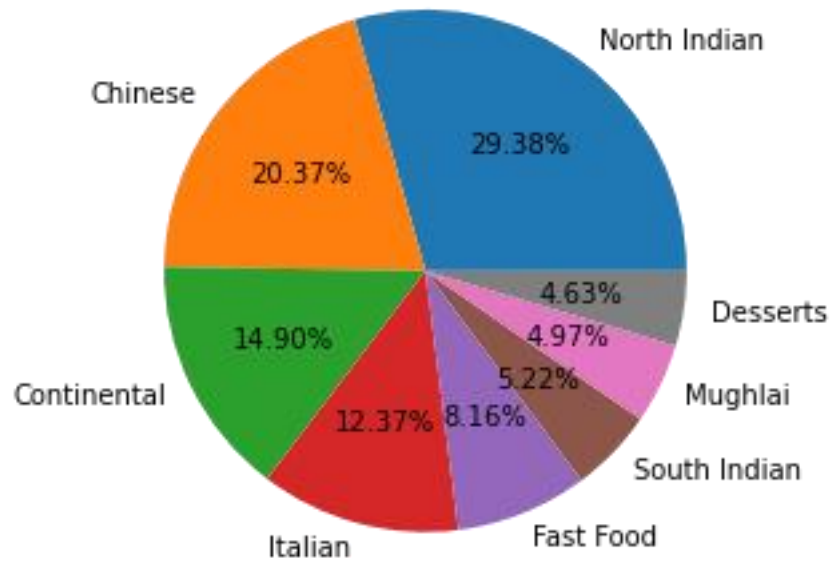
After analyzing pie charts based on Outlet Percentage of both region common top cuisines, the conclusion is:-

- **"North Indian, Chinese, South Indian and Desserts"** cuisines are almost served in the same proportion in both regions.
- **"Continental and Italian"** cuisines are served comparatively higher in the Rest of India.
- **"Fast Food and Mughlai"** cuisines are served comparatively higher in Delhi-NCR Region.

Pie-chart of weightage percentage of Cuisine Delhi-NCR region



Pie-chart of weightage percentage of Cuisine Rest of India Region



Problem:-

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.
2. Find the weighted restaurant rating of each locality and find out the top 10 localities with more weighted restaurant rating?
[Weighted Restaurant Rating = $\frac{\sum (\text{number of votes} * \text{rating})}{\sum (\text{number of votes})}$].

Solution Description:

For each sub-problem:

- Import the libraries.
- Load data in dataframe.
- Extract Indian Restaurant's data.
- Add a column storing region.
- Reset Index

Approach for 2.1.1:- Find average votes with respect to each rating then plot the "bar graph and line graph" to see result.

Solution steps for:-

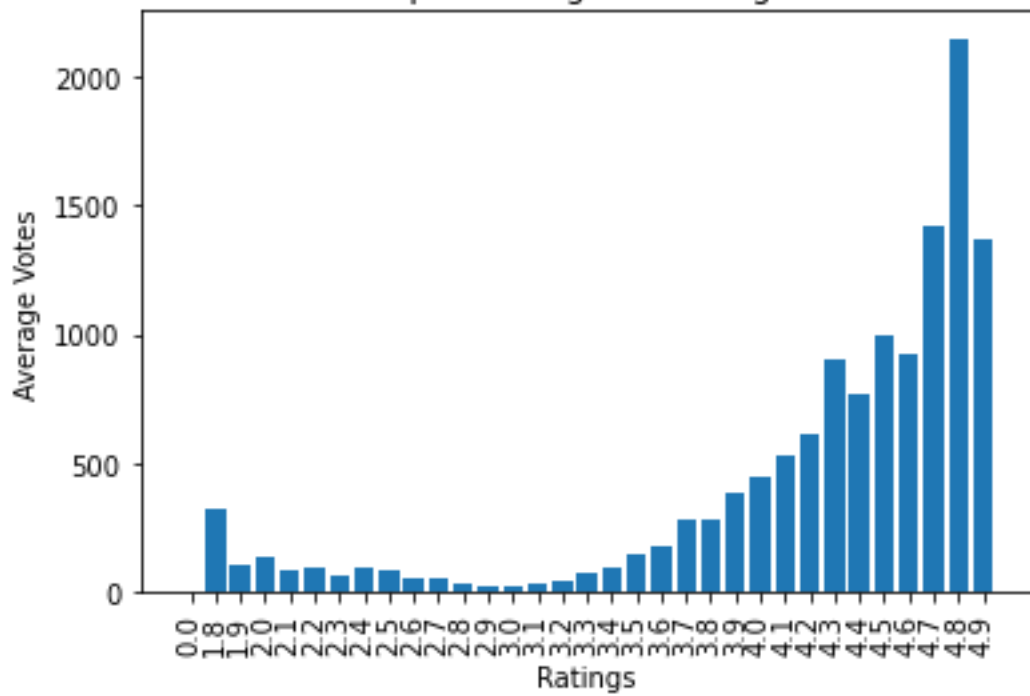
- a) Drop nan values for 'Aggregate Rating and Votes' columns.
- b) Find average votes received for each rating.
- c) Plot the bar graph and curve to visualize the conclusion.

Conclusion:-

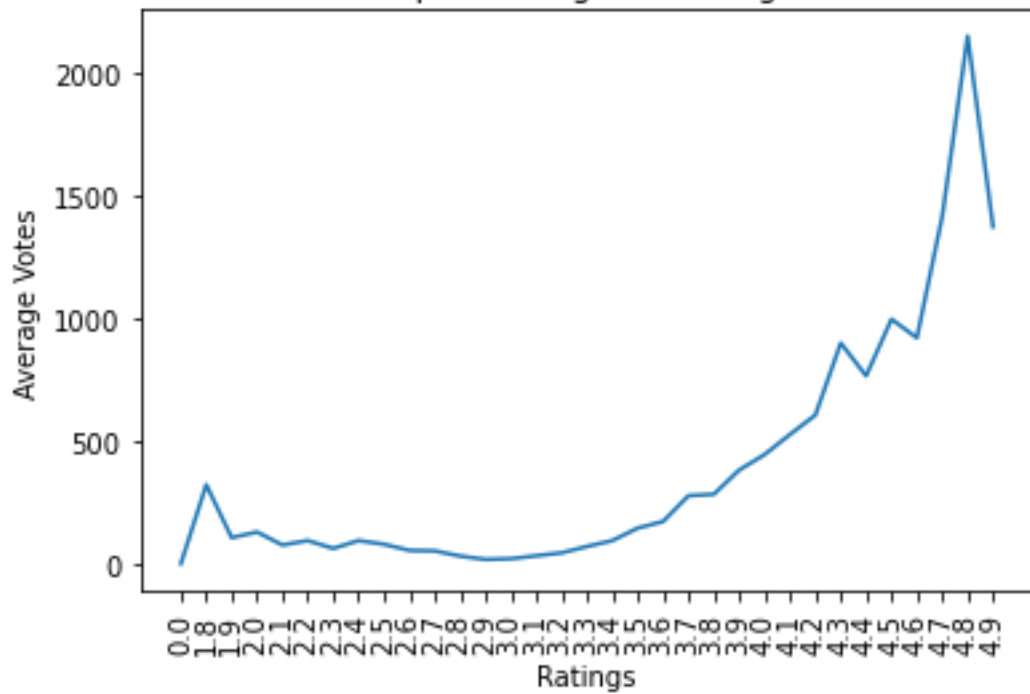
- For rating less than '**3.0**': Number of votes decrease slightly or almost remain same.
- For rating greater than '**3.0**': Number of votes increase significantly, with increase in ratings.
- It can be concluded that the group of restaurants having rating greater than '**4.0**' receive significantly more number of votes.

Graphs:-

Bar Graph - Ratings vs Average Votes



Line Graph - Ratings vs Average Votes



Approach for 2.1.2:- Plot bubble chart between Number of Cuisines Restaurant Serves vs Restaurant's Average weightage rating, using number of restaurants as bubble weights.

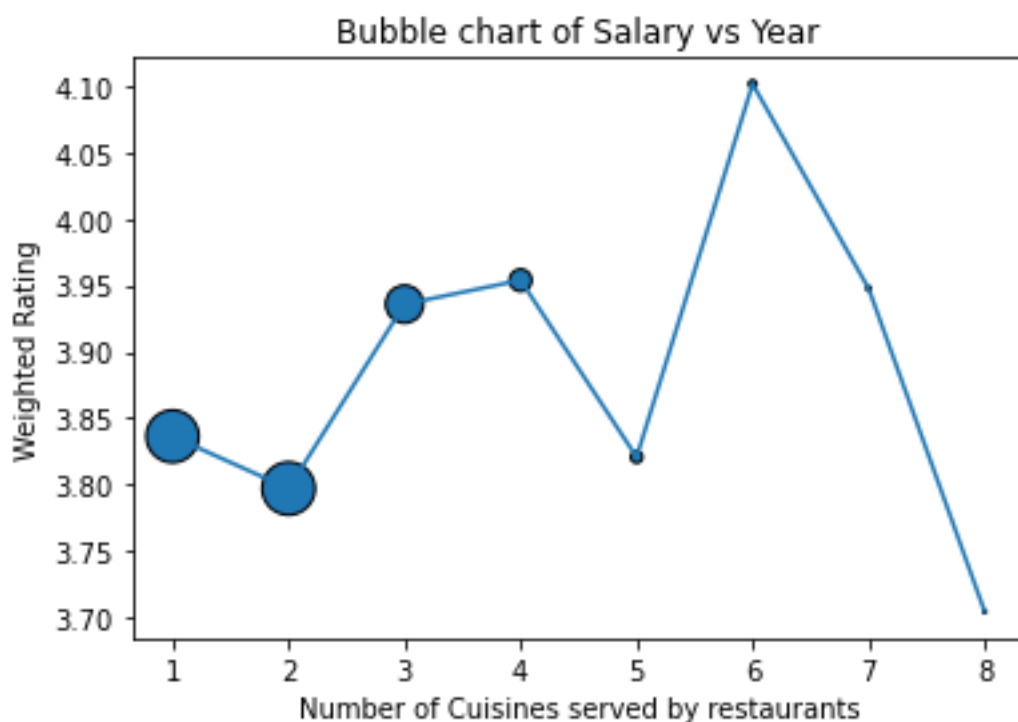
Solution steps for:-

- Drop nan values for 'Cuisines, Aggregate Rating and Votes' columns.
- Make a column containing number of cuisines.
- Group by above column and find average weighted rating.
- Find corpeso for each rating.
- Plot the bubble chart with line curve to visualize the conclusion.

Conclusion:-

- By seeing above chart first thing we can conclude that, there are very less number of outlets serving 5 or more different cuisines.
- Generally restaurant serving 3 or 4 differnt cuisines are more in number and have good rating.

Graphs:-



Approach for 2.1.3:- Find average cost with respect to each rating then plot the "bar graph and line graph" to see result.

Solution steps for:-

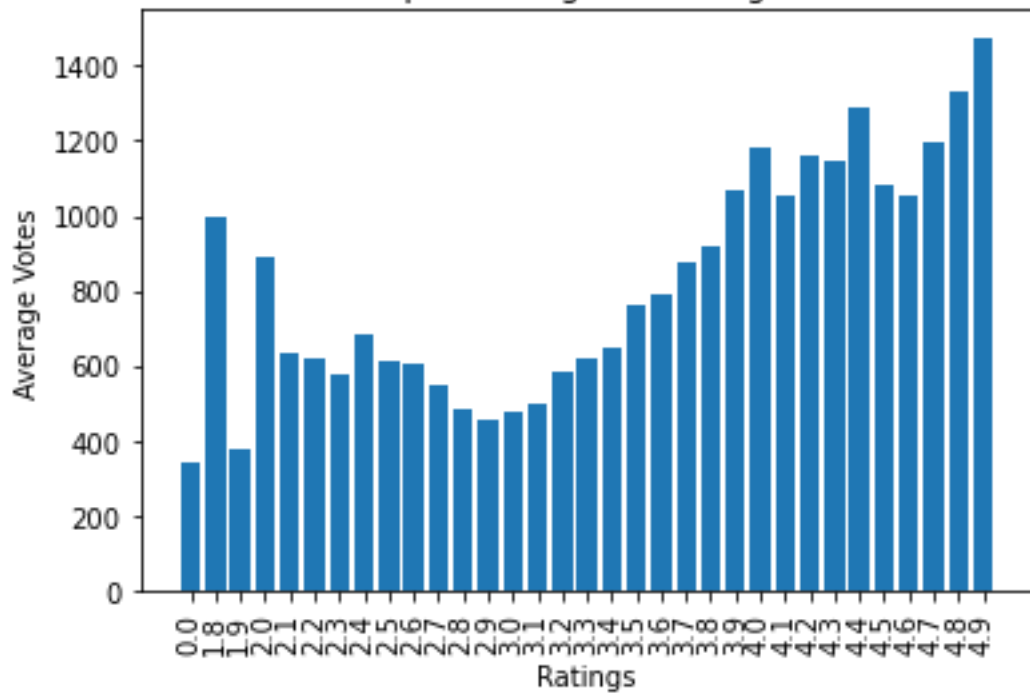
- a) Drop nan values for 'Aggregate Rating and Average Cost for two' columns.
- b) Find average cost received for each rating.
- c) Plot the bar graph and curve to visualize the conclusion.

Conclusion:-

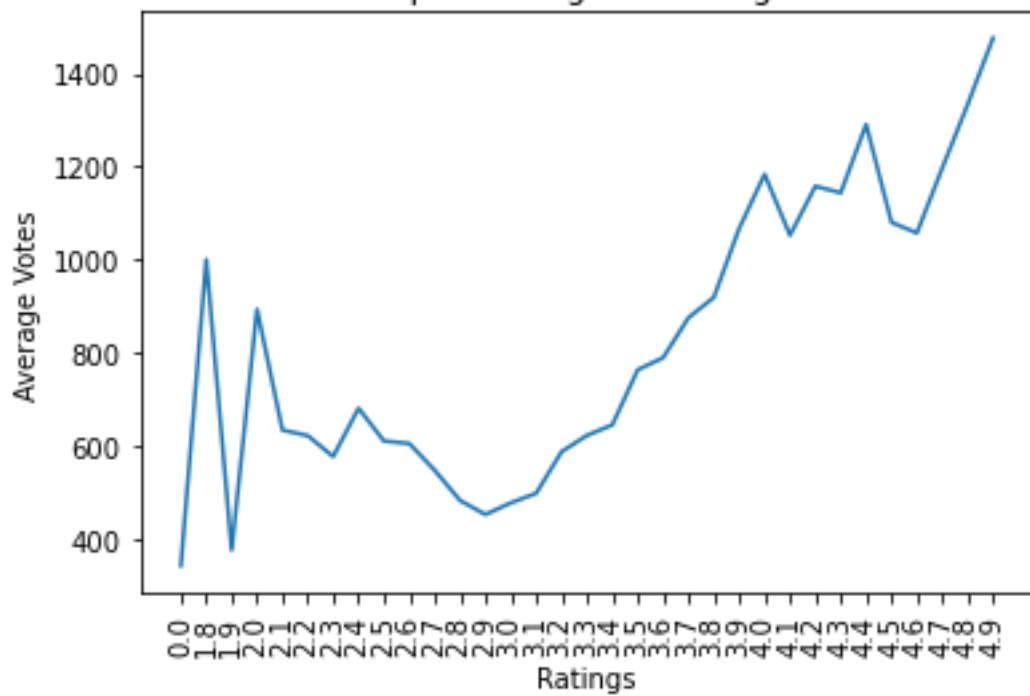
- For rating less than or around '**2.0**': Unable to predict trend.
- For rating between '**2.0**' and '**3.0**': Restaurant's cost generally decreased with increase in rating.
- For rating greater than '**3.0**': Restaurant's cost generally increased with increase in rating.
- It can be concluded that the group of restaurants having rating greater than '**3.9**' are expensive but serves good quality of food service.

Graphs:-

Bar Graph - Ratings vs Average Count



Line Graph - Ratings vs Average Count



Approach for 2.1.4:- Find individual average rating for each cuisine, then plot the bar graph and visualize the result.

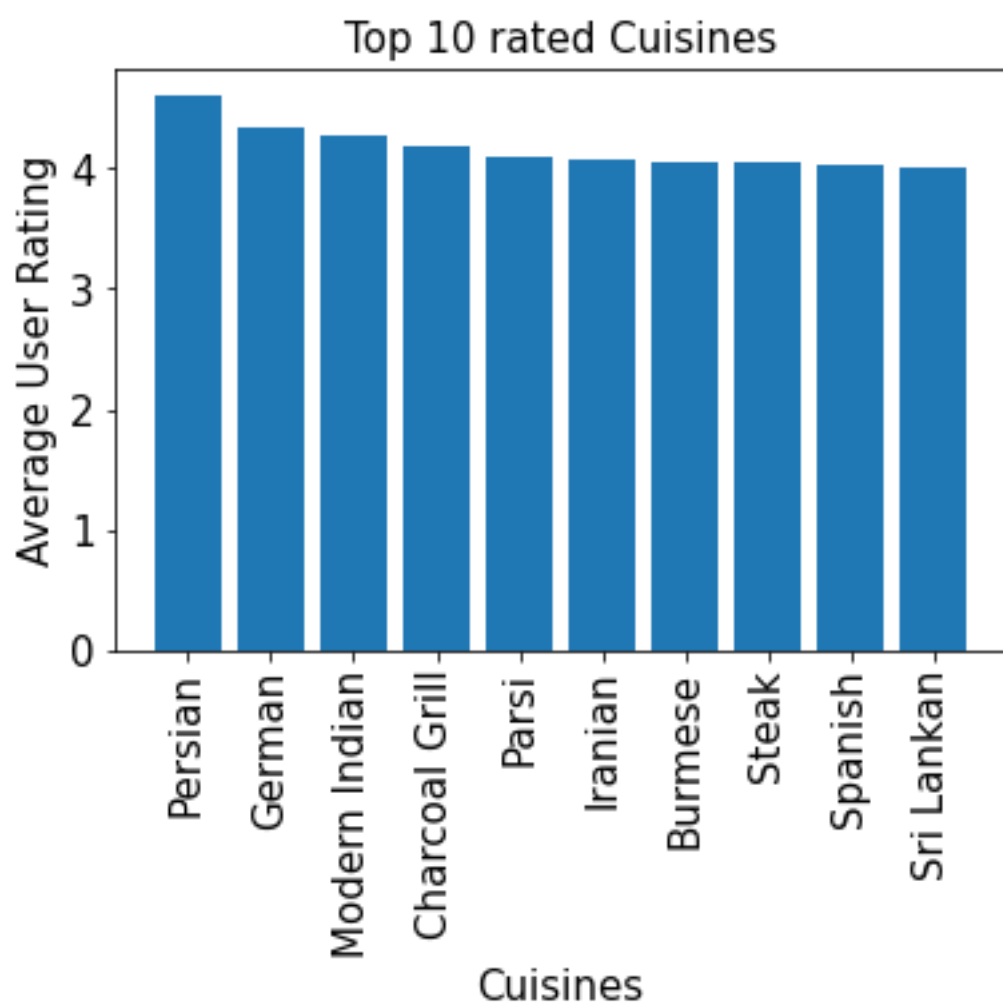
Solution steps for:-

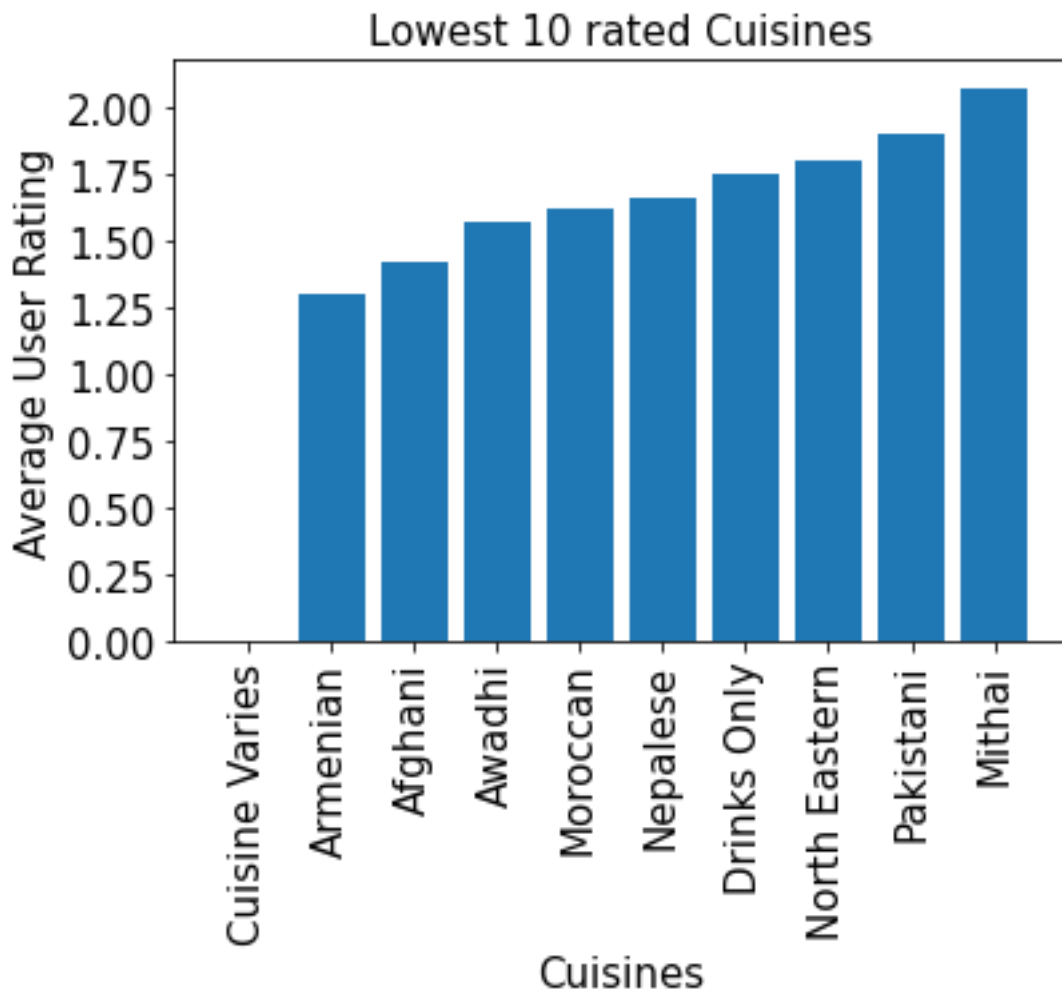
- a) Drop nan values for 'Aggregate Rating and Cuisines columns.
- b) Create a dictionary and store average rating for each cuisine.
- c) Extract top 10 and lowest 10 cuisines accordingly.
- d) Plot the bar graph to visualize the conclusion.

Conclusion:-

- Lowest ten rated cuisines are: '**Cuisine Varies**', '**Armenian**', '**Afghani**', '**Awadhi**', '**Moroccan**', '**Nepalese**', '**Drinks Only**', '**North Eastern**', '**Pakistani**', '**Mithai**'.
- Top ten rated cuisines are: '**Persian**', '**German**', '**Modern Indian**', '**Charcoal Grill**', '**Parsi**', '**Iranian**', '**Burmese**', '**Steak**', '**Spanish**', '**Sri Lankan**'.

Graphs:-





Solution steps for 2.2:-

- Drop nan values for 'Aggregate Rating and Votes' columns.
- Make a column to store product of rating and votes.
- Group by 'Locality Columns' and store corresponding weighted rating

Result:-

- Hotel Clarks Amer 4.9
- Aminabad 4.9
- Friends Colony 4.89

- Powai 4.84
- Kirlampudi Layout 4.82
- Deccan Gymkhana 4.8
- Express Avenue Mall 4.8
- Banjara Hills 4.72
- Sector 5 4.71
- Riverside Mall 4.7

Problem:-

3.) 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.

Solution Description:

For each sub-problem:

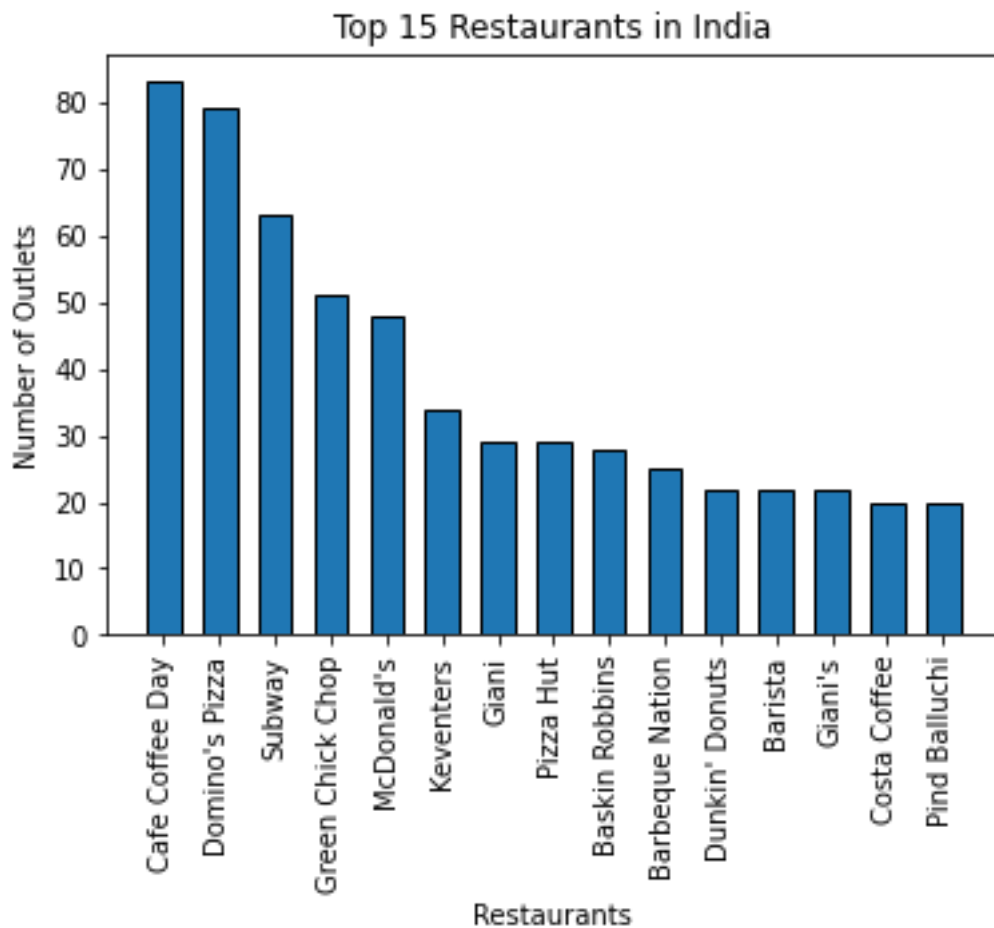
- Import the libraries.
- Load data in dataframe.
- Extract Indian Restaurant's data. (for 3.4 extract data for USA instead).
- Drop nan rows (if required)
- Reset Index. (if required)

Solution steps for 3.1:-

- d) Use value_counts() method to get count wise sorted data of Restaurants.

- e) Extract top 15 results.
- f) Plot the bar graph.

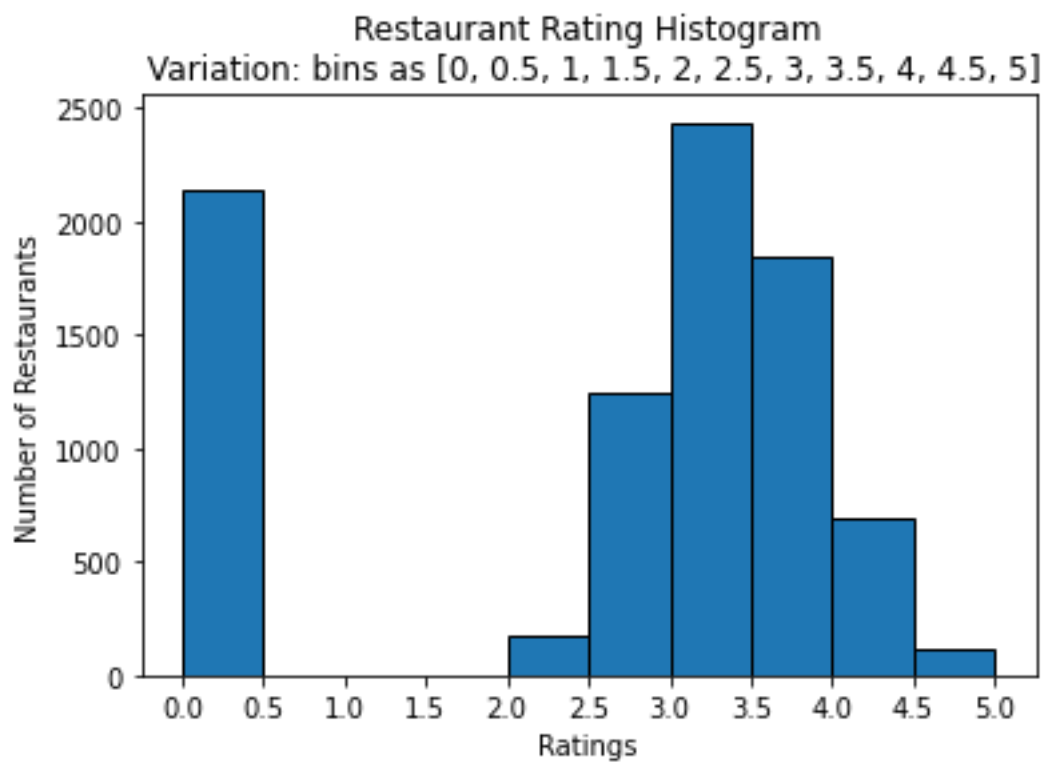
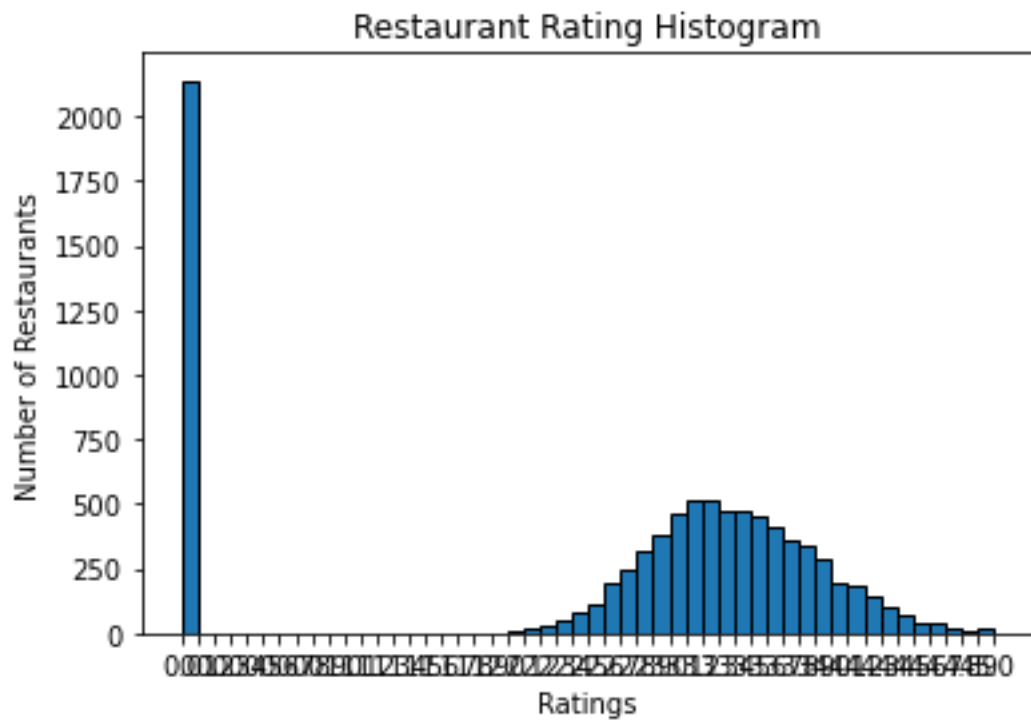
Graphs:-



Solution steps for 3.2:-

- a) Plot the histogram graph for 3 variations.
 - i. Bin size for each rating
 - ii. Bins = [0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5]
 - iii. Auto bins

Graphs:-

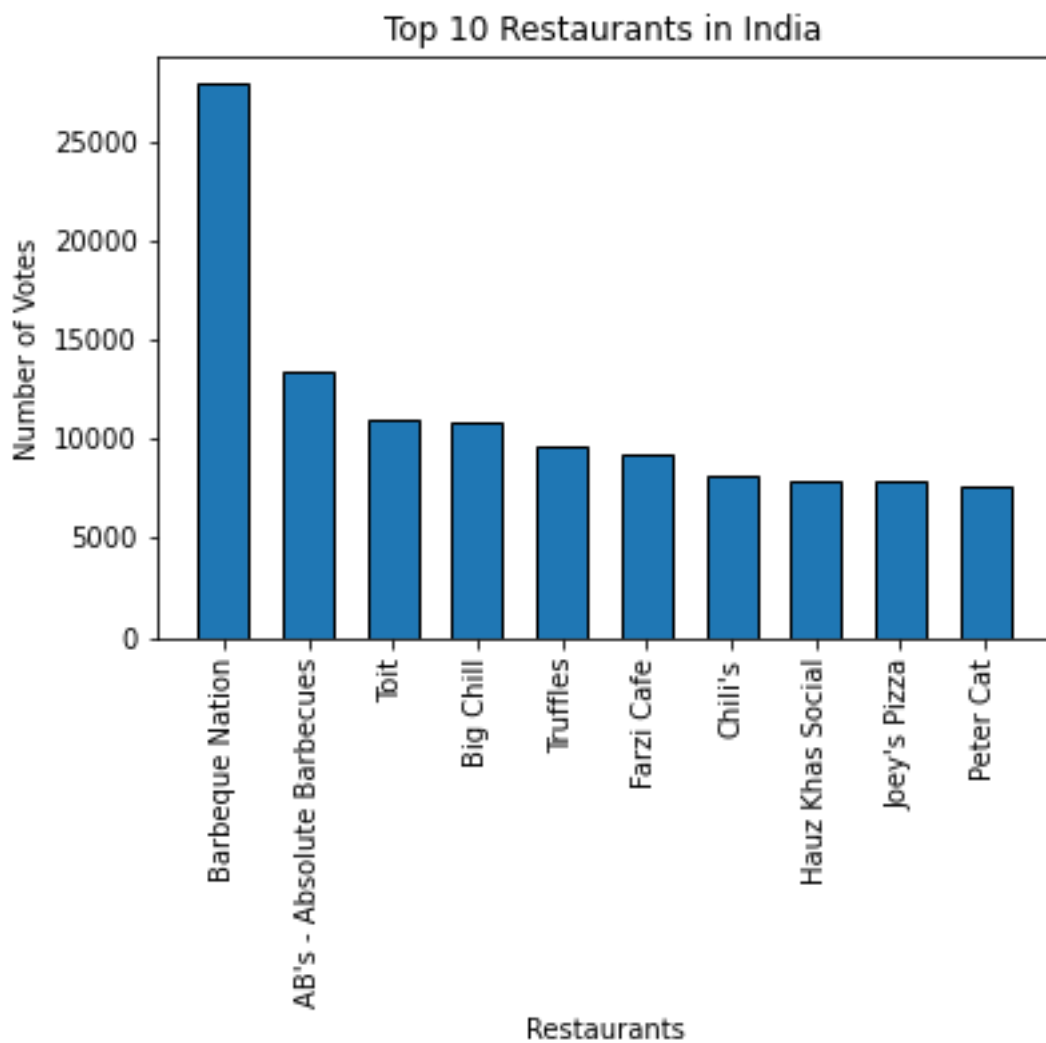




Solution steps for 3.3:-

- Group By Restaurant Name, and store sum of votes for each Restaurant
- Extract top 15 results.
- Plot the bar graph.

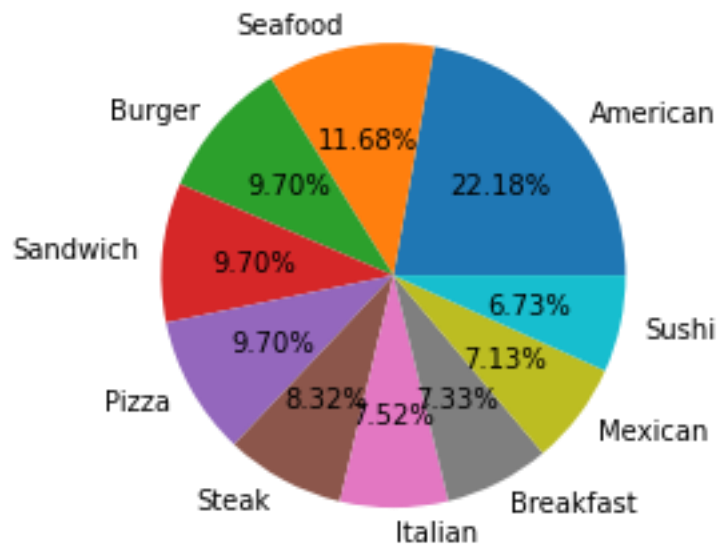
Graphs:-



Solution steps for 3.4:-

- Create a dictionary to store cuisines and their count
- Extract top 10 results.
- Plot the pie graph.

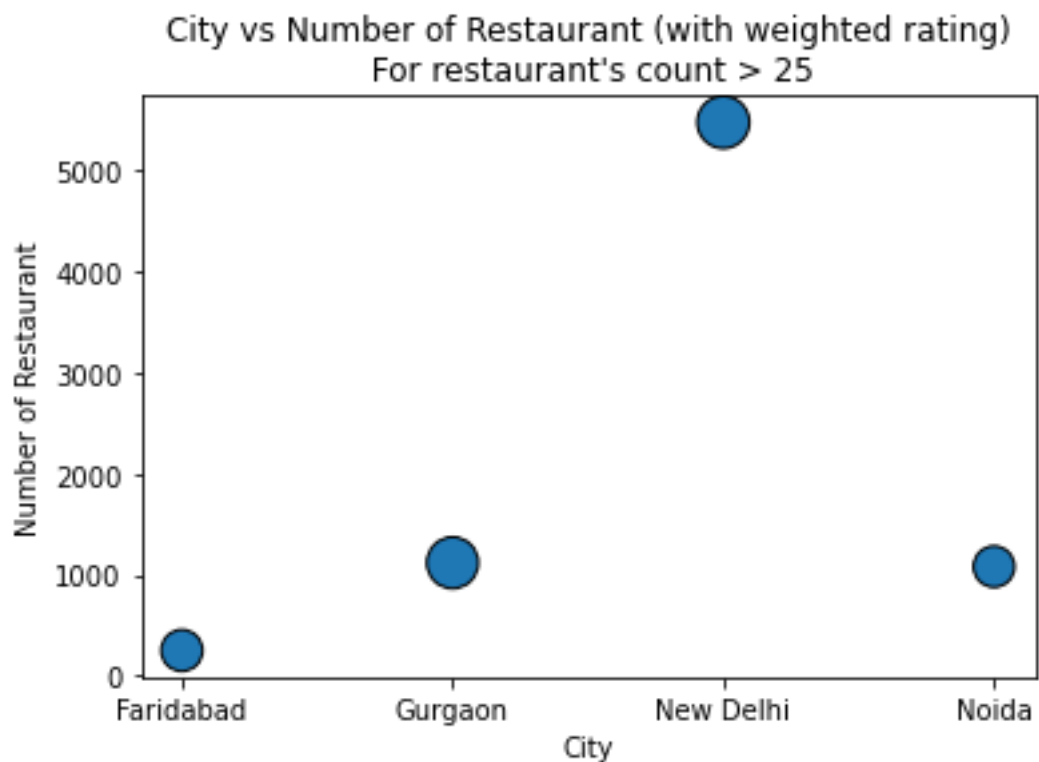
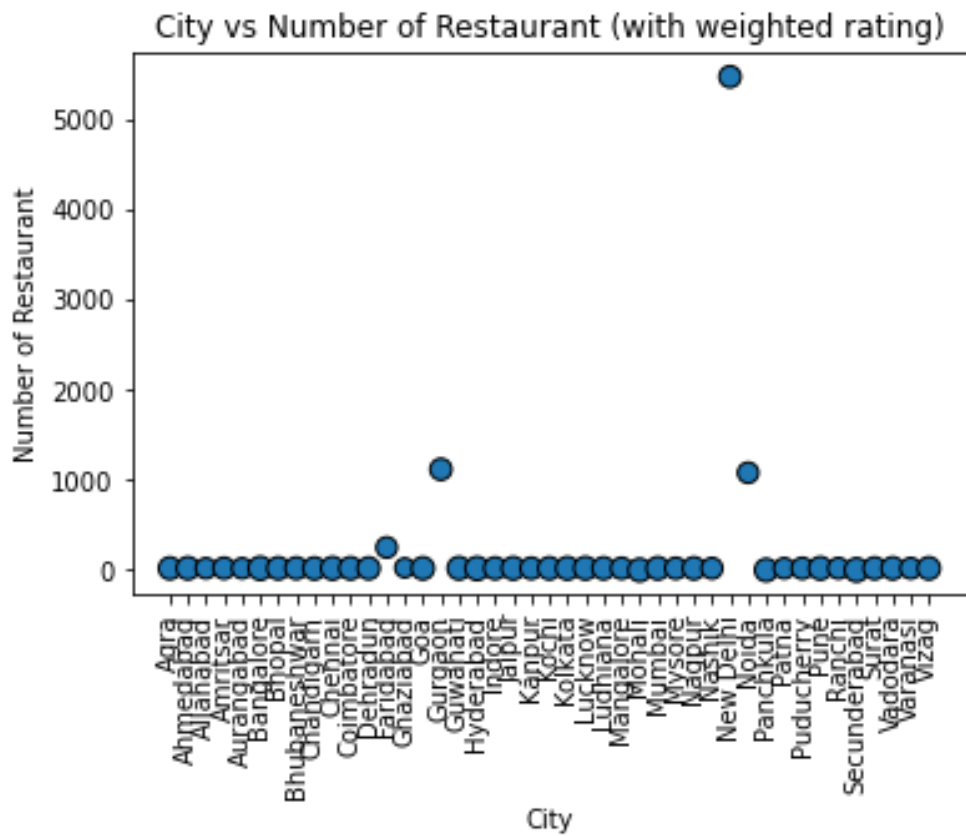
Graphs:-

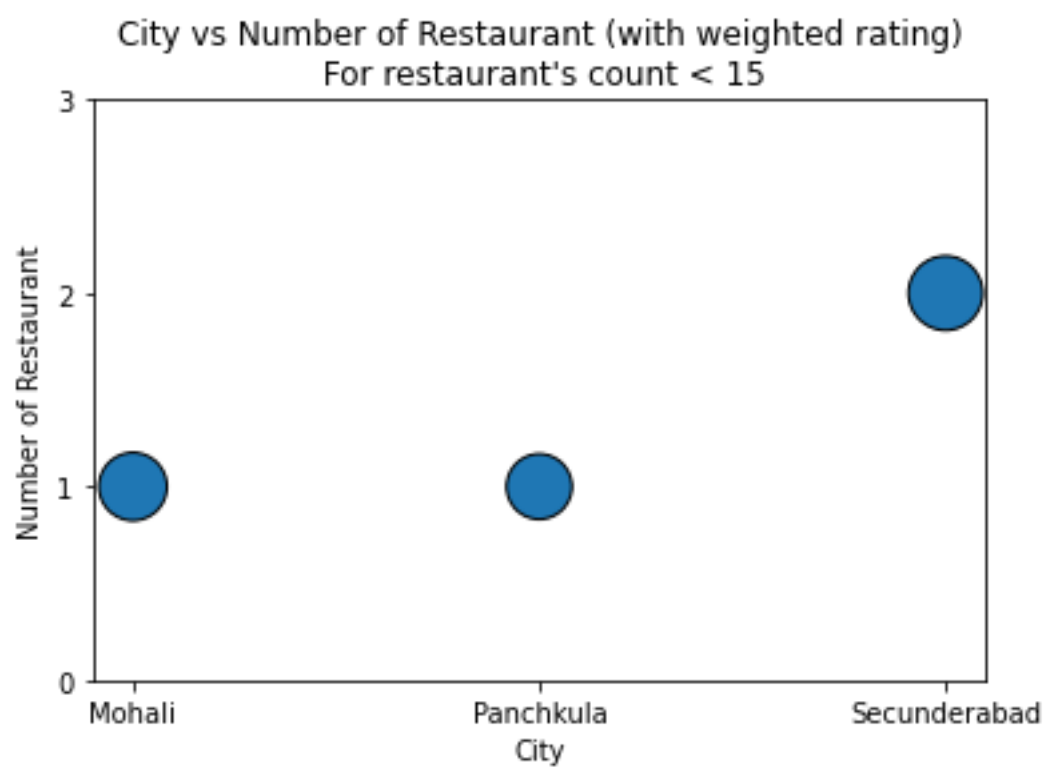


Solution steps for 3.5:-

- Make a column 'vote_rating_product' that contains product of 'vote and rating'.
- Group by City, and store the count and sum of 'vote_rating_product'.
- Get weighted restaurant rating city wise by dividing sum of 'vote_rating_product' by count data.
- Extract top 15 results.
- Plot the bubble graph.

Graphs:-





City vs Number of Restaurant (with weighted rating)
 For restaurant's count ≤ 25 and restaurant's count ≥ 15

