PROJECT – ZOMATO-II

Question-3

Submitted by :-

Aryan Sharma

Q3.) Visualization

Q 3.1.) Plot the bar graph top 15 restaurants have a maximum number of outlets.

Source Code

```
import pandas as pd
import matplotlib.pyplot as plt


data = pd.read_csv('zomato.csv' , encoding = 'latin-1')


data = data[data["Country Code"] == 1]


restaurants = list(data["Restaurant Name"].value_counts().index)[:15]
count  = list(data["Restaurant Name"].value_counts())[:15]
```

```
df = pd.DataFrame(list(zip(restaurants, count)),
                  columns =['Restaurant Name', 'Number of outlets'])


display(df)


plt.style.use("ggplot")
plt.bar(restaurants,count , color = 'red')
plt.xlabel('Restaurant Name')
plt.ylabel('No. of Outlets')
plt.title('Top 15 restaurants and number of outlets')
plt.xticks(rotation = 90)
plt.show()
```

Explanation

- The necessary libraries were included.
- The dataset was loaded and the appropriate encoding was used which happens to be latin-1 in this case .
- The question specifies to analyze the data for India so the data has been filtered by using the Country Code which is 1 for India.
- We then make two lists , restaurants which stores the top-15 restaurant names with the maximum number of outlets and count which stores the number of outlets of these restaurants. This is done using value_counts which gives us restaurant to outlet-count in descending order.
- To show in a tabular form, a dataframe has been made to show only the top-15 restaurant names and the number of outlets they have and displayed.
- A bar-graph is then plotted between the Top-15 restaurant names and no of outlets.

Results

Top – 15 restaurants and their number of outlets

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


Top 15 restaurants and number of outlets
No. of Outlets
0
10
20
30
40
50
60
70
80
Cafe Coffee Day
Domino's Pizza
Subway
Green Chick Chop
McDonald's
Keventers
Giani
Pizza Hut
Baskin Robbins
Barbeque Nation
Giani's
Barista
Dunkin' Donuts
Pind Balluchi
Costa Coffee
Restaurant Name

From the graph it can be seen that Café Coffee Day has the number of outlets – 83 outlets, followed by Domino’s Pizza – 79 outlets and Subway-63 outlets.

Q 3.2.) Plot the histogram of aggregate rating of restaurant(drop the unrated restaurant).

Source Code

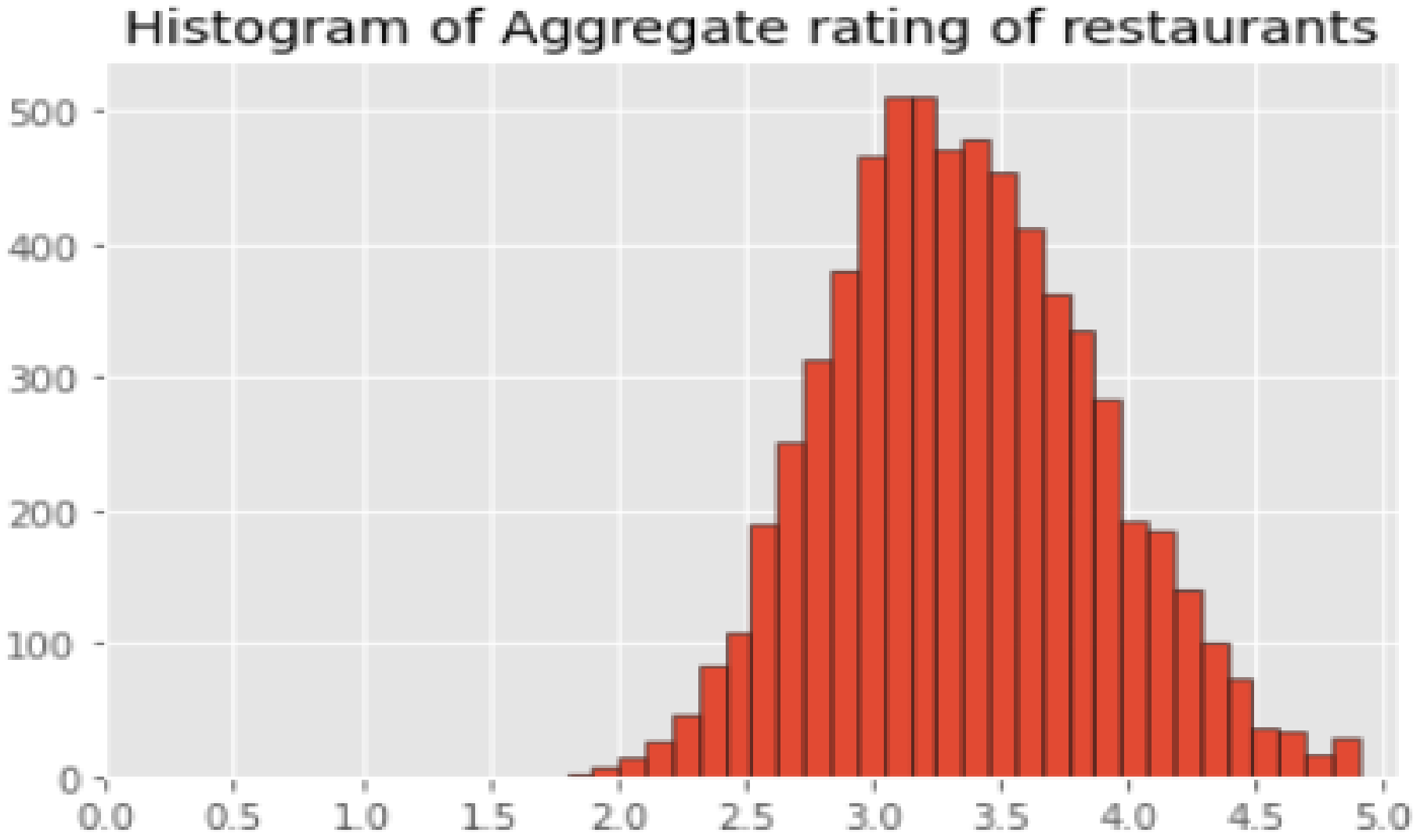
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np


data = pd.read_csv('zomato.csv' , encoding = 'latin-1')
data = data[data["Country Code"] == 1]


data = data[data['Rating text'] != 'Not rated']
data["Aggregate rating"].dropna(inplace = True)
```

```
ratings = list(data["Aggregate rating"])
xaxis = np.arange(0,5.5,0.5)

plt.hist(ratings , bins = 30 , edgecolor = "black")
plt.xticks(xaxis)
plt.title("Histogram of Aggregate rating of restaurants")
plt.show()
```

Explanation

- The necessary libraries were included.
- The dataset was loaded and the appropriate encoding was used which happens to be latin-1 in this case .
- The question specifies to analyze the data for India so the data has been filtered by using the Country Code which is 1 for India.
- **IMPORTANT** – Since in this questions we are analyzing the user rating it is important to get rid of entries for which there is no rating. This has been done by filtering using the “Rating Text” and removing all the data with rating text as “Not rated”.
- From the Aggregate rating column also NAN values are dropped.
- A list is then made named ratings which stores the aggregate rating.
- The histogram is then plotted and 30 bins are used.

Results


Histogram of Aggregate rating of restaurants
500
400
300
200
100
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0

From the histogram it can be clearly seen that most of the ratings lie between 2.5 and 4.5 .

All the restaurants which were “Not Rated” were removed.

Q 3.3.) Plot the bar graph of top 10 restaurants in the data with the highest number of votes.

Source Code

```
import pandas as pd
import matplotlib.pyplot as plt


data = pd.read_csv('zomato.csv' , encoding = 'latin-1')
data = data[data["Country Code"] == 1]


data.sort_values("Votes" , ascending = False,inplace = True)


data["Res_with_Address"] = data["Restaurant Name"]+" , "+data["Address"]
```

```
RestaurantNames = [i.split(" , ")[0] for i in list(data["Res_with_Address"])][:10]

RestaurantNames_with_Address = list(data["Res_with_Address"])[:10]


Address = [i.split(" , ")[1] for i in list(data["Res_with_Address"])][:10]

RestaurantVotes = list(data["Votes"])[:10]


Top10 = pd.DataFrame(list(zip(RestaurantNames,Address,RestaurantVotes)) , columns = ['Restaurant Name',
'Address', 'Votes'])

display(Top10)


plt.style.use("ggplot")

plt.figure(figsize=(10,5))

plt.bar(RestaurantNames,RestaurantVotes , color= 'r')

plt.xticks(rotation = 90)

plt.xlabel("Restaurant Names" , c='k')

plt.ylabel("No. of Votes" , c='k')

plt.title("Top 10 Restaurants vs Votes")
```

```
plt.style.use("ggplot")
plt.figure(figsize=(10,5))
plt.bar(RestaurantNames_with_Address,RestaurantVotes , color= 'b')
plt.xticks(rotation = 90)
plt.xlabel("Restaurant Names with Address" , c='k')
plt.ylabel("No. of Votes" , c='k')
plt.title("Top 10 Unique Restaurants vs Votes")
```

Explanation

- The necessary libraries were included.
- The dataset was loaded and the appropriate encoding was used which happens to be latin-1 in this case .
- The question specifies to analyze the data for India so the data has been filtered by using the Country Code which is 1 for India.
- The data is then sorted based on the number of votes in descending order.
- **IMPORTANT** - There are restaurants which have multiple outlets and thus each have different number of votes. Since its not specified in the question we take into consider two cases : with address and without address. Using the address we can differentiate between the outlets of the same restaurant.
- Now a new column is added in the dataframe named Res_with_Address
- Three lists are made to store the restaurant names , restaurant names with their address and them just the address.
- Another list is made to store the restaurant votes.

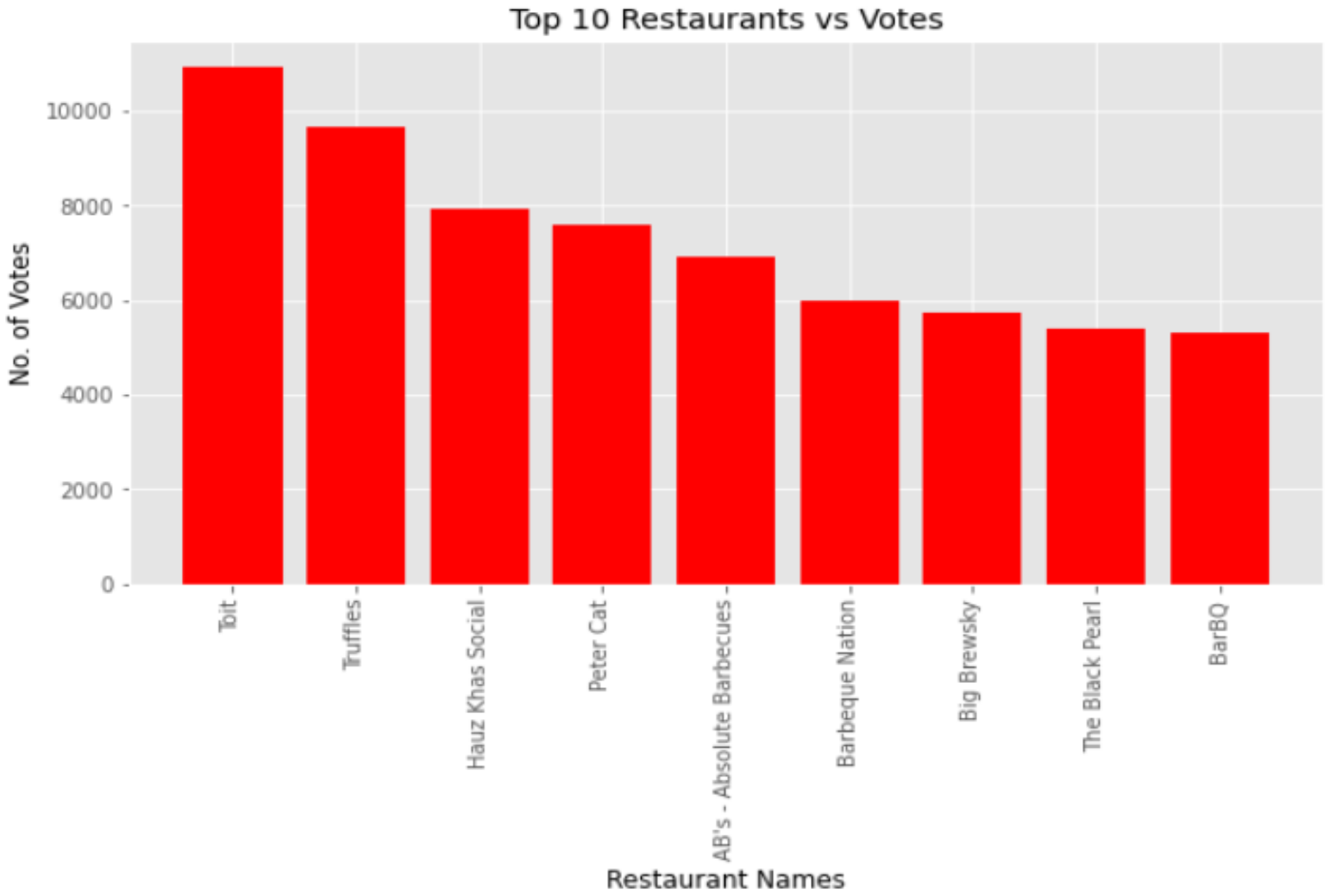
- We then make a dataframe of the top-10 restaurants with their address and Votes and display it.
- Then a bar graph is plotted between Top-10 Restaurants (without considering address)and their votes.
- Another bar graph is plotted between Top-10 Restaurants (with address) and their votes.

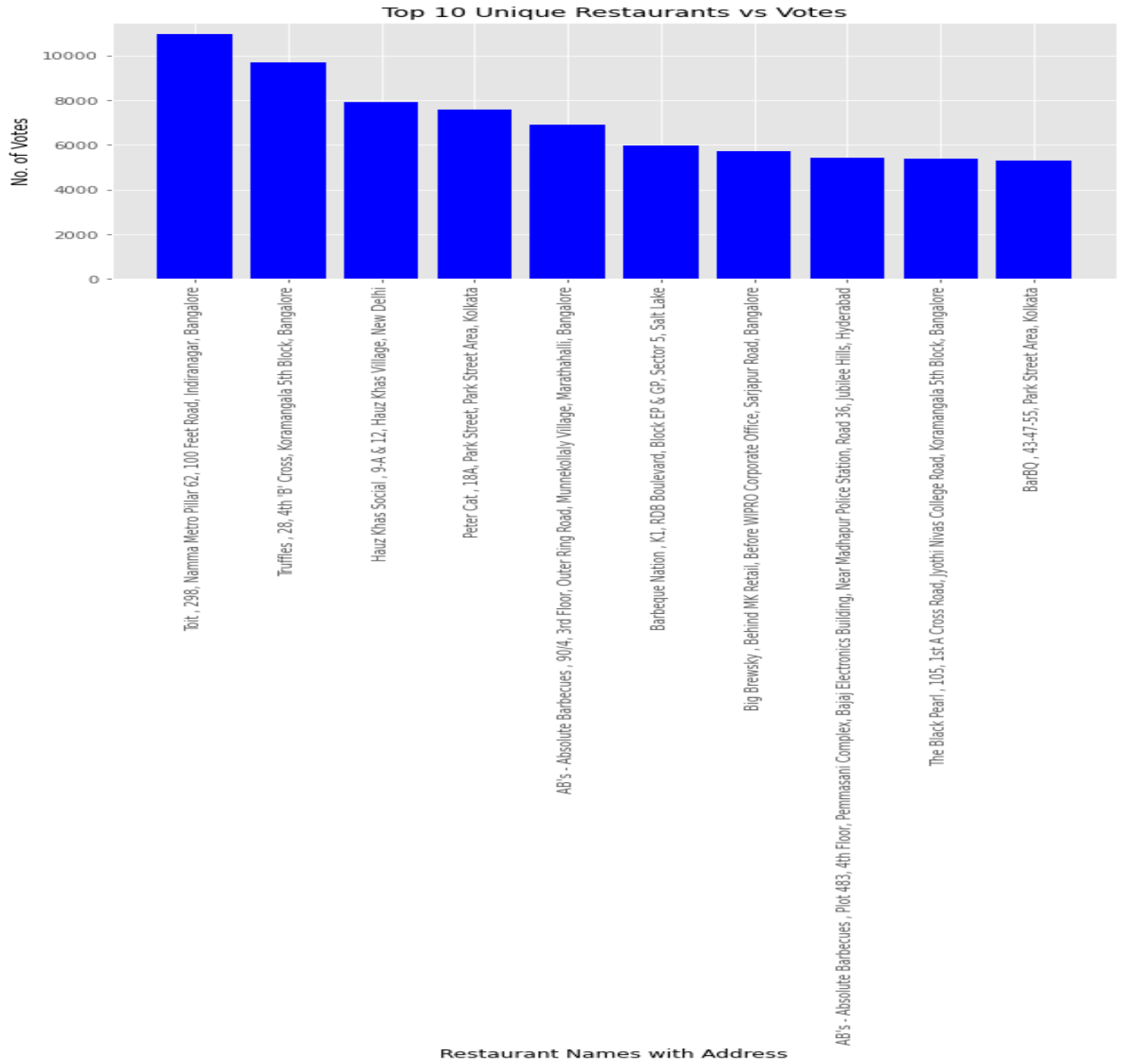
Results

Top-10 Restaurants with address and Votes

Restaurant Name	Address	Votes
Toit	298, Namma Metro Pillar 62, 100 Feet Road, Ind...	10934
Truffles	28, 4th 'B' Cross, Koramangala 5th Block, Bang...	9667
Hauz Khas Social	9-A & 12, Hauz Khas Village, New Delhi	7931
Peter Cat	18A, Park Street, Park Street Area, Kolkata	7574
AB's - Absolute Barbecues	90/4, 3rd Floor, Outer Ring Road, Munnekollaly...	6907
Barbeque Nation	K1, RDB Boulevard, Block EP & GP, Sector 5, Sa...	5966
Big Brewsky	Behind MK Retail, Before WIPRO Corporate Offic...	5705
AB's - Absolute Barbecues	Plot 483, 4th Floor, Pemmasani Complex, Bajaj ...	5434
The Black Pearl	105, 1st A Cross Road, Jyothi Nivas College Ro...	5385
BarBQ	43-47-55, Park Street Area, Kolkata	5288

This bar graphs shows only 9 restaurants. This means that there is a restaurant among these which comes twice in the list of Top-10.
So it is important to distinguish between them based on address as seen in the next graph.




Top 10 Unique Restaurants vs Votes
No. of Votes
0
2000
4000
6000
8000
10000
Toit , 298, Namma Metro Pillar 62, 100 Feet Road, Indiranagar, Bangalore
Truffles , 28, 4th 'B' Cross, Koramangala 5th Block, Bangalore
Hauz Khas Social , 9-A & 12, Hauz Khas Village, New Delhi
Peter Cat , 18A, Park Street, Park Street Area, Kolkata
AB's - Absolute Barbecues , 90/4, 3rd Floor, Outer Ring Road, Munnekollalu Village, Marathahalli, Bangalore
Barbeque Nation , K1, RDB Boulevard, Block EP & GP, Sector 5, Salt Lake
Big Brewsky , Behind MK Retail, Before WIPRO Corporate Office, Sarjapur Road, Bangalore
AB's - Absolute Barbecues , Plot 483, 4th Floor, Pemmasani Complex, Bajaj Electronics Building, Near Madhapur Police Station, Road 36, Jubilee Hills, Hyderabad
The Black Pearl , 105, 1st A Cross Road, Jyothi Nivas College Road, Koramangala 5th Block, Bangalore
BarBQ , 43-47-55, Park Street Area, Kolkata
Restaurant Names with Address

From the graphs we can see that the restaurant Toit has the highest number of votes – 10934 votes.

Another important thing to note is that AB's Absolute Barbeques has two outlets in the Top-10, one in 90/4, 3rd Floor, Outer Ring Road, Munnekollaly and the other in Plot 483, 4th Floor, Pemmasani Complex.

Q 3.4.) Plot the pie graph of top 10 cuisines present in restaurants in the USA.

Source Code

```
import pandas as pd
import matplotlib.pyplot as plt


data = pd.read_csv('zomato.csv' , encoding = 'latin-1')
data = data[data["Country Code"] == 216]
data['Cuisines'].dropna(inplace = True)


cuisine_count = {}


def get_count(cuisines):
    cuisines = str(cuisines)
    if cuisines!="":
        for cuisine in cuisines.split(', '):
            cuisine_count[cuisine] = cuisine_count.get(cuisine,0)+1
    return cuisines
```

```
data['Cuisines'].apply(get_count)

cuisine_count = sorted(cuisine_count.items(), key=lambda x: x[1] , reverse = True)

top10cuisines = [i[0] for i in cuisine_count[:10]]
top10cuisines_count = [i[1] for i in cuisine_count[:10]]

Top10 = pd.DataFrame(list(zip(top10cuisines,top10cuisines_count)) , columns = ['Cuisine', 'Count'])
display(Top10)

plt.pie(top10cuisines_count , labels = top10cuisines , autopct = "%.2f")
plt.title("Top 10 Cuisines in USA")
plt.show()
```

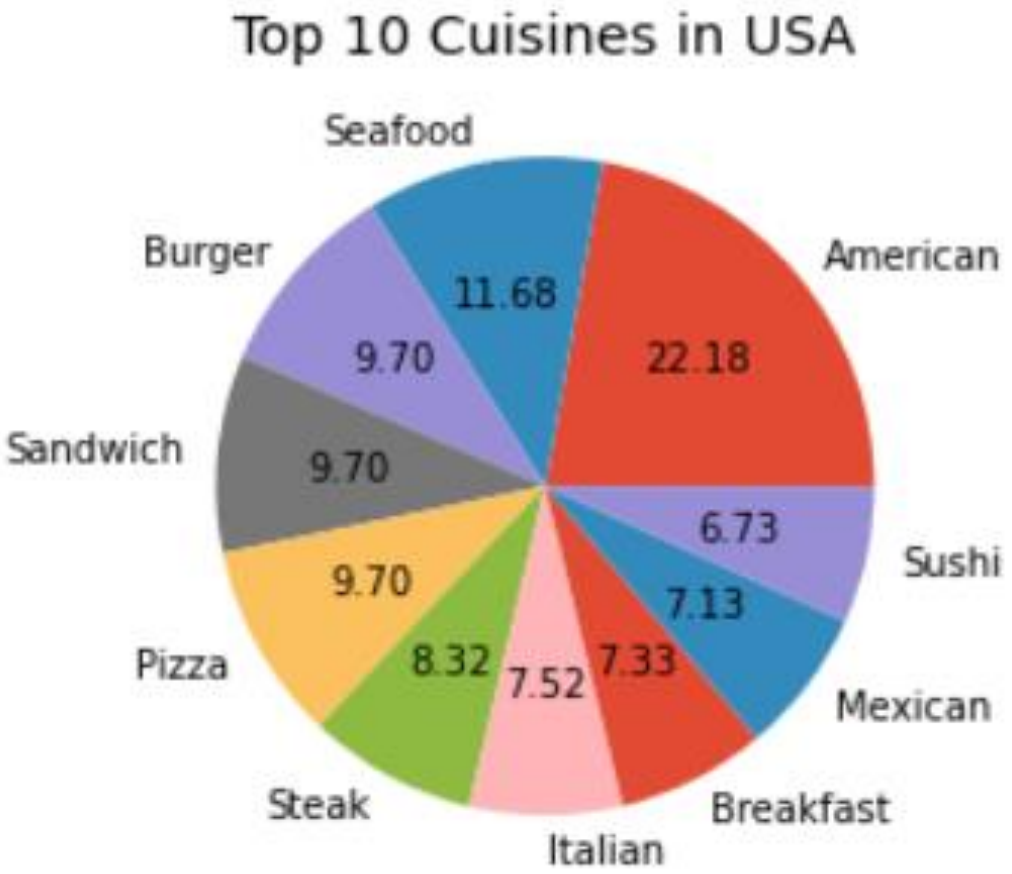
Explanation

- The necessary libraries were included.
- The dataset was loaded and the appropriate encoding was used which happens to be latin-1 in this case .
- The question specifies to analyze the data for USA so the data has been filtered by using the Country Code which is 216 for USA.
- We then make a dictionary named cuisine_count.
- We make a function named get_count which will work on cuisines column an form the dictionary cuisine_count. This function splits on “, “ to get all the cuisines from a restaurant and adds the cuisine name as key to the dictionary and its count as the value.
- This function is applied on the cuisines column of data and the dictionary is formed.
- We then sort in descending order based on the value of dictionary.

- We then make two lists top10cuisines and top10cuisines_count which store top 10 cuisines and their count.
- We make a dataframe just to show the Top-10 cuisines and their count in a tabular form and then display the dataframe.
- Then we make a pir chart of the top-10 cuisines.

Results

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



Rank	Cuisine	Percentage (%)
1	American	22.18
2	Seafood	11.68
3	Burger	9.70
4	Sandwich	9.70
5	Pizza	9.70
6	Steak	8.32
7	Italian	7.52
8	Breakfast	7.33
9	Mexican	7.13
10	Sushi	6.73

From the pie chart it can be concluded that American is the most popular cuisine in USA followed by Seafood , Burger , Sandwich and Pizza.

Q 3.5.) 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.

Source Code

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt


data = pd.read_csv('zomato.csv' , encoding = 'latin-1')

data = data[data["Country Code"] == 1]


city_dict = {}


cities = np.array(data["City"])

ratings = np.array(data["Aggregate rating"])

votes = np.array(data["Votes"])
```

```
for i in range(len(cities)) :
    if votes[i] != 0 :
        if cities[i] not in city_dict :
            city_dict[cities[i]] = [votes[i] * ratings[i] , votes[i] , 1]
        else:
            city_dict[cities[i]][0] += votes[i]*ratings[i]
            city_dict[cities[i]][1] += votes[i]
            city_dict[cities[i]][2] +=1

city_and_rating = []
for key,value in city_dict.items():
    city = key
    weighted_rating = round(value[0]/value[1],2)
    city_and_rating.append([value[2],weighted_rating,key])

city_and_rating.sort(reverse = True)
```

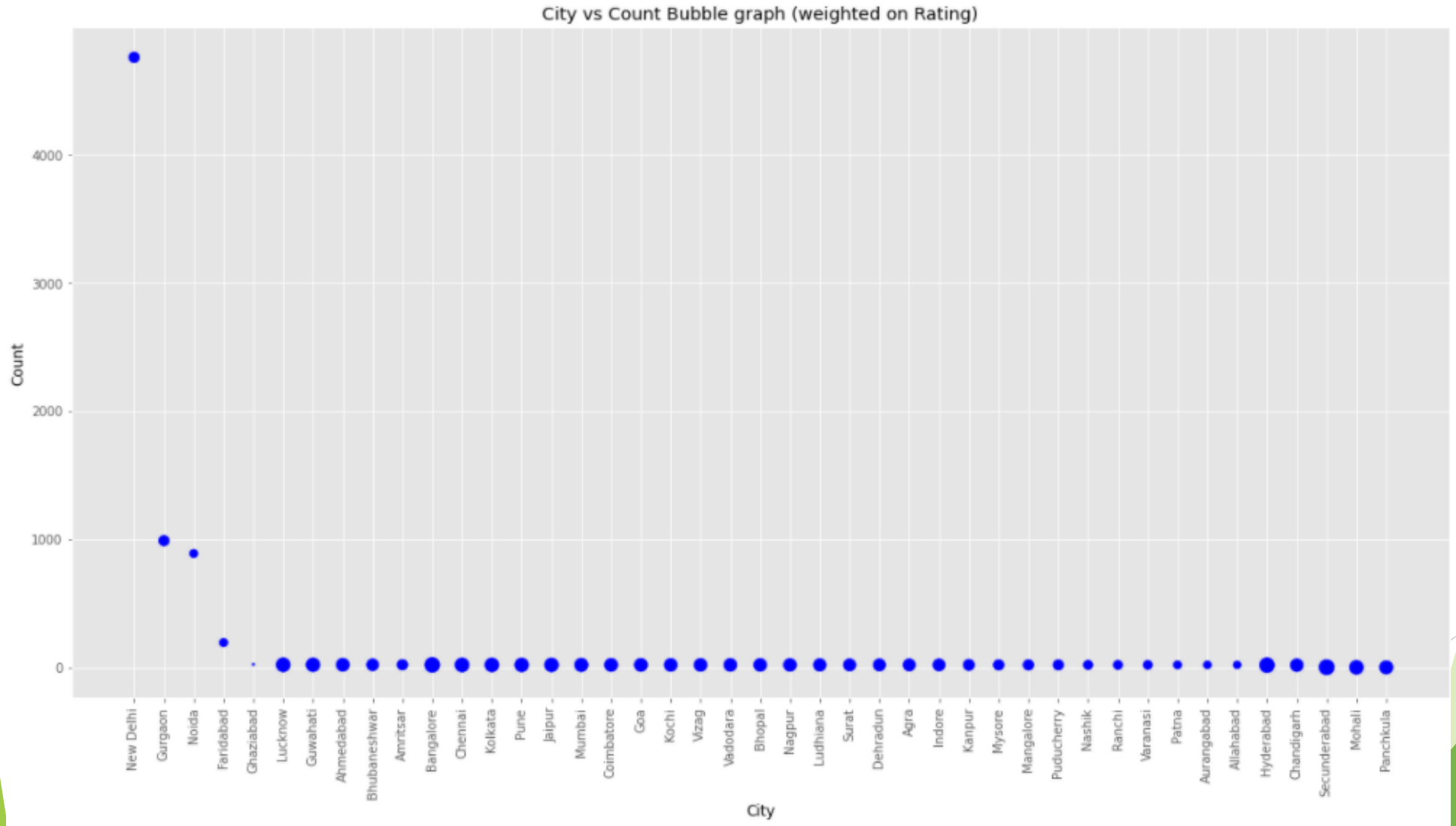
```
CITIES = np.array([i[2] for i in city_and_rating])
RATING = np.array([i[1] for i in city_and_rating])
RATING = RATING-3
COUNT = np.array([i[0] for i in city_and_rating])

plt.figure(figsize=(20,10))
plt.scatter(CITIES , COUNT , s = RATING*100 , color = 'b')
plt.xticks(rotation = 90)
plt.xlabel('City' , color = 'k')
plt.ylabel('Count' , color = 'k')
plt.title('City vs Count Bubble graph (weighted on Rating)')
plt.show()

TOP10_CITIES = CITIES[:10]
TOP10_RATING = RATING[:10]
TOP10_COUNT = COUNT[:10]

plt.figure(figsize=(10,5))
plt.scatter(TOP10_CITIES , TOP10_COUNT , s = TOP10_RATING*100 , color = 'r')
plt.xticks(rotation = 90)
plt.xlabel('City' , color = 'k')
plt.ylabel('Count' , color = 'k')
plt.title('TOP - 10 City vs Count Bubble graph (weighted on Rating)')
plt.show()
```

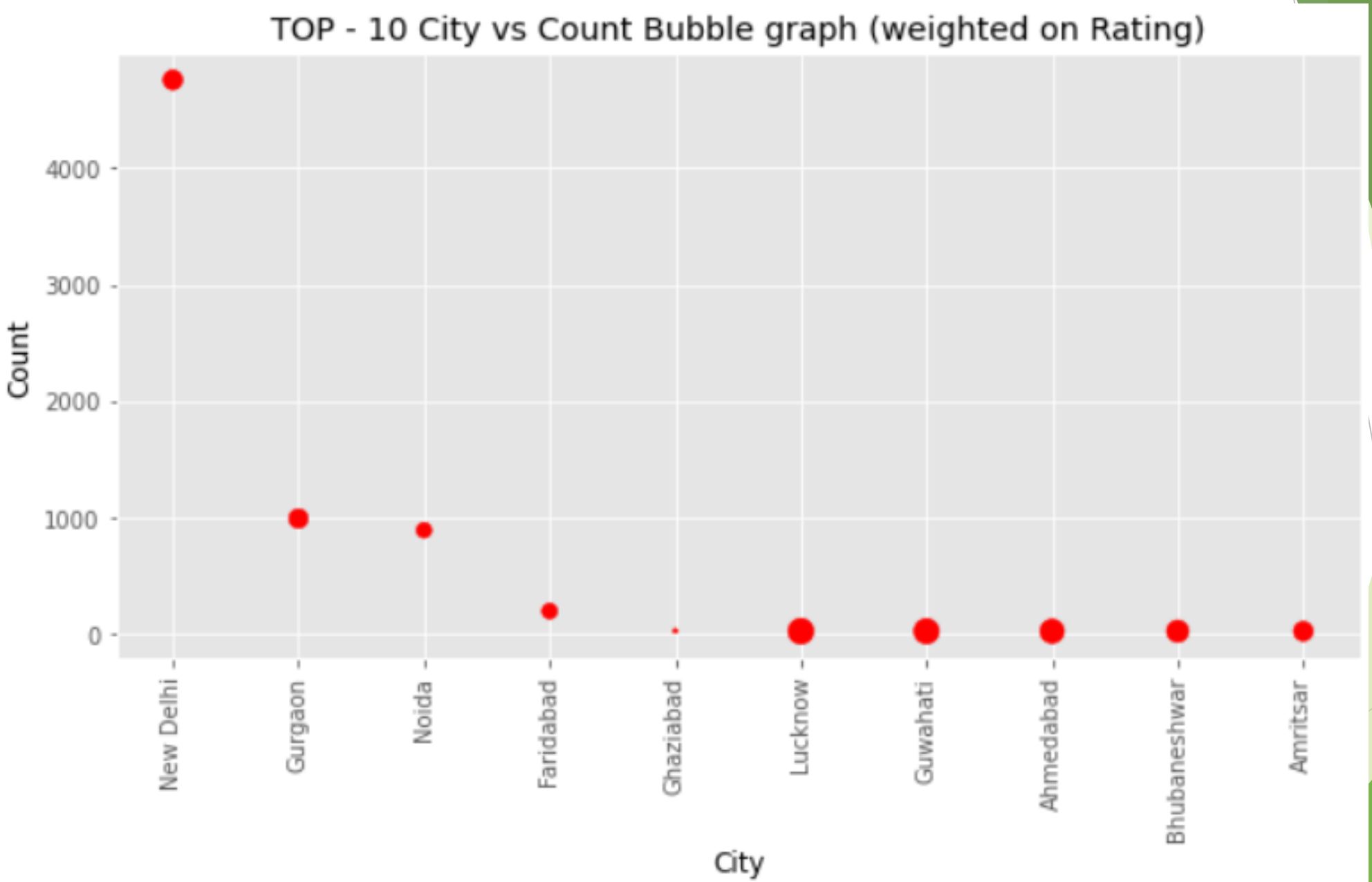
Explanation

- The necessary libraries were included.
- The dataset was loaded and the appropriate encoding was used which happens to be latin-1 in this case .
- The question specifies to analyze the data for India so the data has been filtered by using the Country Code which is 1 for India.
- We make a dictionary named city_dict.
- We make three numpy arrays named cities , ratings and votes.
- We then loop length of cities array times to form the dictionary. We ensure that the number of votes is not 0 . We store the city name as key and votes*rating sum and count as values.
- We then make list of lists named city_and_rating in which we will store the count , weighted rating calculated as value[0]/value[1] of dictionary, and the city name.

- We now sort this list in descending order.
- Now we separate out the Cities , rating and count into 3 lists.
- To make the ratings comparable and difference clear in bubble graph we subtract 3 from each rating.
- We then plot a bubble graph showing all the cities vs count of restaurants and the weights of bubbles as the weighted rating.
- Then we make the lists of the TOP – 10 cities their ratings and count to be used to make the bubble graph for top-10.
- We finally make a bubble graph of TOP - 10 City vs Count Bubble graph (weighted on Rating).

Results


City vs Count Bubble graph (weighted on Rating)
Count
4000
3000
2000
1000
0
New Delhi
Gurgaon
Noida
Faridabad
Ghaziabad
Lucknow
Guwahati
Ahmedabad
Bhubaneshwar
Amritsar
Bangalore
Chennai
Kolkata
Pune
Jaipur
Mumbai
Coimbatore
Goa
Kochi
Vizag
Vadodara
Bhopal
Nagpur
Ludhiana
Surat
Dehradun
Agra
Indore
Kanpur
Mysore
Mangalore
Puducherry
Nashik
Ranchi
Varanasi
Patna
Aurangabad
Allahabad
Hyderabad
Chandigarh
Secunderabad
Mohali
Panchkula
City


TOP - 10 City vs Count Bubble graph (weighted on Rating)
Count
4000
3000
2000
1000
0
New Delhi
Gurgaon
Noida
Faridabad
Ghaziabad
Lucknow
Guwahati
Ahmedabad
Bhubaneshwar
Amritsar
City

From the bubble graph it can be seen that New Delhi has the highest number of restaurants in India.

The relative rating can also be seen from the size of the bubble of the bubble graph.

Based on the graph we can see New Delhi has the highest number of restaurants with decent aggregate rating. Lucknow fewer restaurants but has the highest aggregate rating among all.