

# **ZOMATO API PROJECT – CODING**

## **NINJAS INDIA**

- KESHAV SHARMA

### **PROBLEM 3 :-**

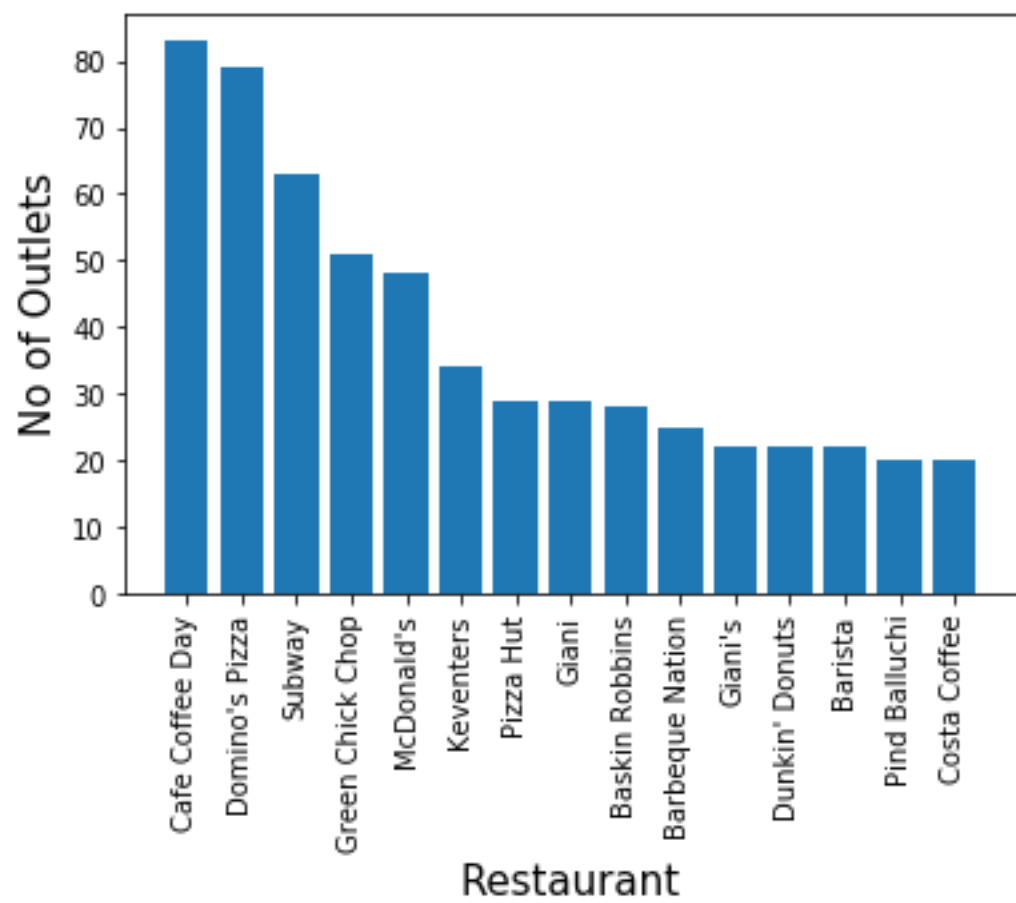
#### Visualization

### **QUESTION 1 :-**

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

### **ANSWER :-**

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 Restaurants 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 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 frequency. 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 needed to filter out those if already appeared.

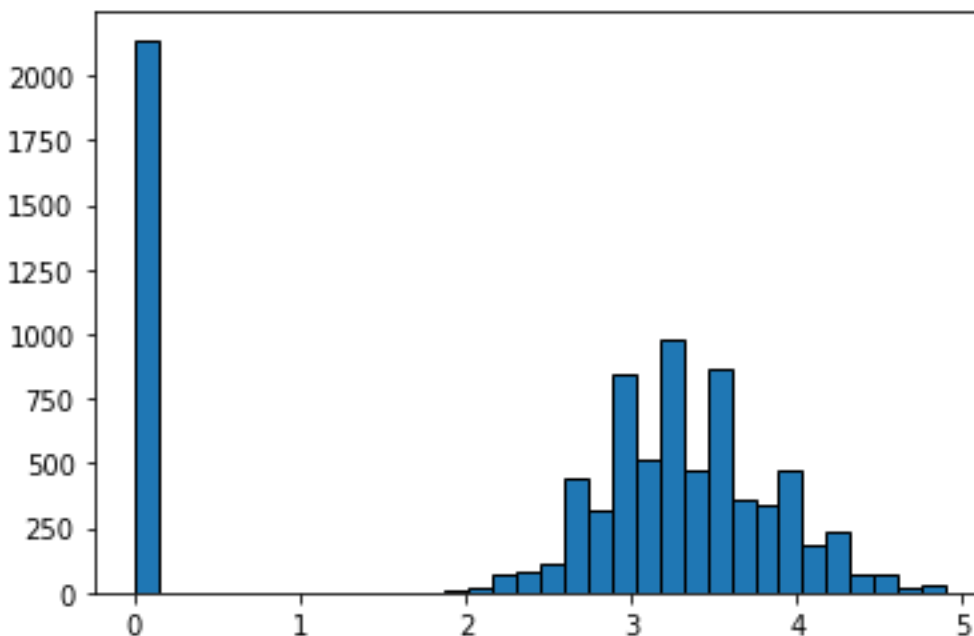


## QUESTION 2:-

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

## ANSWER :-

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 plotting 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.



### QUESTION 3 :-

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

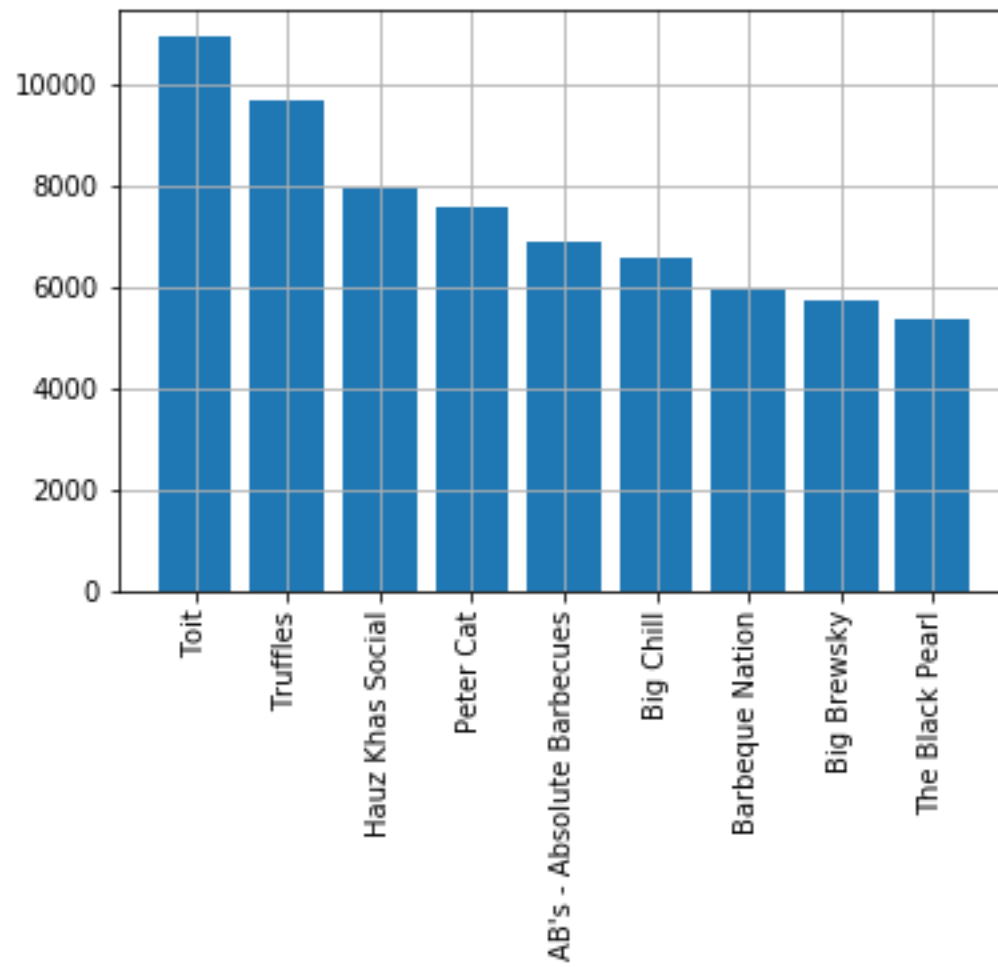
### ANSWER :-

Answer: 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 delimiter 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. From the graph

we can see :

- 1 Toit
- 2 Truffles
- 3 Hauz Khas Social
- 4 Peter Cat
- 5 AB's - Absolute Barbecues
- 6 Big Chill
- 7 Barbeque Nation
- 8 Big Brewsky
- 9 AB's - Absolute Barbecues

- 10 The Black Pearl



#### QUESTION 4 :-

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

#### ANSWER:-

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

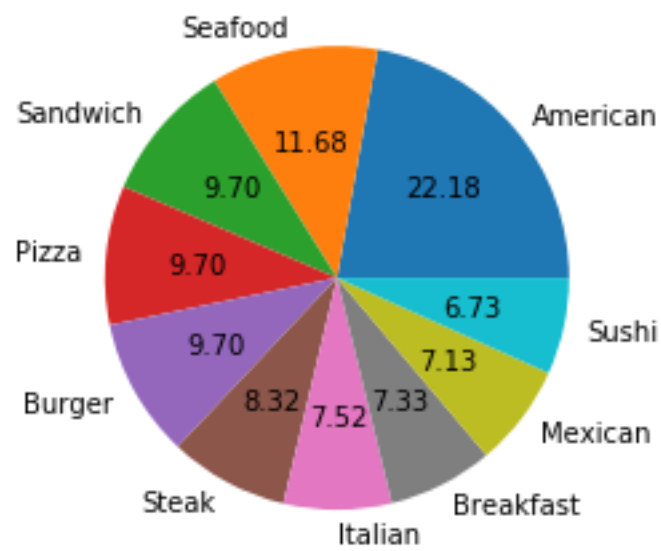
in the cuisines list. After I got all the count stored in the dictionary then I have made one list of list which store values

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. Ploting them in the pie chart we can notice

Rank	Cuisine	%
1	American	22.18
2	Seafood	11.68
3	Sandwich	9.7
4	Pizza	9.7

5	Burger	9.7
6	Steak	8.32
7	Italian	7.52
8	Breakfast	7.33
9	Mexican	7.13
10	Sushi	6.73



### QUESTION 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.

### ANSWER:-

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.

NOTE: I have substracted 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.

Here is the list:

SL	City	Rating
1	New Delhi	3.77



2	Gurgaon	3.74
3	Noida	3.47
4	Faridabad	3.48
5	Ghaziabad	3.04
6	Lucknow	4.32
7	Guwahati	4.27
8	Ahmedabad	4.16
9	Bhubaneswar	3.97
10	Amritsar	3.76
11	Bangalore	4.5
12	Chennai	4.32
13	Kolkata	4.3
14	Pune	4.28
15	Jaipur	4.28
16	Mumbai	4.22
17	Coimbatore	4.17
18	Goa	4.16
19	Kochi	4.14
20	Vizag	4.13

