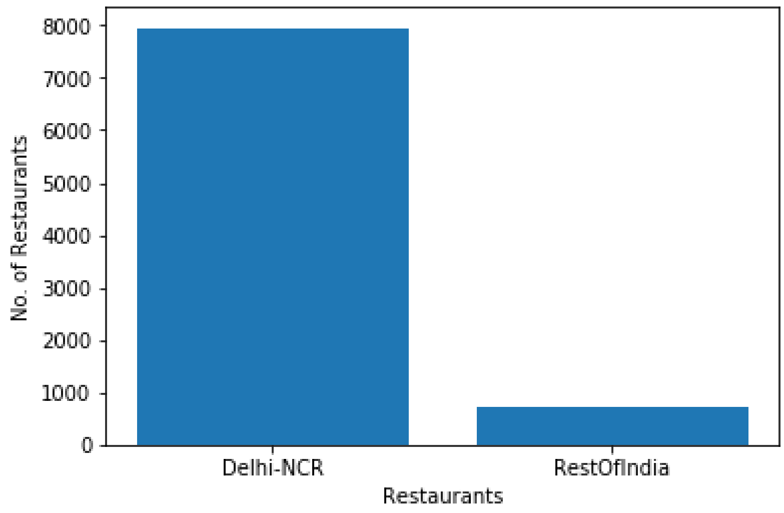
Answer 1.1:



```
Restaurants in Delhi-NCR 7947
Restaurants in RestOfIndia 705
```

Conclusion

Restaurants in Delhi NCR is more than that in RestOfIndia

Justification

- In this code, Reading csv file zomato data and storing it in variable data.
- Creating df which stores a copy of data.
- Renaming column names of df are updated to lower case and spaces replaced with '_'.
- Now restricting data frame df to country id 1 that is of India.
- Cities got filtered and modified among Delhi-NCR and RestOfIndia cities using apply() and update_city().
- Using value count in pandas library, the required count of restaurants in Delhi-NCR and RestOfIndia is calculated and stored in series_city_count.
- From the series_city_count, list of keys and its value is extracted and stored in list key and list value respectively.
- Bar graph is plotted using pyplot using list key and value .

Answer 1.2:

```
Cuisines which are not present in Delhi-NCR but are present in
RestOfIndia
BBQ
Cajun
German
Malwani
Verifying whether above dishes are actually not present via API call
BBQ is present
Cajun is not present
German is present
Malwani is not present
```

Conclusion

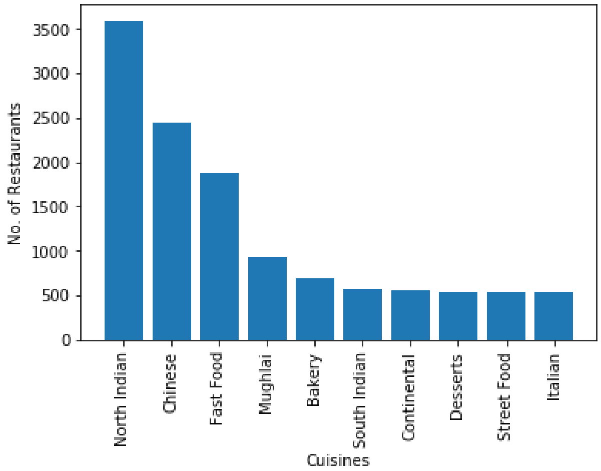
Based on the above inferences, we can conclude that this result from csv is incorrect due to incomplete dataset.

Justification

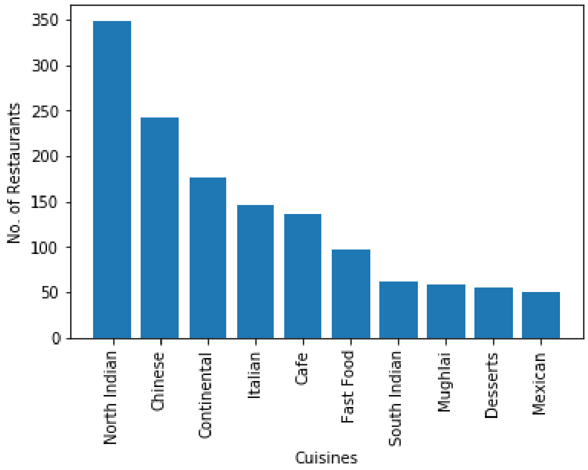
- In this code, Reading csv file zomato data and storing it in variable data.
- Creating df which stores a copy of data.
- Renaming column names of df are updated to lower case and spaces replaced with '_'.
- Now restricting data frame df to country id 1 that is of India.
- Cities got filtered and modified among Delhi-NCR and RestOfIndia cities using apply() and update_city().
- Creating a dictionary dict_city_cuisine of city as key and value as cuisine using group by from pandas libary,then Converting string of values to list and also removing duplicates in values of dict_city_cuisine.
- Creating list of rest_l and ncr_l that stores uniques names of cuisines served in RestOfIndia and Delhi-NCR respectively.
- Filtering list of rest_l and ncr_l according to Cuisines which are not present in Delhi-NCR but are present in RestOfIndia, stored in cuisine_list_csv and printing the same.
- To verify the cuisine actually does not exist, Using requests library we make a request to zomato api to retrieve city_id of Delhi-NCR by coverting json to python using json library.
- Further using city_id to generate the list of cuisines served in Delhi-NCR and storing it in Complete_Cuisine_List.
- To check whether the cuisines present in Delhi NCR or not, comparing the list of cuisines generated by csv file and api call.

Answer 1.3:

```
Delhi-NCR Top 10 Cuisines Vs No. of Restaurants
North Indian 3597
Chinese 2448
Fast Food 1866
Mughlai 933
Bakery 697
South Indian 569
Continental 547
Desserts 542
Street Food 538
Italian 535
```



```
RestOfIndia Top 10 Cuisines Vs No. of Restaurants
North Indian 349
Chinese 242
Continental 177
Italian 147
Cafe 136
Fast Food 97
South Indian 62
Mughlai 59
Desserts 55
Mexican 50
```



Conclusion

Top 10 cuisines served by rest of India nd Delhi-NCR have 8 cuisines in common where north Indian and Chinese served maximum in both these set of cities.

Justification

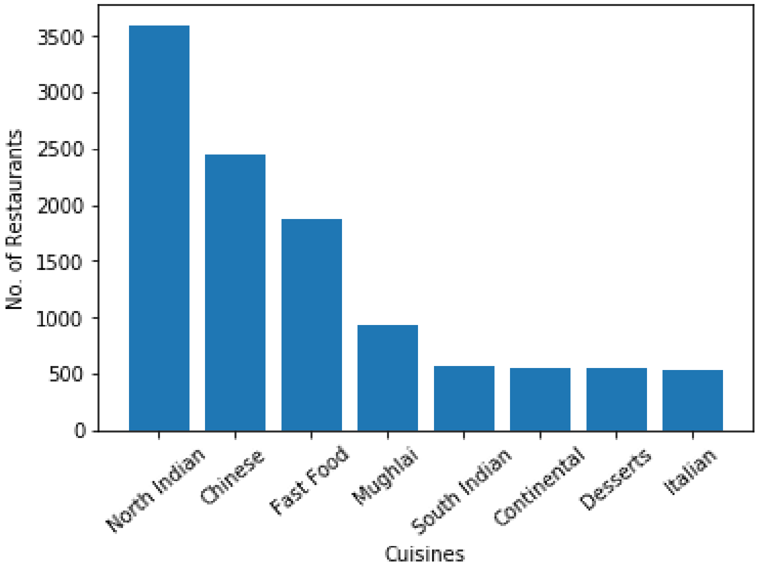
- In this code, Reading csv file zomato data and storing it in var data.
- Creating df which stores a copy of data.
- Renaming column names of df are updated to lower case and spaces replaced with '_' .
- Now restricting data frame df to country id 1 that is of India.
- Cities got filtered and modified among Delhi-NCR and RestOfIndia cities using apply() and update_city().
- Creating a dictionary dict_city_cuisine of city as key and value as cuisine using group by from pandas libary,then Converting string of values to list with stripped elements.
- Now using dict_city_cuisine, creating a dictionary temp_dict stores keys and its values(lists of cuisines(x data) and lists of its frequency of cuisines that is no. of restaurants(y data)) of Delhi-NCR and RestOfIndia.
- Extracting top 10 cuisines for each set of cities.
- Printing and plotting the bar graph of No.of Restaurants vs cuisines.

Answer 1.4:

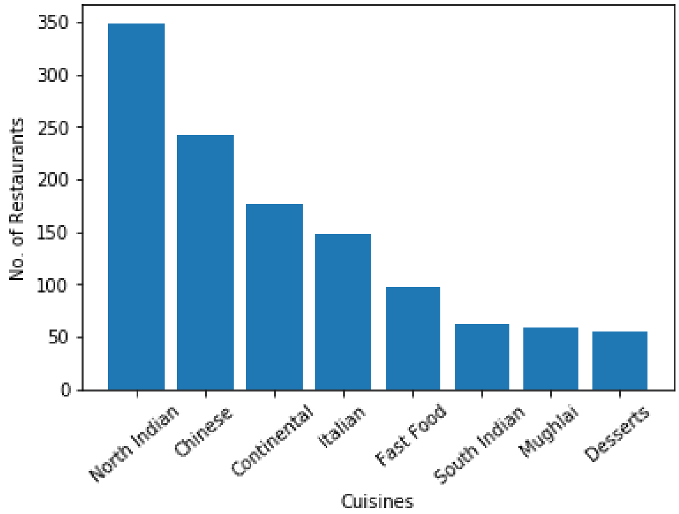
```
Cuisines present in Delhi-NCR but not present in RestOfIndia
Iranian
Bihari
Cuisine Varies
Belgian
Sri Lankan
Assamese
Moroccan
Drinks Only
Turkish
Nepalese
Kashmiri
Persian
Afghani
Pakistani
Sushi
Oriya
Deli
Raw Meats
Naga
South American

Cuisines present in RestOfIndia but not present in Delhi-NCR
BBQ
Cajun
German
Malwani
```

Also from Que1-Part3 above, common cuisines among both region from the set of top 10 cuisines of both region along with the no. of Restaurants are-
Delhi-NCR Region



```
RestOfIndia Region
```



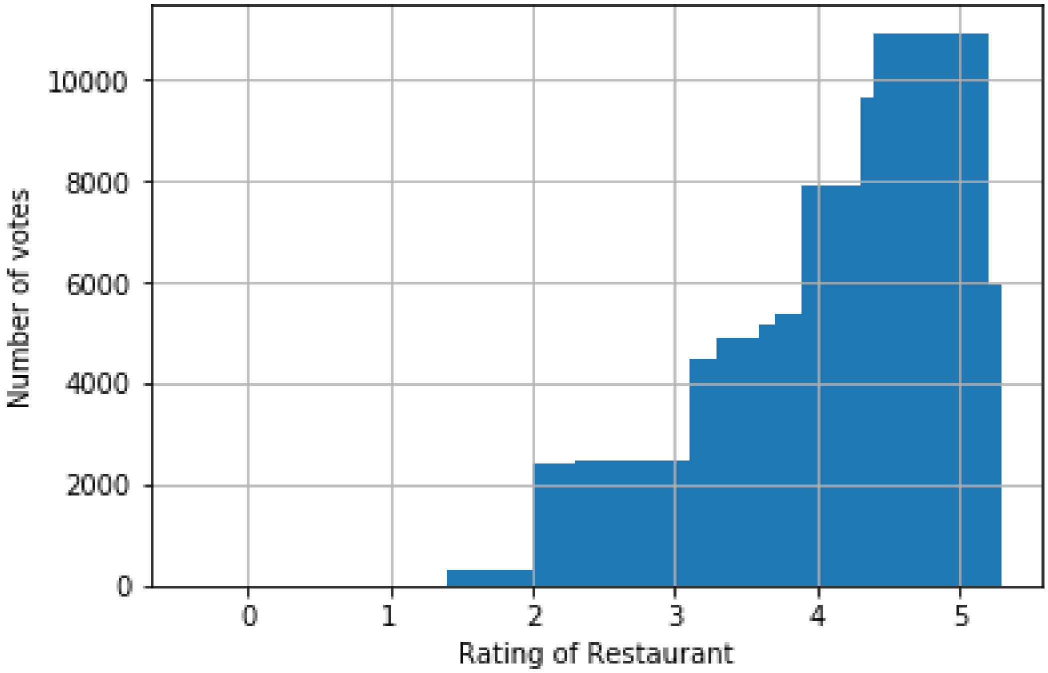
Conclusion

Cuisines present in restOf India and not Delhi NCR are much less than Cuisines present in Delhi NCR and not in restOfIndia.Also,Common cuisines served by rest of india nd Delhi-NCR are 8 cuisnes where north indian and chinese served maximum in both these set of cities among top 10 cuisines served in both the set of cities.

Justification

- In this code, Reading csv file zomato data and storing it in variable data.
- Creating df which stores a copy of data.
- Renaming column names of df are updated to lower case and spaces replaced with ‘_’.
- Now restricting data frame df to country id 1 that is of India.
- Cities got filtered and modified among Delhi-NCR and RestOfIndia cities using apply() and update_city().
- Creating a dictionary dict_city_cuisine of city as key and value as cuisine using group by from pandas libary,then Converting string of values to list and also removing duplicates in values of dict_city_cuisine.
- Creating list of non_ncr_cuisine_list and ncr_cuisine_list that stores uniques names of cuisines served in RestOfIndia and Delhi-NCR respectively.
- Filtering list of non_ncr_cuisine_list and ncr_cuisine_list according to Cuisines which are not present in Delhi-NCR but are present in RestOfIndia and printing the same.
- Also, filtering them according to Cuisines which are not present in RestOfIndia but are present in Delhi-NCR and printing the same.
- Using dictionary temp_dict from ques 1.3 to gererate common cuisines served by Delhi NCR and RestOfIndia.
- Printing and plotting the bar graph of No.of Restaurants vs cuisines**.**

Answer 2.1.1:



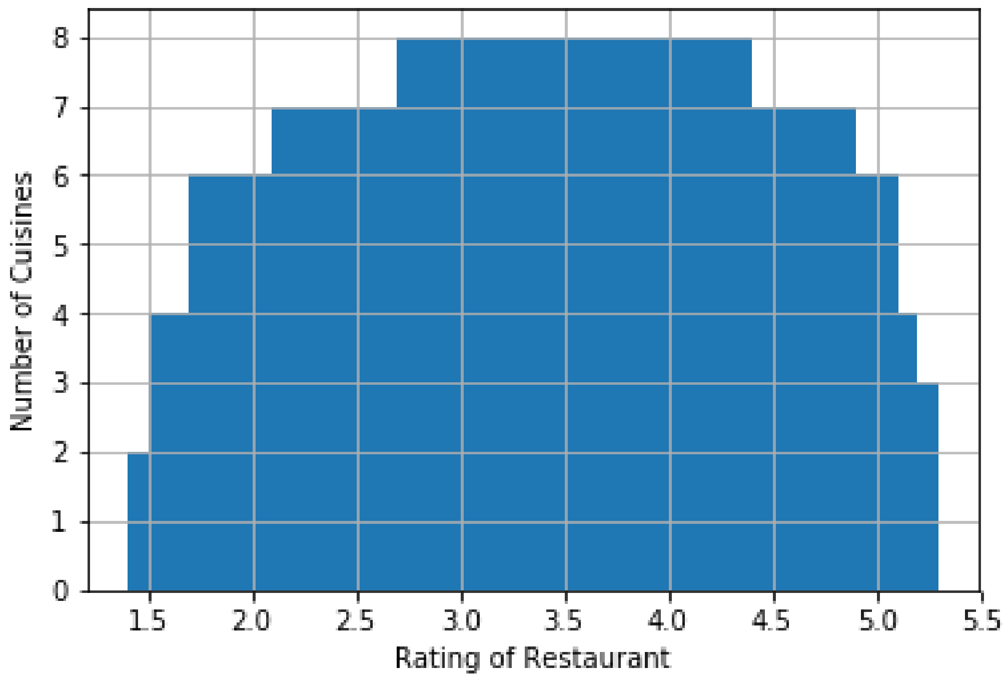
Conclusion

As we can see the rating getting increased with votes, we can conclude that both entities are directly proportional

Justification

- In this code, Reading csv file zomato data and storing it in variable data.
- Creating df which stores a copy of data.
- Renaming column names of df are updated to lower case and spaces replaced with '_'.
- Now restricting data frame df to country id 1 that is of India.
- Now creating list of x and y of aggregate rating and votes respectively extracted from df.
- Plotting bar graph from x and y list with x axis as rating of restaurants and y axis as no. of votes.

Answer 2.1.2:



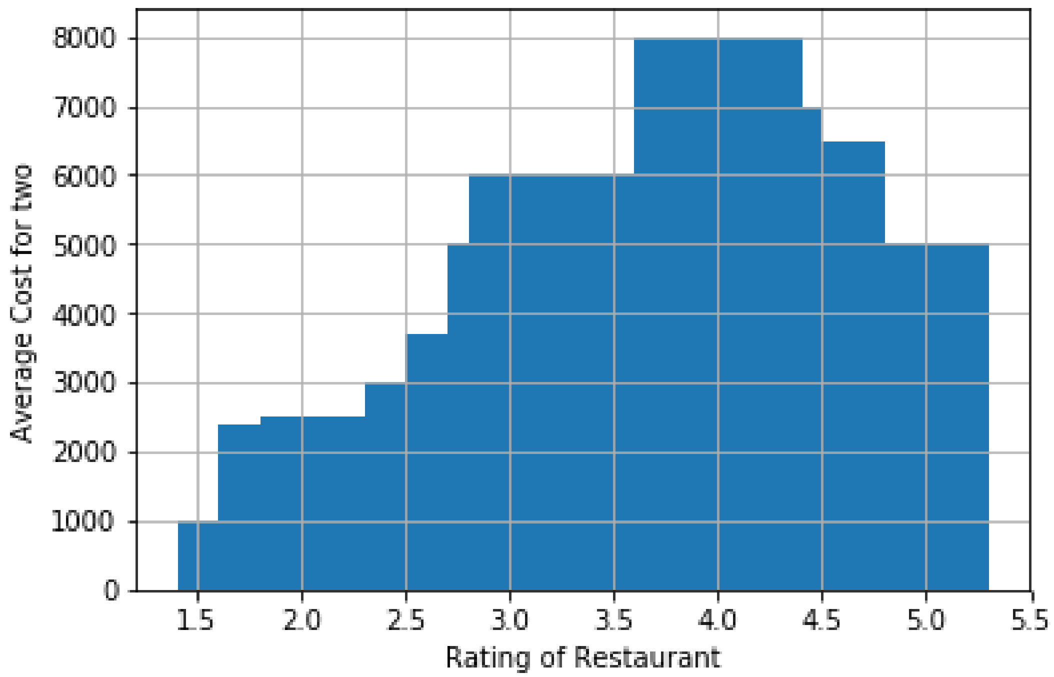
Conclusion

As we can see that highest rated restaurants have less number of cuisines in comparison to restaurants having more cuisines, we can conclude that rating is somewhat inversely proportional here with number of cuisines

Justification

- In this code, Reading csv file zomato data and storing it in variable data.
- Creating df which stores a copy of data.
- Renaming column names of df are updated to lower case and spaces replaced with '_'.
- Now restricting dataframe df to country id 1 that is of India.
- Creating list of x and y of aggregate rating and no. of cuisines respectively.(No. of cuisines are calculated by finding length of cuisines after splitting and operation done row by row using count_cuisines())
- Plotting bar graph from x and y list with x axis as rating of restaurants and y axis as no. of cuisines.

Answer 2.1.3:



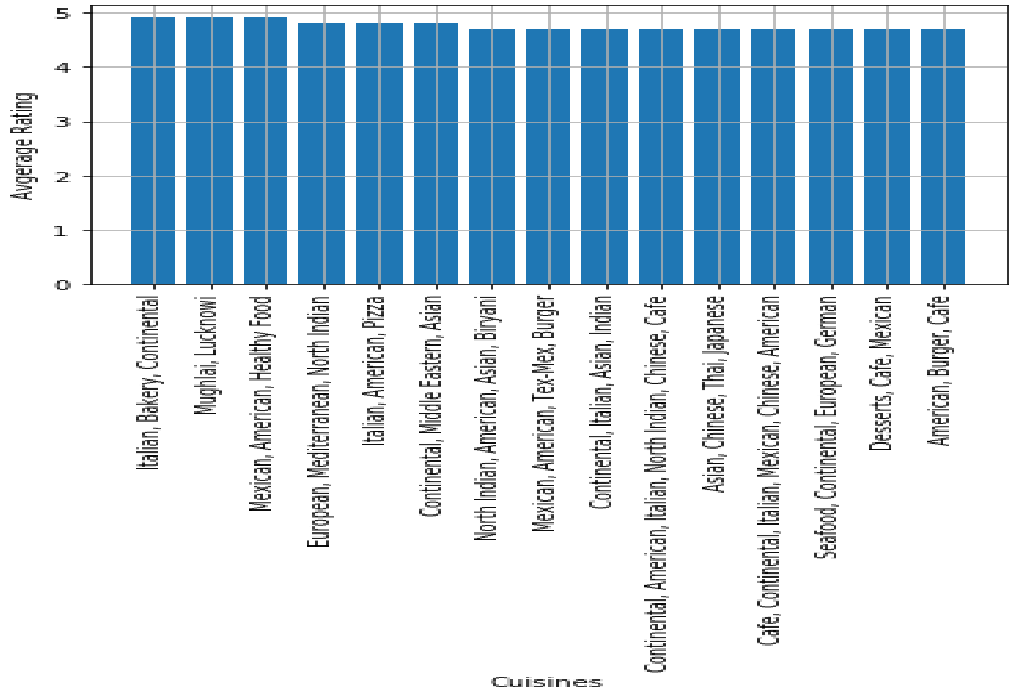
Conclusion

Highest rated restaurants have less average cost in comparison to restaurants having more average cost.

Justification

- In this code, Reading csv file zomato data and storing it in variable data.
- Creating df which stores a copy of data.
- Renaming column names of df are updated to lower case and spaces replaced with '_'.
- Now restricting data frame df to country id 1 that is of India.
- Now creating list of x and y of aggregate rating and average cost for two respectively.
- Plotting bar graph from x and y list with x axis as rating of restaurants and y axis as average cost for two.

Answer 2.1.4:



Conclusion

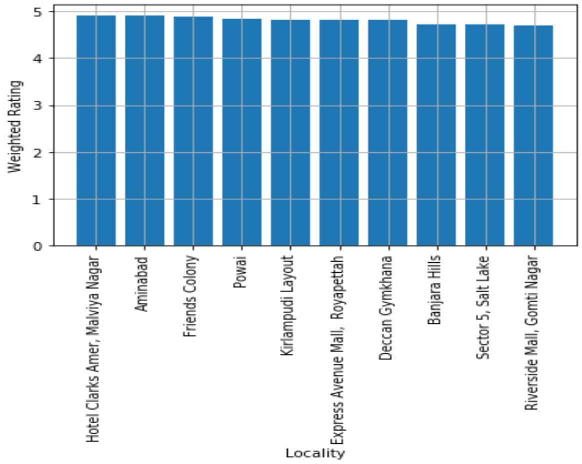
We can infer that serving above dishes leads to highest rating for restaurants.

Justification

- In this code, Reading csv file zomato data and storing it in variable data.
- Creating df which stores a copy of data.
- Renaming column names of df are updated to lower case and spaces replaced with '_'.
- Now restricting data frame df to country id 1 that is of India.
- Grouping cuisines and finding the mean in df and stored in df_grp
- Followed by creating a dataframe df_temp from df_grp containing sorted data according to aggregate_rating in descending order.
- Creating x1 and y1 list from df_temp for plotting with x1 containing list of cuisines and y1 containing list of aggregate rating(average rating of particular cuisine) of top 15 ratings.
- Plotting bar graph from x1 and y1 list and printing the same.

Answer 2.2.1:

```
Top 10 localities with weighted rating in India
Hotel Clarks Amer, Malviya Nagar  :  4.90
Aminabad  :  4.90
Friends Colony  :  4.89
Powai  :  4.84
Kirlampudi Layout  :  4.82
Express Avenue Mall,  Royapettah  :  4.80
Deccan Gymkhana  :  4.80
Banjara Hills  :  4.72
Sector 5, Salt Lake  :  4.71
Riverside Mall, Gomti Nagar  :  4.70
```



Conclusion

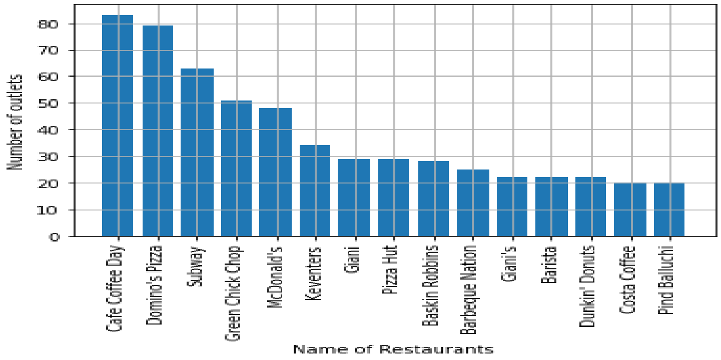
Above extracted data shows top 10 localities with max weighted ratings

Justification

- In this code, Reading csv file zomato data and storing it in variable data.
- Creating df which stores a copy of data.
- Renaming column names of df are updated to lower case and spaces replaced with ‘_’.
- Now restricting data frame df to country id 1 that is of India.
- Creating a list locality of Unique localities
- Using list locality creating a dictionary dict_rating_locality of localities vs weighted ratings.
- Sorting the dictionary in reverse order and fetching top 10 records.
- Creating x and y list for plotting with x containing list of localities and y contaning list of weighted rating
- Plotting bar graph from x and y list and printing the same.

Answer 3.1:

```
Top 15 restaurants in India with respect to 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
Costa Coffee 20
Pind Balluchi 20
```



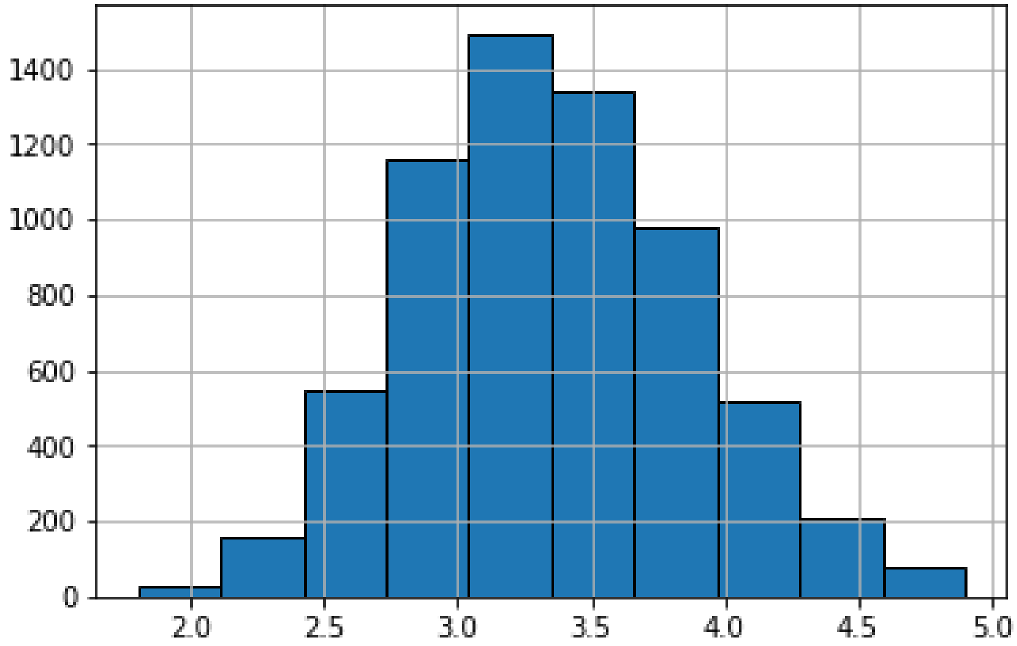
Conclusion

Above extracted data shows Top 15 restaurants in India with respect to number of outlets

Justification

- In this code, Reading csv file zomato data and storing it in variable data.
- Creating df which stores a copy of data.
- Renaming column names of df are updated to lower case and spaces replaced with ‘_’.
- Now restricting data frame df to country id 1 that is of India.
- Calculating the count of number of occurences of a restaurant name(no. of restaurant's outlet) using value count and store it in series_restaurant_count .
- Fetching keys and values from series_restaurant_count of top 15 restaurants with maximum outlets.
- Plotting with x axis containing list of restaurant name and y axis contaning list of its count that is no. of outlets.
- Plotting bar graph from keys and values list and printing the same.

Answer 3.2:



Conclusion

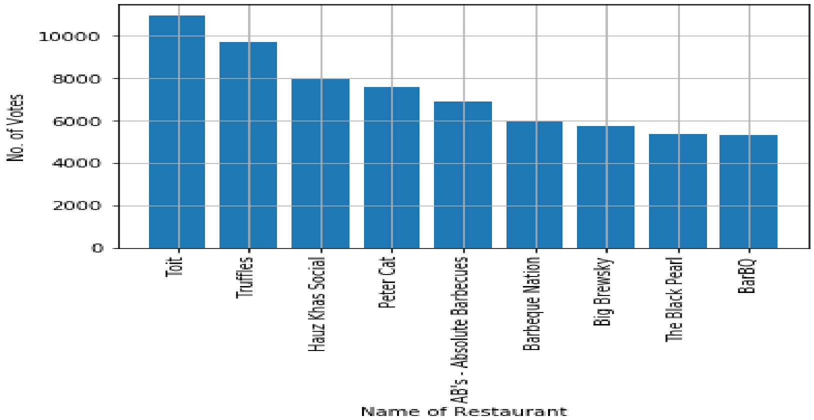
Above extracted data shows Aggregate rating of restaurants in India without considering unrated restaurants. Here, ratings between 3 and 3.5 is maximum while there are v. few restaurants having rating between 4.5 and 5.

Justification

- In this code, Reading csv file zomato data and storing it in variable data.
- Creating df which stores a copy of data.
- Renaming column names of df are updated to lower case and spaces replaced with '_'.
- Now restricting data frame df to country id 1 that is of India.
- Creating list x for plotting containing list of aggregate rating.
- Plotting histogram graph from list x for showing frequency of aggregate ratings

Answer 3.3:

```
51705 Toit 10934
51040 Truffles 9667
308322 Hauz Khas Social 7931
20404 Peter Cat 7574
56618 AB's - Absolute Barbecues 6907
20842 Barbeque Nation 5966
58882 Big Brewsky 5705
94286 AB's - Absolute Barbecues 5434
54162 The Black Pearl 5385
20870 BarBQ 5288
```



Conclusion

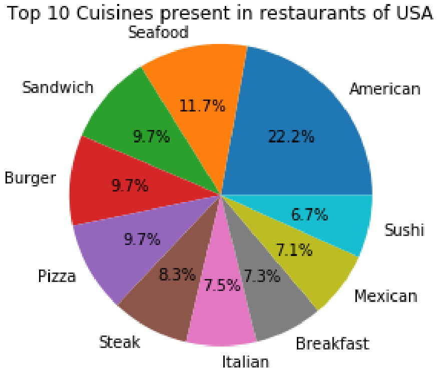
Above extracted data shows top 10 restaurants in the data with the highest number of votes

Justification

- In this code, Reading csv file zomato data and storing it in variable data.
- Creating df which stores a copy of data.
- Renaming column names of df are updated to lower case and spaces replaced with '_'.
- Now restricting data frame df to country id 1 that is of India.
- Grouping data frame wrt restaurant id and applying sum function to it and storing it in df_grp.
- Sorting df_grp accoridng to votes in descending order and storing it in df_temp.
- Creating list x and y by fetching top 10 records from df_temp dataframe.
- Extracting restaurant name with the help of restaurant id and storing it in list z.
- Plotting bar graph from z and y list and printing res_id,res_name,votes.

Answer 3.4:

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



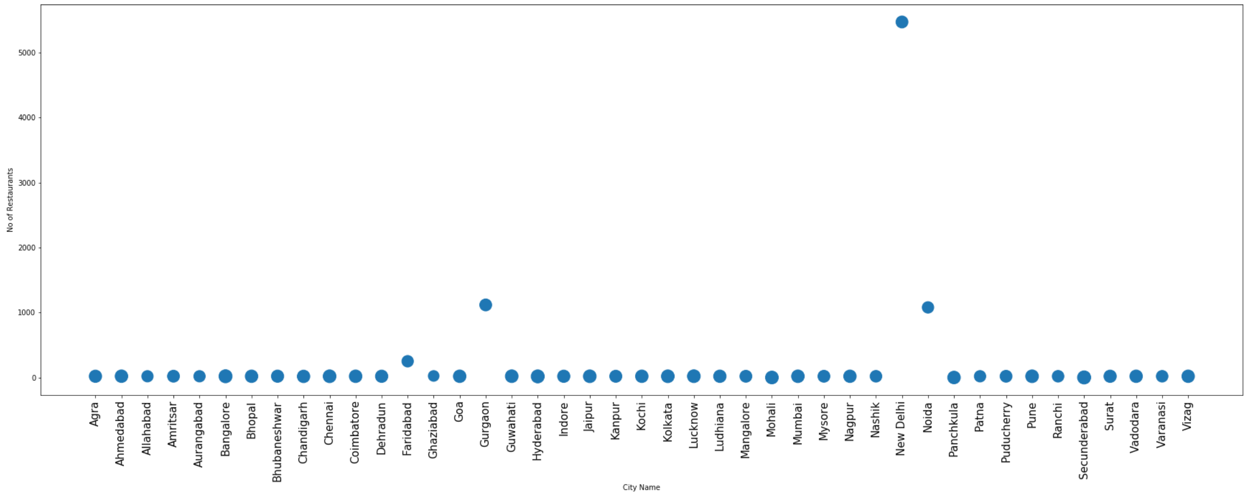
Conclusion

Above extracted data shows top 10 cuisines present in restaurants in the USA

Justification

- In this code, Reading csv file zomato data and storing it in variable data.
- Creating df which stores a copy of data.
- Renaming column names of df are updated to lower case and spaces replaced with ‘_’.
- Now restricting data frame df to country id 216 that is of USA.
- Replacing nan values with None in cuisines column of df.
- Creating data frame df_temp containing df in which cuisines are not none.
- Grouping df_temp wrt to country code and joining + splitting the cuisines too.
- Creating dictionary dict_country_cuisine country as key and value as list of cuisine by grouping the dataframe wrt country
- Extracting list cuisines served in USA with the help of dict_country_cuisine
- Creating data frame df_cuisines for USA Cuisines
- Counting the no. of occurances of cuisines servered in restaurants with the help of value count and stored in series_cuisines_count_usa.
- Using series_cuisines_count_usa, creating a list label as cuisines names and sizes as its counts.
- Plotting pie chart from labels and sizes list and printing the same.

Answer 3.5:



Conclusion

Above extracted data shows 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.

Justification

- In this code, Reading csv file zomato data and storing it in variable data.
- Creating df which stores a copy of data.
- Renaming column names of df are updated to lower case and spaces replaced with '_'.
- Now restricting data frame df to country id 1 that is of India.
- Creating a list city of Unique cities .
- Using list city creating a dictionary dict_rating_city of city vs weighted ratings.
- Finding number of resturants in a City and store it in a dictionary dict_city_resCount.
- Sorting items of dict_city_resCount and dict_rating_city
- From dict_city_rating has data of City VS Weighted Rating and dict_city_resCount has data of City Vs Number of Restaurants, generating list x, y and z.
- Creating list p have each elements of z multiplied by 75(for visual appearance)
- Plotting bubble graph from x, y and p list ,p as scalar and x&y as x axis and y axis.