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

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

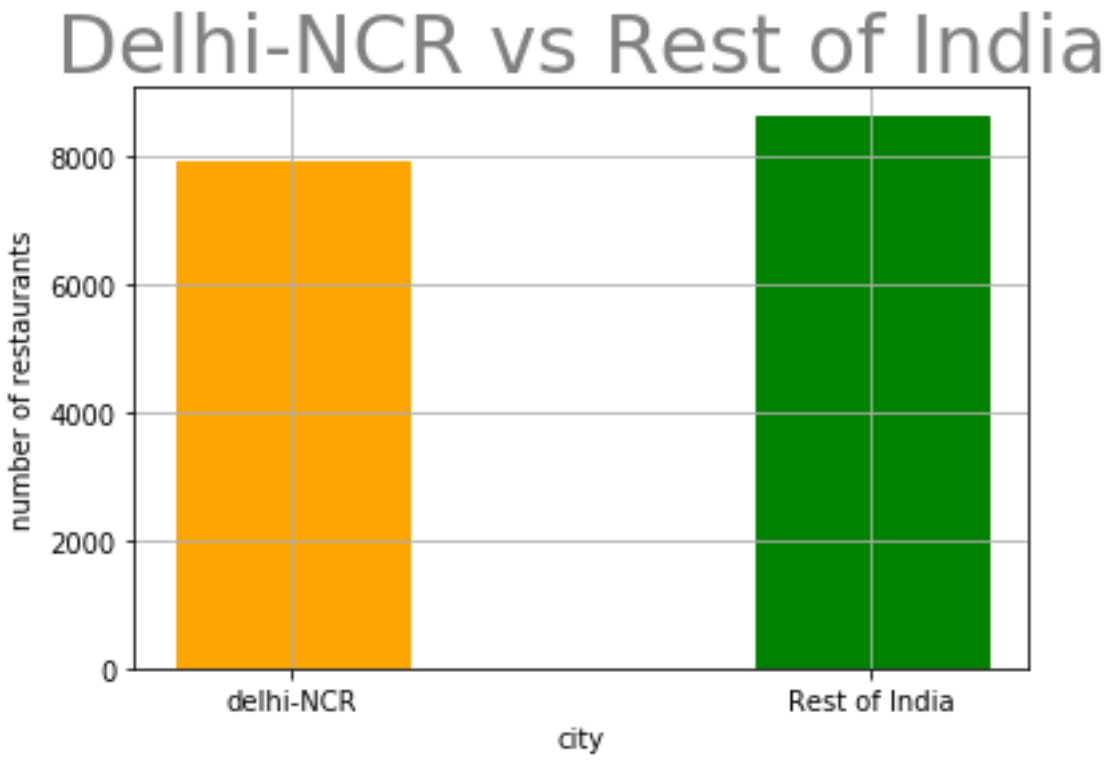
2. Find the cuisines which are not present in restaurant of Delhi NCR but present in rest of India.Check using Zomato API whether this cuisines are actually not served in restaurants of Delhi-NCR or just it due to incomplete dataset.

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

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.

1.

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



code explanation:

We first took all the restaurants belonging to Delhi-NCR in one dataframe and the rest of India in another. Then we plotted a graph between the length of the two dataframes and the regions.

graph explanation:

From the graph, we can see that the number of restaurants in the rest of India is definitely greater than that of Delhi-NCR. Also, the difference between the two numbers is not that significant. Which implies that there are a significant number of restaurants in Delhi-NCR as compared to the rest of India if compared considering the geographical area.

1.

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

```
the cuisines not present in Delhi-NCR are:
Cajun
BBQ
Malwani
German

Themis Barbecue House 4.6 18204820 Netaji Subhash Place
Barbeque Nation 4.2 9561 Netaji Subhash Place
Bengal Bar-Be-Que 3.5 19250749 Pitampura
New Bombay Bar-Be-Que 3.6 5303 Prashant Vihar
UBQ by Barbeque Nation 3.8 19082741 Lawrence Road
UBQ by Barbeque Nation 3.9 18889760 Rohini
Kays Bar-Be-Que 3.8 3079 Ashok Vihar Phase 1
Picnick Scooter Cafe 0 19149978 Punjabi Bagh
7teen BBQ Bistro Bnq 3.8 18595610 Moti Nagar
Pishori Chicken 0 19061375 Kirti Nagar
```

code explanation:

Again we created two different dataframes consisting of data of restaurants present and Delhi-NCR and the rest of India respectively. We then stored the distict cuisines served in the two regions in two sets. For finding the cuisines present in rest of India and not in Delhi-NCR, we simply took the difference of the two sets.

For checking if the dataset is correct or not, we made an API call to zomato API to fetch the restaurants serving one of the missing cuisines(BBQ) near coding ninjas(New Delhi). We found that many restaurants are serving BBQ. Thus we conclude that due to incomplete dataset we are getting an incorrect answer.

1.

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

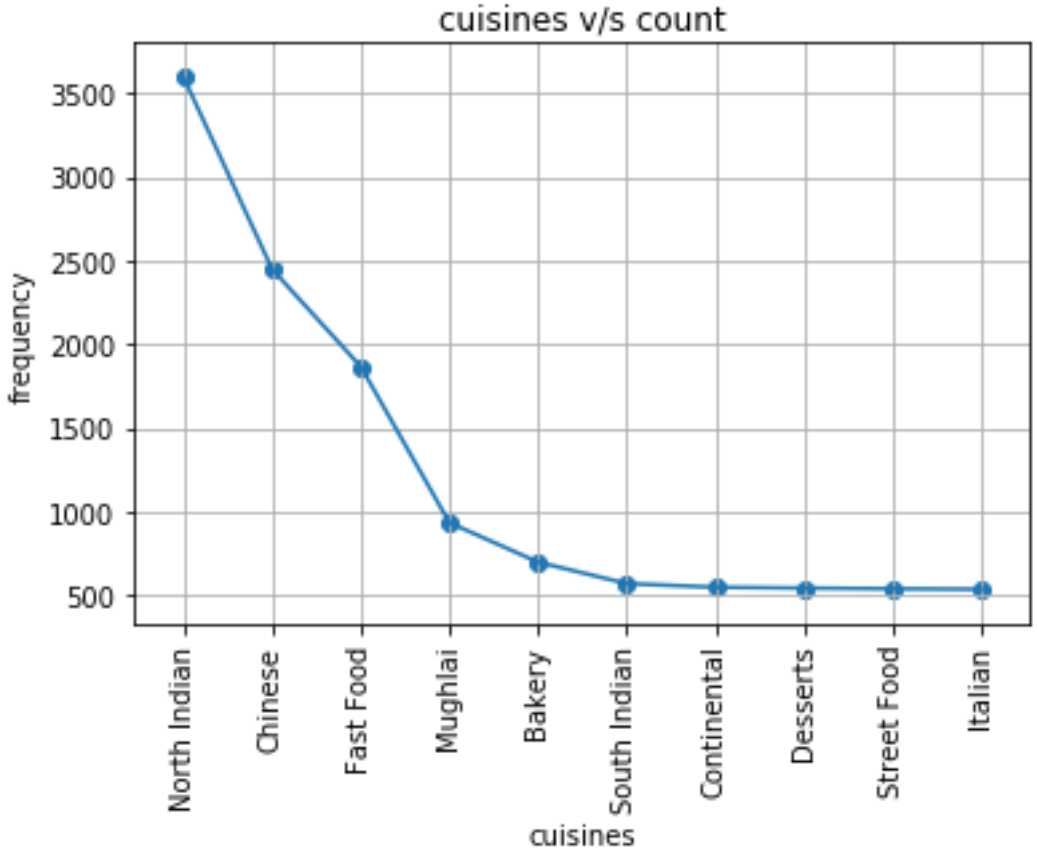
```
TOP 10 CUISINES SERVED IN DELHI-NCR
**********************************
North Indian 3597
Chinese 2448
Fast Food 1866
Mughlai 933
Bakery 697
South Indian 569
Continental 547
Desserts 542
Street Food 538
Italian 535
TOP 10 CUISINES SERVED IN THE REST OF INDIA
******************************************
North Indian 3946
Chinese 2690
Fast Food 1963
Mughlai 992
Bakery 726
Continental 724
Italian 682
South Indian 631
Cafe 627
Desserts 597
```

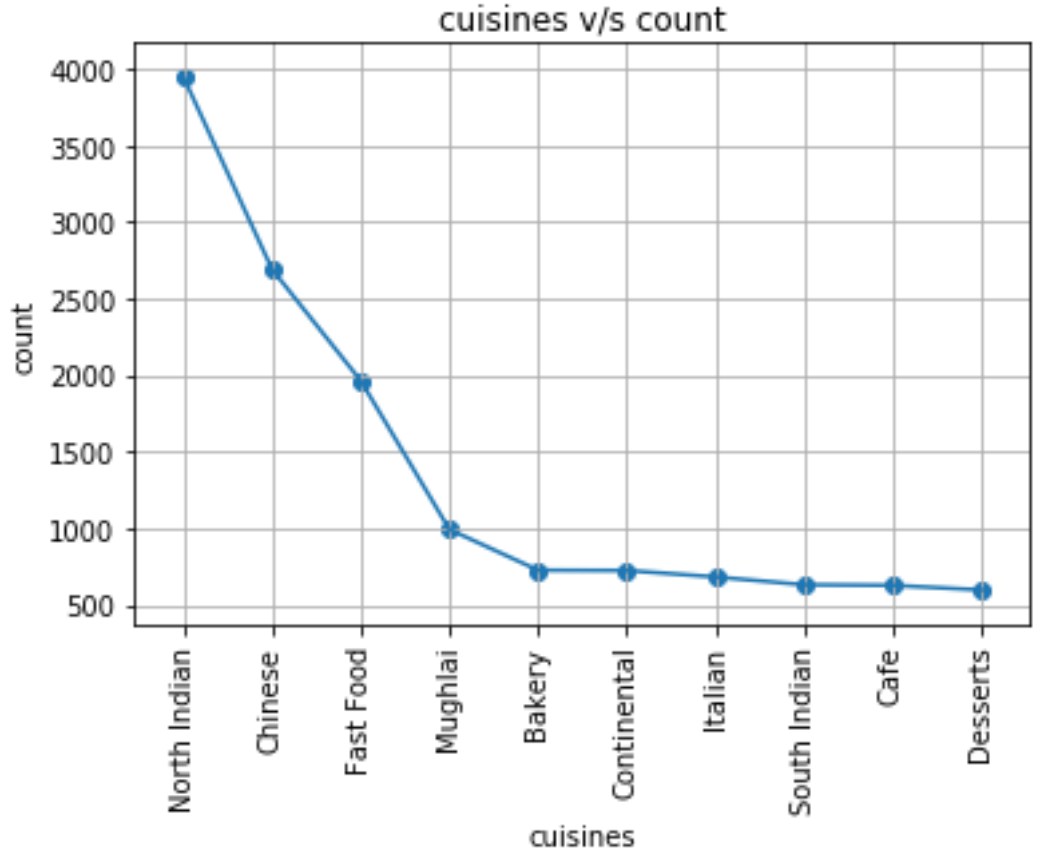
code explanation:

Again we stored the details of the restaurants in Delhi-NCR and in the rest of India in two different datasets respectively. Then we appended all cuisines of each region in two different lists. With the help of a dictionary, we found out the frequency of each cuisine and displayed the top 10 cuisines server in the two regions(Delhi-NCR and the rest of India).

1.

1.4. 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.


cuisines v/s count
frequency
3500
3000
2500
2000
1500
1000
500
North Indian
Chinese
Fast Food
Mughlai
Bakery
South Indian
Continental
Desserts
Street Food
Italian
cuisines


cuisines v/s count
count
4000
3500
3000
2500
2000
1500
1000
500
North Indian
Chinese
Fast Food
Mughlai
Bakery
Continental
Italian
South Indian
Cafe
Desserts
cuisines

graph explanation:

The graph is plotted between the cuisines and From the graph we see that there are more number of restaurants in the rest of India as compared to Delhi-NCR. Although, the curves of both graph as same in shape the scale varies sligthly. Thus, the differenc ebetween the cuisines served in Delhi-NCR an dthe rest of India is that of number of restaurants serving that cuisine, which is more in case of rest of India.

2.

User Rating of a restaurant plays a crucial role in selecting a restaurant or ordering the food from the restaurant. Write a short detail analysis of how the rating is affected by restaurant due following features:

1.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=Σ (number of votes * rating) / Σ (number of votes) .

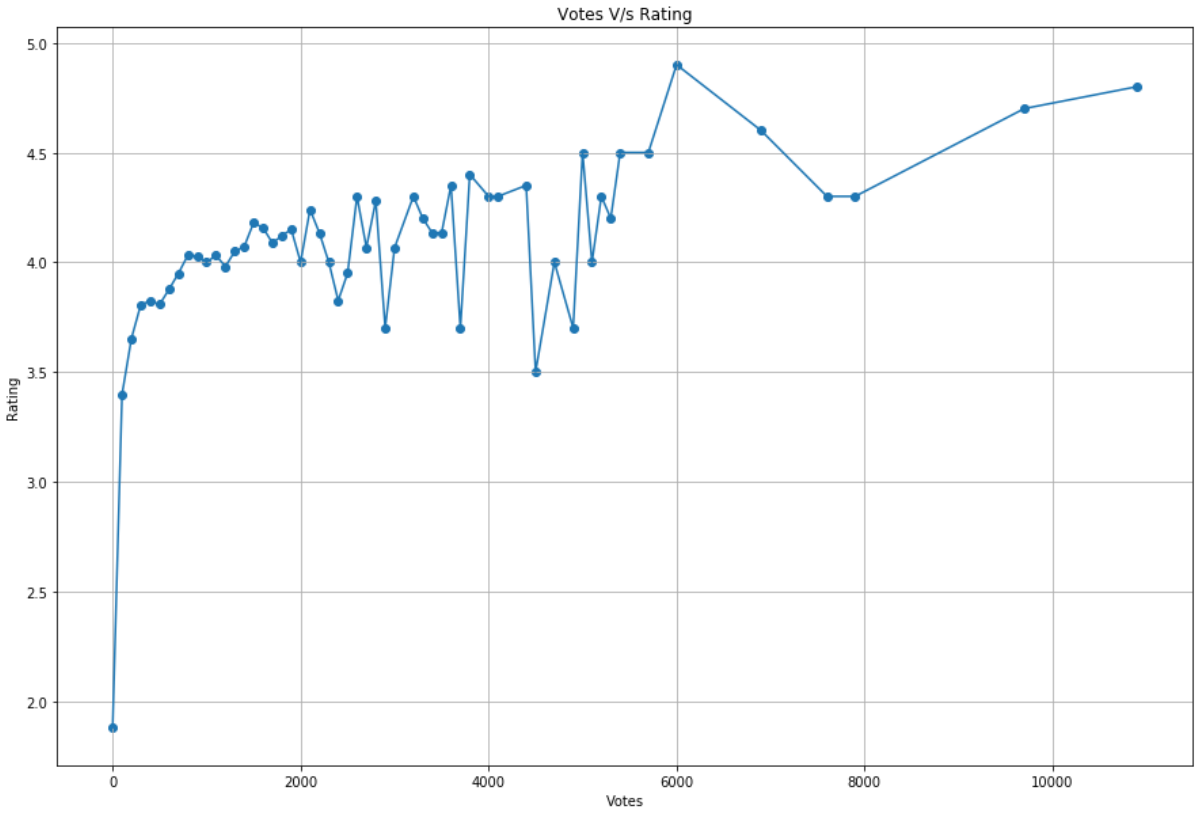
2.

1.1. Number of Votes given Restaurant

Answer:

{100: 3.3971715901948474, 200: 3.6525252525252547, 0: 1.879156693058369
4, 1600: 4.1571428571428575, 1300: 4.05, 800: 4.031372549019607, 700: 3
.949152542372881, 900: 4.026923076923076, 400: 3.823626373626373, 300:
3.806993006993008, 1100: 4.03076923076923, 1000: 3.999999999999999, 500
: 3.807476635514019, 2400: 3.825, 10900: 4.8, 1400: 4.071428571428571,
3500: 4.133333333333333, 9700: 4.7, 5400: 4.5, 6900: 4.6, 600: 3.876388
88888889, 5700: 4.5, 1800: 4.12, 1500: 4.18, 1700: 4.0874999999999995,

1200: 3.98, 3800: 4.4, 2300: 4.0, 2200: 4.133333333333334, 1900: 4.1499999999999995, 4400: 4.35, 2800: 4.279999999999999, 2100: 4.24, 3600: 4.35, 3400: 4.133333333333334, 2600: 4.3, 4500: 3.5, 5300: 4.2, 7600: 4.3, 6000: 4.9, 5100: 4.0, 2000: 4.0, 2700: 4.066666666666667, 3200: 4.3, 2500: 3.95, 3000: 4.066666666666666, 4900: 3.7, 5200: 4.3, 4100: 4.3, 3700: 3.7, 7900: 4.3, 4700: 4.0, 5000: 4.5, 2900: 3.7, 3300: 4.2, 4000: 4.3}



Explanation of code:

We append the Rating into an array and the Votes into another array. We have taken a dictionary in which average rating(values) is stored with the corresponding rounded off votes(keys). Here we have taken the round off of votes to the nearest hundred because there were a lot of votes concentrated at one point and it seemed congested.

Explanation of graph:

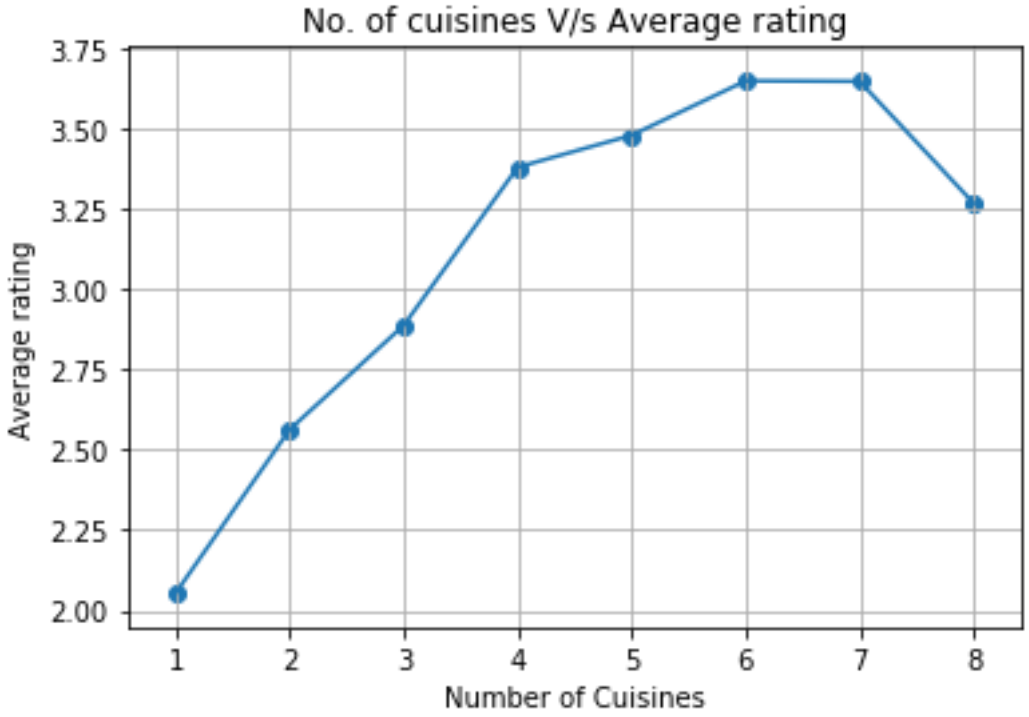
The rating is between or below 3.5 and 4.5 if the votes are below or equal to 2000. We then see that as the rating increases , there is increase in number of votes. The highest rating is given by Votes that are 6000. The rating drops when the Votes are around 8000 and then again the rating rises.

2.

1.2. Restaurant serving more number of cuisines

Answer:

{2: 2.563124403436212, 1: 2.0545219638242873, 3: 2.887052168447519, 4: 3.379166666666665, 6: 3.6478260869565218, 5: 3.478571428571428, 7: 3.646153846153846, 8: 3.266666666666667}



code explanation:

After removing all nan values and taking details of restaurants related to India in a different dataset, we took two dictionaries to store count and the total rating given to a particular number of cuisines served. We used the count dictionary to find the average rating so as to plot an efficient graph.

graph explanation:

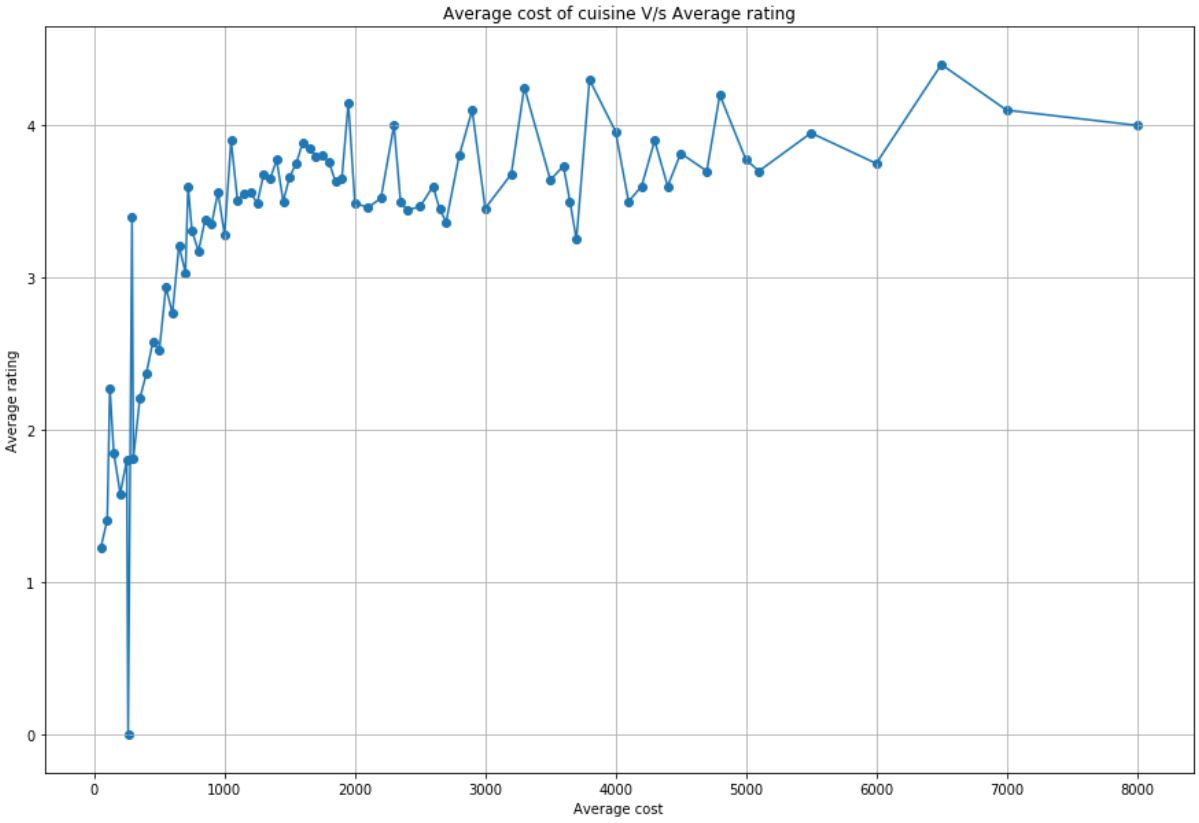
From the graph we see that as the number of cuisines increases the rating seems to increase steeply linearly upto 4 cuisines. From 4-6 number of cuisines the ratings increase less steeply(with lesser slope). For 6-7 number of cuisines the rating remain almost constant but decreases as the number of cuisines reaches 8.

2.

1.3. Average Cost of Restaurant

{850: 3.3777777777777778, 700: 3.0334164588528663, 500: 2.5264277715565497, 400: 2.3710247349823326, 1000: 3.2818840579710153, 2000: 3.4913461538461545, 2500: 3.4666666666666672, 800: 3.1744186046511613, 3600: 3.733333333333333, 550: 2.938461538461538, 1100: 3.5025641025641017, 1500: 3.6639784946236547, 900: 3.3490740740740756, 1800: 3.762121212121212, 300: 1.814576271186441, 1400: 3.7730769230769234, 350: 2.206430155210644, 450: 2.57909090909091, 600: 2.7649230769230746, 950: 3.5619047619047617, 1200: 3.558741258741258, 200: 1.5819793205317574, 100: 1.4090909090

909092, 150: 1.8497175141242936, 650: 3.2090395480225995, 1300: 3.67761
19402985077, 2400: 3.4428571428571426, 250: 1.8040909090909083, 1600: 3
.887179487179487, 2200: 3.5222222222222226, 750: 3.3058823529411767, 17
00: 3.796774193548386, 1250: 3.488235294117647, 1900: 3.649999999999999
5, 2700: 3.3666666666666667, 3500: 3.641666666666666, 5000: 3.775, 2300
: 4.0, 1750: 3.7999999999999994, 3200: 3.6799999999999997, 3000: 3.4552
631578947364, 2100: 3.46, 1850: 3.633333333333333, 4500: 3.816666666666
667, 4000: 3.9599999999999995, 3300: 4.25, 120: 2.275, 1350: 3.65, 1650
: 3.85, 1950: 4.15, 1050: 3.9, 290: 3.4, 50: 1.225, 1450: 3.5, 2800: 3.
8, 260: 0.0, 3650: 3.5, 4800: 4.2, 1550: 3.75, 720: 3.6, 6500: 4.4, 700
0: 4.1, 1150: 3.55, 2900: 4.1, 2600: 3.6, 4700: 3.7, 3700: 3.25, 3800:
4.3, 4100: 3.5, 4300: 3.9, 2650: 3.45, 6000: 3.75, 5100: 3.7, 5500: 3.9
5, 4400: 3.6, 4200: 3.6, 8000: 4.0, 2350: 3.5}



Explanation of code:

We append the Rating into an array and the Cost into another array. We have taken a dictionary in which average rating(values) is stored with the corresponding cost(keys).

Explanation of graph:

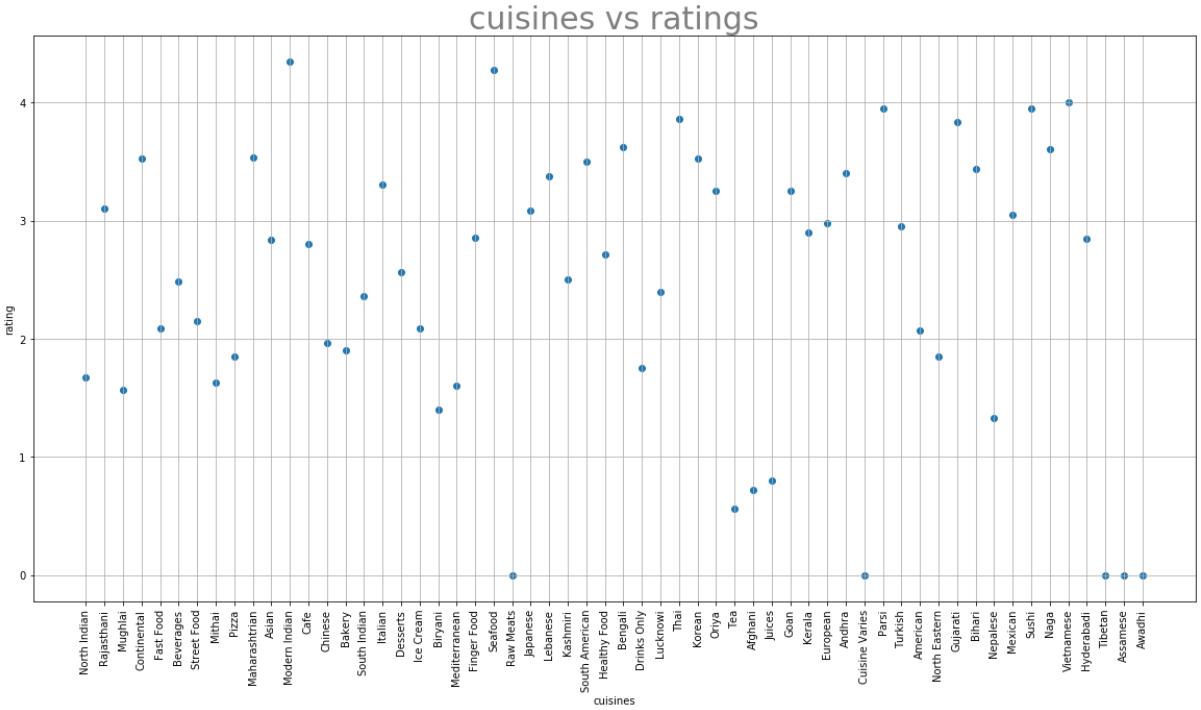
The rating is increasing drastically when cost increases from 0 to 1000. Then the rating increases gradually. This shows that there is an overall increase in the rating if the average cost increases. Although the rating drops a little bit after the cost becomes more than 6500.

2.

1.4. Restaurant serving some specific cuisines

{'North Indian': 1.6723290598290599, 'Rajasthani': 3.1, 'Mughlai': 1.56
4077669902913, 'Continental': 3.525, 'Fast Food': 2.0890804597701145, '

Beverages': 2.4833333333333325, 'Street Food': 2.150675675675676, 'Mithai': 1.6267605633802817, 'Pizza': 1.8499999999999994, 'Maharashtrian': 3.533333333333333, 'Asian': 2.833333333333334, 'Modern Indian': 4.345454545454545, 'Cafe': 2.7989247311827965, 'Chinese': 1.9611764705882349, 'Bakery': 1.9055555555555554, 'South Indian': 2.3576576576576582, 'Italian': 3.306666666666666, 'Desserts': 2.561224489795918, 'Ice Cream': 2.09054054054054, 'Biryani': 1.3981481481481481, 'Mediterranean': 1.6, 'Finger Food': 2.854166666666666, 'Seafood': 4.275, 'Raw Meats': 0.0, 'Japanese': 3.08, 'Lebanese': 3.371428571428571, 'Kashmiri': 2.5, 'South American': 3.5, 'Healthy Food': 2.7133333333333334, 'Bengali': 3.62, 'Drinks Only': 1.75, 'Lucknowi': 2.4, 'Thai': 3.8600000000000003, 'Korean': 3.5250000000000004, 'Oriya': 3.25, 'Tea': 0.5666666666666667, 'Afghani': 0.725, 'Juices': 0.8, 'Goan': 3.25, 'Kerala': 2.9, 'European': 2.9749999999999996, 'Andhra': 3.4, 'Cuisine Varies': 0.0, 'Parsi': 3.95, 'Turkish': 2.95, 'American': 2.0666666666666664, 'North Eastern': 1.85, 'Gujarati': 3.8333333333333335, 'Bihari': 3.4333333333333336, 'Nepalese': 1.3333333333333333, 'Mexican': 3.0499999999999994, 'Sushi': 3.95, 'Naga': 3.6, 'Vietnamese': 4.0, 'Hyderabadi': 2.85, 'Tibetan': 0.0, 'Assamese': 0.0, 'Awadhi': 0.0}



code explanation:

After dropping all nan values and storing the data related to India in a different dataset, we tranverse the Cuisines column and filtered it for restaurants serving only one cuisine. We then store the raing and count of the cuisines in two separate dictionaries. We used the count to take the average rating for each cuisine. And finally plottes a graph between the restaurants serving a particular cuisine and the average rating of the restaurant.

graph explanation:

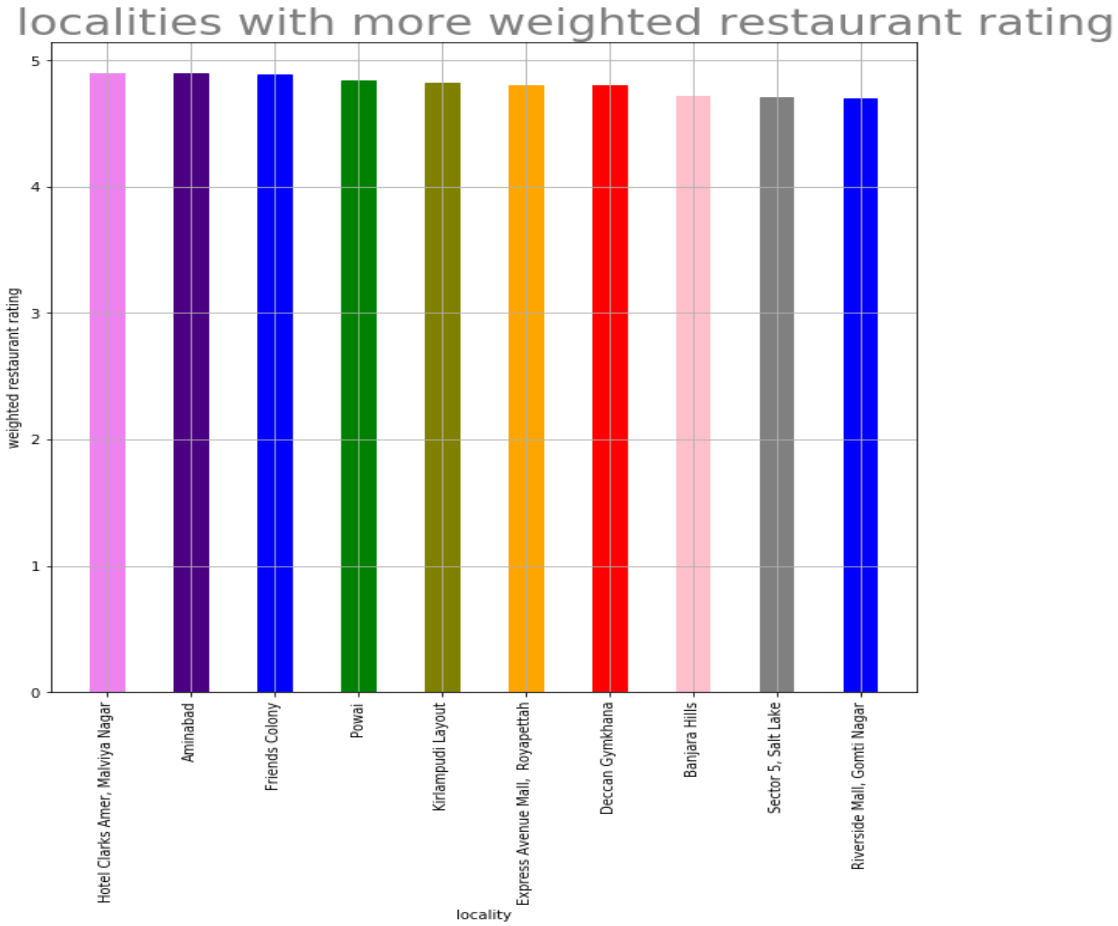
The graph seems scattered randomly everywhere on the plane. We can infer that the highest rating is that of 'Mordern Indian' cuisine and 'Seafood' cuisine. The lowest raing is that of 'raw meat','tibetian','Assamese' and 'Awadhi'

2.

2.1. Find the weighted restaurant rating of each locality and find out the top 10 localities with more weighted restaurant rating? Weighted Restaurant Rating=Σ (number of votes * rating) / Σ (number of votes) .

Answer:

```
Hotel Clarks Amer, Malviya Nagar 4.9
Aminabad 4.9
Friends Colony 4.886916367367881
Powai 4.841868613138686
Kirlampudi Layout 4.820161290322581
Express Avenue Mall,  Royapettah 4.8
Deccan Gymkhana 4.8
Banjara Hills 4.7187617260787995
Sector 5, Salt Lake 4.707022552098202
Riverside Mall, Gomti Nagar 4.7
```



Explanation of code:

We append the Rating into an array and the locality into another array and the votes into another array. We have taken a dictionary in which average rating(values) is stored with the locality(keys). We have taken a second dictionary in which votes(values) is stored with the locality(keys). We have taken a third dictionary in which weighted restaurant rating(values) is

stored with the locality(keys). The weighted rating has been found out using the other two dictionaries.

Explanation of graph:

The weighted restaurant rating of top ten localities seems to be almost the same. The height of the bars are only gradually decreasing.

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).

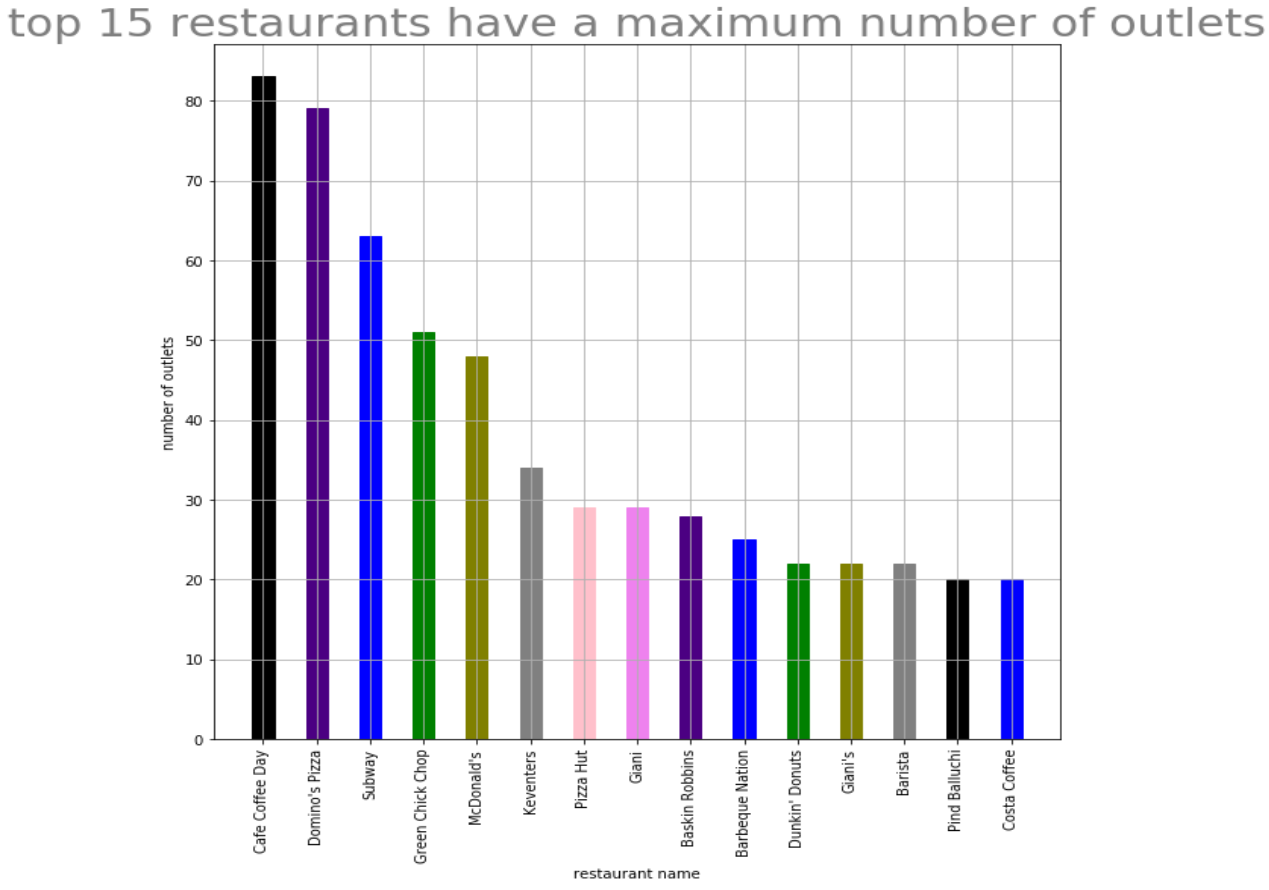
3.Plot the bar graph top 10 restaurants in the data with the highest number of votes.

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.

3.

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



code explanation:

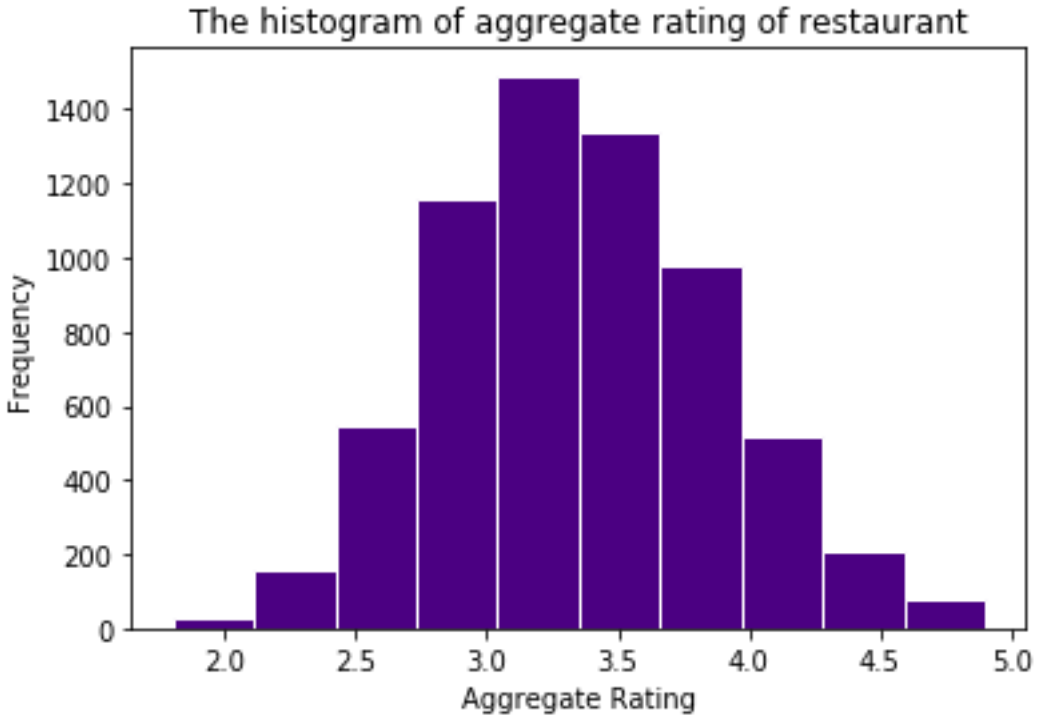
We took the restaurant names along with their frequencies in a dictionary and dislayed the top 15 restaurants.

graph explanation:

From the graph we see that there is a gradual decrease in number of outlets. Cafe Coffe Day has the maximun number of outlets. Costa Coffee has the lowest number of outlets among the top 15 restaurants.

3.

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



Explanation of code:

The aggregate rating has been stored inside an array. That array has been used to plot the histogram.

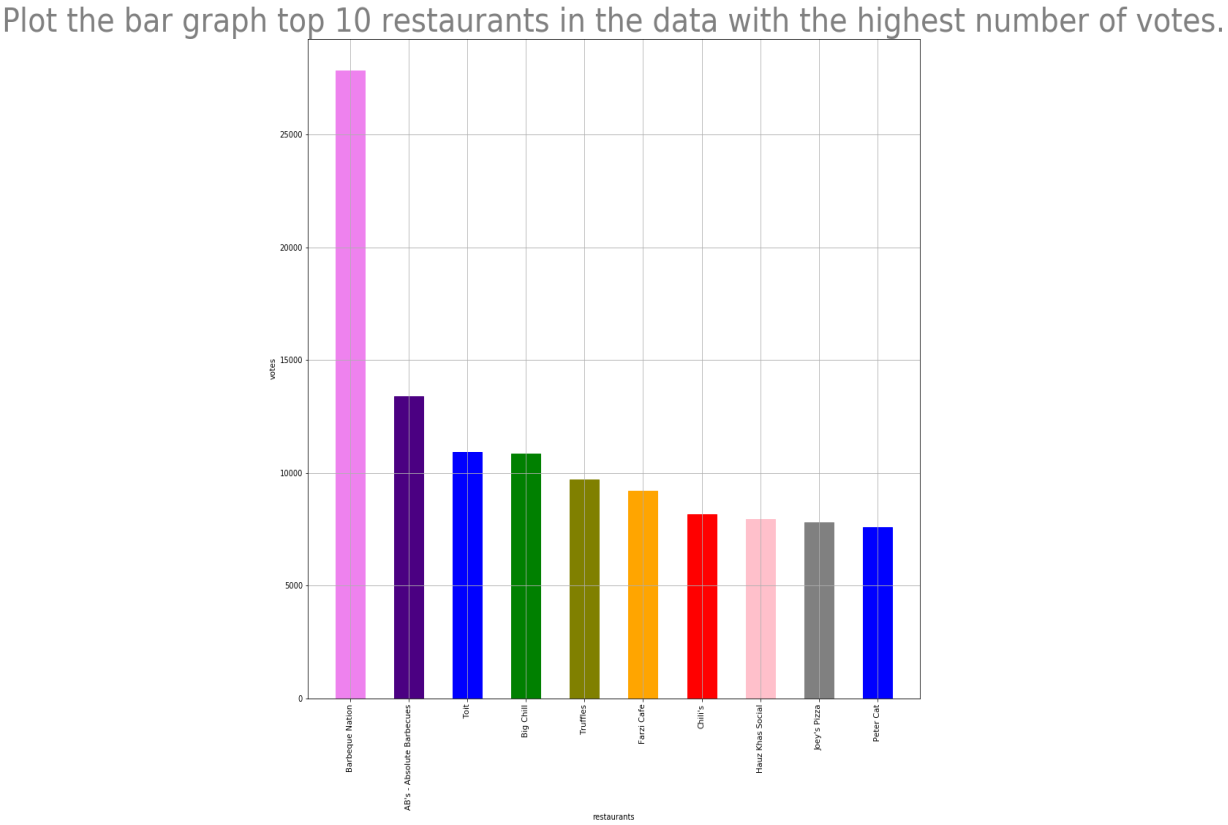
Explanantion of Histogram:

The frequency of the ratings between 2.9 and 3.5 is the highest. The frequency of the rating between 0.5 and 1.8 is zero. The frequency of the ratings are lowest for ratings between 4.5 and 4.9.

3.

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

```
Barbeque Nation 27835
AB's - Absolute Barbecues 13400
Toit 10934
Big Chill 10853
Truffles 9682
Farzi Cafe 9189
Chili's 8156
Hauz Khas Social 7931
Joey's Pizza 7807
Peter Cat 7574
```



code explanation:

We appended the resaurant names and the votes in two separate lists. We maintained a dictionary to store the restaurant name and the corresponding number of votes.

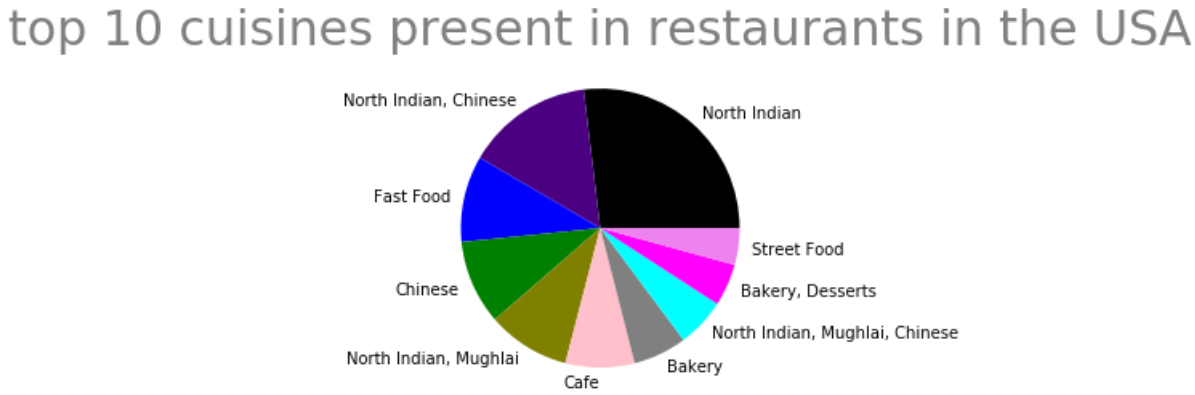
graph explanation:

We see that Barbeque Nation has the maximum number of votes. There is a steady decresea among the rest of the restaurants but the difference between barbeque Nation and AB's is significant.

3.

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

```
North Indian 936
North Indian, Chinese 511
Fast Food 348
Chinese 340
North Indian, Mughlai 334
Cafe 279
Bakery 216
North Indian, Mughlai, Chinese 197
Bakery, Desserts 170
Street Food 148
```



code explanation:

After extracting all details of the restaurants present in USA, we traversed the cuisine column to maintain a dictionary of the frequency of cuisines. And thus plotted a pie chart of the top 10 cuisines present in restaurants in the USA.

graph explanation:

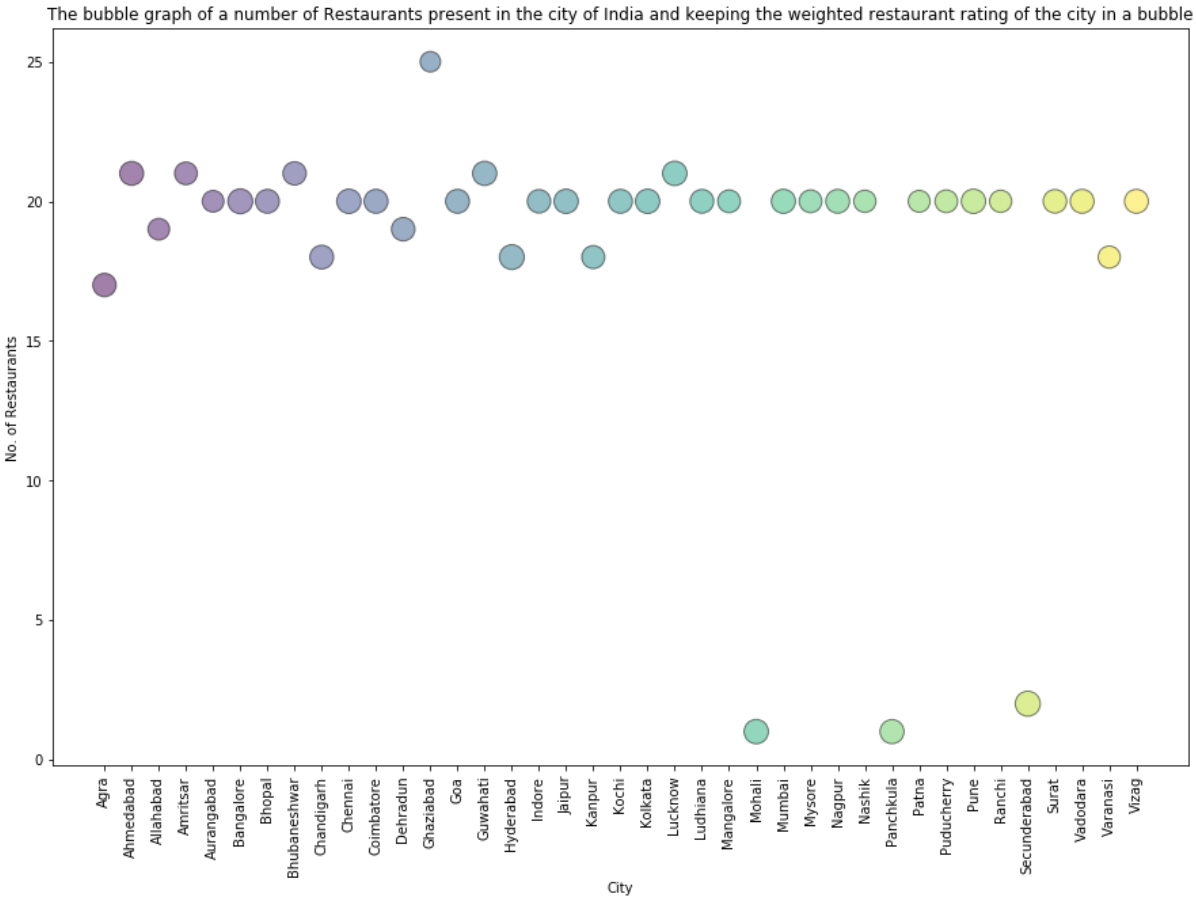
We see that the top cuisine in USA is Mexican followed by American. All other cuisines in the top 10 list have more or less the same significance.

3.

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.

```
City  Number of restaurants  Weighted rating
Agra   17   3.980493273542602
Ahmedabad   21   4.163215392140877
Allahabad   19   3.4136262099776618
Amritsar   21   3.7645839017735336
```

Aurangabad	20	3.427314814814815
Bangalore	20	4.497423148890671
Bhopal	20	4.1259297879735835
Bhubaneshwar	21	3.9677586613245337
Chandigarh	18	4.106690515324375
Chennai	20	4.319682253114281
Coimbatore	20	4.174994074425219
Dehradun	19	4.008618504435994
Ghaziabad	25	3.0378698224852068
Goa	20	4.158438203735593
Guwahati	21	4.265733279613215
Hyderabad	18	4.487772943857468
Indore	20	3.9983462424115555
Jaipur	20	4.281507965242579
Kanpur	18	3.8553828439838798
Kochi	20	4.137380191693291
Kolkata	20	4.295151705424619
Lucknow	21	4.3235268346111715
Ludhiana	20	4.083607130373995
Mangalore	20	3.761117471948706
Mohali	1	4.3
Mumbai	20	4.2166077381553695
Mysore	20	3.76120244246125
Nagpur	20	4.10865162037037
Nashik	20	3.5872553699284
Panchkula	1	4.2
Patna	20	3.455065789473683
Puducherry	20	3.73706191588785
Pune	20	4.2839957553540415
Ranchi	20	3.5736026563364693
Secunderabad	2	4.548054679284963
Surat	20	4.017698391055359
Vadodara	20	4.131989247311829
Varanasi	18	3.556420233463035
Vizag	20	4.131187061336588



Explanation for code:

We have taken an array and stored the number of restaurants present in Cities . We append the Rating into an array and the City into another array and the votes into another array. We have taken a dictionary in which average rating(values) is stored with the City(keys). We have taken a second dictionary in which votes(values) is stored with the City(keys). We have taken a third dictionary in which weighted restaurant rating(values) is stored with the City(keys). The weighted rating has been found out using the other two dictionaries. We have multiplied the weighted rating with 75 to get relatively larger bubbles for better visualisation. Putting the arrays of number of Restaurants present in the city(y axis),city name(x axis) and weighted restaurant rating(bubble size) into the bubble plot. We have taken the cities which have number of restaurants less than 50 because the rest of the points less.

Explanation for graph:

We see that the number of restaurants between 15 and 25 is more. Ghaziabad has the restaurants around 25. Panchkula, Secunderabad and Mohali have less than 5 restaurants. The Weighted restaurant rating seems to be almost the same that is between 3.5 and 4.5.