

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.
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?
 1. $\text{Weighted Restaurant Rating} = \frac{\sum (\text{number of votes} * \text{rating})}{\sum (\text{number of votes})}$.
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.

ANSWERS

1.1

CODES

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

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

df_start=pd.read_csv(r'C:\Users\nikhi\OneDrive\Desktop\zomato.csv',engine='python')
df_start['City'].dropna(inplace=True)
df_start = df_start[df_start['Country Code'] == 1]
df_start['City'].replace("New Delhi","Delhi-NCR",inplace=True)
df_start['City'].replace("Ghaziabad","Delhi-NCR",inplace=True)
df_start['City'].replace("Noida","Delhi-NCR",inplace=True)
df_start['City'].replace("Gurgaon","Delhi-NCR",inplace=True)
df_start['City'].replace("Faridabad","Delhi-NCR",inplace=True)
df_start1 = df_start[df_start['City'] != 'Delhi-NCR']
#print(df_start['Country Code'])

Delhi=df_start.groupby('City')['City'].count().sort_values(ascending=False)
k = ['Delhi-NCR','Rest of India']
l = []

India = df_start1.groupby('Country Code')['Country Code'].count().sort_values(ascending=False)
l.append(Delhi['Delhi-NCR'])
l.append(India[1])

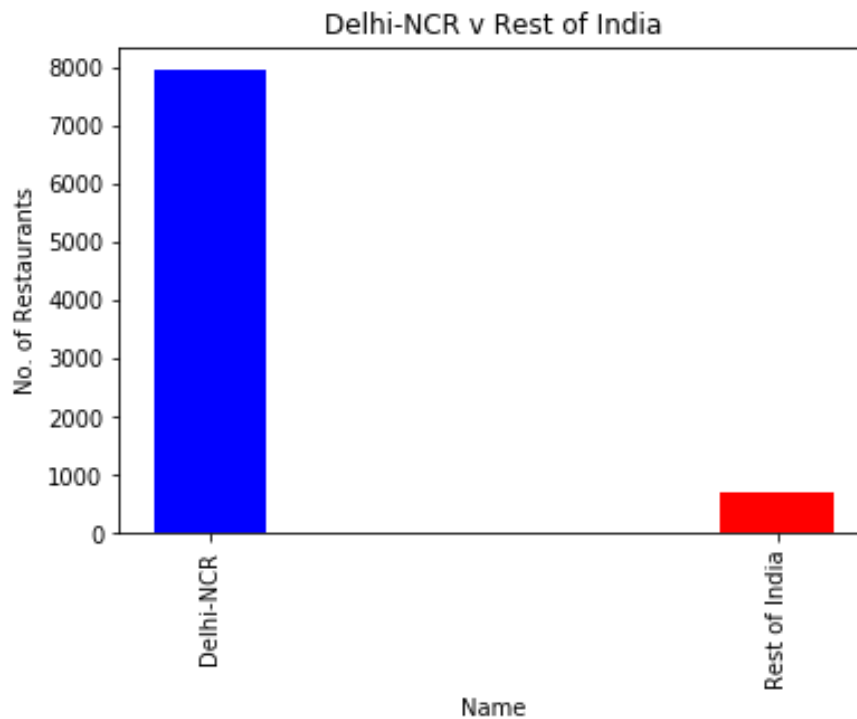
print('Delhi-NCR',l[0])
print('Rest of India',l[1])

plt.bar(k,l,width = 0.2,color = ["blue","red"])
plt.xticks(rotation = 90)
plt.xlabel('Name')
plt.ylabel('No. of Restaurants')
plt.title('Delhi-NCR v Rest of India')
plt.show()
```

ANSWER

Delhi-NCR 7947

Rest of India 705



JUSTIFICATION

1. First we have to read zomato.csv file in CSV format and we have to select India by selecting 'Country Code' == 1.
2. Than we have to replace some cities name with Delhi-NCR as mention in the question.
3. Than we have to count No. of restaurants in Delhi-NCR and Rest_of_India.
4. We plot the bar graph b/w No. of Restaurants and Delhi-NCR v Rest of India.

1.2

CODE

##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.

```
from collections import Counter
```

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
import csv
```

```
import pandas as pd
```

```
import operator
```

```

df_start=pd.read_csv(r'C:\Users\nikhi\OneDrive\Desktop\zomato.csv',engine='python')
df_start['City'].dropna(inplace=True)
df_start = df_start[df_start['Country Code'] == 1]
df_start['City'].replace("New Delhi","Delhi-NCR",inplace=True)
df_start['City'].replace("Ghaziabad","Delhi-NCR",inplace=True)
df_start['City'].replace("Noida","Delhi-NCR",inplace=True)
df_start['City'].replace("Gurgaon","Delhi-NCR",inplace=True)
df_start['City'].replace("Faridabad","Delhi-NCR",inplace=True)
df_start2 = df_start[df_start['City'] == 'Delhi-NCR']

def sep_Cuisines(Cuisines):
    return [i.strip() for i in Cuisines.split(",")]

delhi = { }
rest_of_india = { }
df_start2["Cuisines"]= df_start2["Cuisines"].apply(sep_Cuisines)
for i in df_start2["Cuisines"]:
    for j in i:
        delhi[j] = 1
df_start3 = df_start[df_start['City'] != 'Delhi-NCR']
df_start3["Cuisines"]= df_start3["Cuisines"].apply(sep_Cuisines)
for i in df_start3["Cuisines"]:
    for j in i:
        rest_of_india[j] = 1
delhi1 = []
rest_of_india1 = []
for key in delhi.keys():
    delhi1.append(key)
for key in rest_of_india.keys():
    rest_of_india1.append(key)
#print(a)

```

```
#print(b)

for i in rest_of_india1:
    if(i in delhi1):
        continue
    else:
        print(i)

print("\n\nBBQ is served")
```

ANSWER

German
Malwani
BBQ
Cajun

BBQ is served

JUSTIFICATION

1. First we have to read zomato.csv file in CSV format and we have to select India by selecting 'Country Code' == 1.
2. Than we have to replace some cities name with Delhi-NCR as mention in the question.
3. Than we have to separate all cuisines name in the list for Delhi-NCR and Rest_of_India.
4. Than we have to compare cuisines of Delhi-NCR and Rest_of_India.

1.3

CODE

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

```
from collections import Counter

import matplotlib.pyplot as plt

import numpy as np

import csv

import pandas as pd

import operator
```

```
df_start=pd.read_csv(r'C:\Users\nikhi\OneDrive\Desktop\zomato.csv',engine='python')

df_start['City'].dropna(inplace=True)
```

```

df_start = df_start[df_start['Country Code'] == 1]
df_start['City'].replace("New Delhi", "Delhi-NCR", inplace=True)
df_start['City'].replace("Ghaziabad", "Delhi-NCR", inplace=True)
df_start['City'].replace("Noida", "Delhi-NCR", inplace=True)
df_start['City'].replace("Gurgaon", "Delhi-NCR", inplace=True)
df_start['City'].replace("Faridabad", "Delhi-NCR", inplace=True)
df_start2 = df_start[df_start['City'] == 'Delhi-NCR']

```

```

def sep_Cuisines(Cuisines):
    return [i.strip() for i in Cuisines.split(",")]

```

```

delhi = { }
rest_of_india = { }
df_start2["Cuisines"] = df_start2["Cuisines"].apply(sep_Cuisines)
for i in df_start2["Cuisines"]:
    for j in i:
        if(j in delhi):
            delhi[j] += 1
        else:
            delhi[j] = 1
df_start3 = df_start[df_start['City'] != 'Delhi-NCR']
df_start3["Cuisines"] = df_start3["Cuisines"].apply(sep_Cuisines)
for i in df_start3["Cuisines"]:
    for j in i:
        if(j in rest_of_india):
            rest_of_india[j] += 1
        else:
            rest_of_india[j] = 1

```

```

sorted_delhi = dict( sorted(delhi.items(), key=operator.itemgetter(1), reverse=True))
sorted_rest_of_india = dict( sorted(rest_of_india.items(), key=operator.itemgetter(1), reverse=True))

```

```
#print(sorted_delhi)
#print(sorted_rest_of_india)

#print(rest_of_india)
delhi1 = []
x1 = []
rest_of_india1 = []
x2 = []
count = 0
print('Delhi-NCR\n')
for key in sorted_delhi:
    if(count == 10):
        break
    count += 1
    print(key)
    delhi1.append(key)
    x1.append(sorted_delhi[key])

count = 0
print("\n\nRest of India\n")
for key in sorted_rest_of_india:
    if(count == 10):
        break
    count += 1
    print(key)
    rest_of_india1.append(key)
    x2.append(sorted_rest_of_india[key])
plt.bar(rest_of_india1,x2,width = 0.8,color = "Red")
plt.xticks(rotation = 90)
plt.xlabel('Cuisine name')
plt.ylabel('Count')
plt.title('Rest_of_India')
```

```
plt.show()
```

```
plt.bar(delhi1,x1,width = 0.8,color = "Blue")
```

```
plt.xticks(rotation = 90)
```

```
plt.xlabel('Cuisine name')
```

```
plt.ylabel('Count')
```

```
plt.title('Delhi')
```

```
plt.show()
```

```
#print(delhi1)
```

```
#print(rest_of_india1)
```

ANSWER

Delhi-NCR

North Indian

Chinese

Fast Food

Mughlai

Bakery

South Indian

Continental

Desserts

Street Food

Italian

Rest of India

North Indian

Chinese

Continental

Italian

Cafe

Fast Food

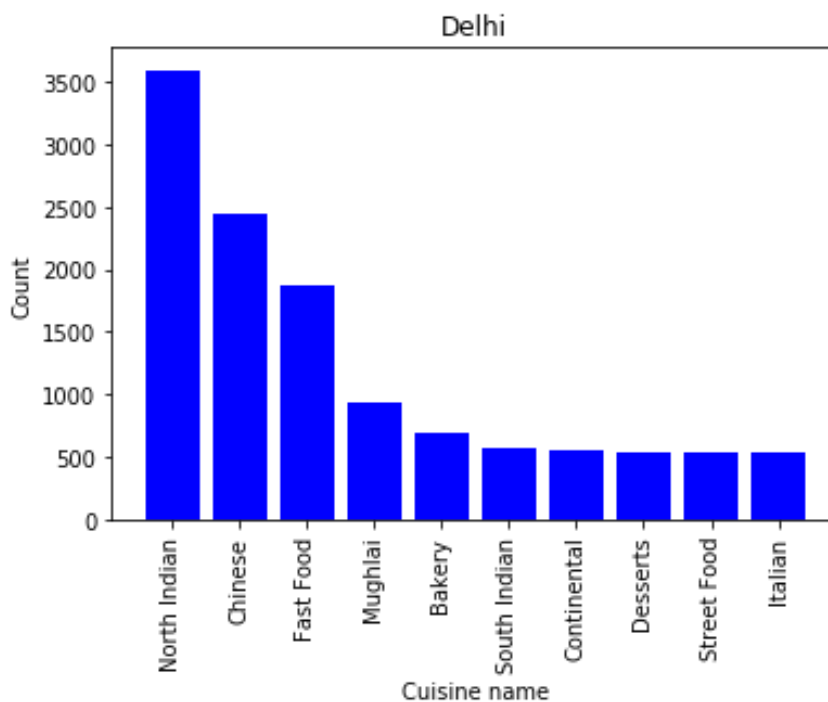
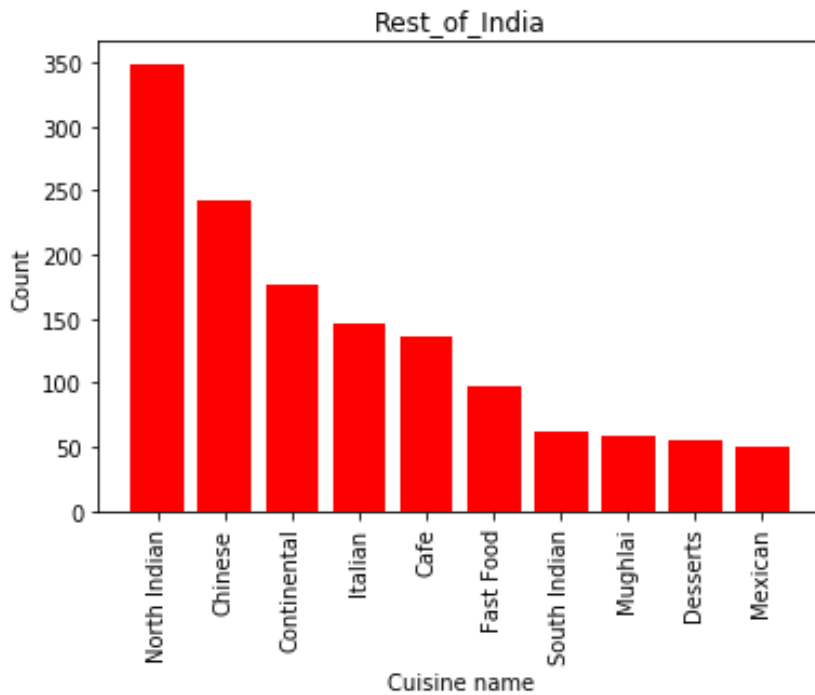
South Indian

Mughlai

Desserts

Mexican

PLOT



JUSTIFICATION

1. First we have to read zomato.csv file in CSV format and we have to select India by selecting 'Country Code' == 1.
2. Than we have to replace some cities name with Delhi-NCR as mention in the question.
3. Than we have to separate all cuisines name in the list for Delhi-NCR and Rest_of_India.
4. And we have to count each cuisine separately for both and we have to fetch the top 10 cuisine from both Delhi-NCR and Rest_of_India.
5. Than we have to plot Bar Graph for Dehi-NCR and Rest_of_india.

1.4

##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.

```
from collections import Counter
```

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
import csv
```

```
import pandas as pd
```

```
import operator
```

```
df_start=pd.read_csv(r'C:\Users\nikhi\OneDrive\Desktop\zomato.csv',engine='python')
```

```
df_start['City'].dropna(inplace=True)
```

```
df_start = df_start[df_start['Country Code'] == 1]
```

```
df_start['City'].replace("New Delhi", "Delhi-NCR", inplace=True)
```

```
df_start['City'].replace("Ghaziabad", "Delhi-NCR", inplace=True)
```

```
df_start['City'].replace("Noida", "Delhi-NCR", inplace=True)
```

```
df_start['City'].replace("Gurgaon", "Delhi-NCR", inplace=True)
```

```
df_start['City'].replace("Faridabad", "Delhi-NCR", inplace=True)
```

```
df_start2 = df_start[df_start['City'] == 'Delhi-NCR']
```

```
def sep_Cuisines(Cuisines):
```

```
    return [i.strip() for i in Cuisines.split(",")]
```

```
delhi = { }
```

```
rest_of_india = { }
```

```
df_start2["Cuisines"] = df_start2["Cuisines"].apply(sep_Cuisines)
```

```
for i in df_start2["Cuisines"]:
```

```
    for j in i:
```

```
        if(j in delhi):
```

```
            delhi[j] += 1
```

```
        else:
```

```

        delhi[j] = 1
df_start3 = df_start[df_start['City'] != 'Delhi-NCR']
df_start3["Cuisines"] = df_start3["Cuisines"].apply(sep_Cuisines)
for i in df_start3["Cuisines"]:
    for j in i:
        if(j in rest_of_india):
            rest_of_india[j] += 1
        else:
            rest_of_india[j] = 1

```

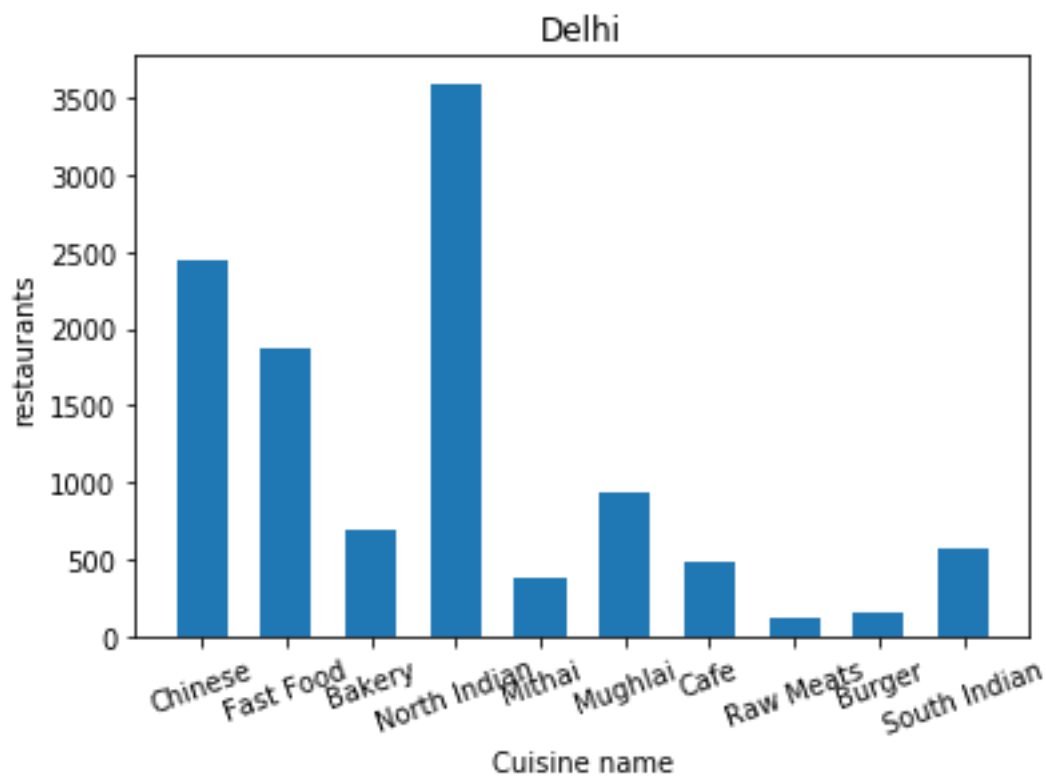
```

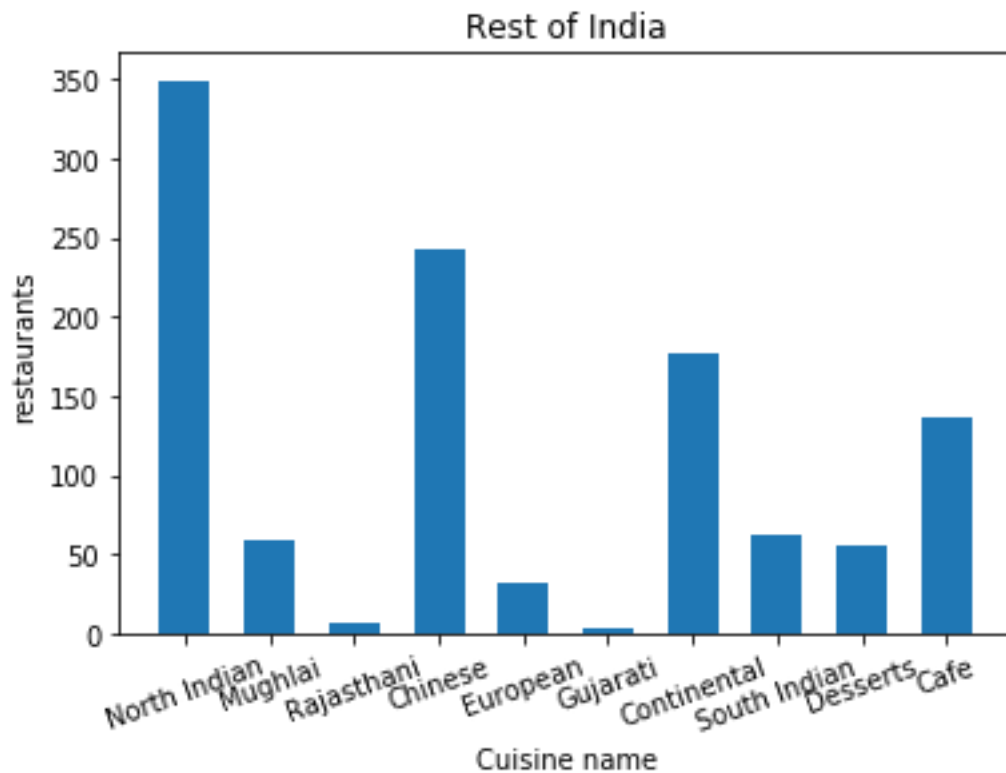
x1 = []
x2 = []
y1 = []
y2 = []
count = 0
for i in delhi:
    if(count == 10):
        break
    count += 1
    x1.append(i)
    y1.append(delhi[i])
count = 0
for i in rest_of_india:
    if(count == 10):
        break
    count += 1
    x2.append(i)
    y2.append(rest_of_india[i])
plt.bar(x1,y1,width = 0.6)
plt.xticks(rotation = 20)

```

```
plt.xlabel('Cuisine name')
plt.ylabel('restaurants')
plt.title('Delhi')
plt.show()
plt.bar(x2,y2,width = 0.6)
plt.xticks(rotation = 20)
plt.xlabel('Cuisine name')
plt.ylabel('restaurants')
plt.title('Rest of India')
plt.show()
```

PLOT





JUSTIFICATION

1. First we have to read zomato.csv file in CSV format and we have to select India by selecting 'Country Code' == 1.
2. Than we have to replace some cities name with Delhi-NCR as mention in the question.
3. Than we have to separate all cuisines name in the list for Delhi-NCR and Rest_of_India.
4. And we have to count each cuisine separately for both Delhi-NCR and Rest_of_India.
5. Than we have to plot Bar Graph for Dehi-NCR and Rest_of_india.

Short detailed analysis

As we can see in both the Bar Graph of Delhi-NCR and Rest_of_India, Many cuisines serve.

By comparing both the graph we can say that in Delhi-NCR the people likes North Indian, Fast Food, Chinese, etc. From this we can say that people in Delhi-NCR basically likes North Indian and other foreign type dishes like Chinese, Italian, etc.

And Rest_Of_India we can see that people likes North Indian food and other continental food same as Delhi-NCR so we can predict that people in India like North Indian and other foreign dishes.

2.1.1

CODES

##2.1.1 Number of Votes given Restaurant

```
from collections import Counter
```

```
import matplotlib.pyplot as plt
```

```

import numpy as np

import csv

import pandas as pd

import operator


df_start=pd.read_csv(r'C:\Users\nikhi\OneDrive\Desktop\zomato.csv',engine='python')
df_start['City'].dropna(inplace=True)
df_start = df_start[df_start['Country Code'] == 1]
df_start['City'].replace("New Delhi","Delhi-NCR",inplace=True)
df_start['City'].replace("Ghaziabad","Delhi-NCR",inplace=True)
df_start['City'].replace("Noida","Delhi-NCR",inplace=True)
df_start['City'].replace("Gurgaon","Delhi-NCR",inplace=True)
df_start['City'].replace("Faridabad","Delhi-NCR",inplace=True)
df_start2 = df_start[df_start['City'] == 'Delhi-NCR']


number=df_start.groupby(['Restaurant Name','Rating text'])['Votes'].max()[0:10]


Name=number.index
Rate=number.values
#print(Name,Rate)

a = []
b = []
c = []

for i in Name:
    a.append(i[0])
    b.append(i[1])

for i in Rate:
    c.append(i)

#print(a,b,c)

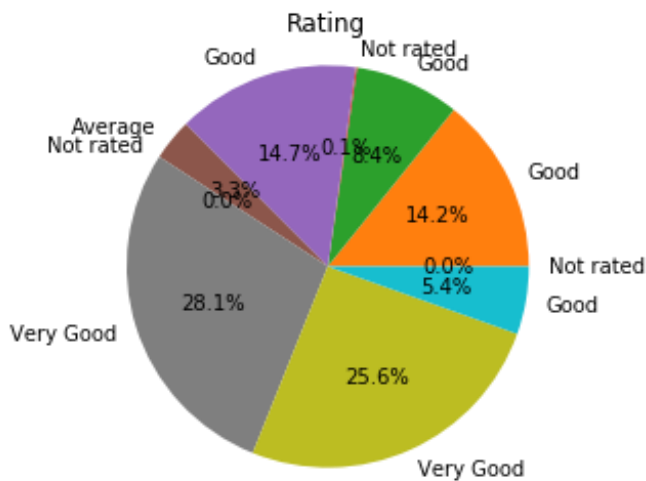
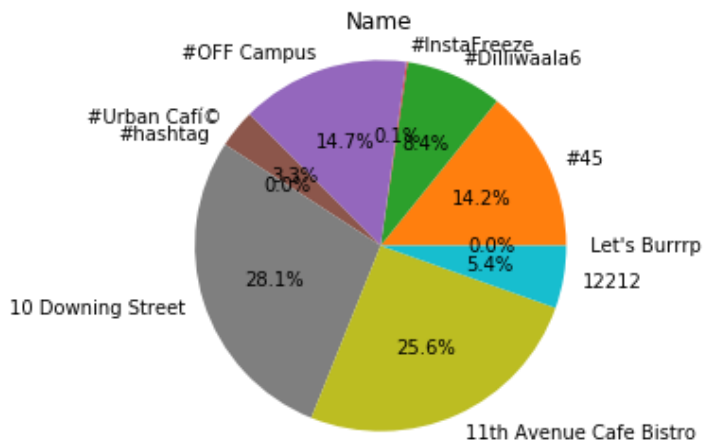
plt.pie(c,labels = a,autopct = '% 1.1f%%')

plt.axis("equal")

```

```
plt.title('Name')
plt.show()
plt.pie(c, labels = b, autopct = '%1.1f%%')
plt.axis("equal")
plt.title('Rating')
plt.show()
```

PLOT



JUSTIFICATION

1. First we have to read zomato.csv file in CSV format and we have to select India by selecting 'Country Code' == 1.
2. Than we have to replace some cities name with Delhi-NCR as mention in the question.
3. Than we use groupby to 'Restaurant Name' and 'Rating text' with 'Votes' and make separate them in list
4. Than we have to plot Pie Chart for Name and Rating.

Short detailed analysis

So by examine the pie chart we can say that those restaurants whose votes are less we clearly see that for that restaurants the rating is also less.

So we can say that more the votes more will be the rating.

2.1.2

CODE

```
## 2.1.2 Restaurant serving more number of cuisines.

from collections import Counter

import matplotlib.pyplot as plt

import numpy as np

import csv

import pandas as pd

import operator

df_start=pd.read_csv(r'C:\Users\nikhi\OneDrive\Desktop\zomato.csv',engine='python')

df_start['City'].dropna(inplace=True)

df_start = df_start[df_start['Country Code'] == 1]

df_start['City'].replace("New Delhi","Delhi-NCR",inplace=True)

df_start['City'].replace("Ghaziabad","Delhi-NCR",inplace=True)

df_start['City'].replace("Noida","Delhi-NCR",inplace=True)

df_start['City'].replace("Gurgaon","Delhi-NCR",inplace=True)

df_start['City'].replace("Faridabad","Delhi-NCR",inplace=True)

#df_start2 = df_start[df_start['City'] == 'Delhi-NCR']

def sep_Cuisines(Cuisines):

    a = [i.strip() for i in Cuisines.split(",")]

    return len(a)

df_start["Cuisines"]= df_start["Cuisines"].apply(sep_Cuisines)

number=df_start.groupby(['Restaurant Name','Rating

text'])['Cuisines'].max().sort_values(ascending=False)[0:10]

Name=number.index

Rate=number.values

#print(Name,Rate)

a = []

b = []
```



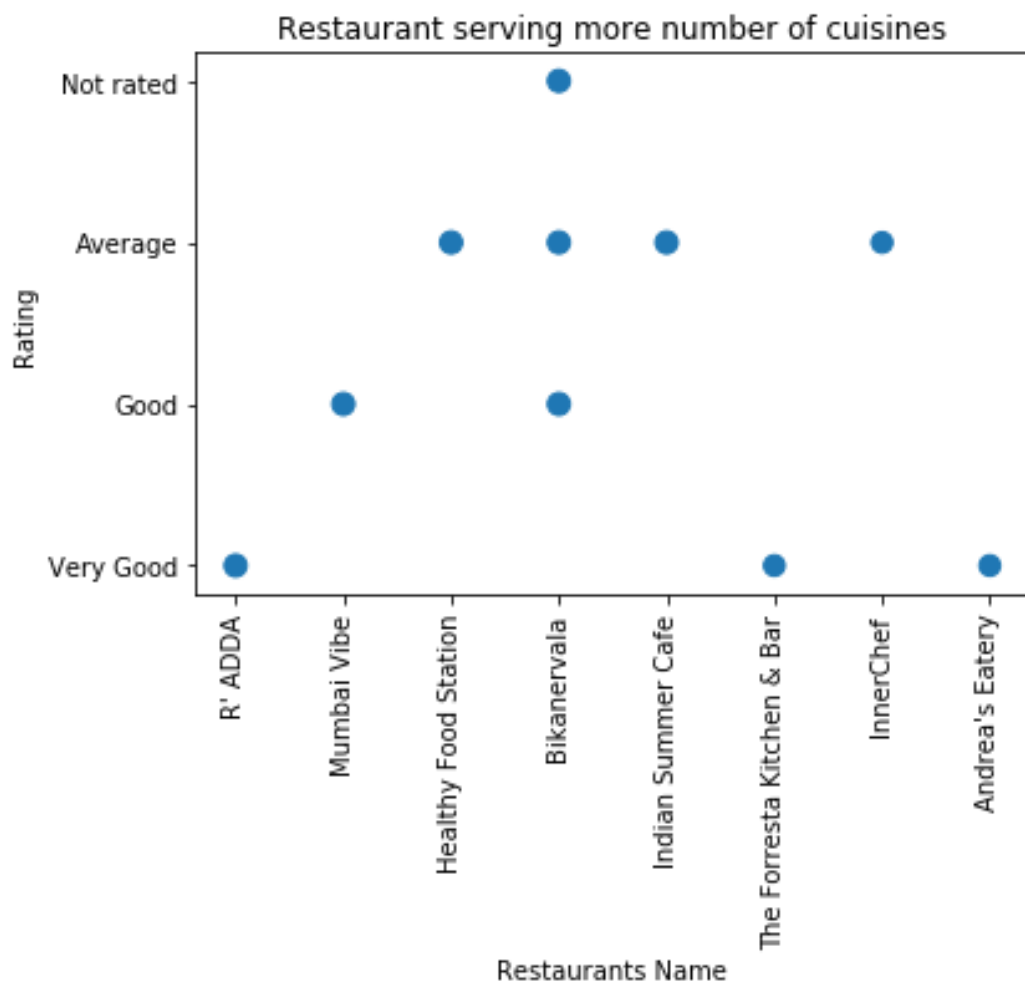
```

c = []
for i in Name:
    a.append(i[0])
    b.append(i[1])
for i in Rate:
    c.append(i*10)
plt.scatter(a,b,s=c)
plt.xticks(rotation = 90)
plt.xlabel('Restaurants Name')
plt.ylabel('Rating')
plt.title('Restaurant serving more number of cuisines')

plt.show()

```

PLOT



JUSTIFICATION

1. First we have to read zomato.csv file in CSV format and we have to select India by selecting 'Country Code' == 1.
2. Than we have to replace some cities name with Delhi-NCR as mention in the question.
3. Than we use groupby to 'Restaurant Name' and 'Rating text' with 'Cuisines' and make separate them in list
4. Than we have to plot Scatter Graph for Restaurant serving more number of cuisines.

Short detailed analysis

So by examine the Scatter Graph we can say that those restaurants having more no. of cuisines or less no. of cuisines there is no effect occur on their ratings.

No. of cuisines will not affect the rating of restaurants.

So we can say that Restaurant Rating will not affect by number of cuisines.

2.1.3

CODE

2.1.3 Average Cost of Restaurant

```
from collections import Counter
```

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
import csv
```

```
import pandas as pd
```

```
import operator
```

```
df_start=pd.read_csv(r'C:\Users\nikhi\OneDrive\Desktop\zomato.csv',engine='python')
```

```
df_start['City'].dropna(inplace=True)
```

```
df_start = df_start[df_start['Country Code'] == 1]
```

```
df_start['City'].replace("New Delhi", "Delhi-NCR", inplace=True)
```

```
df_start['City'].replace("Ghaziabad", "Delhi-NCR", inplace=True)
```

```
df_start['City'].replace("Noida", "Delhi-NCR", inplace=True)
```

```
df_start['City'].replace("Gurgaon", "Delhi-NCR", inplace=True)
```

```
df_start['City'].replace("Faridabad", "Delhi-NCR", inplace=True)
```

```
number=df_start.groupby(['Restaurant Name','Rating text'])['Average Cost for two'].max()[0:10]
```

```
Name=number.index
```

```
Rate=number.values
```

```
#print(Name,Rate)
```

```
a = []
```

```
b = []
```

```
c = []
```

```
for i in Name:
```

```
    a.append(i[0])
```

```
    b.append(i[1])
```

```
for i in Rate:
```

```
    c.append(i/2)
```

```
plt.scatter(a,b,s=c)
```

```
plt.xticks(rotation = 90)
```

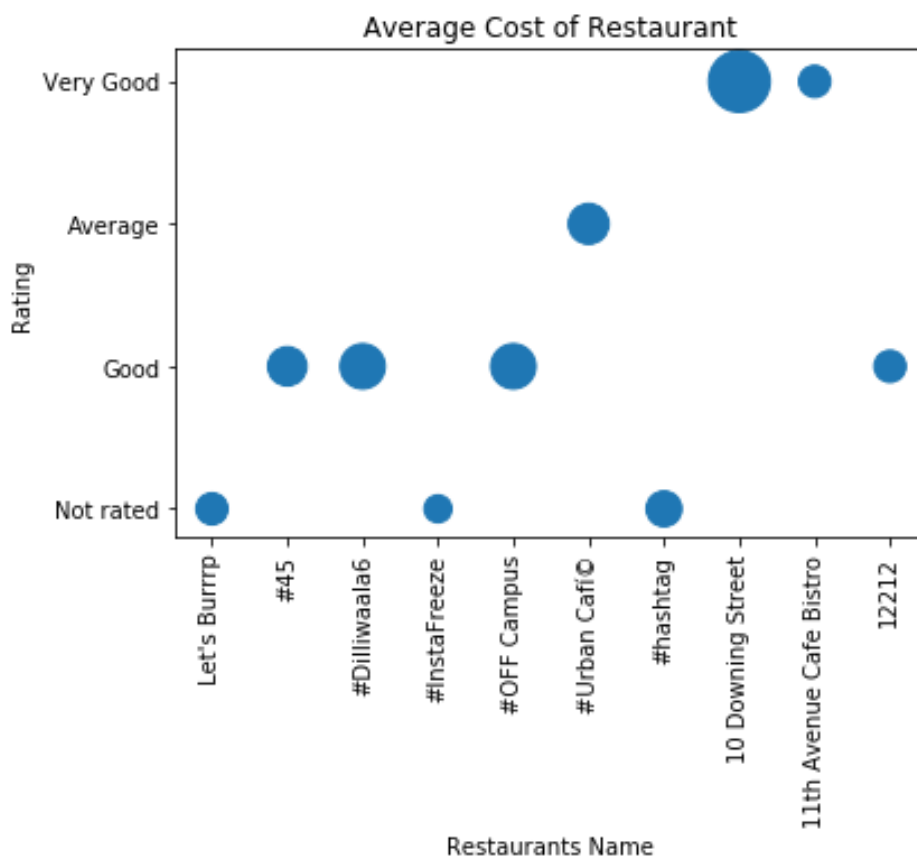
```
plt.xlabel('Restaurants Name')
```

```
plt.ylabel('Rating')
```

```
plt.title('Average Cost of Restaurant')
```

```
plt.show()
```

PLOT



JUSTIFICATION

1. First we have to read zomato.csv file in CSV format and we have to select India by selecting 'Country Code' == 1.
2. Than we have to replace some cities name with Delhi-NCR as mention in the question.
3. Than we use groupby to 'Restaurant Name' and 'Rating text' with 'Average Cost for two' and make separate them in list
4. Than we have to plot Scatter Graph for Average Cost of Restaurant.

Short detailed analysis

So by examine the Scatter Graph we can say that those restaurants Average Cost for two is more or less will not affect the rating.

But those having Average Cost for two is more those restaurants counts are more so in some cases it will effect or in some cases it will not.

As we can see that there are some restaurants having more average cost of two but they are at low ratings.

2.1.4

##2.1.4 Restaurant serving some specific cuisines.

```
from collections import Counter
```

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
import csv
```

```
import pandas as pd
```

```
import operator
```

```
df_start=pd.read_csv(r'C:\Users\nikhi\OneDrive\Desktop\zomato.csv',engine='python')
```

```
df_start['City'].dropna(inplace=True)
```

```
df_start = df_start[df_start['Country Code'] == 1]
```

```
df_start['City'].replace("New Delhi","Delhi-NCR",inplace=True)
```

```
df_start['City'].replace("Ghaziabad","Delhi-NCR",inplace=True)
```

```
df_start['City'].replace("Noida","Delhi-NCR",inplace=True)
```

```
df_start['City'].replace("Gurgaon","Delhi-NCR",inplace=True)
```

```
df_start['City'].replace("Faridabad","Delhi-NCR",inplace=True)
```

```
#df_start2 = df_start[df_start['City'] == 'Delhi-NCR']
```

```
df_start['Cuisines'].dropna(inplace=True)
```

```
def sep_Cuisines(Cuisines):
```

```
    a = [i.strip() for i in Cuisines.split(",")]
```

```
    return len(a)
```

```
df_start["Cuisines"] = df_start["Cuisines"].apply(sep_Cuisines)
```

```
#print(df_start["Cuisines"])
```

```
df_start = df_start[(df_start['Cuisines'] == 1)|(df_start['Cuisines'] == 2)]
```

```
number=df_start.groupby(['Restaurant Name','Rating  
text'])['Cuisines'].max().sort_values(ascending=False)[0:10]
```

```
Name=number.index
```

```
Rate=number.values
```

```
#print(Name,Rate)
```

```
a = []
```

```
b = []
```

```
c = []
```

```
for i in Name:
```

```
    a.append(i[0])
```

```
    b.append(i[1])
```

```
for i in Rate:
```

```
    c.append(i*10)
```

```
plt.scatter(a,b,s=c)
```

```
plt.xticks(rotation = 90)
```

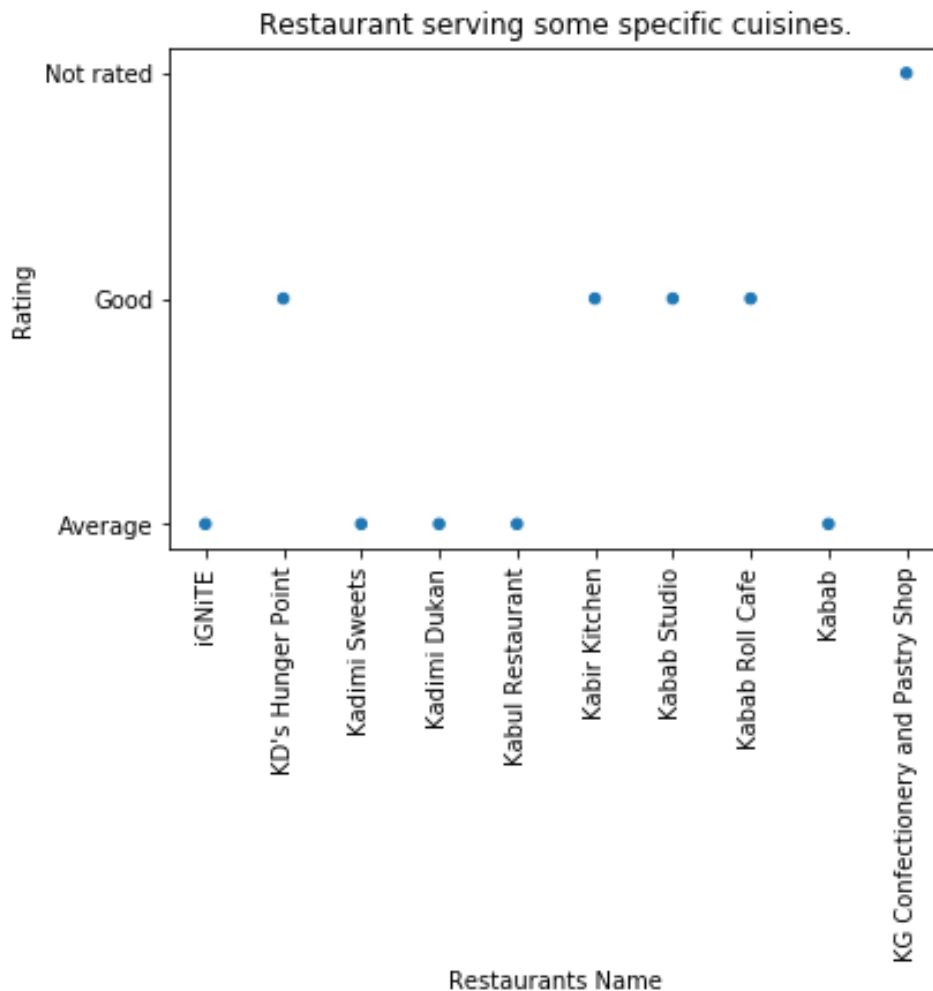
```
plt.xlabel('Restaurants Name')
```

```
plt.ylabel('Rating')
```

```
plt.title('Restaurant serving some specific cuisines.')
```

```
plt.show()
```

PLOT



JUSTIFICATION

1. First we have to read zomato.csv file in CSV format and we have to select India by selecting 'Country Code' == 1.
2. Than we have to replace some cities name with Delhi-NCR as mention in the question.
3. Than we use groupby to 'Restaurant Name' and 'Rating text' with 'Cuisines' and make separate them in list
4. Than we have to plot Scatter Graph for Restaurant serving some specific cuisines.

Short detailed analysis

So by examine the Scatter Graph we can say that those restaurants serving some specific cuisines having less rating.

As those having less cuisine type those restaurants ratings are less.

So we can say that more the cuisines more will be the rating.

2.2.1

CODE

2.2.1 Weighted Restaurant Rating= Σ (number of votes * rating) / Σ (number of votes) .

```
from collections import Counter
```

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
import csv
```

```
import pandas as pd
```

```
import operator
```

```
df_start=pd.read_csv(r'C:\Users\nikhi\OneDrive\Desktop\zomato.csv',engine='python')
```

```
df_start['City'].dropna(inplace=True)
```

```
df_start = df_start[df_start['Country Code'] == 1]
```

```
df_start['City'].replace("New Delhi","Delhi-NCR",inplace=True)
```

```
df_start['City'].replace("Ghaziabad","Delhi-NCR",inplace=True)
```

```
df_start['City'].replace("Noida","Delhi-NCR",inplace=True)
```

```
df_start['City'].replace("Gurgaon","Delhi-NCR",inplace=True)
```

```
df_start['City'].replace("Faridabad","Delhi-NCR",inplace=True)
```

```
Locality_dict = { }
```

```
Localities = np.array(df_start['Locality'])
```

```
ratings = np.array(df_start['Aggregate rating'])
```

```
votes = np.array(df_start['Votes'])
```

```
for i in range(len(Localities)):
```

```
    if votes[i]!=0:
```

```
        if Localities[i] not in Locality_dict:
```

```
            Locality_dict[Localities[i]] = [votes[i]*ratings[i],votes[i]]
```

```
        else:
```

```
            Locality_dict[Localities[i]][0] += votes[i]*ratings[i]
```

```
            Locality_dict[Localities[i]][1] += votes[i]
```

```
ratings = []
```

```
for key,value in Locality_dict.items():
```

```
    Locality = key
```

```

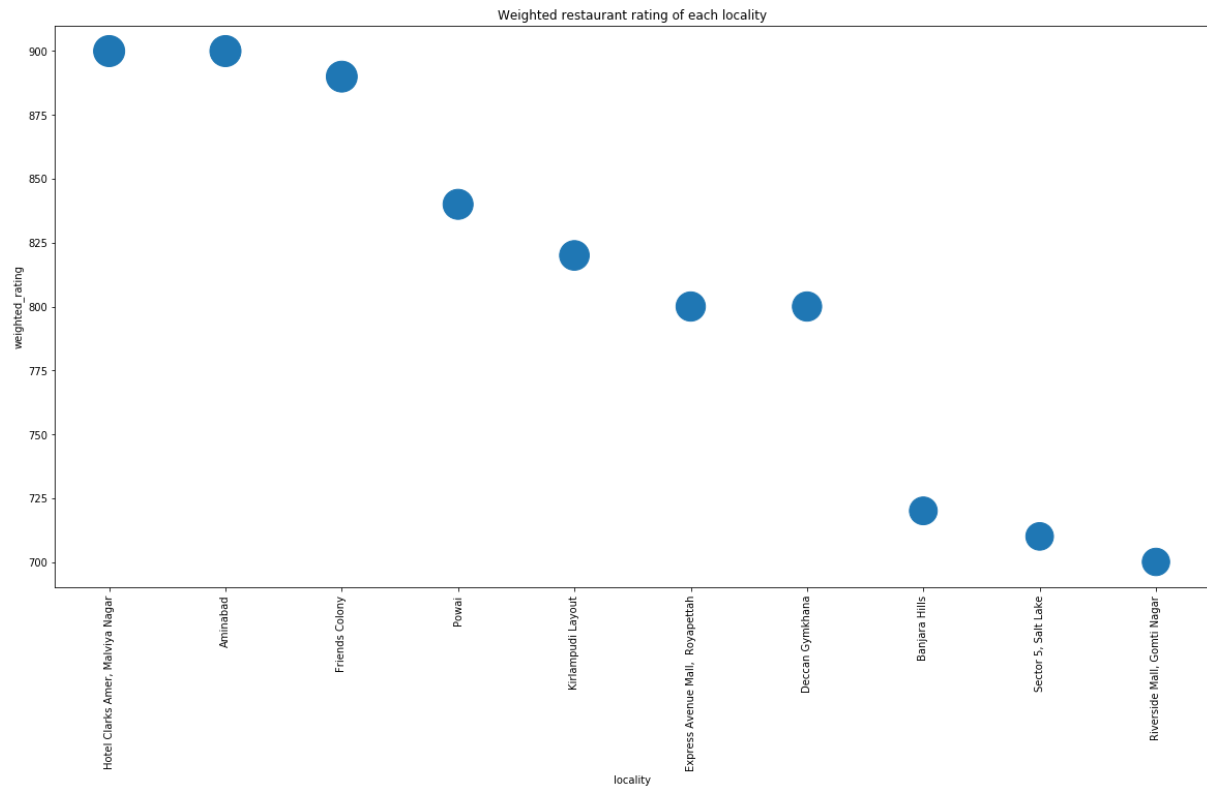
weighted_rating = round(value[0]/value[1],2)
ratings.append([weighted_rating,key])
ratings.sort(reverse=True)
Localities1 = [ i[1] for i in ratings][0:10]
ratings1 = [ (i[0]-4)*1000 for i in ratings][0:10]
for i in range(10):
    print(Localities1[i], "%.2f" %ratings1[i])
sizes = np.arange(10,0,1)
plt.rcParams['figure.figsize'] = (20,10)
plt.scatter(Localities1,ratings1,s=ratings1)
plt.xticks(rotation = 90)
plt.xlabel('locality')
plt.ylabel('weighted_rating')
plt.title('Weighted restaurant rating of each locality')
plt.show()

```

ANSWER

Hotel Clarks Amer, Malviya Nagar 900.00
 Aminabad 900.00
 Friends Colony 890.00
 Powai 840.00
 Kirlampudi Layout 820.00
 Express Avenue Mall, Royapettah 800.00
 Deccan Gymkhana 800.00
 Banjara Hills 720.00
 Sector 5, Salt Lake 710.00
 Riverside Mall, Gomti Nagar 700.00

PLOT



JUSTIFICATION

1. First we have to read zomato.csv file in CSV format and we have to select India by selecting 'Country Code' == 1.
2. Than we have to replace some cities name with Delhi-NCR as mention in the question.
3. Than we have to calculate Weighted Restaurant Rating= $\frac{\sum (\text{number of votes} * \text{rating})}{\sum (\text{number of votes})}$ by using loop.
4. Than we plot Scatter Graph for Weighted restaurant rating of each locality.

3.1

CODE

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

```
from collections import Counter
```

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
import csv
```

```
import pandas as pd
```

```
import operator
```

```
df_start=pd.read_csv(r'C:\Users\nikhi\OneDrive\Desktop\zomato.csv',engine='python')
```

```
df_start['City'].dropna(inplace=True)
```

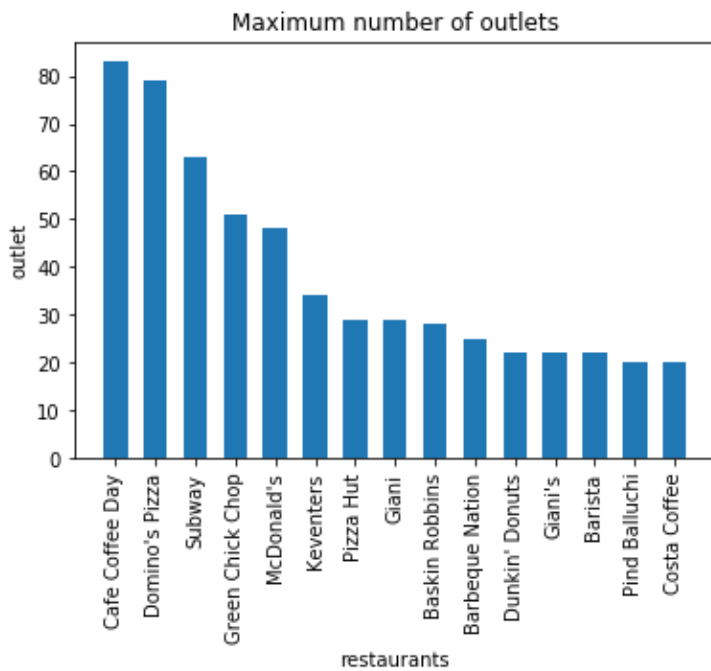
```
df_start = df_start[df_start['Country Code'] == 1]
```

```

df_start['City'].replace("New Delhi","Delhi-NCR",inplace=True)
df_start['City'].replace("Ghaziabad","Delhi-NCR",inplace=True)
df_start['City'].replace("Noida","Delhi-NCR",inplace=True)
df_start['City'].replace("Gurgaon","Delhi-NCR",inplace=True)
df_start['City'].replace("Faridabad","Delhi-NCR",inplace=True)
#df_start2 = df_start[df_start['City'] == 'Delhi-NCR']
a = { }
for ind in df_start.index:
    if(df_start['Restaurant Name'][ind] in a):
        a[df_start['Restaurant Name'][ind]] += 1
    else:
        a[df_start['Restaurant Name'][ind]] = 1
sorted_a = dict( sorted(a.items(), key=operator.itemgetter(1),reverse=True))
k = []
l = []
count = 0
for i in sorted_a:
    if(count == 15):
        break
    count += 1
    k.append(i)
    l.append(sorted_a[i])
plt.bar(k,l,width = 0.6)
plt.xticks(rotation = 90)
plt.xlabel('restaurants')
plt.ylabel('outlet')
plt.title('Maximum number of outlets')
plt.show()

```

PLOT



JUSTIFICATION

1. First we have to read zomato.csv file in CSV format and we have to select India by selecting 'Country Code' == 1.
2. Than we have to replace some cities name with Delhi-NCR as mention in the question.
3. Than we calculate the outlets of restaurants by using dict and sort accordingly
4. Than we plot Bar Graph for Weighted Maximum number of outlets.

3.1.2

CODE

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

```
from collections import Counter
```

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
import csv
```

```
import pandas as pd
```

```
import operator
```

```
df_start=pd.read_csv(r'C:\Users\nikhi\OneDrive\Desktop\zomato.csv',engine='python')
```

```
df_start['City'].dropna(inplace=True)
```

```
df_start = df_start[df_start['Country Code'] == 1]
```

```

df_start['City'].replace("New Delhi","Delhi-NCR",inplace=True)
df_start['City'].replace("Ghaziabad","Delhi-NCR",inplace=True)
df_start['City'].replace("Noida","Delhi-NCR",inplace=True)
df_start['City'].replace("Gurgaon","Delhi-NCR",inplace=True)
df_start['City'].replace("Faridabad","Delhi-NCR",inplace=True)
#df_start2 = df_start[df_start['City'] == 'Delhi-NCR']

df_start["Rating text"].replace("Not rated", np.nan, inplace = True)

df_start.dropna(inplace=True, subset=["Rating text"])

number=df_start.groupby(['Restaurant Name'])['Aggregate
rating'].max().sort_values(ascending=False)

Name=number.index

Rate=number.values

k = []
l = []

count = 0

for i in Rate:
    l.append(i)

for i in Name:
    k.append(i)

plt.hist(l,alpha = 0.5,bins = 5,edgecolor = 'black',color = "Red")

plt.xticks(rotation = 90)

plt.xlabel('Aggregate rating')

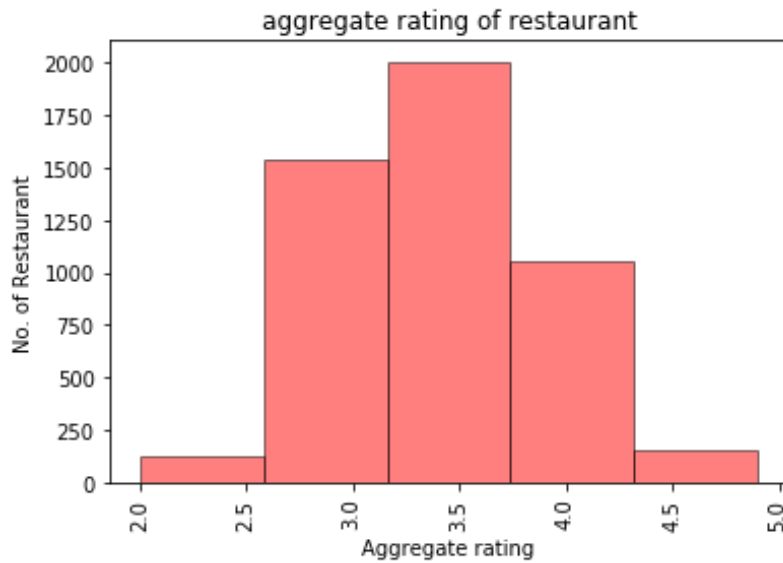
plt.ylabel('No. of Restaurant')

plt.title('aggregate rating of restaurant')

plt.show()

```

PLOT



JUSTIFICATION

1. First we have to read zomato.csv file in CSV format and we have to select India by selecting 'Country Code' == 1.
2. Than we have to replace some cities name with Delhi-NCR as mention in the question.
3. Than we groupby Restaurant Name with Aggregate Rating and separate them in list
4. Than we plot Histogram Graph for Weighted aggregate rating of restaurant.

3.1.3

CODE

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

```
from collections import Counter
```

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
import csv
```

```
import pandas as pd
```

```
import operator
```

```
df_start=pd.read_csv(r'C:\Users\nikhi\OneDrive\Desktop\zomato.csv',engine='python')
```

```
df_start['City'].dropna(inplace=True)
```

```
df_start = df_start[df_start['Country Code'] == 1]
```

```
df_start['City'].replace("New Delhi", "Delhi-NCR", inplace=True)
```

```
df_start['City'].replace("Ghaziabad", "Delhi-NCR", inplace=True)
```

```
df_start['City'].replace("Noida", "Delhi-NCR", inplace=True)
```

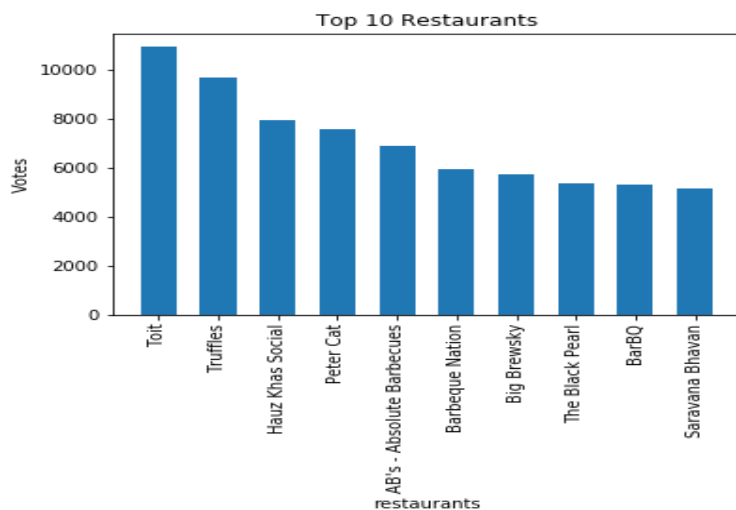
```
df_start['City'].replace("Gurgaon", "Delhi-NCR", inplace=True)
```

```

df_start['City'].replace("Faridabad", "Delhi-NCR", inplace=True)
#df_start2 = df_start[df_start['City'] == 'Delhi-NCR']
df_start["Rating text"].replace("Not rated", np.nan, inplace = True)
df_start.dropna(inplace=True, subset=["Rating text"])
number=df_start.groupby(['Restaurant Name'])['Votes'].max().sort_values(ascending=False)[0:10]
Name=number.index
Rate=number.values
#print(Name)
#print(Rate)
k = []
l = []
count = 0
for i in Rate:
    l.append(i)
for i in Name:
    k.append(i)
plt.bar(k,l,width = 0.6)
plt.xticks(rotation = 90)
plt.xlabel('restaurants')
plt.ylabel('Votes')
plt.title('Top 10 Restaurants')
plt.show()

```

PLOT



JUSTIFICATION

1. First we have to read zomato.csv file in CSV format and we have to select India by selecting 'Country Code' == 1.
2. Than we have to replace some cities name with Delhi-NCR as mention in the question.
3. Than we groupby Restaurant Name with Votes and separate them in list
4. Than we plot Bar Graph for Top 10 Restaurants.

3.1.4

CODE

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

```
from collections import Counter
```

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
import csv
```

```
import pandas as pd
```

```
import operator
```

```
df_start=pd.read_csv(r'C:\Users\nikhi\OneDrive\Desktop\zomato.csv',engine='python')
```

```
df_start['City'].dropna(inplace=True)
```

```
df_start = df_start[df_start['Country Code'] == 216]
```

```
df_start['Cuisines'].dropna(inplace=True)
```

```
d = { }
```

```
def sep_Cuisines(Cuisines):
```

```
    a = [i.strip() for i in Cuisines.split(",")]
```

```
    for i in a:
```

```
        if(i in d):
```

```
            d[i] += 1
```

```
        else:
```

```
            d[i] = 1
```

```
df_start["Cuisines"]= df_start["Cuisines"].apply(sep_Cuisines)
```

```
sorted_d = dict( sorted(d.items(), key=operator.itemgetter(1),reverse=True))
```

```
a = []
```

```
b = []
```

```

count = 0
for i in sorted_d:
    if(count == 10):
        break
    count += 1
    a.append(i)
    b.append(sorted_d[i])
#print(a,b)
plt.pie(b,labels = a,autopct = "%.1f")
plt.axis("equal")
plt.title('Top 10 cuisines present in restaurants in the USA')
plt.show()

```

PLOT



JUSTIFICATION

1. First we have to read zomato.csv file in CSV format and we have to select USA by selecting 'Country Code' == 216.
2. Than we drop all NA of Cuisines and split them.
3. Than we put all top cuisines in dict and separate them in list.
4. Than we Pie Chart for Top 10 cuisines present in restaurants in the USA.

3.1.5

CODE

##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.

```
from collections import Counter
```

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
import csv
```

```
import pandas as pd
```

```
import operator
```

```
df_start=pd.read_csv(r'C:\Users\nikhi\OneDrive\Desktop\zomato.csv',engine='python')
```

```
df_start['City'].dropna(inplace=True)
```

```
df_start = df_start[df_start['Country Code'] == 1]
```

```
df_start['City'].replace("New Delhi", "Delhi-NCR",inplace=True)
```

```
df_start['City'].replace("Ghaziabad", "Delhi-NCR",inplace=True)
```

```
df_start['City'].replace("Noida", "Delhi-NCR",inplace=True)
```

```
df_start['City'].replace("Gurgaon", "Delhi-NCR",inplace=True)
```

```
df_start['City'].replace("Faridabad", "Delhi-NCR",inplace=True)
```

```
city_dict = { }
```

```
cities = np.array(df_start['City'])
```

```
ratings = np.array(df_start['Aggregate rating'])
```

```
votes = np.array(df_start['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_ratings = []
```

```
for key,value in city_dict.items():
```

```
    city = key
```

```
    weighted_rating = round(value[0]/value[1],2)
```

```
    city_ratings.append([weighted_rating,key,value[2]])
```

```
city_ratings.sort(reverse=True)
```

```
cities = [ i[1] for i in city_ratings]
```

```
ratings = [ (i[0] - 3)*300 for i in city_ratings]
```

```
counts = [ i[2] for i in city_ratings]
```

```
#print(counts)
```

```
#print(cities)
```

```
#print(ratings)
```

```
sizes = np.arange(10,0,1)
```

```
plt.rcParams['figure.figsize'] = (20,10)
```

```
plt.scatter(cities,counts,s=ratings)
```

```
plt.xticks(rotation = 90)
```

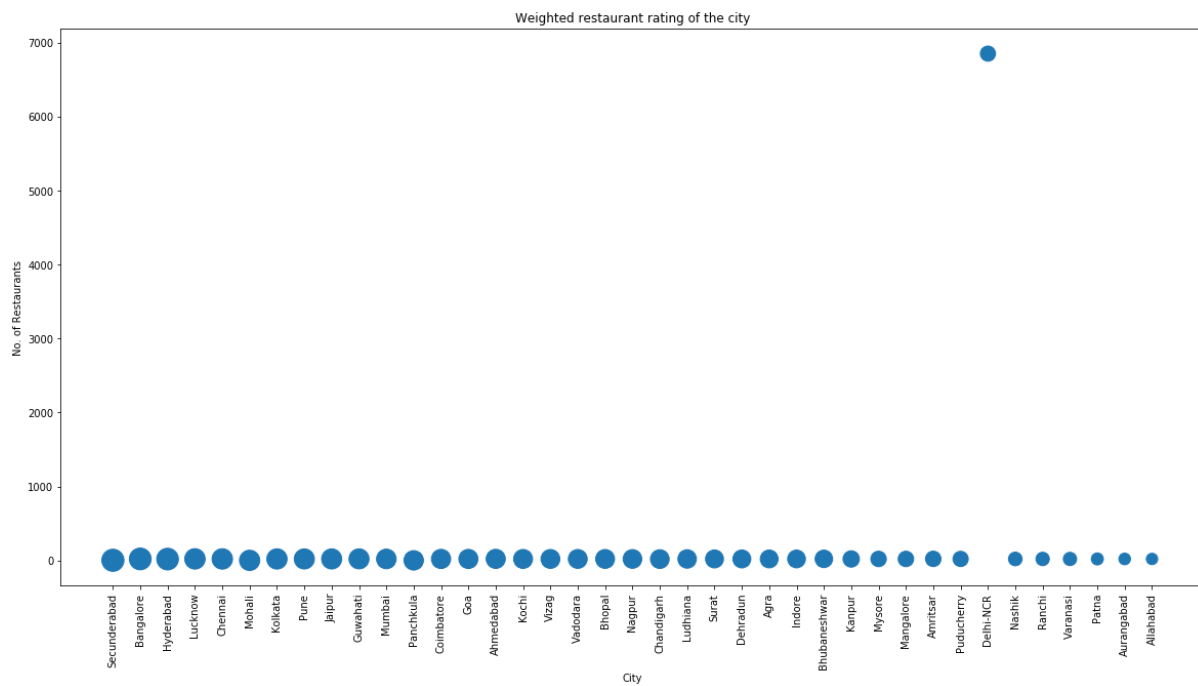
```
plt.xlabel('City')
```

```
plt.ylabel('No. of Restaurants')
```

```
plt.title('Weighted restaurant rating of the city')
```

```
plt.show()
```

PLOT



JUSTIFICATION

1. First we have to read zomato.csv file in CSV format and we have to select INDIA by selecting 'Country Code' == 1.
2. Than we have to replace some cities name with Delhi-NCR as mention in the question.
3. Than we calculate Weighted Rating of Restaurants.
4. Than we Scatter Graph for Weighted restaurant rating of the city.