

## Question-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.

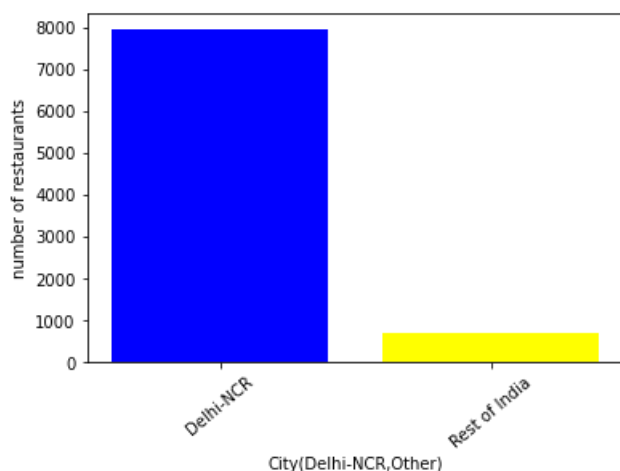
1. Plot the bar graph of number of restaurants present in Delhi NCR vs Rest of India.

In [20]:

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

d= pd.read_csv('zomato.csv',encoding="ISO-8859-1")# reading the file and encoding it to convert it
to readle format
d=d[d['Country Code']==1] # Only selecting Indian Cities
df=d.copy()
delhi=[] #creating an empty list to store Delhi-NCR cities data
other=[] #creating an empty list to store Rest of India cities
for i in df['City']:
    if 'New Delhi'in i:
        delhi.append(i)
    elif 'Ghaziabad' in i:
        delhi.append(i)
    elif 'Noida' in i:
        delhi.append(i)
    elif 'Gurgaon' in i:
        delhi.append(i)
    elif 'Faridabad' in i:
        delhi.append(i)
    else:
        other.append(i)

delhi=len(delhi) #getiing the count of list
other=len(other) #getiing the count of list
plt.bar(['Delhi-NCR','Rest of India'],[delhi,other],color=['blue','yellow']) #ploting a bar graph
based on count values
plt.xlabel('City(Delhi-NCR,Other)')
plt.ylabel('number of restaurants')
plt.xticks(rotation=40)
plt.show()
```



1. 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.

In [21]:

```
import pandas as pd
import numpy as np
```

```

import numpy as np
import matplotlib.pyplot as plt
import requests as r
d= pd.read_csv('zomato.csv',encoding="ISO-8859-1")# reading the file and encoding it to convert it
to readle format
d=d[d['Country Code']==1] # Only selecting Indian Cities
df=d.copy()
data_d=df[(df.City=='New Delhi')|(df.City=='Ghaziabad')|(df.City=='Noida')|(df.City=='Gurgaon')|(d
f.City=='Faridabad')] #getting data for only Delhi_ncr cities
data_o=df[(df.City!='New Delhi')&(df.City!='Ghaziabad')&(df.City!='Noida')&(df.City!='Gurgaon')&(d
f.City!='Faridabad')] #getting data for rest of the cities in india.
data_d=data_d.Cuisines #extracting Cuisines present in delhi_ncr
data_o=data_o.Cuisines #extracting Cuisines present in other cities
d=[] #creating an empty list to store Delhi-NCR cities data
o=[] #creating an empty list to store Rest of India cities
for i in data_d:
    for j in i.split(','): #here we have multiple cuisines in a single Cuisines columns so i have
used spilt() and strip() to split and iterate each Cuisines seprately
        d.append(j.strip())
for i in data_o:
    for j in i.split(','): #here we have multiple cuisines in a single Cuisines columns so i have
used spilt() and strip() to split and iterate each Cuisines seprately
        o.append(j.strip())
result=(set(o)-set(d)) #using set minus operator to get the cuisines which is not present in
delhi_ncr
for i in result:
    print(i)

```

Malwani  
 Cajun  
 BBQ  
 German

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

In [22]:

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import requests as r
from collections import Counter
d= pd.read_csv('zomato.csv',encoding="ISO-8859-1")# reading the file and encoding it to convert it
to readle format
d=d[d['Country Code']==1] # Only selecting Indian Cities
df=d.copy()
data_d=df[(df.City=='New Delhi')|(df.City=='Ghaziabad')|(df.City=='Noida')|(df.City=='Gurgaon')|(d
f.City=='Faridabad')] #getting data for only Delhi_ncr cities
data_o=df[(df.City!='New Delhi')&(df.City!='Ghaziabad')&(df.City!='Noida')&(df.City!='Gurgaon')&(d
f.City!='Faridabad')] #getting data for rest of the cities in india.
data_d=data_d.Cuisines #extracting Cuisines present in delhi_ncr
data_o=data_o.Cuisines #extracting Cuisines present in other cities
delhi=[]
other=[]
for i in data_d:
    for j in i.split(','): #here we have multiple cuisines in a single Cuisines columns so i have
used spilt() and strip() to split and iterate each Cuisines seprately
        delhi.append(j.strip())
for i in data_o:
    for j in i.split(','): #here we have multiple cuisines in a single Cuisines columns so i have
used spilt() and strip() to split and iterate each Cuisines seprately
        other.append(j.strip())
dic_delhi={} # declaring an empty dic to store cuisines as key and count(cuisins) as values
dic_other={} # declaring an empty dic to store cuisines as key and count(cuisins) as values
for i in delhi:
    dic_delhi[i]=dic_delhi.get(i,0)+1
for i in other:
    dic_other[i]=dic_other.get(i,0)+1
od_delhi=dict(Counter(dic_delhi).most_common(10)) # getting top 10 cuisines name
od_other=dict(Counter(dic_other).most_common(10)) # getting top 10 cuisines name
print('Delhi_NCR CUSINES')
print("")
for i in od_delhi:
    print(i)
print("")

```

```

print('OTHERS CITIES')
print("")
for i in od_other:
    print(i)

```

Delhi\_NCR CUSINES

North Indian  
 Chinese  
 Fast Food  
 Mughlai  
 Bakery  
 South Indian  
 Continental  
 Desserts  
 Street Food  
 Italian

OTHERS CITIES

North Indian  
 Chinese  
 Continental  
 Italian  
 Cafe  
 Fast Food  
 South Indian  
 Mughlai  
 Desserts  
 Mexican

1. 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.

In [37]:

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import requests as r
from collections import Counter

d= pd.read_csv('zomato.csv',encoding="ISO-8859-1")# reading the file and encoding it to convert it to readable format
d=d[d['Country Code']==1] # Only selecting Indian Cities
df=d.copy()
data_d=df[(df.City=='New Delhi')|(df.City=='Ghaziabad')|(df.City=='Noida')|(df.City=='Gurgaon')|(df.City=='Faridabad')] #getting data for only Delhi_ncr cities
data_o=df[(df.City!='New Delhi')&(df.City!='Ghaziabad')&(df.City!='Noida')&(df.City!='Gurgaon')&(df.City!='Faridabad')] #getting data for rest of the cities in india.
data_d=data_d.Cuisines #extracting Cuisines present in delhi_ncr
data_o=data_o.Cuisines #extracting Cuisines present in other cities
delhi=[]
other=[]
for i in data_d:
    for j in i.split(','): #here we have multiple cuisines in a single Cuisines columns so i have used split() and strip() to split and iterate each Cuisines separately
        delhi.append(j.strip())
for i in data_o:
    for j in i.split(','): #here we have multiple cuisines in a single Cuisines columns so i have used split() and strip() to split and iterate each Cuisines separately
        other.append(j.strip())

dic_delhi={} # declaring an empty dic to store cuisines as key and count(cuisins) as values
dic_other={} # declaring an empty dic to store cuisines as key and count(cuisins) as values
for i in delhi:
    dic_delhi[i]=dic_delhi.get(i,0)+1
for i in other:
    dic_other[i]=dic_other.get(i,0)+1
od_delhi=dict(Counter(dic_delhi).most_common(10)) # getting top 10 cuisines name
od_other=dict(Counter(dic_other).most_common(10)) # getting top 10 cuisines name
x1=list(od_delhi.keys())
y1=list(od_delhi.values())
x2=list(od_other.keys())
y2=list(od_other.values())

```

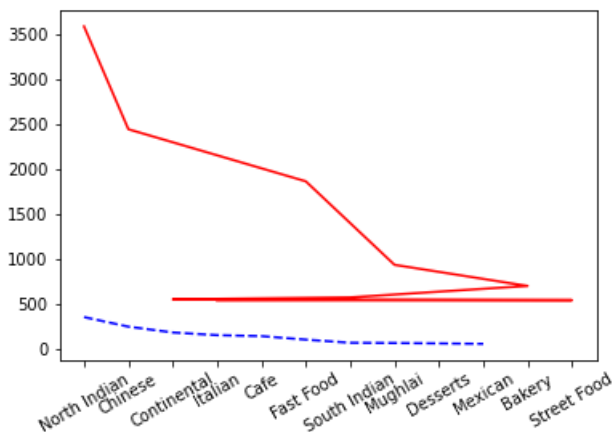
```

y2=list(x2_cuisine.values())

plt.plot(x2,y2,'b--')
plt.plot(x1,y1,'r-')
plt.xticks(rotation=30)
print('the top 10 cuisines most commonly served in delhi-ncr are:',x1)
print('top top 10 cusines most commonly served in other states are:',x2)
print("Among top 10 the cuisine which are only served in delhi_ncr= ",set(x1)-set(x2))
print("Among top 10 the cuisine which are only served in other states= ",set(x2)-set(x1))
plt.show()

```

the top 10 cuisines most commonly served in delhi-ncr are: ['North Indian', 'Chinese', 'Fast Food', 'Mughlai', 'Bakery', 'South Indian', 'Continental', 'Desserts', 'Street Food', 'Italian']  
top top 10 cusines most commonly served in other states are: ['North Indian', 'Chinese', 'Continental', 'Italian', 'Cafe', 'Fast Food', 'South Indian', 'Mughlai', 'Desserts', 'Mexican']  
Among top 10 the cuisine which are only served in delhi\_ncr= {'Bakery', 'Street Food'}  
Among top 10 the cuisine which are only served in other states= {'Cafe', 'Mexican'}



## Question 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.

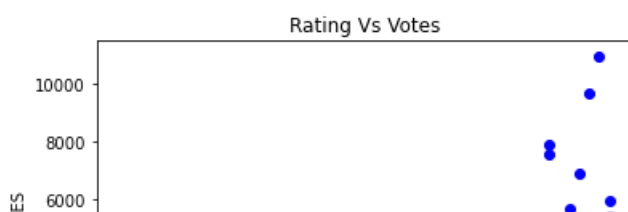
a. Number of Votes given Restaurant

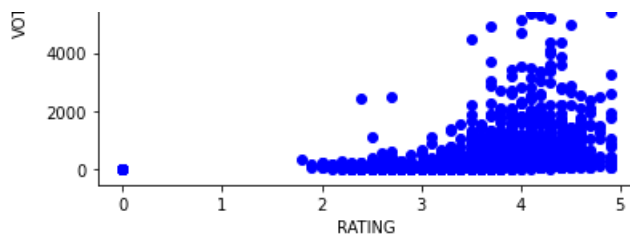
In [24]:

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.pyplot as plt
import requests as r
d= pd.read_csv('zomato.csv',encoding="ISO-8859-1")# reading the file and encoding it to convert it to readable format
d=d[d['Country Code']==1] # Only selecting Indian Cities
df=d.copy()
rating=df['Aggregate rating'] # extracting value of user rating and storing it in a variable
votes=df['Votes'] # extracting the number of votes and storing it into a variable
plt.plot(rating,votes,'bo') #plotting the graph between the rating and number of votes
plt.xlabel('RATING')
plt.ylabel('VOTES')
plt.title("Rating Vs Votes")
plt.show()

```

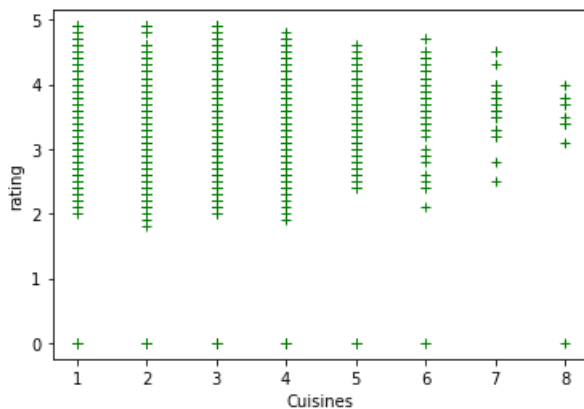




b. Restaurant serving more number of cuisines.

In [25]:

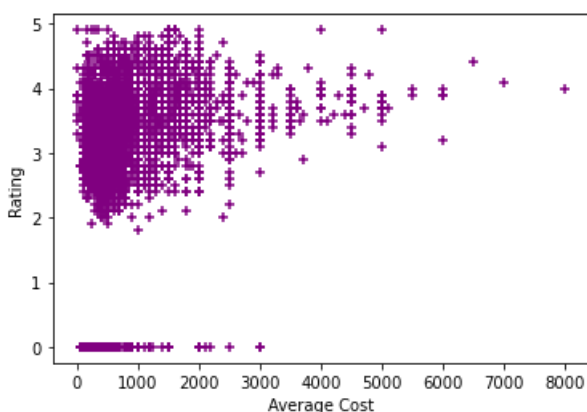
```
cusine=df.Cuisines
def f(s):
    s=len(s.split(','))
    return s
cusine=df.Cuisines.apply(f) # as we have mutiple cuisines name in one columns so i have have hanle
the string throug apply funtion of pandas
plt.plot(cusine,rating,'g+') #plotting the graph between cuisines and rating
plt.xlabel('Cuisines')
plt.ylabel('rating')
plt.show()
```



c. Average Cost of Restaurant

In [26]:

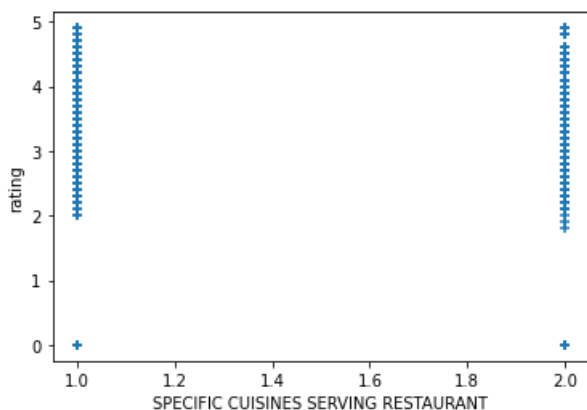
```
Cost=df['Average Cost for two'] # extracting the average cost of two values and storing it into a
variable
plt.scatter(Cost,rating,marker='+',color='purple',edgecolor='Black') #plotting a scatter chat cost
as x-aixs and rating as y-axis
plt.xlabel('Average Cost')
plt.ylabel('Rating')
plt.show()
```



d. Restaurant serving some specific cuisines.

In [27]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.pyplot as plt
import requests as r
d= pd.read_csv('zomato.csv',encoding="ISO-8859-1")# reading the file and encoding it to convert it
to readle format
d=d[d['Country Code']==1] # Only selecting Indian Cities
df=d.copy()
bool=cuisine<=2
sp=cuisine[bool]
spr=rating[bool]
plt.scatter(sp,spr,marker='+',edgecolor='black')
plt.ylabel('rating')
plt.xlabel('SPECIFIC CUISINES SERVING RESTAURANT')
plt.show()
```



1. Find the weighted restaurant rating of each locality and find out the top 10 localities with more weighted restaurant rating?

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

In [28]:

```
data=df.copy()
data['Weighted_Restaurant']=data['Votes']*data['Aggregate rating']
data=data.groupby('Locality').sum()
data.Weighted_Restaurant=data.Weighted_Restaurant/data.Votes
data.dropna(subset=['Weighted_Restaurant'],inplace=True)
print(data.sort_values('Weighted_Restaurant').loc[:, 'Weighted_Restaurant'].iloc[-10:])
```

```
Locality
Riverside Mall, Gomti Nagar      4.700000
Sector 5, Salt Lake              4.707023
Banjara Hills                   4.718762
Deccan Gymkhana                 4.800000
Express Avenue Mall, Royapettah 4.800000
Kirlampudi Layout               4.820161
Powai                           4.841869
Friends Colony                  4.886916
Hotel Clarks Amer, Malviya Nagar 4.900000
Aminabad                        4.900000
Name: Weighted_Restaurant, dtype: float64
```

### Question 3

#### Visualization

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

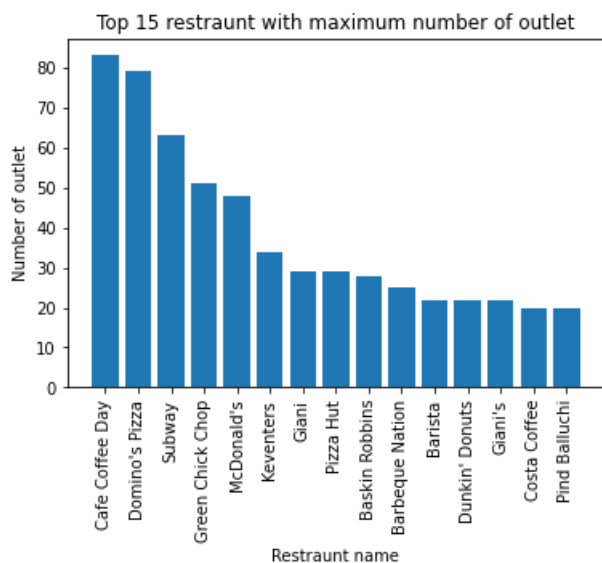
In [29]:

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.pyplot as plt
import requests as r

d= pd.read_csv('zomato.csv',encoding="ISO-8859-1")# reading the file and encoding it to convert it
to readle format
d=d[d['Country Code']==1] # Only selecting Indian Cities
df=d.copy()
x=df['Restaurant Name'].value_counts()[:15].index #for x axis I am taking the top 15 restraunt nam
e.
y=df['Restaurant Name'].value_counts()[:15]# y axis represents the number of outlet the restraunt
has.
plt.bar(x,y)
plt.xticks(rotation=90)
plt.xlabel('Restraunt name') # declaring the label for x axis
plt.ylabel("Number of outlet") # declaring the label for y axis
plt.title('Top 15 restraunt with maximum number of outlet')
plt.show()

```



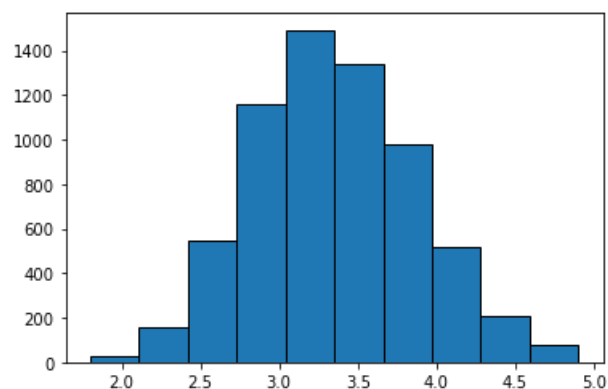
1. Plot the histogram of aggregate rating of restaurant( drop the unrated restaurant).

In [30]:

```

dat=d.copy()
dat=dat[dat['Aggregate rating']!=0] #removing the unrated restaurant
plt.hist(dat['Aggregate rating'],edgecolor='black') #ploting the histogram
plt.show()

```



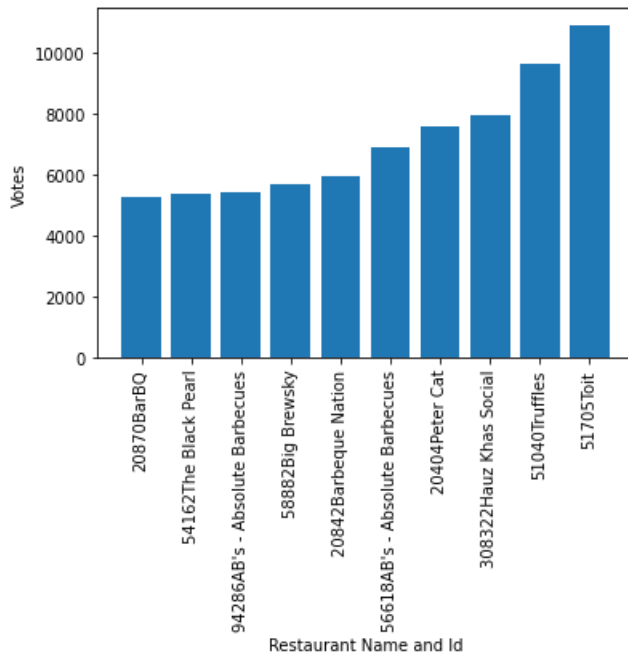
1. Plot the bar graph top 10 restaurants in the data with the highest number of votes.

In [33]:

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import requests as r
k= pd.read_csv('zomato.csv',encoding="ISO-8859-1")# reading the file and encoding it to convert it
to readle format
k=k[k['Country Code']==1] # Only selecting Indian Cities
da=k.copy()
da.sort_values('Votes',inplace=True)
y=da.Votes.iloc[-10:]
plt.bar(np.array(da['Restaurant ID'].iloc[-10:],dtype=str)+ np.array(da['Restaurant Name'].iloc[-10
:]),y)#considering restraunt name with id as x-axis
plt.xticks(rotation=90)
plt.xlabel('Restaurant Name and Id')
plt.ylabel('Votes')
plt.show()

```



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

In [34]:

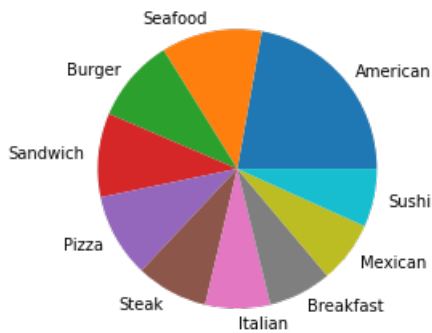
```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import requests as r
from collections import Counter
d= pd.read_csv('zomato.csv',encoding="ISO-8859-1")# reading the file and encoding it to convert it
to readle format
d=d[d['Country Code']==216] # Only selecting USA Cities
df=d.copy()
df.drop(df.index[df.Cuisines.isnull()], inplace=True)
cus_usa=df.Cuisines

cus=[]
for i in cus_usa:
    for j in i.split(','):
        cus.append(j.strip())
cus_d={}
for i in cus:
    cus_d[i]=cus_d.get(i,0)+1
od=dict(Counter(cus_d).most_common(10)) ## for getting top 10 cusiines
x=list(od)
y=list(od.values())
plt.pie(y,labels=x)
plt.show()
for i in range(len(x)):#iterating to print the cusiines name
    print(x[i],y[i])

```





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

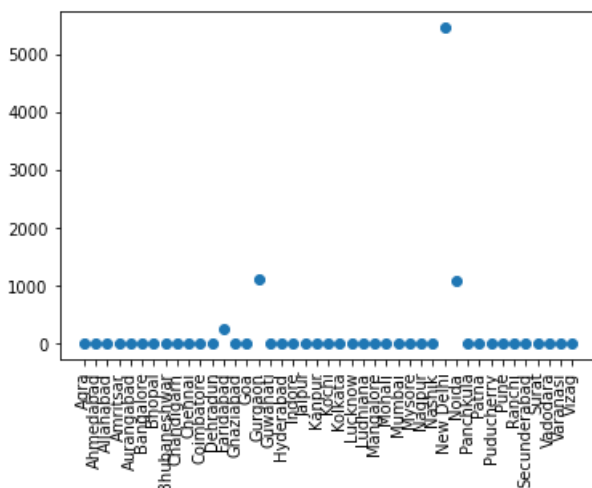
In [ ]:

1. 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.

In [19]:

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

s= pd.read_csv('zomato.csv',encoding="ISO-8859-1")# reading the file and encoding it to convert it
to readle format
s=s[s['Country Code']==1] # Only selecting Indian Cities
fl=s.copy()
rat=fl['Aggregate rating']
fl['Weighted_Restaurant']=fl['Votes']*fl['Aggregate rating'] #using the formula given in the above
question to calculate Weighted_Restaurant
f=fl.groupby('City').sum()
f['wr']=f['Weighted_Restaurant']/f['Votes']
f=f.wr
no=f.groupby('City').count().Address
plt.scatter(f.index,no)
plt.xticks(rotation=90)
plt.show()
```



In [ ]:

In [ ]: