In [4]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [6]:
```
df = pd.read_csv("resturant.csv")

df.head()
```

Out[6]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	...	Currency	Has Table booking	Has Online delivery	Is delivering now	Switch to order menu	Price range
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City	Century City Mall, Poblacion, Makati City, Mak...	121.027535	14.565443	French, Japanese, Desserts	...	Botswana Pula(P)	Yes	No	No	No	3
1	6304287	Izakaya Kikufuji	162	Makati City	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...	Little Tokyo, Legaspi Village, Makati City	Little Tokyo, Legaspi Village, Makati City, Ma...	121.014101	14.553708	Japanese	...	Botswana Pula(P)	Yes	No	No	No	3
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City	Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal...	Edsa Shangri-La, Ortigas, Mandaluyong City	Edsa Shangri-La, Ortigas, Mandaluyong City, Ma...	121.056831	14.581404	Seafood, Asian, Filipino, Indian	...	Botswana Pula(P)	Yes	No	No	No	4
3	6318506	Ooma	162	Mandaluyong City	Third Floor, Mega Fashion Hall, SM Megamall, O...	SM Megamall, Ortigas, Mandaluyong City	SM Megamall, Ortigas, Mandaluyong City, Mandal...	121.056475	14.585318	Japanese, Sushi	...	Botswana Pula(P)	No	No	No	No	4
4	6314302	Sambo Kojin	162	Mandaluyong City	Third Floor, Mega Atrium, SM Megamall, Ortigas...	SM Megamall, Ortigas, Mandaluyong City	SM Megamall, Ortigas, Mandaluyong City, Mandal...	121.057508	14.584450	Japanese, Korean	...	Botswana Pula(P)	Yes	No	No	No	4

5 rows × 21 columns

In [8]:
```
df.columns
```

Out[8]:
```
Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'Address',
       'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuisines',
       'Average Cost for two', 'Currency', 'Has Table booking',
       'Has Online delivery', 'Is delivering now', 'Switch to order menu',
       'Price range', 'Aggregate rating', 'Rating color', 'Rating text',
       'Votes'],
      dtype='object')
```

In [10]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9551 entries, 0 to 9550
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   Restaurant ID         9551 non-null   int64  
 1   Restaurant Name       9551 non-null   object 
 2   Country Code          9551 non-null   int64  
 3   City                  9551 non-null   object 
 4   Address               9551 non-null   object 
 5   Locality              9551 non-null   object 
 6   Locality Verbose      9551 non-null   object 
 7   Longitude             9551 non-null   float64
 8   Latitude              9551 non-null   float64
 9   Cuisines              9542 non-null   object 
 10  Average Cost for two  9551 non-null   int64  
 11  Currency              9551 non-null   object 
 12  Has Table booking     9551 non-null   object 
 13  Has Online delivery   9551 non-null   object 
 14  Is delivering now     9551 non-null   object 
 15  Switch to order menu  9551 non-null   object 
 16  Price range           9551 non-null   int64  
 17  Aggregate rating      9551 non-null   float64
 18  Rating color          9551 non-null   object 
 19  Rating text           9551 non-null   object 
 20  Votes                 9551 non-null   int64  
dtypes: float64(3), int64(5), object(13)
memory usage: 1.5+ MB
```

In [12]:
```
df.dtypes
```

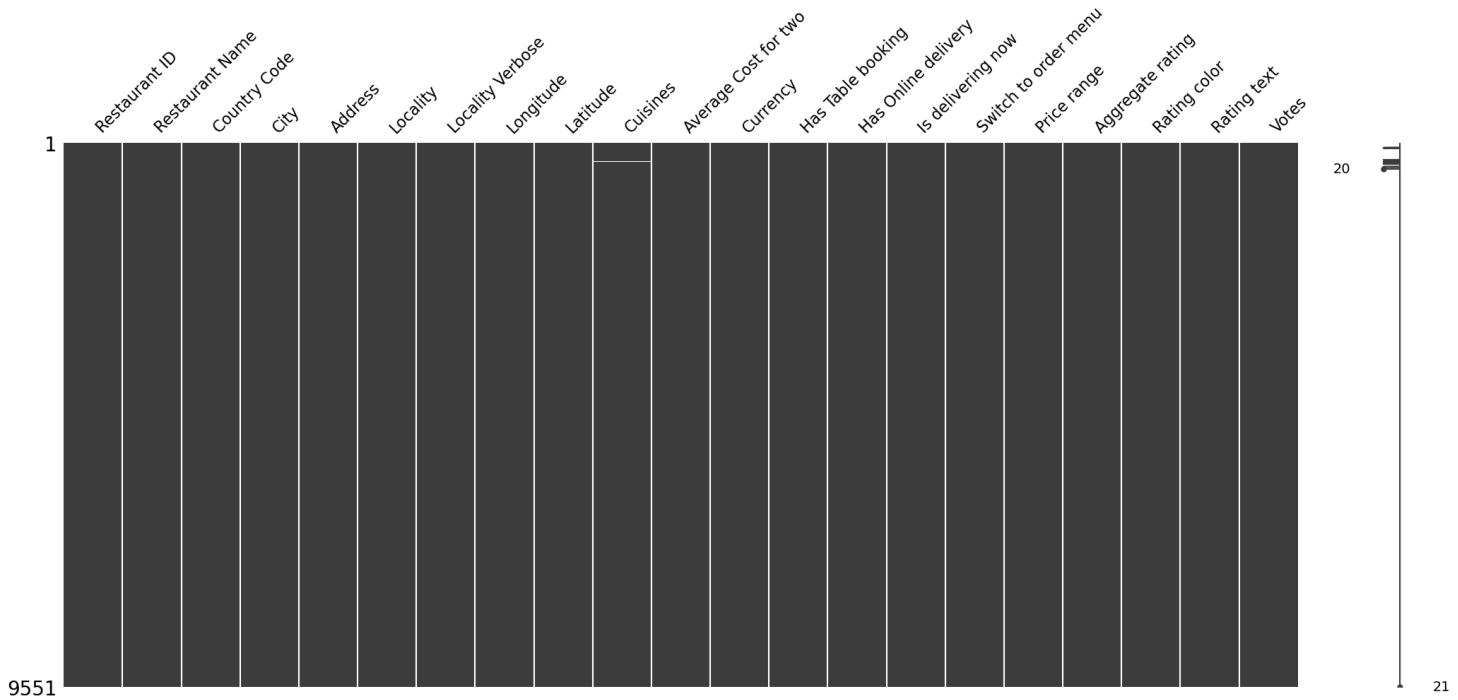
Out[12]:
```
Restaurant ID             int64
Restaurant Name          object
Country Code              int64
City                     object
Address                  object
Locality                 object
Locality Verbose         object
Longitude               float64
Latitude                float64
Cuisines                 object
Average Cost for two      int64
Currency                 object
Has Table booking        object
Has Online delivery      object
Is delivering now        object
Switch to order menu     object
Price range               int64
Aggregate rating        float64
Rating color             object
Rating text              object
Votes                     int64
dtype: object
```

In [14]:
```
df.describe()
```

Out[14]:

	Restaurant ID	Country Code	Longitude	Latitude	Average Cost for two	Price range	Aggregate rating	Votes
count	9.551000e+03	9551.000000	9551.000000	9551.000000	9551.000000	9551.000000	9551.000000	9551.000000
mean	9.051128e+06	18.365616	64.126574	25.854381	1199.210763	1.804837	2.666370	156.909748
std	8.791521e+06	56.750546	41.467058	11.007935	16121.183073	0.905609	1.516378	430.169145
min	5.300000e+01	1.000000	-157.948486	-41.330428	0.000000	1.000000	0.000000	0.000000
25%	3.019625e+05	1.000000	77.081343	28.478713	250.000000	1.000000	2.500000	5.000000
50%	6.004089e+06	1.000000	77.191964	28.570469	400.000000	2.000000	3.200000	31.000000
75%	1.835229e+07	1.000000	77.282006	28.642758	700.000000	2.000000	3.700000	131.000000
max	1.850065e+07	216.000000	174.832089	55.976980	800000.000000	4.000000	4.900000	10934.000000

In [16]:
```
import missingno as msno
msno.matrix(df)
plt.show()
```



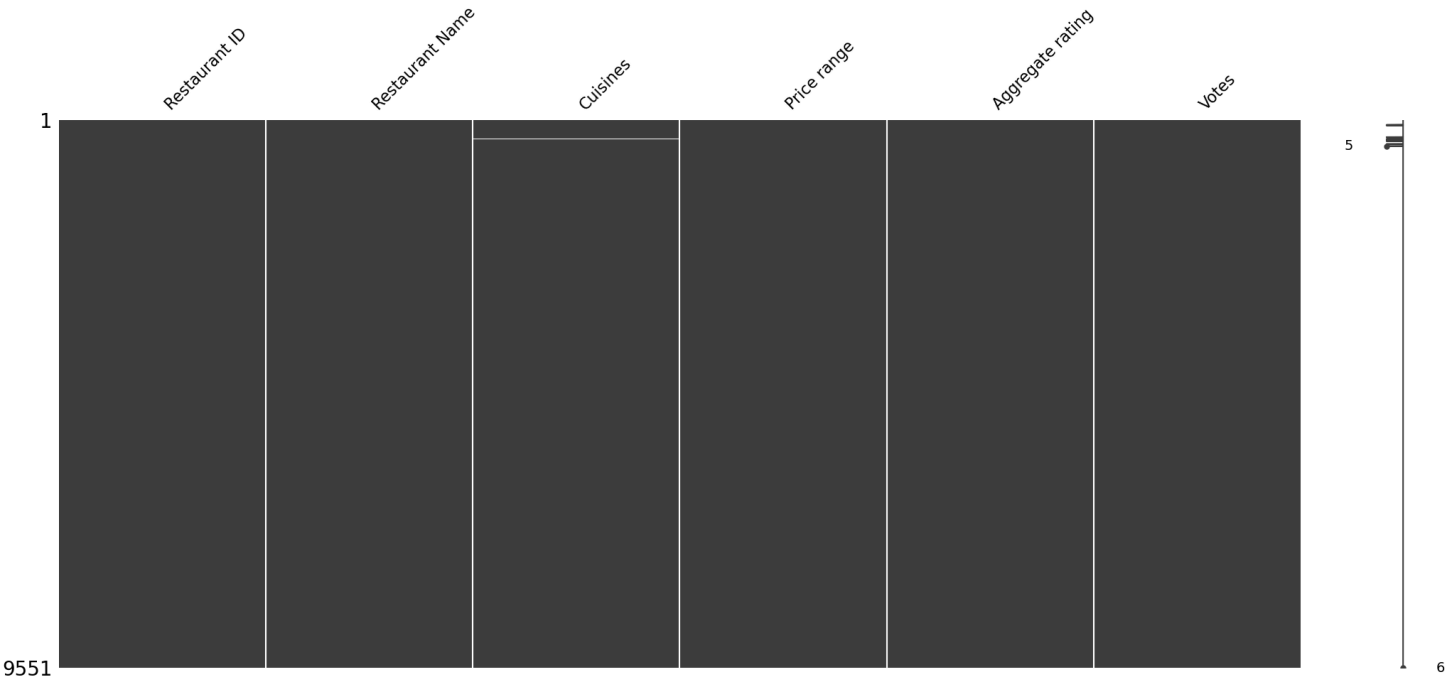
In [18]:
```
#create a refined dataframe
dfR = df[['Restaurant ID','Restaurant Name','Cuisines','Price range','Aggregate rating','Votes']]
dfR
```

Out[18]:

	Restaurant ID	Restaurant Name	Cuisines	Price range	Aggregate rating	Votes
0	6317637	Le Petit Souffle	French, Japanese, Desserts	3	4.8	314
1	6304287	Izakaya Kikufuji	Japanese	3	4.5	591
2	6300002	Heat - Edsa Shangri-La	Seafood, Asian, Filipino, Indian	4	4.4	270
3	6318506	Ooma	Japanese, Sushi	4	4.9	365
4	6314302	Sambo Kojin	Japanese, Korean	4	4.8	229
...	...	...	...	...	...	...
9546	5915730	Namlı Gurme	Turkish	3	4.1	788
9547	5908749	Ceviz A acl	World Cuisine, Patisserie, Cafe	3	4.2	1034
9548	5915807	Huqqa	Italian, World Cuisine	4	3.7	661
9549	5916112	A k Kahve	Restaurant Cafe	4	4.0	901
9550	5927402	Walter's Coffee Roastery	Cafe	2	4.0	591

9551 rows × 6 columns

In [20]:
```
msno.matrix(dfR)
plt.show()
```



In [22]:
```
#handle missing values
dfR.isna().sum()
```

Out[22]:
```
Restaurant ID       0
Restaurant Name     0
Cuisines            9
Price range         0
Aggregate rating    0
Votes               0
dtype: int64
```

In [24]:
```
dfR = dfR.dropna()
dfR.isna().sum()
```

Out[24]:
```
Restaurant ID       0
Restaurant Name     0
Cuisines            0
Price range         0
Aggregate rating    0
Votes               0
dtype: int64
```

In [26]:
```
dfR.duplicated().sum()
```

Out[26]:
```
0
```

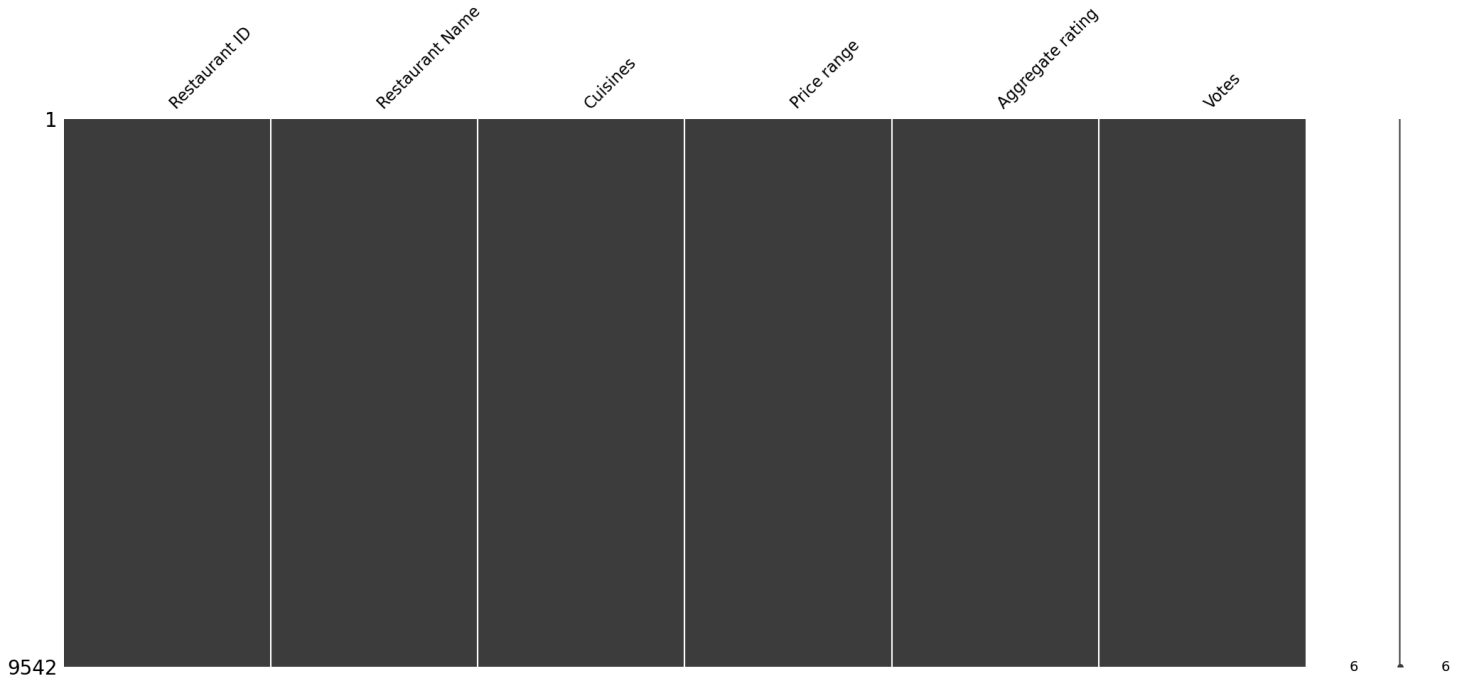
In [28]:
```
dfR['Restaurant Name'].duplicated().sum()
```

Out[28]:
```
2105
```

In [30]:
```
dfR['Restaurant Name'].value_counts()
```

Out[30]:
```
Restaurant Name
Cafe Coffee Day             83
Domino's Pizza              79
Subway                      63
Green Chick Chop            51
McDonald's                  48
                            ..
The Town House Cafe          1
The G.T. Road                1
The Darzi Bar & Kitchen      1
Smoke On Water               1
Walter's Coffee Roastery     1
Name: count, Length: 7437, dtype: int64
```

In [32]:
```
msno.matrix(dfR)
plt.show()
```



In [34]:
```
#sorting the restaurants by name and rating
dfR = dfR.sort_values(by=['Restaurant Name','Aggregate rating'],ascending=False)
dfR.head()
```

Out[34]:

	Restaurant ID	Restaurant Name	Cuisines	Price range	Aggregate rating	Votes
3120	18222559	{Niche} - Cafe & Bar	North Indian, Chinese, Italian, Continental	3	4.1	492
9334	7100938	wagamama	Japanese, Asian	4	3.7	131
9454	6401789	tashas	Cafe, Mediterranean	4	4.1	374
4659	18361747	t Lounge by Dilmah	Cafe, Tea, Desserts	2	3.6	34
9385	6113857	sketch Gallery	British, Contemporary	4	4.5	148

In [36]:
```
dfR[dfR["Restaurant Name"]=="Cafe Coffee Day"].head()
```

Out[36]:

	Restaurant ID	Restaurant Name	Cuisines	Price range	Aggregate rating	Votes
6430	5595	Cafe Coffee Day	Cafe	1	3.6	58
8432	594	Cafe Coffee Day	Cafe	1	3.6	125
3946	305736	Cafe Coffee Day	Cafe	1	3.5	35
5877	8828	Cafe Coffee Day	Cafe	1	3.5	50
3001	596	Cafe Coffee Day	Cafe	1	3.4	277

In [38]:
```
#removing duplicate entries of same restaurant name
dfR = dfR.drop_duplicates('Restaurant Name',keep='first')
dfR
```

Out[38]:

	Restaurant ID	Restaurant Name	Cuisines	Price range	Aggregate rating	Votes
3120	18222559	{Niche} - Cafe & Bar	North Indian, Chinese, Italian, Continental	3	4.1	492
9334	7100938	wagamama	Japanese, Asian	4	3.7	131
9454	6401789	tashas	Cafe, Mediterranean	4	4.1	374
4659	18361747	t Lounge by Dilmah	Cafe, Tea, Desserts	2	3.6	34
9385	6113857	sketch Gallery	British, Contemporary	4	4.5	148
...	...	...	...	...	...	...
6998	18336489	#OFF Campus	Cafe, Continental, Italian, Fast Food	2	3.7	216
2613	18311951	#InstaFreeze	Ice Cream	1	0.0	2
9148	18378803	#Dilliwaala6	North Indian	3	3.7	124
2459	3100446	#45	Cafe	2	3.6	209
9523	6000871	ukura a Sofras\	Kebab, Izgara	3	4.4	296

7437 rows × 6 columns

In [40]:
```
dfR['Restaurant Name'].value_counts()
```

Out[40]:
```
Restaurant Name
{Niche} - Cafe & Bar        1
French Omelette             1
Four Queens Dairy Cream     1
Fourteen Eleven Tea Cafe    1
Fozzie's Pizzaiolo          1
                           ..
Pizza Man                   1
Pizza Point                 1
Pizza Station               1
Pizza Street                1
 ukura  a Sofras\           1
Name: count, Length: 7437, dtype: int64
```

In [42]:
```
dfR = dfR[dfR['Aggregate rating']>3.9]
dfR
```

Out[42]:

	Restaurant ID	Restaurant Name	Cuisines	Price range	Aggregate rating	Votes
3120	18222559	{Niche} - Cafe & Bar	North Indian, Chinese, Italian, Continental	3	4.1	492
9454	6401789	tashas	Cafe, Mediterranean	4	4.1	374
9385	6113857	sketch Gallery	British, Contemporary	4	4.5	148
1837	18418247	feel ALIVE	North Indian, American, Asian, Biryani	3	4.7	69
440	17616266	Zunzi's	International, Mediterranean, Sandwich	2	4.5	796
...	...	...	...	...	...	...
2484	18233317	145 Kala Ghoda	Fast Food, Beverages, Desserts	3	4.2	1606
2292	2100784	11th Avenue Cafe Bistro	Cafe, American, Italian, Continental	2	4.1	377
751	2600031	10 Downing Street	North Indian, Chinese	3	4.0	257
351	17057397	'Ohana	Hawaiian	3	4.5	1151
9523	6000871	ukura a Sofras\	Kebab, Izgara	3	4.4	296

1236 rows × 6 columns

In [44]:
```
#splitting cuisines into list
dfR['Cuisines'] = dfR['Cuisines'].str.split(', ')
dfR
```

```
C:\Users\sguha\AppData\Local\Temp\ipykernel_1432\3621809458.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  dfR['Cuisines'] = dfR['Cuisines'].str.split(', ')
```

Out[44]:

	Restaurant ID	Restaurant Name	Cuisines	Price range	Aggregate rating	Votes
3120	18222559	{Niche} - Cafe & Bar	[North Indian, Chinese, Italian, Continental]	3	4.1	492
9454	6401789	tashas	[Cafe, Mediterranean]	4	4.1	374
9385	6113857	sketch Gallery	[British, Contemporary]	4	4.5	148
1837	18418247	feel ALIVE	[North Indian, American, Asian, Biryani]	3	4.7	69
440	17616266	Zunzi's	[International, Mediterranean, Sandwich]	2	4.5	796
...	...	...	...	...	...	...
2484	18233317	145 Kala Ghoda	[Fast Food, Beverages, Desserts]	3	4.2	1606
2292	2100784	11th Avenue Cafe Bistro	[Cafe, American, Italian, Continental]	2	4.1	377
751	2600031	10 Downing Street	[North Indian, Chinese]	3	4.0	257
351	17057397	'Ohana	[Hawaiian]	3	4.5	1151
9523	6000871	ukura a Sofras\	[Kebab, Izgara]	3	4.4	296

1236 rows × 6 columns

In [46]:
```
dfR = dfR.explode('Cuisines')
dfR
```

Out[46]:

	Restaurant ID	Restaurant Name	Cuisines	Price range	Aggregate rating	Votes
3120	18222559	{Niche} - Cafe & Bar	North Indian	3	4.1	492
3120	18222559	{Niche} - Cafe & Bar	Chinese	3	4.1	492
3120	18222559	{Niche} - Cafe & Bar	Italian	3	4.1	492
3120	18222559	{Niche} - Cafe & Bar	Continental	3	4.1	492
9454	6401789	tashas	Cafe	4	4.1	374
...	...	...	...	...	...	...
751	2600031	10 Downing Street	North Indian	3	4.0	257
751	2600031	10 Downing Street	Chinese	3	4.0	257
351	17057397	'Ohana	Hawaiian	3	4.5	1151
9523	6000871	ukura a Sofras\	Kebab	3	4.4	296
9523	6000871	ukura a Sofras\	Izgara	3	4.4	296

2971 rows × 6 columns

In [48]:
```
dfR['Cuisines'].value_counts()
```

Out[48]:
```
Cuisines
North Indian    270
Italian         237
Chinese         200
Continental     199
Cafe            177
                ...
Durban            1
Iranian           1
Awadhi            1
Persian           1
Irish             1
Name: count, Length: 128, dtype: int64
```

In [50]:
```
restoXcuisines = pd.crosstab(dfR['Restaurant Name'], dfR['Cuisines'])
restoXcuisines
```

Out[50]:

Cuisines	Afghani	African	American	Andhra	Arabian	Argentine	Asian	Asian Fusion	Australian	Awadhi	...	Teriyaki	Tex-Mex	Thai	Tibetan	Turkish	Turkish Pizza	Vegetarian	Vietnamese
Restaurant Name																			
ukura a Sofrasl	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
'Ohana	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
10 Downing Street	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
11th Avenue Cafe Bistro	0	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
145 Kala Ghoda	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Zunzi's	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
feel ALIVE	0	0	1	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0
sketch Gallery	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
tashas	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
{Niche} - Cafe & Bar	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0

1236 rows × 128 columns

In [52]:
```
dfR['Restaurant Name'].sample(20, random_state=194)
```

Out[52]:
```
2099                          Indian Grill Room
388                                Native Cafe
2145                             Sandburg Shakes
2317                Sahib  s Barbeque by Ohri  s
1151                                Souza Lobo
2142                                Cafe Sante
2407                             India Restaurant
637                              Sheroes Hangout
1858                            Boombox Brewstreet
2393    Aangan - Downtown Multicuisine Restaurant
264                               Bandit Burrito
7067                                    Pa Pa Ya
9352                                Tipu Sultan
9141                               Baker Street
9469                            The Belgian Triple
354                              Caf   Tu Tu Tango
1149                               Martin's Corner
2366                                Dunkin Donuts
81                       Sainte Marie Gastronomia
6714                               Showstopper
Name: Restaurant Name, dtype: object
```

In [54]:
```
from sklearn.metrics import jaccard_score
print(jaccard_score(restoXcuisines.loc["Olive Bistro"].values,
                    restoXcuisines.loc["Rose Cafe"].values))
```

```
0.3333333333333333
```

In [56]:
```
from scipy.spatial.distance import pdist, squareform

jaccardDist = pdist(restoXcuisines.values, metric='jaccard')
jaccardMatrix = squareform(jaccardDist)
jaccardSim = 1 - jaccardMatrix
dfJaccard = pd.DataFrame(
    jaccardSim,
    index=restoXcuisines.index,
    columns=restoXcuisines.index)

dfJaccard
```

Out[56]:

Restaurant Name	ukura a Sofrasl	'Ohana	10 Downing Street	11th Avenue Cafe Bistro	145 Kala Ghoda	19 Flavours Biryani	1918 Bistro & Grill	2 Dog	22nd Parallel	3 Wise Monkeys	...	Zoe	Zoeys Pizzeria	Zolocrust - Hotel Clarks Amer	Zombie Burger + Drink Lab	Zuka Choco-la	Zunzi's	feel ALIVE	sketch Gallery	tash
Restaurant Name																				
ukura a Sofrasl	1.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.00	0.0	0.0	0.000000	0.00	0.000000	0.0	0.
'Ohana	0.0	1.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.00	0.0	0.0	0.000000	0.00	0.000000	0.0	0.
10 Downing Street	0.0	0.0	1.0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.00	0.0	0.0	0.000000	0.00	0.200000	0.0	0.
11th Avenue Cafe Bistro	0.0	0.0	0.0	1.000000	0.0	0.0	0.0	0.166667	0.0	0.0	...	0.0	0.00	0.4	0.0	0.000000	0.00	0.142857	0.0	0.
145 Kala Ghoda	0.0	0.0	0.0	0.000000	1.0	0.0	0.0	0.000000	0.0	0.0	...	0.2	0.00	0.0	0.2	0.333333	0.00	0.000000	0.0	0.
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
Zunzi's	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.200000	0.0	0.0	...	0.0	0.25	0.0	0.0	0.000000	1.00	0.000000	0.0	0.
feel ALIVE	0.0	0.0	0.2	0.142857	0.0	0.0	0.0	0.166667	0.0	0.0	...	0.0	0.00	0.0	0.0	0.000000	0.00	1.000000	0.0	0.
sketch Gallery	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.00	0.0	0.0	0.000000	0.00	0.000000	1.0	0.
tashas	0.0	0.0	0.0	0.200000	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.00	0.0	0.0	0.000000	0.25	0.000000	0.0	1.
{Niche} - Cafe & Bar	0.0	0.0	0.5	0.333333	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.00	0.4	0.0	0.000000	0.00	0.142857	0.0	0.

1236 rows × 1236 columns

In [58]:
```
dfR['Restaurant Name'].sample(20)
```

Out[58]:
```
7540        Varq - The Taj Mahal Hotel
8209        New Town Cafe - Park Plaza
1402                               Zoe
504     Ceviche Tapas Bar & Restaurant
7036                     Cafe Kazbaah
9204                 The Centre Court
429                     Marukame Udon
6921                        Rose Cafe
2116                         Udaipuri
760       Manohar Dairy And Restaurant
9316                         Baduzzi
2382            Grand Hotel Restaurant
9480                      Crawdaddy's
6850                     Tossin Pizza
1212                      Big Wong XL
7036                     Cafe Kazbaah
9310                  The Garden Shed
9541                  Emirgan S _ti
1154              The Fisherman's Wharf
66                  Skye - Hotel Unique
Name: Restaurant Name, dtype: object
```

In [60]:
```
resto = 'Ooma'

sim = dfJaccard.loc[resto].sort_values(ascending=False)
sim = pd.DataFrame({'Restaurant Name': sim.index, 'simScore': sim.values})
sim = sim[(sim['Restaurant Name']!= resto) & (sim['simScore']>=0.7)].head(5)
RestoRec = pd.merge(sim,dfR[['Restaurant Name','Aggregate rating']],how='inner',on='Restaurant Name')
FinalRestoRec = RestoRec.sort_values('Aggregate rating',ascending=False).drop_duplicates('Restaurant Name',keep='first')
```

In [62]:
```
FinalRestoRec
```

Out[62]:

	Restaurant Name	simScore	Aggregate rating
8	Sushi Masa	1.0	4.9
2	Roka	1.0	4.6
4	Nobu	1.0	4.4
0	Nagai	1.0	4.3
6	Osaka	1.0	4.2

In [ ]:
```
```