

Restaurant Recommendation System based on user preference.

```
In [4]: # Import Libraries
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import jaccard_score
from scipy.spatial.distance import pdist, squareform
pd.reset_option('display.max_rows')
```

```
In [5]: import warnings
warnings.filterwarnings("ignore")
```

```
In [9]: # Creating the Dataframe
file_path = r'D:\Software\New Project\Internship\Cognifyz\Predict Restaurant Ratings\Dataset .csv'
df = pd.read_csv(file_path)
df.head()
```

```
Out[9]:
```

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	...	Curr
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	...	Bots Pt
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	...	Bots Pt
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	...	Bots Pt
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	...	Bots Pt
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	...	Bots Pt

5 rows × 21 columns

```
In [10]: df.columns
```

```
Out[10]: 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 [11]: dfRS = df[['Restaurant ID', 'Restaurant Name', 'Cuisines', 'Aggregate rating', 'Votes']]
dfRS
```

Out[11]:

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

9551 rows × 5 columns

Data Cleaning

In [12]:

```
# Gathering information of every columns

# Columns Description
def dataDesc():
    listItem = []
    for col in dfRS.columns :
        listItem.append(
            [col,
             dfRS[col].dtype,
             dfRS[col].isna().sum(),
             round(dfRS[col].isna().sum()/len(dfRS)*100,2),
             dfRS[col].nunique(),
             list(dfRS[col].drop_duplicates().sample(2).values)]
        )
    descData = pd.DataFrame(data = listItem,
                             columns = ['Column','Data Type', 'Missing Value',
                                         'Pct Missing Value', 'Num Unique', 'Unique Sample'])

    return descData

dataDesc()
```

Out[12]:

	Column	Data Type	Missing Value	Pct Missing Value	Num Unique	Unique Sample
0	Restaurant ID	int64	0	0.00	9551	[1820, 18381643]
1	Restaurant Name	object	0	0.00	7446	[Narang's Bake 'n' Cake, Indo Chinese]
2	Cuisines	object	9	0.09	1825	[Lebanese, Mediterranean, Arabian, Cafe, North...
3	Aggregate rating	float64	0	0.00	33	[4.2, 4.5]
4	Votes	int64	0	0.00	1012	[348, 797]

In [13]:

```
dfRS = dfRS.dropna()
```

In [14]:

```
dfRS
```

Out [14]:

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

9542 rows × 5 columns

In [15]:

```
# Renaming the Columns
dfRS = dfRS.rename(columns={'Restaurant ID': 'restaurant_id'})
dfRS = dfRS.rename(columns={'Restaurant Name': 'restaurant_name'})
dfRS = dfRS.rename(columns={'Cuisines': 'cuisines'})
dfRS = dfRS.rename(columns={'Aggregate rating': 'aggregate_rating'})
dfRS = dfRS.rename(columns={'Votes': 'votes'})
```

In [16]:

```
dfRS
```

Out [16]:

	restaurant_id	restaurant_name	cuisines	aggregate_rating	votes
0	6317637	Le Petit Souffle	French, Japanese, Desserts	4.8	314
1	6304287	Izakaya Kikufuji	Japanese	4.5	591
2	6300002	Heat - Edsa Shangri-La	Seafood, Asian, Filipino, Indian	4.4	270
3	6318506	Ooma	Japanese, Sushi	4.9	365
4	6314302	Sambo Kojin	Japanese, Korean	4.8	229
...
9546	5915730	Naml\ Gurme	Turkish	4.1	788
9547	5908749	Ceviz Aac\	World Cuisine, Patisserie, Cafe	4.2	1034
9548	5915807	Huqqa	Italian, World Cuisine	3.7	661
9549	5916112	Ak Kahve	Restaurant Cafe	4.0	901
9550	5927402	Walter's Coffee Roastery	Cafe	4.0	591

9542 rows × 5 columns

In [18]:

```
# Check for Duplicates
dfRS.duplicated().sum()
```

Out [18]: 0

In [19]:

```
dfRS['restaurant_name'].duplicated().sum()
```

Out [19]: 2105

In [20]:

```
dfRS['restaurant_name'].value_counts()
```

Out [20]:

```
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 [21]:

```
dfRS = dfRS.sort_values(by=['restaurant_name','aggregate_rating'],ascending=False)
```

```
In [22]: dfRS[dfRS['restaurant_name']=="Domino's Pizza"].head()
```

Out[22]:

	restaurant_id	restaurant_name	cuisines	aggregate_rating	votes	
	3031	143	Domino's Pizza	Pizza, Fast Food	3.7	336
	1844	5065	Domino's Pizza	Pizza, Fast Food	3.6	146
	2448	15078	Domino's Pizza	Pizza, Fast Food	3.6	86
	7618	18263236	Domino's Pizza	Pizza, Fast Food	3.6	24
	8437	384	Domino's Pizza	Pizza, Fast Food	3.6	547

```
In [23]: # Dropping duplicaes only keeping first Value.
dfRS = dfRS.drop_duplicates('restaurant_name',keep='first')
dfRS
```

Out[23]:

	restaurant_id	restaurant_name	cuisines	aggregate_rating	votes	
	9523	6000871	ukuraa Sofras	Kebab, Izgara	4.4	296
	3120	18222559	{Niche} - Cafe & Bar	North Indian, Chinese, Italian, Continental	4.1	492
	9334	7100938	wagamama	Japanese, Asian	3.7	131
	9454	6401789	tashas	Cafe, Mediterranean	4.1	374
	4659	18361747	t Lounge by Dilmah	Cafe, Tea, Desserts	3.6	34

	8692	18317511	#Urban Caf	North Indian, Chinese, Italian	3.3	49
	6998	18336489	#OFF Campus	Cafe, Continental, Italian, Fast Food	3.7	216
	2613	18311951	#InstaFreeze	Ice Cream	0.0	2
	9148	18378803	#Dilliwaala6	North Indian	3.7	124
	2459	3100446	#45	Cafe	3.6	209

7437 rows × 5 columns

```
In [24]: dfRS['restaurant_name'].value_counts()
```

Out[24]:

restaurant_name	
ukuraa Sofras	1
French Toast	1
Fourteen Eleven Tea Cafe	1
Fozzie's Pizzaiolo	1
Frasers	1
..	
Pizza Street	1
Pizza Treat	1
Pizza Yum	1
Pizza Bessa	1
#45	1
Name: count, Length: 7437, dtype: int64	

```
In [26]: dfRS = dfRS[dfRS['aggregate_rating']>=4.0]
dfRS
```

Out[26]:

	restaurant_id	restaurant_name	cuisines	aggregate_rating	votes	
	9523	6000871	ukuraa Sofras	Kebab, Izgara	4.4	296
	3120	18222559	{Niche} - Cafe & Bar	North Indian, Chinese, Italian, Continental	4.1	492
	9454	6401789	tashas	Cafe, Mediterranean	4.1	374
	9385	6113857	sketch Gallery	British, Contemporary	4.5	148
	1837	18418247	feel ALIVE	North Indian, American, Asian, Biryani	4.7	69

	1468	18408054	19 Flavours Biryani	Mughlai, Hyderabad	4.1	84
	2484	18233317	145 Kala Ghoda	Fast Food, Beverages, Desserts	4.2	1606
	2292	2100784	11th Avenue Cafe Bistro	Cafe, American, Italian, Continental	4.1	377
	751	2600031	10 Downing Street	North Indian, Chinese	4.0	257
	351	17057397	'Ohana	Hawaiian	4.5	1151

1236 rows × 5 columns

```
In [27]: # Split Cuisines into list
dfRS['cuisines'] = dfRS['cuisines'].str.split(', ')
dfRS
```

Out[27]:

	restaurant_id	restaurant_name	cuisines	aggregate_rating	votes
9523	6000871	ukuraa Sofras	[Kebab, Izgara]	4.4	296
3120	18222559	{Niche} - Cafe & Bar	[North Indian, Chinese, Italian, Continental]	4.1	492
9454	6401789	tashas	[Cafe, Mediterranean]	4.1	374
9385	6113857	sketch Gallery	[British, Contemporary]	4.5	148
1837	18418247	feel ALIVE	[North Indian, American, Asian, Biryani]	4.7	69
...
1468	18408054	19 Flavours Biryani	[Mughlai, Hyderabadi]	4.1	84
2484	18233317	145 Kala Ghoda	[Fast Food, Beverages, Desserts]	4.2	1606
2292	2100784	11th Avenue Cafe Bistro	[Cafe, American, Italian, Continental]	4.1	377
751	2600031	10 Downing Street	[North Indian, Chinese]	4.0	257
351	17057397	'Ohana	[Hawaiian]	4.5	1151

1236 rows × 5 columns

```
In [28]: # Exploding 'cuisines'
dfRS = dfRS.explode('cuisines')
dfRS
```

Out[28]:

	restaurant_id	restaurant_name	cuisines	aggregate_rating	votes
9523	6000871	ukuraa Sofras	Kebab	4.4	296
9523	6000871	ukuraa Sofras	Izgara	4.4	296
3120	18222559	{Niche} - Cafe & Bar	North Indian	4.1	492
3120	18222559	{Niche} - Cafe & Bar	Chinese	4.1	492
3120	18222559	{Niche} - Cafe & Bar	Italian	4.1	492
...
2292	2100784	11th Avenue Cafe Bistro	Italian	4.1	377
2292	2100784	11th Avenue Cafe Bistro	Continental	4.1	377
751	2600031	10 Downing Street	North Indian	4.0	257
751	2600031	10 Downing Street	Chinese	4.0	257
351	17057397	'Ohana	Hawaiian	4.5	1151

2971 rows × 5 columns

```
In [29]: # Cuisines Check
dfRS['cuisines'].value_counts()
```

Out[29]:

```
cuisines
North Indian    270
Italian         237
Chinese         200
Continental     199
Cafe            177
...
Pub Food        1
Durban          1
Irish           1
Persian         1
Sunda           1
Name: count, Length: 128, dtype: int64
```

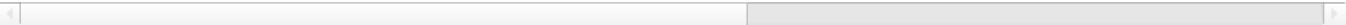
```
In [30]: # Cross Tabulate Restaurant Name and Cuisines
xTabRestoCuisines = pd.crosstab(dfRS['restaurant_name'],
                                dfRS['cuisines'])
```

```
In [31]: xTabRestoCuisines
```


Out[35]:

restaurant_name	'Ohana	10 Downing Street	11th Avenue Cafe Bistro	145 Kala Ghoda	19 Flavours Biryani	1918 Bistro & Grill	2 Dog	22nd Parallel	3 Wise Monkeys	38 Barracks	...	Zoeys Pizzeria	Zoloc...
restaurant_name													
'Ohana	1.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	...	0.0	
10 Downing Street	0.0	1.0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.200000	...	0.0	
11th Avenue Cafe Bistro	0.0	0.0	1.000000	0.0	0.0	0.0	0.166667	0.0	0.0	0.333333	...	0.0	
145 Kala Ghoda	0.0	0.0	0.000000	1.0	0.0	0.0	0.000000	0.0	0.0	0.000000	...	0.0	
19 Flavours Biryani	0.0	0.0	0.000000	0.0	1.0	0.0	0.000000	0.0	0.0	0.000000	...	0.0	
...	
feel ALIVE	0.0	0.2	0.142857	0.0	0.0	0.0	0.166667	0.0	0.0	0.600000	...	0.0	
sketch Gallery	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	...	0.0	
tashas	0.0	0.0	0.200000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	...	0.0	
{Niche} - Cafe & Bar	0.0	0.5	0.333333	0.0	0.0	0.0	0.000000	0.0	0.0	0.333333	...	0.0	
ukura Sofras	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	...	0.0	

1236 rows × 1236 columns



In [36]:

```
# Resto Names Sample
dfRS['restaurant_name'].sample(20)
```

Out[36]:

849

Kalsang Friends Corner

1468

19 Flavours Biryani

2293

Café Riverrun

588

Farzi Cafe

9203

Falafel Lovers

2434

Royal Sky

2142

Cafe Sante

515

Columbia Restaurant

1837

feel ALIVE

4

Sambo Kojin

3115

The Town House Cafe

2451

Hot Breads

820

Basil With A Twist

55

Talho Capixaba

2425

Free Spirit

9509

Butter Boutique

158

Lucianos Italian Restaurant

432

Duke's Waikiki

9169

Agent Jack's Bar

184

Zoeys Pizzeria

Name: restaurant_name, dtype: object

Final Recommendation System

In [37]:

```
# Input Initial Restaurant Name
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)

# Merge The Rating
RestoRec = pd.merge(sim,dfRS[['restaurant_name','aggregate_rating']],how='inner',on='restaurant_name')
FinalRestoRec = RestoRec.sort_values('aggregate_rating',ascending=False).drop_duplicates('restaurant_name',keep=
```

In [38]:

FinalRestoRec

Out[38]:

	restaurant_name	simScore	aggregate_rating
0	Sushi Masa	1.0	4.9
2	Nobu	1.0	4.4
4	Ichiban	1.0	4.3
8	Osaka	1.0	4.2
6	Guppy	1.0	4.1

Conclusion:

The above Data will show up to top 5 recommended restaurants with the best rating, the rating is also curated only 4 and above, so the recommendation system provide good rating objectively.

THANK YOU!!!

GitHub Link: <https://github.com/anujtiwari21?tab=repositories>

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js