```
# 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')
```

```
import warnings
warnings.filterwarnings("ignore")
```

```
# Creating a Dataframe
df = pd.read_csv("/content/Dataset .csv")
df.head()
```

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	...	Curren
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	...	Botswa Pula(
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	...	Botswa Pula(
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	...	Botswa Pula(
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	...	Botswa Pula(
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	...	Botswa Pula(

5 rows × 21 columns

```
df.columns
```

```
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')
```

```
dfRS = df[['Restaurant ID','Restaurant Name','Cuisines','Aggregate rating','Votes']]
dfRS
```

	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

```
# 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()
```

	Column	Data Type	Missing Value	Pct Missing Value	Num Unique	Unique Sample
0	Restaurant ID	int64	0	0.00	9551	[619, 305153]
1	Restaurant Name	object	0	0.00	7446	[Latitude 28, South Indian]
2	Cuisines	object	9	0.09	1825	[North Indian, European, Mediterranean, Asian,...
3	Aggregate rating	float64	0	0.00	33	[1.8, 4.2]
4	Votes	int64	0	0.00	1012	[778, 199]

```
dfRS
```

	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 ı Gurme	Turkish	4.1	788
9547	5908749	Ceviz A��ac ı	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

```
# 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'})
```

```
dfRS
```

	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 ı Gurme	Turkish	4.1	788
9547	5908749	Ceviz A��ac ı	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

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

```
np.int64(0)
```

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

```
np.int64(2105)
```

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

restaurant_name	count
Cafe Coffee Day	83
Domino's Pizza	79
Subway	63
Green Chick Chop	51
McDonald's	48
...	...
HotMess	1
Healthy Routes	1
Fuji Japanese Restaurant	1
Fa Yian	1
Unplugged Courtyard	1

7446 rows × 1 columns

dtype: int64

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

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

	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

```
# Dropping duplicaes only keeping first Value.
dfRS = dfRS.drop_duplicates('restaurant_name',keep='first')
dfRS
```

	restaurant_id	restaurant_name	cuisines	aggregate_rating	votes
9523	6000871	ukura a 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

7446 rows × 5 columns

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

restaurant_name	count
#45	1
ukura a Sofras	1
{Niche} - Cafe & Bar	1
wagamama	1
tashas	1
...	...
eastWEST - Radisson Hotel	1
feel ALIVE	1
hug!	1
iGNiTE	1
iKitchen	1

7446 rows × 1 columns

dtype: int64

```
dfRS = dfRS[dfRS['aggregate_rating']>=4.0]
dfRS
```

	restaurant_id	restaurant_name	cuisines	aggregate_rating	votes
9523	6000871	ukura a 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

1238 rows × 5 columns

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

	restaurant_id	restaurant_name	cuisines	aggregate_rating	votes
9523	6000871	ukura a 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

1238 rows × 5 columns

```
# Exploding 'cuisines'
dfRS = dfRS.explode('cuisines')
dfRS
```

	restaurant_id	restaurant_name	cuisines	aggregate_rating	votes
9523	6000871	ukura a Sofras	Kebab	4.4	296
9523	6000871	ukura a 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

2973 rows × 5 columns

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

```
                count
```

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

```
Italian         237
```

```
xTabRestoCuisines
```

```
Cafe            177
...             ...
restaurant_name   1
```

cuisines	Afghani	African	American	Andhra	Arabian	Argentine	Asian	Asian Fusion	Australian	Awadhi	...	Teriyaki	Tex-Mex
'Ohana	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
11th Avenue Cafe Bistro	0	0	1	0	0	0	0	0	0	0	...	0	0
145 Kala Ghoda	0	0	0	0	0	0	0	0	0	0	...	0	0
19 Flavours Biryani	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
sketch Gallery	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
{Niche} - Cafe & Bar	0	0	0	0	0	0	0	0	0	0	...	0	0
Sofras	0	0	0	0	0	0	0	0	0	0	...	0	0

1236 rows × 128 columns

```
# Checking on restaurant name value
xTabRestoCuisines.loc['feel ALIVE'].values
```

```
array([0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])
```

```
# Resto Names Sample
dfRS['restaurant_name'].sample(20, random_state=101)
```

	restaurant_name
3518	Asian Haus
9545	Baltazar
367	Ethos Vegan Kitchen
6701	AMPM Caf�� & Bar
2392	Paragon
120	Transmetropolitan
2900	Giani's di Hatti
1861	Downtown - Diners & Living Beer Cafe
9360	Roseleaf Bar Cafe
2438	Cappuccino Blast
1438	7 Degrees Brauhaus
784	Food Fever
579	Via Delhi
376	Tu-Do Vietnamese Restaurant
4088	Tian - Asian Cuisine Studio - ITC Maurya
153	Boise Fry Company
172	Ting's Red Lantern
6460	Locale
9513	The Sizzle
839	Sree Annapoorna

dtype: object

```
# Measure Similarity
print(jaccard_score(xTabRestoCuisines.loc["Olive Bistro"].values,
                    xTabRestoCuisines.loc["Rose Cafe"].values))
```

```
0.3333333333333333
```

```
# Create Similarity Value DF
jaccardDist = pdist(xTabRestoCuisines.values, metric='jaccard')
jaccardMatrix = squareform(jaccardDist)
jaccardSim = 1 - jaccardMatrix
dfJaccard = pd.DataFrame(
    jaccardSim,
    index=xTabRestoCuisines.index,
    columns=xTabRestoCuisines.index)

dfJaccard
```

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	Zolocru - Hot Clar Am
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	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	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	0
19 Flavours	0.0	0.0	0.000000	0.0	1.0	0.0	0.000000	0.0	0.0	0.000000	...	0.0	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	0

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

	restaurant_name
655	Brick Kitchen
9464	Marble
1419	Cafe Soul Garden
2346	Taruveda Bistro
2492	Joey's Pizza