

Restaurant Recommendation

Task :

Restaurant Recommendation

Objective :

Create a restaurant recommendation system based on user preferences.

Importing Libraries

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import jaccard_score
from scipy.spatial.distance import pdist, squareform
```

Loading Dataset

```
In [2]: dataset = pd.read_csv('Dataset .csv')
dataset.head()
```

	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	Aggregate rating	Rating color	Rating text	Votes
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenue...	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	4.8	Dark Green	Excellent	314
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	4.5	Dark Green	Excellent	591
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, Mandal...	121.056831	14.581404	Seafood, Asian, Filipino, Indian	...	Botswana Pula(P)	Yes	No	No	No	4	4.4	Green	Very Good	270
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	4.9	Dark Green	Excellent	365
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	4.8	Dark Green	Excellent	229

5 rows × 21 columns

```
In [3]: refined_dataset = dataset[['Restaurant ID','Restaurant Name','Cuisines','Price range','Aggregate rating','Votes']]
refined_dataset
```

	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	Namul Gurme	Turkish	3	4.1	788
9547	5908749	Ceviz Afi	World Cuisine, Pastry, Cafe	3	4.2	1034
9548	5915807	Huqqa	Italian, World Cuisine	4	3.7	661
9549	5916112	Af Kahve	Restaurant Cafe	4	4.0	901
9550	5927402	Walter's Coffee Roastery	Cafe	2	4.0	591

9551 rows × 6 columns

```
In [4]: #handle missing values
refined_dataset.isna().sum()
```

refined_dataset = refined_dataset.dropna()
refined_dataset.isna().sum()
refined_dataset.duplicated().sum()
refined_dataset['Restaurant Name'].duplicated().sum()
refined_dataset['Restaurant Name'].value_counts()

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 [5]: #sorting
refined_dataset = refined_dataset.sort_values(by=['Restaurant Name','Aggregate rating'],ascending=False)
refined_dataset.head()
```

	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	Namul Gurme	Turkish	3	4.1	788
9547	5908749	Ceviz Afi	World Cuisine, Pastry, Cafe	3	4.2	1034
9548	5915807	Huqqa	Italian, World Cuisine	4	3.7	661
9549	5916112	Af Kahve	Restaurant Cafe	4	4.0	901
9550	5927402	Walter's Coffee Roastery	Cafe	2	4.0	591

9551 rows × 6 columns

```
In [6]: #handle missing values
refined_dataset.isna().sum()
```

refined_dataset = refined_dataset.dropna()
refined_dataset.isna().sum()
refined_dataset.duplicated().sum()
refined_dataset['Restaurant Name'].duplicated().sum()
refined_dataset['Restaurant Name'].value_counts()

	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	Namul Gurme	Turkish	3	4.1	788
9547	5908749	Ceviz Afi	World Cuisine, Pastry, Cafe	3	4.2	1034
9548	5915807	Huqqa	Italian, World Cuisine	4	3.7	661
9549	5916112	Af Kahve	Restaurant Cafe	4	4.0	901
9550	5927402	Walter's Coffee Roastery	Cafe	2	4.0	591

9551 rows × 6 columns

```
In [7]: #removing duplicate entries
refined_dataset = refined_dataset.drop_duplicates(['Restaurant Name'],keep='first')
refined_dataset
```

	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	Namul Gurme	Turkish	3	4.1	788
9547	5908749	Ceviz Afi	World Cuisine, Pastry, Cafe	3	4.2	1034
9548	5915807	Huqqa	Italian, World Cuisine	4	3.7	661
9549	5916112	Af Kahve	Restaurant Cafe	4	4.0	901
9550	5927402	Walter's Coffee Roastery	Cafe	2	4.0	591

9551 rows × 6 columns

```
In [8]: #sorting
refined_dataset = refined_dataset.sort_values(by=['Restaurant Name','Aggregate rating'],ascending=False)
refined_dataset.head()
```

	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	Namul Gurme	Turkish	3	4.1	788
9547	5908749	Ceviz Afi	World Cuisine, Pastry, Cafe	3	4.2	1034
9548	5915807	Huqqa	Italian, World Cuisine	4	3.7	661
9549	5916112	Af Kahve	Restaurant Cafe	4	4.0	901
9550	5927402	Walter's Coffee Roastery	Cafe	2	4.0	591

9551 rows × 6 columns

```
In [9]: #removing duplicate entries
refined_dataset = refined_dataset.drop_duplicates(['Restaurant Name'],keep='first')
refined_dataset
```

	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	Namul Gurme	Turkish	3	4.1	788
9547	5908749	Ceviz Afi	World Cuisine, Pastry, Cafe	3	4.2	1034
9548	5915807	Huqqa	Italian, World Cuisine	4	3.7	661
9549	5916112	Af Kahve	Restaurant Cafe	4	4.0	901
9550	5927402	Walter's Coffee Roastery	Cafe	2	4.0	591

9551 rows × 6 columns

```
In [10]: #removing duplicate entries
refined_dataset = refined_dataset.drop_duplicates(['Restaurant Name'],keep='first')
refined_dataset
```

	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	Namul Gurme	Turkish	3	4.1	788
9547	5908749	Ceviz Afi	World Cuisine, Pastry, Cafe	3	4.2	1034
9548	5915807	Huqqa	Italian, World Cuisine	4	3.7	661
9549	5916112	Af Kahve	Restaurant Cafe	4	4.0	901
9550	5927402	Walter's Coffee Roastery	Cafe	2	4.0	591

9551 rows × 6 columns

```
In [11]: #removing duplicate entries
refined_dataset = refined_dataset.drop_duplicates(['Restaurant Name'],keep='first')
refined_dataset
```

	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	Namul Gurme	Turkish	3	4.1	788
9547	5908749	Ceviz Afi	World Cuisine, Pastry, Cafe	3	4.2	1034
9548	5915807	Huqqa	Italian, World Cuisine	4	3.7	661
9549	5916112	Af Kahve	Restaurant Cafe	4	4.0	901
9550	5927402	Walter's Coffee Roastery	Cafe	2	4.0	591

9551 rows × 6 columns

```
In [12]: #removing duplicate entries
refined_dataset = refined_dataset.drop_duplicates(['Restaurant Name'],keep='first')
refined_dataset
```

	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	Namul Gurme	Turkish	3	4.1	788
9547	5908749	Ceviz Afi	World Cuisine, Pastry, Cafe	3	4.2	1034
9548	5915807	Huqqa	Italian, World Cuisine	4	3.7	661
9549	5916112	Af Kahve	Restaurant Cafe	4	4.0	901
9550	5927402	Walter's Coffee Roastery	Cafe	2	4.0	591

9551 rows × 6 columns

```
In [13]: #removing duplicate entries
refined_dataset = refined_dataset.drop_duplicates(['Restaurant Name'],keep='first')
refined_dataset
```

	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	Namul Gurme	Turkish	3	4.1	788
9547	5908749	Ceviz Afi	World Cuisine, Pastry, Cafe	3	4.2	1034
9548	5915807	Huqqa	Italian, World Cuisine	4	3.7	661
9549	5916112	Af Kahve	Restaurant Cafe	4	4.0	901
9550	5927402	Walter's Coffee Roastery	Cafe	2	4.0	591

9551 rows × 6 columns

```
In [14]: #removing duplicate entries
refined_dataset = refined_dataset.drop_duplicates(['Restaurant Name'],keep='first')
refined_dataset
```

	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	Namul Gurme	Turkish	3	4.1	788
9547	5908749	Ceviz Afi	World Cuisine, Pastry, Cafe	3	4.2	1034
9548	5915807	Huqqa	Italian, World Cuisine	4	3.7	661
9549	5916112	Af Kahve	Restaurant Cafe	4	4.0	901
9550	5927402	Walter's Coffee Roastery	Cafe	2	4.0	591

9551 rows × 6 columns

```
In [15]: #removing duplicate entries
refined_dataset = refined_dataset.drop_duplicates(['Restaurant Name'],keep='first')
refined_dataset
```

out[11]:

Cuisines	
North Indian	270
Italian	237
Chinese	200
Continental	199
Cafe	177
...	
Pub Food	1