

Task: Customer Preference Analysis

Import Libraries

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
```

Load Dataset

```
In [2]: df=pd.read_csv('D:\Intern\Cognifyz Intern\Dataset .csv')
```

Data characteristics

```
In [3]: df.head(3)
```

Out[3]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude
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
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
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

3 rows × 21 columns



Checking Null Values

```
In [4]: df.isnull().sum()
```

```
Out[4]: Restaurant ID      0
Restaurant Name      0
Country Code      0
City      0
Address      0
Locality      0
Locality Verbose      0
Longitude      0
Latitude      0
Cuisines      9
Average Cost for two      0
Currency      0
Has Table booking      0
Has Online delivery      0
Is delivering now      0
Switch to order menu      0
Price range      0
Aggregate rating      0
Rating color      0
Rating text      0
Votes      0
dtype: int64
```

Analyze the relationship between the type of cuisine and the restaurant's rating

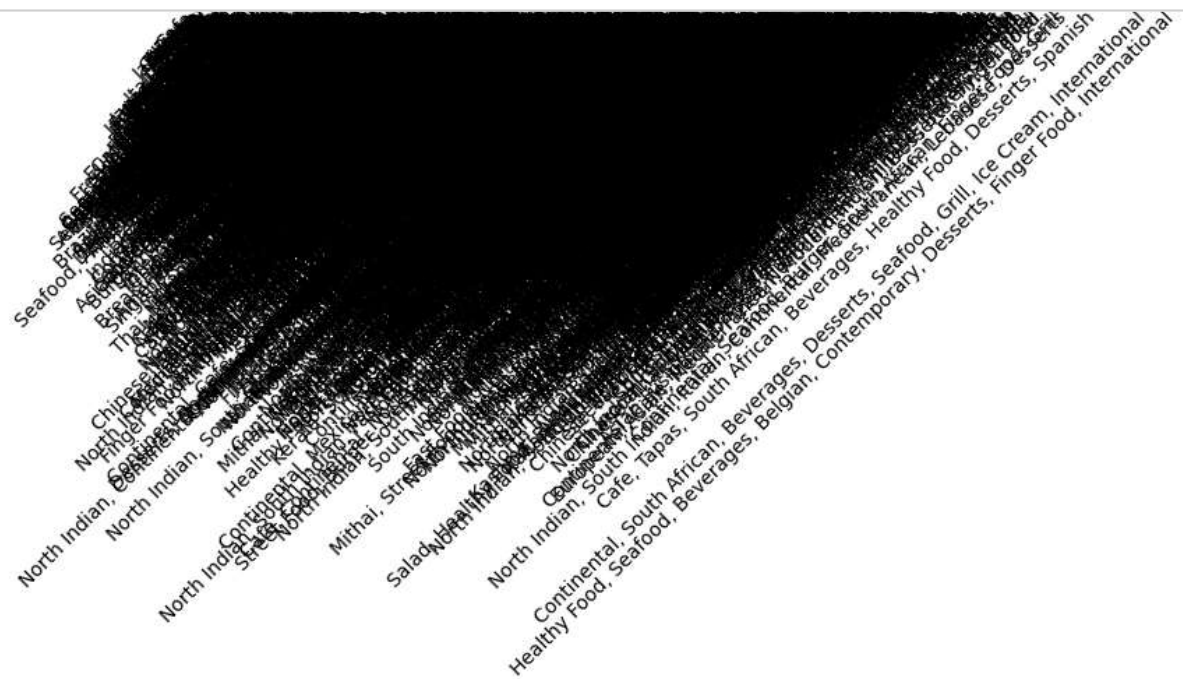
```
In [6]: print(df['Rating text'].describe())
print(df['Cuisines'].value_counts())
```

```
count      9551
unique        6
top      Average
freq      3737
Name: Rating text, dtype: object
Cuisines
North Indian      936
North Indian, Chinese      511
Chinese      354
Fast Food      354
North Indian, Mughlai      334
...
Bengali, Fast Food      1
North Indian, Rajasthani, Asian      1
Chinese, Thai, Malaysian, Indonesian      1
Bakery, Desserts, North Indian, Bengali, South Indian      1
Italian, World Cuisine      1
Name: count, Length: 1825, dtype: int64
```

```
In [12]: rating_scores = {
    "Excellent": 5,
    "Good": 4,
    "Average": 3,
    "Poor": 2
}

df["Rating score"] = df["Rating text"].map(rating_scores)
```

```
In [15]: sns.countplot(
    x = "Cuisines",
    data=df
)
plt.xticks(rotation=45)
plt.show()
```



Identify the most popular cuisines among customers based on the number of votes

```
In [19]: cuisine_votes = df.groupby('Cuisines')['Votes'].sum()

cuisine_votes_sorted = cuisine_votes.sort_values(ascending=False)

print("Top 5 Most Popular Cuisines:")
print(cuisine_votes_sorted.head(5))
```

```
Top 5 Most Popular Cuisines:
Cuisines
North Indian, Mughlai    53747
North Indian             46241
North Indian, Chinese   42012
Cafe                     30657
Chinese                  21925
Name: Votes, dtype: int64
```

Determine if there are any specific cuisines that tend to receive higher ratings

```
In [21]: rating_counts = df['Rating text'].value_counts()
print(rating_counts)
```

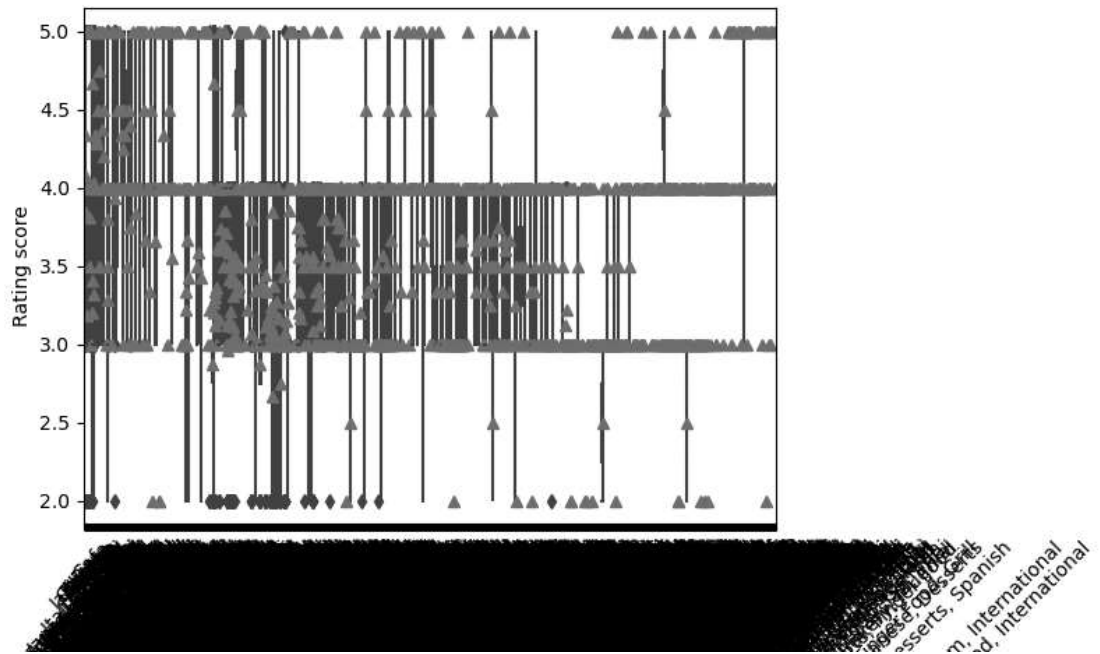
```
Rating text
Average      3737
Not rated    2148
Good         2100
Very Good    1079
Excellent     301
Poor          186
Name: count, dtype: int64
```

```
In [22]: cuisine_ratings_counts = df.groupby('Cuisines')['Rating text'].value_counts()
print(cuisine_ratings_counts)
```

```
Cuisines
Afghani
Afghani, Mughlai, Chinese
Afghani, North Indian
Afghani, North Indian, Pakistani, Arabian
..
Western, Asian, Cafe
Western, Fusion, Fast Food
World Cuisine
World Cuisine, Mexican, Italian
World Cuisine, Patisserie, Cafe
Name: count, Length: 2616, dtype: int64

Rating text
Not rated    3
Average      1
Not rated    1
Not rated    1
Not rated    1
..
Very Good    1
Average      1
Excellent    1
Very Good    1
Very Good    1
```

```
In [26]: sns.boxplot(
    x = "Cuisines",
    y = "Rating score",
    showmeans=True,
    data=df
)
plt.xticks(rotation=45)
plt.show()
```



Statistical Test (Kruskal-Wallis Test)

```
In [29]: from scipy.stats import kruskal

H, pval = kruskal(*[group for _, group in cuisine_ratings_counts.groupby('Cuisine')])

if pval < 0.05:
    print("There is a statistically significant difference in ratings across cuisines")
else:
    print("There is no statistically significant difference in ratings across cuisines")
```

There is a statistically significant difference in ratings across cuisines (p-value: 0.0019416211356509865)

In []: