PROJECT OVERVIEW
RESTAURANT DATA ANALYSIS PROJECT

Restaurant datasets encompass various attributes such as names, locations, cuisine types, ratings, review counts, price ranges, and operating hours. They are sourced from online review platforms, food delivery apps, and restaurant websites. Analyses can include descriptive summaries, sentiment analysis, predictive modeling, and geospatial mapping. These insights help restaurant owners improve services, understand customer preferences, and conduct market research. Common challenges involve ensuring data quality, maintaining privacy, and integrating diverse data sources effectively.

```
# importing necessary libraries for data analysis and visualization
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Loading the restaurant dataset into a dataframe
dataset = pd.read_csv('/content/Dataset (1).csv')
```

```
# Displaying the first few rows of the dataset to understand the structure
dataset.head()
```

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	...	Currency	Ta book
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)	
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)	
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)	
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)	
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)	

5 rows × 21 columns

```
# Displaying nubers of rows and columns in the dataset
dataset.shape
```

```
(9551, 21)
```

```
# Displaying summary of the datset
dataset.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
```

```
# Checking for missing values
dataset.isnull().sum()
```

	0
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

```
# ststical summary of the dataset
dataset.describe()
```

	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

```
# Data cleaning
# Handling missing values
dataset['Cuisines'].fillna('Unknown',inplace=True)
```

```
# Checking for missing values
dataset.isnull().sum()
```

	0
Restaurant ID	0
Restaurant Name	0
Country Code	0
City	0
Address	0
Locality	0
Locality Verbose	0
Longitude	0
Latitude	0
Cuisines	0
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

LEVEL 2

TASK 1= RESTAURANT RATING

Analyze the distribution of aggregate rating and dedtermine the most common rating range

```
Average_rating_common_range = dataset['Aggregate rating'].value_counts()
print(Average_rating_common_range)
```

```
Aggregate rating
0.0    2148
3.2     522
```

```
3.1     519
3.4     498
3.3     483
3.5     480
3.0     468
3.6     458
3.7     427
3.8     400
2.9     381
3.9     335
2.8     315
4.1     274
4.0     266
2.7     250
4.2     221
2.6     191
4.3     174
4.4     144
2.5     110
4.5      95
2.4      87
4.6      78
4.9      61
2.3      47
4.7      42
2.2      27
4.8      25
2.1      15
2.0       7
1.9       2
1.8       1
Name: count, dtype: int64
```

Calculate the average number of votes recieved by reataurant

```
average_votes = dataset['Votes'].mean()
print(average_votes)
```

```
156.909747670401
```

TASK 2 - CUISINES COMBINATION

Identigy the most common combination of cuisnes in the dataset

```
most_common_cuisines_combination = dataset['Cuisines'].value_counts()
print(most_common_cuisines_combination.head())
```

```
Cuisines
North Indian             936
North Indian, Chinese    511
Chinese                  354
Fast Food                354
North Indian, Mughlai    334
Name: count, dtype: int64
```

Determine if certain cuisines combination tend to have higher rating

```
higher_rating = dataset.groupby('Cuisines')['Aggregate rating'].count().sort_values(ascending = False)
print(higher_rating.head())
```

```
Cuisines
North Indian             936
North Indian, Chinese    511
Fast Food                354
Chinese                  354
North Indian, Mughlai    334
Name: Aggregate rating, dtype: int64
```

TASK 4 - RESTAURANT CHAINS

Identify if there are any restaurant chains present in thhe datset

```
restaurant = dataset['Restaurant Name'].value_counts()
restaurant_chains = restaurant[restaurant > 1]
print(restaurant_chains)
```

```
Restaurant Name
Cafe Coffee Day       83
Domino's Pizza        79
Subway                63
Green Chick Chop      51
McDonald's            48
                      ..
Town Hall              2
Halki Aanch            2
Snack Junction         2
Delhi Biryani Hut      2
Beliram Degchiwala     2
Name: count, Length: 734, dtype: int64
```

Analyze the rating and popularity of different restaurant chains

```
# filter the datasset to include only restaurant chains
chain_data = dataset[dataset['Restaurant Name'].isin(restaurant_chains.index)]

#Group by 'Restaurant Name' and calculate average rating and total votes
chains_stats = chain_data.groupby('Restaurant Name').agg({'Aggregate rating':'mean','Votes':'sum'}).reset_index()

# Sort by average rating and total votes for analysis
chains_stats = chains_stats.sort_values(by=['Aggregate rating','Votes'],ascending=[False,False])

# Displaying the top reataurant chains by rating and popularity
print(chains_stats.head())
```

```
               Restaurant Name  Aggregate rating  Votes
629           Talaga Sampireun             4.900   5514
8      AB's Absolute Barbecues             4.850   3151
589           Silantro Fil-Mex             4.850   1364
7    AB's - Absolute Barbecues             4.825  13400
449         Naturals Ice Cream             4.800   3094
```