

```
In [1]: '''COGNIFY TECHNOLOGIES INTERNSHIP '''
```

```
Out[1]: 'COGNIFY TECHNOLOGIES INTERNSHIP '
```

```
In [2]: '''LEVEL 1 , TASK 1'''
'''Task: Data Exploration and Preprocessing

-->Explore the dataset and identify the number
of rows and columns.

-->Check for missing values in each column and
handle them accordingly.

-->Perform data type conversion if necessary.
Analyze the distribution of the target variable
("Aggregate rating") and identify any class
imbalances.'''
```

```
Out[2]: 'LEVEL 1 , TASK 1'
```

```
In [3]: #Importing requeried libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sb
```

```
In [4]: #Reading the dataset
data = pd.read_csv("Dataset.csv")
```

```
In [5]: #Displaying the dataset
print(data)
```

	Restaurant ID	Restaurant Name	Country Code	City \
0	6317637	Le Petit Souffle	162	Makati City
1	6304287	Izakaya Kikufuji	162	Makati City
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City
3	6318506	Ooma	162	Mandaluyong City
4	6314302	Sambo Kojin	162	Mandaluyong City
...	...	...	...	...
9546	5915730	Naml\ Gurme	208	istanbul
9547	5908749	Ceviz Aac\	208	istanbul
9548	5915807	Huqqa	208	istanbul
9549	5916112	Ak Kahve	208	istanbul
9550	5927402	Walter's Coffee Roastery	208	istanbul

	Address \
0	Third Floor, Century City Mall, Kalayaan Avenu...
1	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...
2	Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal...
3	Third Floor, Mega Fashion Hall, SM Megamall, O...
4	Third Floor, Mega Atrium, SM Megamall, Ortigas...
...	...
9546	Kemanke Karamustafa Pa Mahallesi, R\ht\m ...
9547	Kouyolu Mahallesi, Muhittin st\nda Cadd...
9548	Kuru_e_me Mahallesi, Muallim Naci Caddesi, N...
9549	Kuru_e_me Mahallesi, Muallim Naci Caddesi, N...
9550	Cafea Mahallesi, Bademalt\ Sokak, No 21/B, ...

	Locality \
0	Century City Mall, Poblacion, Makati City
1	Little Tokyo, Legaspi Village, Makati City
2	Edsa Shangri-La, Ortigas, Mandaluyong City
3	SM Megamall, Ortigas, Mandaluyong City
4	SM Megamall, Ortigas, Mandaluyong City
...	...
9546	Karak_y
9547	Kouyolu
9548	Kuru_e_me
9549	Kuru_e_me
9550	Moda

	Locality Verbose	Longitude \
0	Century City Mall, Poblacion, Makati City, Mak...	121.027535
1	Little Tokyo, Legaspi Village, Makati City, Ma...	121.014101
2	Edsa Shangri-La, Ortigas, Mandaluyong City, Ma...	121.056831
3	SM Megamall, Ortigas, Mandaluyong City, Mandal...	121.056475
4	SM Megamall, Ortigas, Mandaluyong City, Mandal...	121.057508
...	...	...
9546	Karak_y, istanbul	28.977392
9547	Kouyolu, istanbul	29.041297
9548	Kuru_e_me, istanbul	29.034640
9549	Kuru_e_me, istanbul	29.036019
9550	Moda, istanbul	29.026016

	Latitude	Cuisines	...	Currency \
0	14.565443	French, Japanese, Desserts	...	Botswana Pula(P)
1	14.553708	Japanese	...	Botswana Pula(P)
2	14.581404	Seafood, Asian, Filipino, Indian	...	Botswana Pula(P)
3	14.585318	Japanese, Sushi	...	Botswana Pula(P)
4	14.584450	Japanese, Korean	...	Botswana Pula(P)
...	...	...	...	...
9546	41.022793	Turkish	...	Turkish Lira(TL)
9547	41.009847	World Cuisine, Patisserie, Cafe	...	Turkish Lira(TL)
9548	41.055817	Italian, World Cuisine	...	Turkish Lira(TL)
9549	41.057979	Restaurant Cafe	...	Turkish Lira(TL)
9550	40.984776	Cafe	...	Turkish Lira(TL)

	Has Table booking	Has Online delivery	Is delivering now	\
0	Yes	No	No	
1	Yes	No	No	
2	Yes	No	No	
3	No	No	No	
4	Yes	No	No	
...	...	...	...	
9546	No	No	No	
9547	No	No	No	
9548	No	No	No	
9549	No	No	No	
9550	No	No	No	

	Switch to order menu	Price range	Aggregate rating	Rating color	\
0	No	3	4.8	Dark Green	
1	No	3	4.5	Dark Green	
2	No	4	4.4	Green	
3	No	4	4.9	Dark Green	
4	No	4	4.8	Dark Green	
...	...	...	...	...	
9546	No	3	4.1	Green	
9547	No	3	4.2	Green	
9548	No	4	3.7	Yellow	
9549	No	4	4.0	Green	
9550	No	2	4.0	Green	

	Rating text	Votes
0	Excellent	314
1	Excellent	591
2	Very Good	270
3	Excellent	365
4	Excellent	229
...	...	...
9546	Very Good	788
9547	Very Good	1034
9548	Good	661
9549	Very Good	901
9550	Very Good	591

[9551 rows x 21 columns]

In [6]: `data.head()`

Out[6]:

	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...	121.0275
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...	121.0141
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...	121.0568
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...	121.0564
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...	121.0575

5 rows × 21 columns



In [7]: data.tail()

Out[7]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Local Verbc
9546	5915730	Naml Gurme	208	istanbul	Kemanke Karamustafa Pa a Mahallesi, Rihlm ...	Karak_y	Karak stanl
9547	5908749	Ceviz Aacl	208	istanbul	Ko uyolu Mahallesi, Muhittin st_nda Cadd...	Ko uyolu	Ko uyc stanl
9548	5915807	Huqqa	208	istanbul	Kuru_e me Mahallesi, Muallim Naci Caddesi, N...	Kuru_e me	Kuru_e n stanl
9549	5916112	A k Kahve	208	istanbul	Kuru_e me Mahallesi, Muallim Naci Caddesi, N...	Kuru_e me	Kuru_e n stanl
9550	5927402	Walter's Coffee Roastery	208	istanbul	Cafea a Mahallesi, Bademalt Sokak, No 21/B, ...	Moda	Mo stanl

5 rows × 21 columns

```
In [8]: """DATASET EXPLORATION AND PREPROCESSING"""
# Get the number of rows and columns
num_rows, num_columns = data.shape

In [9]: print(f"Number of rows: {num_rows}")
print(f"Number of columns: {num_columns}")

Number of rows: 9551
Number of columns: 21

In [10]: #Identifying the missed values in dataset
values_missed = data.isnull().sum()
values_missed
```

```
Out[10]: 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
```

```
In [11]: #Removing the null valued data
data.dropna(inplace=True)
```

```
In [13]: # Get the data types of all columns
data_types = data.dtypes

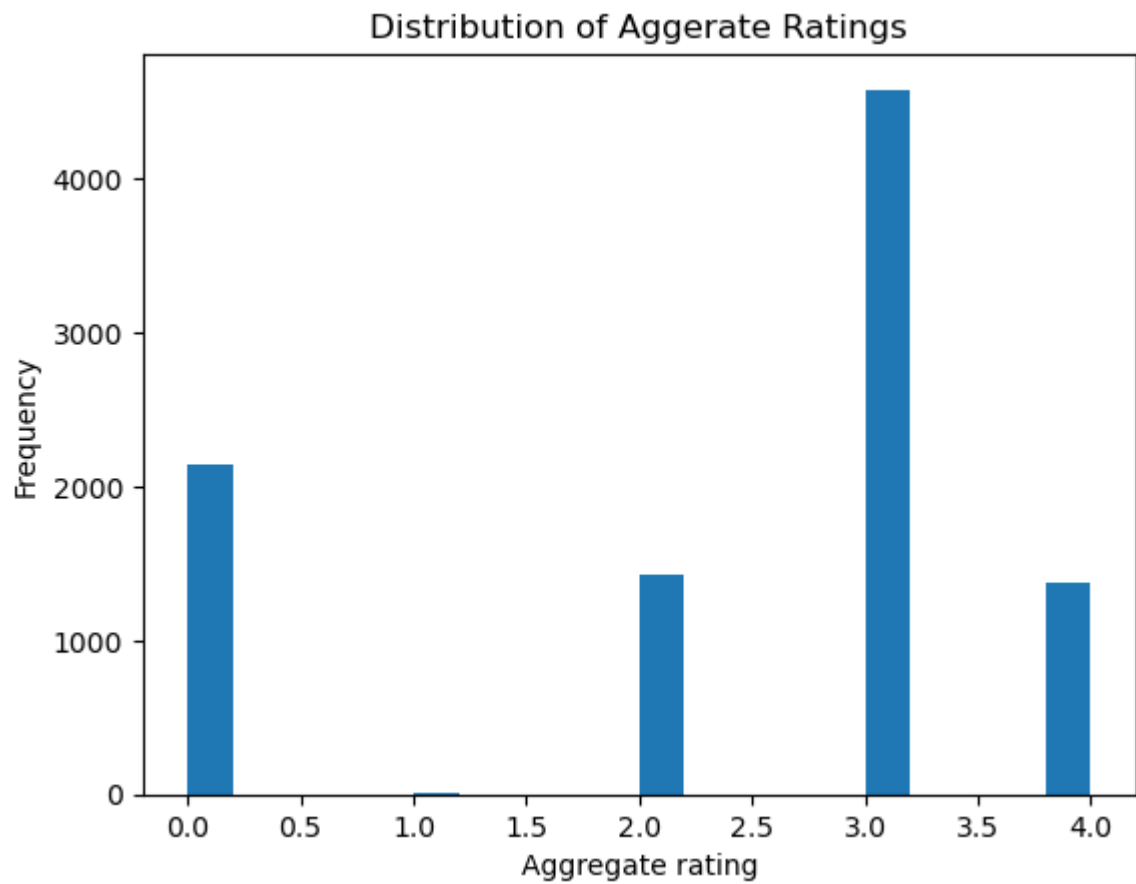
# Print the data types
print("Data types of columns:")
print(data_types)
```

```
Data types of columns:
Restaurant ID      int64
Restaurant Name    object
Country Code      int64
City              object
Address           object
Locality          object
Locality Verbose  object
Longitude         float64
Latitude          float64
Cuisines          object
Average Cost for two int64
Currency          object
Has Table booking  object
Has Online delivery object
Is delivering now  object
Switch to order menu object
Price range       int64
Aggregate rating  float64
Rating color      object
Rating text       object
Votes            int64
dtype: object
```

```
In [14]: #Converting datatype
data['Aggregate rating'] = data['Aggregate rating'].astype('int64')
```

```
In [15]: #By using matplotlib
plt.hist(data["Aggregate rating"],bins=20)
plt.xlabel("Aggregate rating")
```

```
plt.ylabel("Frequency")  
plt.title("Distribution of Aggerate Ratings")  
plt.show()
```



In [ ]:

In [ ]:

In [ ]:

In [ ]: