Task: Feature Engineering

Import Libraries

In [2]:
```python
import pandas as pd
```

Load Dataset

In [3]:
```python
df=pd.read_csv('D:\Intern\Cognifyz Intern\Dataset .csv')
```

Data Characteristics

In [4]: `df.head()`

Out[4]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitu
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 [5]: `df.tail()`

Out[5]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Loc Ver
9546	5915730	Naml\ Gurme	208	��stanbul	Kemanke�� Karamustafa Pa��a Mahallesi, R\ht\m ...	Karak�_y	Karak ��st
9547	5908749	Ceviz A��ac\	208	��stanbul	Ko��uyolu Mahallesi, Muhittin ��st�_nda�� Cadd...	Ko��uyolu	Ko��u ��st
9548	5915807	Huqqa	208	��stanbul	Kuru�_e��me Mahallesi, Muallim Naci Caddesi, N...	Kuru�_e��me	Kuru�_e� ��st
9549	5916112	A���k Kahve	208	��stanbul	Kuru�_e��me Mahallesi, Muallim Naci Caddesi, N...	Kuru�_e��me	Kuru�_e� ��st
9550	5927402	Walter's Coffee Roastery	208	��stanbul	Cafea��a Mahallesi, Bademalt\ Sokak, No 21/B, ...	Moda	M ��st

5 rows × 21 columns

◀ ▬▬▬▬▬▬▬▬ ▶

In [6]: `df.describe()`

Out[6]:

	Restaurant ID	Country Code	Longitude	Latitude	Average Cost for two	Price range	Aggr
count	9.551000e+03	9551.000000	9551.000000	9551.000000	9551.000000	9551.000000	9551.0
mean	9.051128e+06	18.365616	64.126574	25.854381	1199.210763	1.804837	2.6
std	8.791521e+06	56.750546	41.467058	11.007935	16121.183073	0.905609	1.5
min	5.300000e+01	1.000000	-157.948486	-41.330428	0.000000	1.000000	0.0
25%	3.019625e+05	1.000000	77.081343	28.478713	250.000000	1.000000	2.5
50%	6.004089e+06	1.000000	77.191964	28.570469	400.000000	2.000000	3.2
75%	1.835229e+07	1.000000	77.282006	28.642758	700.000000	2.000000	3.7
max	1.850065e+07	216.000000	174.832089	55.976980	800000.000000	4.000000	4.9

◀ ▬▬▬▬▬▬▬▬▬▬▬▬ ▶

Extract Feature form Existing columns (Restaurant name and Address)

In [8]:
```python
df['Length of name']=df['Restaurant Name'].str.len()
df['Lenght of address']=df['Address'].str.len()
```

In [9]:
```python
print(df['Length of name'])
print(df['Lenght of address'])
```

```
0        16
1        16
2        22
3         4
4        11
         ..
9546     11
9547     12
9548      5
9549     11
9550     24
Name: Length of name, Length: 9551, dtype: int64
0        71
1        67
2        56
3        70
4        64
        ...
9546    103
0547     77
```

Create new features like "Has Table Booking" or "Has Online Delivery" by encoding categorical variables

In [17]:
```python
booking_column = 'Has Table booking'
delivery_column = 'Has Online delivery'


booking_dummies = pd.get_dummies(df[booking_column], prefix='Booking')
df = pd.concat([df, booking_dummies], axis=1)

delivery_dummies = pd.get_dummies(df[delivery_column], prefix='Delivery')
df = pd.concat([df, delivery_dummies], axis=1)
```

In [18]:
```python
print(booking_dummies)
```

```
      Booking_No  Booking_Yes
0          False         True
1          False         True
2          False         True
3           True        False
4          False         True
...          ...          ...
9546        True        False
9547        True        False
9548        True        False
9549        True        False
9550        True        False

[9551 rows x 2 columns]
```

In [19]:
```python
print(delivery_dummies)
```

```
      Delivery_No  Delivery_Yes
0           True         False
1           True         False
2           True         False
3           True         False
4           True         False
...          ...           ...
9546        True         False
9547        True         False
9548        True         False
9549        True         False
9550        True         False

[9551 rows x 2 columns]
```