1. IMPORT LIBRARIES

In [ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

2. Load The Dataset

In [6]:
```
df=pd.read_excel('/Swiggy_Data.xlsx')
print(df.info())  #
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 197430 entries, 0 to 197429
Data columns (total 10 columns):
 #   Column           Non-Null Count   Dtype
---  ------           --------------   -----
 0   State            197430 non-null  object
 1   City             197430 non-null  object
 2   Order Date       197430 non-null  datetime64[ns]
 3   Restaurant Name  197430 non-null  object
 4   Location         197430 non-null  object
 5   Category         197430 non-null  object
 6   Dish Name        197430 non-null  object
 7   Price (INR)      197430 non-null  float64
 8   Rating           197430 non-null  float64
 9   Rating Count     197430 non-null  int64
dtypes: datetime64[ns](1), float64(2), int64(1), object(6)
memory usage: 15.1+ MB
None
```

3. Understand the Data (Basic EDA)

In [11]:
```
print(df.shape)  # show number of rows and columns
```

```
(197430, 10)
```

In [13]:
```
print(df.info())  #shows data types, null values, memory usage
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 197430 entries, 0 to 197429
Data columns (total 10 columns):
 #   Column           Non-Null Count   Dtype
---  ------           --------------   -----
 0   State            197430 non-null  object
 1   City             197430 non-null  object
 2   Order Date       197430 non-null  datetime64[ns]
 3   Restaurant Name  197430 non-null  object
 4   Location         197430 non-null  object
 5   Category         197430 non-null  object
 6   Dish Name        197430 non-null  object
 7   Price (INR)      197430 non-null  float64
 8   Rating           197430 non-null  float64
 9   Rating Count     197430 non-null  int64
dtypes: datetime64[ns](1), float64(2), int64(1), object(6)
memory usage: 15.1+ MB
None
```

In [12]:
```
print(df.describe)   ## Summary statistics → only for numeric columns
```

```
<bound method NDFrame.describe of             State       City Order Date           Restaurant Name  \
0       Karnataka  Bengaluru 2025-06-29  Anand Sweets & Savouries
1       Karnataka  Bengaluru 2025-04-03     Srinidhi Sagar Deluxe
2       Karnataka  Bengaluru 2025-01-15     Srinidhi Sagar Deluxe
3       Karnataka  Bengaluru 2025-04-17     Srinidhi Sagar Deluxe
4       Karnataka  Bengaluru 2025-03-13     Srinidhi Sagar Deluxe
...           ...        ...        ...                      ...
197425     Sikkim    Gangtok 2025-01-25            Mama's Kitchen
197426     Sikkim    Gangtok 2025-07-02            Mama's Kitchen
197427     Sikkim    Gangtok 2025-03-25            Mama's Kitchen
197428     Sikkim    Gangtok 2025-03-26            Mama's Kitchen
197429     Sikkim    Gangtok 2025-03-27            Mama's Kitchen

                   Location     Category  \
0       Rajarajeshwari Nagar        Snack
1                    Kengeri  Recommended
2                    Kengeri  Recommended
3                    Kengeri  Recommended
4                    Kengeri  Recommended
...                      ...          ...
197425               Gangtok        Momos
197426               Gangtok        Momos
197427               Gangtok        Momos
197428               Gangtok        Momos
197429               Gangtok        Momos

                                          Dish Name  Price (INR)  \
0                                 Butter Murukku-200gm        133.9
1                                           Badam Milk         52.0
2                                      Chow Chow Bath        117.0
3                                         Kesari Bath         65.0
4                                          Mix Raitha        130.0
...                                               ...          ...
197425  Soya cheese chilli momo                   ...        112.0
197426  Kurkure momo fried                        ...        140.0
197427                               Chilli cheese momo        126.0
197428                                  Veg Momos (8 Pc)         85.0
197429                                         Soya Momo        100.0

        Rating  Rating Count
0          4.0             0
```

```
1          4.5            25
2          4.7            48
3          4.6            65
4          4.0             0
...        ...           ...
197425     4.4             0
197426     4.4             0
197427     4.4             0
197428     4.4             0
197429     4.4             0

[197430 rows x 10 columns]>
```

In [14]:
```
print(df.columns)   # check columns
```

```
Index(['State', 'City', 'Order Date', 'Restaurant Name', 'Location',
       'Category', 'Dish Name', 'Price (INR)', 'Rating', 'Rating Count'],
      dtype='object')
```

4. Data Cleaning

In [15]:
```
#check missing values
print(df.isnull().sum())
```

```
State              0
City               0
Order Date         0
Restaurant Name    0
Location           0
Category           0
Dish Name          0
Price (INR)        0
Rating             0
Rating Count       0
dtype: int64
```

In [16]:
```
df.dropna()  # drop missing rows
```

Out[16]:

	State	City	Order Date	Restaurant Name	Location	Category	Dish Name	Price (INR)	Rating	Rating Count
0	Karnataka	Bengaluru	2025-06-29	Anand Sweets & Savouries	Rajarajeshwari Nagar	Snack	Butter Murukku-200gm	133.9	4.0	0
1	Karnataka	Bengaluru	2025-04-03	Srinidhi Sagar Deluxe	Kengeri	Recommended	Badam Milk	52.0	4.5	25
2	Karnataka	Bengaluru	2025-01-15	Srinidhi Sagar Deluxe	Kengeri	Recommended	Chow Chow Bath	117.0	4.7	48
3	Karnataka	Bengaluru	2025-04-17	Srinidhi Sagar Deluxe	Kengeri	Recommended	Kesari Bath	65.0	4.6	65
4	Karnataka	Bengaluru	2025-03-13	Srinidhi Sagar Deluxe	Kengeri	Recommended	Mix Raitha	130.0	4.0	0
...	...	...	...	...	...	...	...	...	...	...
197425	Sikkim	Gangtok	2025-01-25	Mama's Kitchen	Gangtok	Momos	Soya cheese chilli momo ...	112.0	4.4	0
197426	Sikkim	Gangtok	2025-07-02	Mama's Kitchen	Gangtok	Momos	Kurkure momo fried ...	140.0	4.4	0
197427	Sikkim	Gangtok	2025-03-25	Mama's Kitchen	Gangtok	Momos	Chilli cheese momo	126.0	4.4	0
197428	Sikkim	Gangtok	2025-03-26	Mama's Kitchen	Gangtok	Momos	Veg Momos (8 Pc)	85.0	4.4	0
197429	Sikkim	Gangtok	2025-03-27	Mama's Kitchen	Gangtok	Momos	Soya Momo	100.0	4.4	0

197430 rows × 10 columns

4.2 Fill missing values with(default,mean,median)

```
df['column_name'] = df['column_name'].fillna(df['column_name'].median())
```

4.3 Remove Duplicates (if required)

```
df = df.drop_duplicates()
```

4.4 Fix Incorrect Data Types

In [29]:
```
# Convert to datetime
df['Order Date'] = pd.to_datetime(df['Order Date'])
print(df['Order Date'])
```

```
0        2025-06-29
1        2025-04-03
2        2025-01-15
3        2025-04-17
4        2025-03-13
            ...
197425   2025-01-25
197426   2025-07-02
197427   2025-03-25
197428   2025-03-26
197429   2025-03-27
Name: Order Date, Length: 197430, dtype: datetime64[ns]
```

4.5 Handle Outliers(if required)

Using IQR method

```
Q1 = df['numeric_column'].quantile(0.25)

Q3 = df['numeric_column'].quantile(0.75)
```

```
IQR = Q3 - Q1

lower = Q1 - 1.5 * IQR
upper = Q3 + 1.5 * IQR
```

Filter outliers

```
df = df[(df['numeric_column'] >= lower) & (df['numeric_column'] <= upper)]
```

Separate Numerical & Categorical Columns

In [22]:
```
# Numerical columns
num_cols = df.select_dtypes(include=['int64', 'float64']).columns
print("Numerical Columns:")
print(num_cols)

print()

# Categorical columns
cat_cols = df.select_dtypes(include=['object', 'category']).columns
print("Categorical Columns:")
print(cat_cols)
```

```
Numerical Columns:
Index(['Price (INR)', 'Rating', 'Rating Count'], dtype='object')

Categorical Columns:
Index(['State', 'City', 'Restaurant Name', 'Location', 'Category',
       'Dish Name'],
      dtype='object')
```

Data Manipulation

In [30]:
```
# make the column name in snake case
df.columns = df.columns.str.lower().str.replace(' ', '_')
print(df.columns)
```

```
Index(['state', 'city', 'order_date', 'restaurant_name', 'location',
       'category', 'dish_name', 'price_(inr)', 'rating', 'rating_count'],
      dtype='object')
```

In [31]:
```
# rename specific columns
df.rename(columns = {'price_(inr)': 'price'}, inplace = True)
print(df.columns)
```

```
Index(['state', 'city', 'order_date', 'restaurant_name', 'location',
       'category', 'dish_name', 'price', 'rating', 'rating_count'],
      dtype='object')
```

Extract (Year, Month, Day, Week, DayOfWeek, Hour)

In [32]:
```
# Time columns
df['order_year'] = df['order_date'].dt.year
df['order_month'] = df['order_date'].dt.to_period('M').astype(str)
df['order_week'] = df['order_date'].dt.isocalendar().week
df['day_of_week'] = df['order_date'].dt.day_name()
df['hour'] = df['order_date'].dt.hour
print(df.columns)
```

```
Index(['state', 'city', 'order_date', 'restaurant_name', 'location',
       'category', 'dish_name', 'price', 'rating', 'rating_count',
       'order_year', 'order_month', 'order_week', 'day_of_week', 'hour'],
      dtype='object')
```

In [39]:
```
# Weekend flag
df['weekend'] = df['day_of_week'].isin(['Saturday', 'Sunday'])
```

In [40]:
```
df.columns
```

Out[40]:
```
Index(['state', 'city', 'order_date', 'restaurant_name', 'location',
       'category', 'dish_name', 'price', 'rating', 'rating_count',
       'order_year', 'order_month', 'order_week', 'day_of_week', 'hour',
       'weekend'],
      dtype='object')
```

Saved cleaned csv file

In [42]:
```
df.to_csv('cleaned_swiggy_data.csv', index=False)
print('DataFrame saved as cleaned_swiggy_data.csv')
```

```
DataFrame saved as cleaned_swiggy_data.csv
```

In [44]:
```
df
```

Out[44]:

	state	city	order_date	restaurant_name	location	category	dish_name	price	rating	rating_count	order_ye
0	Karnataka	Bengaluru	2025-06-29	Anand Sweets & Savouries	Rajarajeshwari Nagar	Snack	Butter Murukku-200gm	133.9	4.0	0	20
1	Karnataka	Bengaluru	2025-04-03	Srinidhi Sagar Deluxe	Kengeri	Recommended	Badam Milk	52.0	4.5	25	20
2	Karnataka	Bengaluru	2025-01-15	Srinidhi Sagar Deluxe	Kengeri	Recommended	Chow Chow Bath	117.0	4.7	48	20
3	Karnataka	Bengaluru	2025-04-17	Srinidhi Sagar Deluxe	Kengeri	Recommended	Kesari Bath	65.0	4.6	65	20
4	Karnataka	Bengaluru	2025-03-13	Srinidhi Sagar Deluxe	Kengeri	Recommended	Mix Raitha	130.0	4.0	0	20
...	...	...	...	...	...	...	...	...	...	...	
197425	Sikkim	Gangtok	2025-01-25	Mama's Kitchen	Gangtok	Momos	Soya cheese chilli momo ...	112.0	4.4	0	20
197426	Sikkim	Gangtok	2025-07-02	Mama's Kitchen	Gangtok	Momos	Kurkure momo fried ...	140.0	4.4	0	20
197427	Sikkim	Gangtok	2025-03-25	Mama's Kitchen	Gangtok	Momos	Chilli cheese momo	126.0	4.4	0	20
197428	Sikkim	Gangtok	2025-03-26	Mama's Kitchen	Gangtok	Momos	Veg Momos (8 Pc)	85.0	4.4	0	20
197429	Sikkim	Gangtok	2025-03-27	Mama's Kitchen	Gangtok	Momos	Soya Momo	100.0	4.4	0	20

197430 rows × 16 columns

In [ ]: