

----- Created By
Ankit Kumar -----

Exploratory Data Analysis on Zomato Sales

In [2]: `# pip install ipython`

In [3]: `from IPython.display import Image`

In [4]: `Image('https://restaurantindia.s3.ap-south-1.amazonaws.com/s3fs-public/2023-02/z`

Out[4]:



In [5]: `import IPython
IPython.display.Image('https://www.equentis.com/blog/wp-content/uploads/2024/06/`

Out[5]:



Project Objective:

-To Analyze Zomato Sales Data to find Hidden insights for different Dept and managers for better decision making

Python Exploratory Data Analysis (EDA) on Zomato Sales Data

Project Objective

To analyze **Zomato Sales Data** using **Python-based Exploratory Data Analysis (EDA)** techniques in order to uncover **hidden insights, trends, and patterns** that can support **different departments and managers** in making **data-driven and informed business decisions**.

Project Overview

This project focuses on performing an in-depth **Exploratory Data Analysis (EDA)** using Python to understand the underlying structure, relationships, and trends within Zomato's sales data.

The analysis aims to provide **valuable insights** that help enhance **sales performance, customer satisfaction, marketing effectiveness, and operational efficiency**.

By applying **data cleaning, transformation, statistical analysis, and visualization techniques**, the project translates **raw data into meaningful business intelligence** that supports **strategic decision-making** across various departments.

Key Objectives

Sales Performance Analysis

- Examine total and average sales across different time periods.
- Identify top-performing restaurants, cuisines, and regions.
- Discover seasonal or time-based sales trends.

Customer Behavior Analysis

- Study ordering frequency, average order values, and repeat customers.
- Identify top customer segments and preferences.
- Analyze peak ordering times and delivery demand patterns.

Operational Insights

- Evaluate delivery performance, order completion, and cancellation rates.
- Identify bottlenecks or inefficiencies in delivery operations.
- Suggest improvements to enhance order fulfillment and reduce delays.

Marketing & Promotions Analysis

- Measure the impact of discounts, offers, and campaigns on sales.
- Identify which promotions attract the most customers.
- Provide data-backed recommendations for future marketing strategies.

Financial Insights & Forecasting

- Analyze revenue distribution across restaurants and regions.
- Detect profit trends and revenue leakage points.
- Forecast future sales based on historical data patterns.

Tools and Technologies Used

Programming Language

- Python

Libraries and Packages

- **Data Handling:** pandas, numpy
- **Data Visualization:** matplotlib, seaborn, plotly
- **Statistical Analysis:** scipy, statsmodels
- **Data Cleaning & Preprocessing:** pandas, missingno, regex

Development Environment

- Jupyter Notebook / Google Colab / VS Code

EDA Methodology

Data Import & Understanding

- Load Zomato Sales dataset (CSV/Excel/SQL).
- Explore dataset structure, features, and metadata.

Data Cleaning & Preprocessing

- Handle missing values, duplicates, and inconsistent data types.
- Detect and treat outliers for accurate analysis.

Data Exploration & Visualization

- Analyze distributions, correlations, and variable relationships.
- Create visualizations (bar charts, histograms, heatmaps, scatter plots, etc.) to identify key insights.

Feature Analysis & Trend Identification

- Perform grouping, aggregation, and comparative analysis.
- Detect hidden trends and business patterns.

Insights Generation & Recommendations

- Summarize key findings relevant to each department.
- Provide actionable insights for improving operations, sales, and marketing.

Expected Outcomes

- Comprehensive understanding of **sales patterns**, **customer preferences**, and **operational challenges**.
- **Actionable insights** for optimizing sales, improving customer retention, and enhancing service quality.
- **Data-driven recommendations** to support departmental and managerial decision-making.
- **Interactive visualizations and dashboards** for clear communication of insights.
- A **reproducible Python-based analysis workflow** for future data updates.

Beneficiary Departments

- **Sales Department:** Performance monitoring and growth optimization.
- **Marketing Department:** Campaign evaluation and target segmentation.
- **Operations Department:** Delivery and logistics improvement.

- **Finance Department:** Revenue and profitability tracking.
- **Executive Management:** Strategic planning and business forecasting.

Step : 1 Load Important Modules

```
In [6]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os
import random
import warnings
warnings.filterwarnings('ignore')
print('Module Loaded Successfully!!')
```

Module Loaded Successfully!!

Step 2: Load dataset

```
In [7]: file_path = r"C:\Users\Ankit\Downloads\zomato_data.xlsx"
df=pd.read_excel(file_path)
print('done')
```

done

Step 3: EDA

```
In [8]: # shape
df.shape
```

Out[8]: (9551, 29)

```
In [9]: r, c = df.shape
print(f'''This Dataset Contains
Row: {r}
Columns : {c}''')
```

This Dataset Contains
Row: 9551
Columns : 29

```
In [10]: # size
df.size
```

Out[10]: 276979

```
In [11]: # head to check top 5 rows sample
df.head()
```

Out[11]:

	RestaurantID	RestaurantName	CountryCode	City	Address	Locality	Locality
0	18435314	Punjabi's Veg Grill	1	New Delhi	13/288 , 14 Block Gurudwra, Geeta Colony, New ...	Geeta Colony	Geeta N
1	18378015	Tasty Tandoor	1	New Delhi	726/2, Jheel Khuranja, Geeta Colony, New Delhi	Geeta Colony	Geeta N
2	18424905	Taste of Spice	1	New Delhi	C-222, Lajpat Nagar 1, New Delhi	Lajpat Nagar 1	Lajpat N
3	18180072	Kolkata Bengali Dhaba	1	New Delhi	Gali 7, Mahipalpur, New Delhi	Mahipalpur	Ma N
4	18415377	Sunil Punjabi Dhaba	1	New Delhi	Main Vasant Kunj Road, Mahipalpur, New Delhi	Mahipalpur	Ma N

5 rows × 29 columns



In [12]: `# tail to check top 5 rows sample`
`df.tail()`

Out[12]:

	RestaurantID	RestaurantName	CountryCode	City	Address	Localit
9546	7101042	The Hangar	148	Wellington City	171-177 Willis Street, Te Aro, Wellington City	Te Aro
9547	7100502	Fidel's	148	Wellington City	234 Cuba Street, Te Aro, Wellington City	Te Aro
9548	6900992	Mughal E Azam	215	Birmingham	Stratford Road, Sparkhill, Birmingham B11 4DA	Sparkhill
9549	5800590	The Commons	191	Colombo	39 A, Flower Road, Cinnamon Gardens, Colombo 07	Cinnamon Gardens, Colombo 07
9550	6001980	Timboo Cafe	208	Ankara	Armada AVM, Kat -1, Eskişehir Yolu, No 6, Yenimahalle	Armada AVM, Yenimahalle

5 rows × 29 columns



```
In [13]: # all columns
df.columns
```

```
Out[13]: Index(['RestaurantID', 'RestaurantName', 'CountryCode', 'City', 'Address',
               'Locality', 'LocalityVerbose', 'Longitude', 'Latitude', 'Cuisines',
               'Currency', 'Has_Table_booking', 'Has_Online_delivery',
               'Is_delivering_now', 'Switch_to_order_menu', 'Price_range', 'Votes',
               'Average_Cost_for_two', 'Rating', 'Datekey_Opening', 'Unnamed: 20',
               'Cuisines 1', 'Cuisines 2', 'Cuisines 3', 'Cuisines 4', 'Cuisines 5',
               'Cuisines 6', 'Cuisines 7', 'Cuisines 8'],
              dtype='object')
```

```
In [14]: # df info
df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9551 entries, 0 to 9550
Data columns (total 29 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   RestaurantID          9551 non-null   int64
 1   RestaurantName        9551 non-null   object
 2   CountryCode          9551 non-null   int64
 3   City                  9551 non-null   object
 4   Address               9551 non-null   object
 5   Locality              9551 non-null   object
 6   LocalityVerbose       9551 non-null   object
 7   Longitude             9551 non-null   float64
 8   Latitude              9551 non-null   float64
 9   Cuisines              9542 non-null   object
10   Currency              9551 non-null   object
11   Has_Table_booking     9551 non-null   object
12   Has_Online_delivery   9551 non-null   object
13   Is_delivering_now     9551 non-null   object
14   Switch_to_order_menu  9551 non-null   object
15   Price_range           9551 non-null   int64
16   Votes                 9551 non-null   int64
17   Average_Cost_for_two  9551 non-null   int64
18   Rating                9551 non-null   float64
19   Datekey_Opening       9551 non-null   object
20   Unnamed: 20           0 non-null      float64
21   Cuisines 1            9542 non-null   object
22   Cuisines 2            6148 non-null   object
23   Cuisines 3            2704 non-null   object
24   Cuisines 4            864 non-null    object
25   Cuisines 5            280 non-null    object
26   Cuisines 6            116 non-null    object
27   Cuisines 7            42 non-null     object
28   Cuisines 8            14 non-null     object
dtypes: float64(4), int64(5), object(20)
memory usage: 2.1+ MB

```

```

In [15]: # checking missing values
df.isna().sum().sort_values(ascending=False)

```



```
Out[15]: Unnamed: 20      9551
         Cuisines 8      9537
         Cuisines 7      9509
         Cuisines 6      9435
         Cuisines 5      9271
         Cuisines 4      8687
         Cuisines 3      6847
         Cuisines 2      3403
         Cuisines 1         9
         Cuisines         9
         Price_range       0
         Datekey_Opening   0
         Rating            0
         Average_Cost_for_two 0
         Votes             0
         RestaurantID      0
         RestaurantName    0
         Is_delivering_now  0
         Has_Online_delivery 0
         Has_Table_booking  0
         Currency          0
         Latitude          0
         Longitude         0
         LocalityVerbose   0
         Locality          0
         Address           0
         City              0
         CountryCode       0
         Switch_to_order_menu 0
         dtype: int64
```

```
In [16]: # missing values in %
         (df.isna().mean().sort_values(ascending=False))*100
```

```
Out[16]: Unnamed: 20      100.000000
          Cuisines 8      99.853418
          Cuisines 7      99.560255
          Cuisines 6      98.785467
          Cuisines 5      97.068370
          Cuisines 4      90.953827
          Cuisines 3      71.688828
          Cuisines 2      35.629777
          Cuisines 1       0.094231
          Cuisines       0.094231
          Price_range      0.000000
          Datekey_Opening  0.000000
          Rating           0.000000
          Average_Cost_for_two 0.000000
          Votes           0.000000
          RestaurantID     0.000000
          RestaurantName   0.000000
          Is_delivering_now 0.000000
          Has_Online_delivery 0.000000
          Has_Table_booking 0.000000
          Currency         0.000000
          Latitude         0.000000
          Longitude        0.000000
          LocalityVerbose  0.000000
          Locality         0.000000
          Address          0.000000
          City             0.000000
          CountryCode      0.000000
          Switch_to_order_menu 0.000000
          dtype: float64
```

```
In [17]: temp_df= (df.isna().mean().sort_values(ascending= False))>=0.3
          temp_df= temp_df.reset_index()
```

```
In [18]: unwanted_cols = list(temp_df[temp_df[0] == True]['index'].values)
```

```
In [19]: unwanted_cols
```

```
Out[19]: ['Unnamed: 20',
          'Cuisines 8',
          'Cuisines 7',
          'Cuisines 6',
          'Cuisines 5',
          'Cuisines 4',
          'Cuisines 3',
          'Cuisines 2']
```

```
In [20]: temp_df[temp_df[0] == True]['index'].values
```

```
Out[20]: array(['Unnamed: 20', 'Cuisines 8', 'Cuisines 7', 'Cuisines 6',
                'Cuisines 5', 'Cuisines 4', 'Cuisines 3', 'Cuisines 2'],
               dtype=object)
```

```
In [21]: unwanted_cols = list(temp_df[temp_df[0] == True]['index'].values)
```

```
In [22]: unwanted_cols
```

```
Out[22]: ['Unnamed: 20',  
          'Cuisines 8',  
          'Cuisines 7',  
          'Cuisines 6',  
          'Cuisines 5',  
          'Cuisines 4',  
          'Cuisines 3',  
          'Cuisines 2']
```

```
In [23]: # drop unwanted_cols  
df.drop(unwanted_cols,axis = 1)
```

Out[23]:

	RestaurantID	RestaurantName	CountryCode	City	Address	Localit
0	18435314	Punjabi's Veg Grill	1	New Delhi	13/288 , 14 Block Gurudwra, Geeta Colony, New ...	Gee Color
1	18378015	Tasty Tandoor	1	New Delhi	726/2, Jheel Khuranja, Geeta Colony, New Delhi	Gee Color
2	18424905	Taste of Spice	1	New Delhi	C-222, Lajpat Nagar 1, New Delhi	Lajp Nagar
3	18180072	Kolkata Bengali Dhaba	1	New Delhi	Gali 7, Mahipalpur, New Delhi	Mahipalpur
4	18415377	Sunil Punjabi Dhaba	1	New Delhi	Main Vasant Kunj Road, Mahipalpur, New Delhi	Mahipalpur
...
9546	7101042	The Hangar	148	Wellington City	171-177 Willis Street, Te Aro, Wellington City	Te Aro
9547	7100502	Fidel's	148	Wellington City	234 Cuba Street, Te Aro, Wellington City	Te Aro
9548	6900992	Mughal E Azam	215	Birmingham	Stratford Road, Sparkhill, Birmingham B11 4DA	Sparkhill
9549	5800590	The Commons	191	Colombo	39 A, Flower Road, Cinnamon Gardens, Colombo 07	Cinnamon Gardens Colombo C
9550	6001980	Timboo Cafe	208	Ankara	Armada AVM, Kat	Armada AVM

RestaurantID	RestaurantName	CountryCode	City	Address	Localities
				-1, Eskişehir Yolu, No 6, Yenimahalle	Sırtı, Yenimahalle

9551 rows × 21 columns

```
In [24]: # drop unwanted_cols
df1 = df.drop(unwanted_cols,axis = 1)
```

```
In [25]: df
```

Out[25]:

	RestaurantID	RestaurantName	CountryCode	City	Address	Localit
0	18435314	Punjabi's Veg Grill	1	New Delhi	13/288 , 14 Block Gurudwra, Geeta Colony, New ...	Gee Color
1	18378015	Tasty Tandoor	1	New Delhi	726/2, Jheel Khuranja, Geeta Colony, New Delhi	Gee Color
2	18424905	Taste of Spice	1	New Delhi	C-222, Lajpat Nagar 1, New Delhi	Lajp Nagar
3	18180072	Kolkata Bengali Dhaba	1	New Delhi	Gali 7, Mahipalpur, New Delhi	Mahipalpur
4	18415377	Sunil Punjabi Dhaba	1	New Delhi	Main Vasant Kunj Road, Mahipalpur, New Delhi	Mahipalpur
...
9546	7101042	The Hangar	148	Wellington City	171-177 Willis Street, Te Aro, Wellington City	Te Aro
9547	7100502	Fidel's	148	Wellington City	234 Cuba Street, Te Aro, Wellington City	Te Aro
9548	6900992	Mughal E Azam	215	Birmingham	Stratford Road, Sparkhill, Birmingham B11 4DA	Sparkhill
9549	5800590	The Commons	191	Colombo	39 A, Flower Road, Cinnamon Gardens, Colombo 07	Cinnamon Gardens Colombo C
9550	6001980	Timboo Cafe	208	Ankara	Armada AVM, Kat	Armada AVM

RestaurantID	RestaurantName	CountryCode	City	Address	Locali
				-1, SÍ_Ôôí_tí_z EskiÔôehir Yenimahal Yolu, No 6, Yen...	

9551 rows × 29 columns

```
In [26]: df1.isna().mean().sort_values(ascending = False)
```

```
Out[26]: Cuisines_1      0.000942
Cuisines          0.000942
Has_Table_booking 0.000000
Datekey_Opening   0.000000
Rating            0.000000
Average_Cost_for_two 0.000000
Votes            0.000000
Price_range       0.000000
Switch_to_order_menu 0.000000
Is_delivering_now 0.000000
Has_Online_delivery 0.000000
RestaurantID      0.000000
RestaurantName     0.000000
Latitude          0.000000
Longitude         0.000000
LocalityVerbose   0.000000
Locality          0.000000
Address           0.000000
City              0.000000
CountryCode       0.000000
Currency          0.000000
dtype: float64
```

```
In [27]: # fill missing values
df2 = df1.fillna('others')
```

```
In [28]: df2.isna().mean().sort_values(ascending = False)
```

```
Out[28]: RestaurantID      0.0
         Has_Table_booking  0.0
         Datekey_Opening    0.0
         Rating             0.0
         Average_Cost_for_two 0.0
         Votes              0.0
         Price_range        0.0
         Switch_to_order_menu 0.0
         Is_delivering_now   0.0
         Has_Online_delivery 0.0
         Currency            0.0
         RestaurantName      0.0
         Cuisines            0.0
         Latitude            0.0
         Longitude           0.0
         LocalityVerbose     0.0
         Locality            0.0
         Address             0.0
         City                0.0
         CountryCode         0.0
         Cuisines 1          0.0
         dtype: float64
```

```
In [29]: # num_cols
         df2.select_dtypes('number')
```

```
Out[29]:
```

	RestaurantID	CountryCode	Longitude	Latitude	Price_range	Votes	Average_
0	18435314	1	77.276769	28.650775	1	0	
1	18378015	1	77.275052	28.658216	1	0	
2	18424905	1	77.241312	28.578311	1	0	
3	18180072	1	77.123932	28.543587	1	0	
4	18415377	1	77.129706	28.541369	1	0	
...
9546	7101042	148	174.773933	-41.290801	3	171	
9547	7100502	148	174.774134	-41.295970	3	242	
9548	6900992	215	-1.858529	52.443963	3	32	
9549	5800590	191	79.858105	6.908536	3	209	
9550	6001980	208	32.809247	39.913206	3	134	

9551 rows × 8 columns



```
In [30]: # num_cols
         df2.select_dtypes('number').columns
```

```
Out[30]: Index(['RestaurantID', 'CountryCode', 'Longitude', 'Latitude', 'Price_range',
               'Votes', 'Average_Cost_for_two', 'Rating'],
              dtype='object')
```



```
In [31]: # num_cols
num_cols = df2.select_dtypes('number').columns
```

```
In [32]: # object
cat_cols = df2.select_dtypes('object').columns
```

```
In [33]: cat_cols
```

```
Out[33]: Index(['RestaurantName', 'City', 'Address', 'Locality', 'LocalityVerbose',
               'Cuisines', 'Currency', 'Has_Table_booking', 'Has_Online_delivery',
               'Is_delivering_now', 'Switch_to_order_menu', 'Datekey_Opening',
               'Cuisines 1'],
              dtype='object')
```

```
In [34]: df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9551 entries, 0 to 9550
Data columns (total 21 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   RestaurantID                          9551 non-null   int64
1   RestaurantName                        9551 non-null   object
2   CountryCode                          9551 non-null   int64
3   City                                 9551 non-null   object
4   Address                              9551 non-null   object
5   Locality                             9551 non-null   object
6   LocalityVerbose                      9551 non-null   object
7   Longitude                            9551 non-null   float64
8   Latitude                             9551 non-null   float64
9   Cuisines                             9551 non-null   object
10  Currency                             9551 non-null   object
11  Has_Table_booking                    9551 non-null   object
12  Has_Online_delivery                 9551 non-null   object
13  Is_delivering_now                   9551 non-null   object
14  Switch_to_order_menu                9551 non-null   object
15  Price_range                         9551 non-null   int64
16  Votes                              9551 non-null   int64
17  Average_Cost_for_two                 9551 non-null   int64
18  Rating                              9551 non-null   float64
19  Datekey_Opening                     9551 non-null   object
20  Cuisines 1                          9551 non-null   object
dtypes: float64(3), int64(5), object(13)
memory usage: 1.5+ MB
```

```
In [35]: df2.drop_duplicates('RestaurantID').shape
```

```
Out[35]: (9551, 21)
```

```
In [36]: df2.shape
```

```
Out[36]: (9551, 21)
```

```
In [37]: # Because both shape are same hence no duplicates records
```

```
In [38]: # starts analysis
```

```
In [39]: df2.describe()
```

Out[39]:

	RestaurantID	CountryCode	Longitude	Latitude	Price_range	Votes
count	9.551000e+03	9551.000000	9551.000000	9551.000000	9551.000000	9551.000000
mean	9.051128e+06	18.365616	64.126574	25.854381	1.804837	156.909748
std	8.791521e+06	56.750546	41.467058	11.007935	0.905609	430.169145
min	5.300000e+01	1.000000	-157.948486	-41.330428	1.000000	0.000000
25%	3.019625e+05	1.000000	77.081343	28.478713	1.000000	5.000000
50%	6.004089e+06	1.000000	77.191964	28.570469	2.000000	31.000000
75%	1.835229e+07	1.000000	77.282006	28.642758	2.000000	131.000000
max	1.850065e+07	216.000000	174.832089	55.976980	4.000000	10934.000000



In [40]:

df2.describe().round(2)

Out[40]:

	RestaurantID	CountryCode	Longitude	Latitude	Price_range	Votes	Average
count	9551.00	9551.00	9551.00	9551.00	9551.00	9551.00	
mean	9051128.35	18.37	64.13	25.85	1.80	156.91	
std	8791521.28	56.75	41.47	11.01	0.91	430.17	
min	53.00	1.00	-157.95	-41.33	1.00	0.00	
25%	301962.50	1.00	77.08	28.48	1.00	5.00	
50%	6004089.00	1.00	77.19	28.57	2.00	31.00	
75%	18352291.50	1.00	77.28	28.64	2.00	131.00	
max	18500652.00	216.00	174.83	55.98	4.00	10934.00	



In [41]:

```
# Check corr
corr = df2.corr(numeric_only=True).round(2)
```

In [42]:

corr

Out[42]:

	RestaurantID	CountryCode	Longitude	Latitude	Price_range	Votes
RestaurantID	1.00	0.15	-0.23	-0.05	-0.13	-0.15
CountryCode	0.15	1.00	-0.70	0.02	0.24	0.15
Longitude	-0.23	-0.70	1.00	0.04	-0.08	-0.09
Latitude	-0.05	0.02	0.04	1.00	-0.17	-0.02
Price_range	-0.13	0.24	-0.08	-0.17	1.00	0.31
Votes	-0.15	0.15	-0.09	-0.02	0.31	1.00
Average_Cost_for_two	-0.00	0.04	0.05	-0.11	0.08	0.00
Rating	-0.29	0.32	-0.15	-0.02	0.46	0.00

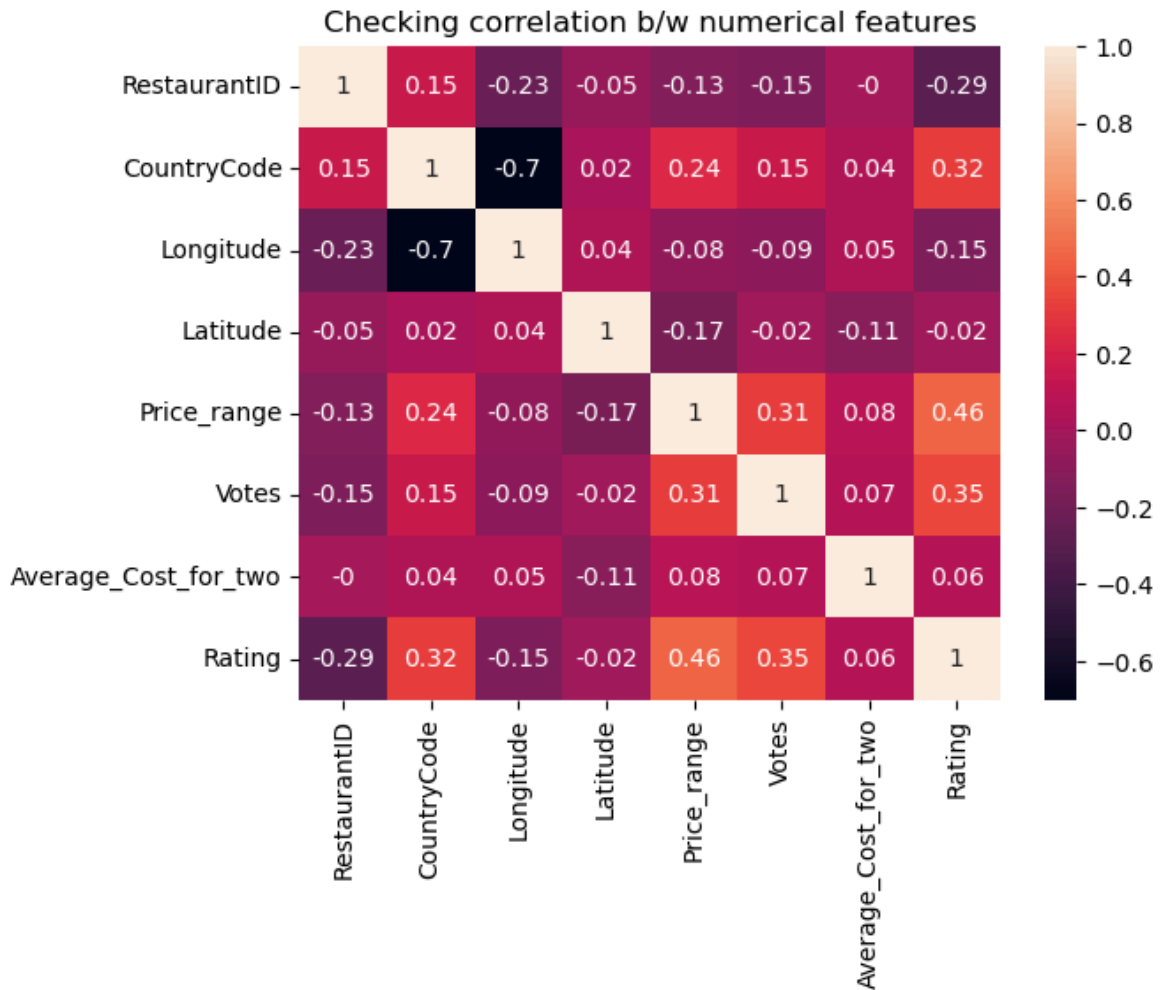
In [43]: `import seaborn as sns`
`print('done')`

done

In [44]: `plt.title('Checking correlation b/w numerical features')`
`sns.heatmap(corr, annot=True)`

`plt.savefig('correlation_heatmap.png', dpi=700, bbox_inches='tight')`

`plt.show()`



```
In [45]: df2.apply(lambda col: df2[col].str.upper() if isinstance(col,str) else col )
```

Out[45]:

	RestaurantID	RestaurantName	CountryCode	City	Address	Localit
0	18435314	Punjabi's Veg Grill	1	New Delhi	13/288 , 14 Block Gurudwra, Geeta Colony, New ...	Gee Color
1	18378015	Tasty Tandoor	1	New Delhi	726/2, Jheel Khuranja, Geeta Colony, New Delhi	Gee Color
2	18424905	Taste of Spice	1	New Delhi	C-222, Lajpat Nagar 1, New Delhi	Lajp Nagar
3	18180072	Kolkata Bengali Dhaba	1	New Delhi	Gali 7, Mahipalpur, New Delhi	Mahipalpur
4	18415377	Sunil Punjabi Dhaba	1	New Delhi	Main Vasant Kunj Road, Mahipalpur, New Delhi	Mahipalpur
...
9546	7101042	The Hangar	148	Wellington City	171-177 Willis Street, Te Aro, Wellington City	Te Aro
9547	7100502	Fidel's	148	Wellington City	234 Cuba Street, Te Aro, Wellington City	Te Aro
9548	6900992	Mughal E Azam	215	Birmingham	Stratford Road, Sparkhill, Birmingham B11 4DA	Sparkhill
9549	5800590	The Commons	191	Colombo	39 A, Flower Road, Cinnamon Gardens, Colombo 07	Cinnamon Gardens Colombo C
9550	6001980	Timboo Cafe	208	Ankara	Armada AVM, Kat	Armada AVM

RestaurantID	RestaurantName	CountryCode	City	Address	Localities
				-1, Eskişehir Yolu, No 6, Yenimahalle	Sırtı, Yenimahalle

9551 rows × 21 columns

```
In [46]: df2 = df2.apply(lambda col:col.str.upper() if col.dtype == 'object' else col )
```

In [47]: df2

Out[47]:

	RestaurantID	RestaurantName	CountryCode	City	Address	
0	18435314	PUNJABI'S VEG GRILL	1	NEW DELHI	13/288 , 14 BLOCK GURUDWRA, GEETA COLONY, NEW ...	(
1	18378015	TASTY TANDOOR	1	NEW DELHI	726/2, JHEEL KHURANJA, GEETA COLONY, NEW DELHI	(
2	18424905	TASTE OF SPICE	1	NEW DELHI	C-222, LAJPAT NAGAR 1, NEW DELHI	N
3	18180072	KOLCATA BENGALI DHABA	1	NEW DELHI	GALI 7, MAHIPALPUR, NEW DELHI	MAH
4	18415377	SUNIL PUNJABI DHABA	1	NEW DELHI	MAIN VASANT KUNJ ROAD, MAHIPALPUR, NEW DELHI	MAH
...
9546	7101042	THE HANGAR	148	WELLINGTON CITY	171-177 WILLIS STREET, TE ARO, WELLINGTON CITY	
9547	7100502	FIDEL'S	148	WELLINGTON CITY	234 CUBA STREET, TE ARO, WELLINGTON CITY	
9548	6900992	MUGHAL E AZAM	215	BIRMINGHAM	STRATFORD ROAD, SPARKHILL, BIRMINGHAM B11 4DA	SP.
9549	5800590	THE COMMONS	191	COLOMBO	39 A, FLOWER ROAD, CINNAMON GARDENS, COLOMBO 07	CINI G/ COLO
9550	6001980	TIMBOO CAFE	208	ANKARA	ARMADA AVM, KAT -1, ESKİÖZEHİR YOLU, NO 6, YEN...	A SÍ_ÜÇ YENİM

9551 rows × 21 columns

```
In [48]: df2.columns = [i.upper() for i in df2.columns]
```

```
In [49]: df2
```


Out[49]:

	RESTAURANTID	RESTAURANTNAME	COUNTRYCODE		CITY	ADDRESS
0	18435314	PUNJABI'S VEG GRILL	1	NEW DELHI	13/288 , 14 BLOCK GURUDWARA GEETA COLONY NEW ..	
1	18378015	TASTY TANDOOR	1	NEW DELHI	726/2, JHEEL KHURANJA GEETA COLONY NEW DELH	
2	18424905	TASTE OF SPICE	1	NEW DELHI	C-222, LAJPAT NAGAR 1 NEW DELH	
3	18180072	KOLCATA BENGALI DHABA	1	NEW DELHI	GALI 7 MAHIPALPUR NEW DELH	
4	18415377	SUNIL PUNJABI DHABA	1	NEW DELHI	MAIN VASANT KUNJ ROAD MAHIPALPUR NEW DELH	
...	
9546	7101042	THE HANGAR	148	WELLINGTON CITY	171-177 WILLIS STREET, TE ARO WELLINGTON CITY	
9547	7100502	FIDEL'S	148	WELLINGTON CITY	234 CUBA STREET, TE ARO WELLINGTON CITY	
9548	6900992	MUGHAL E AZAM	215	BIRMINGHAM	STRATFORD ROAD SPARKHILL BIRMINGHAM B11 4DA	
9549	5800590	THE COMMONS	191	COLOMBO	39 A, FLOWER ROAD CINNAMON GARDENS COLOMBO 07	
9550	6001980	TIMBOO CAFE	208	ANKARA	ARMADA AVM, KAT -1 ESKİÖEHİF YOLU, NO 6 YEN.	

9551 rows × 21 columns

```
In [50]: df2.sample(3)
```

Out[50]:

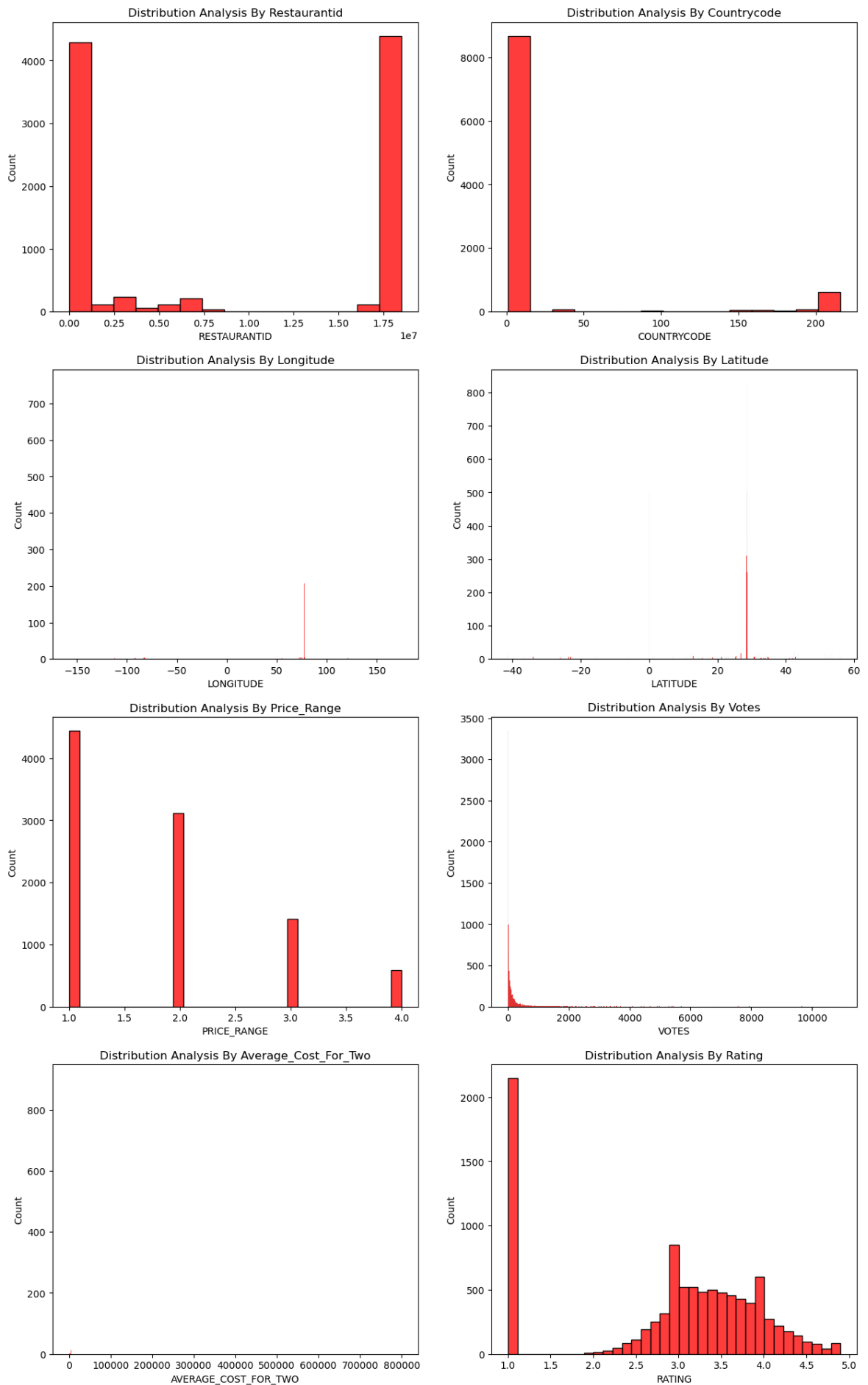
	RESTAURANTID	RESTAURANTNAME	COUNTRYCODE	CITY	ADDRESS	LOCAL
3940	306015	STANDARD SWEETS	1	NEW DELHI	3510, CHAWRI BAZAR, NEW DELHI	CHA' BA:
1316	18306530	CIRCUS	1	NEW DELHI	D-14, 3RD FLOOR, SOUTH EXTENSION 2, NEW DELHI	SOI EXTENS
3641	312860	SHANGHAI CHINESE FOOD	1	NEW DELHI	MAHAVIR SWAMI PARK, OPPOSITE ADITYA ARCADE, PR...	PR VIT

3 rows × 21 columns



Univariate Analysis

```
In [51]: plt.figure(figsize = (15,25))
for i,j in enumerate(num_cols):
    plt.subplot(4,2,i+1)
    plt.title(f'Distribution analysis by {j}'.title())
    sns.histplot(data = df2,x = j.upper(),color = 'r')
plt.show()
```

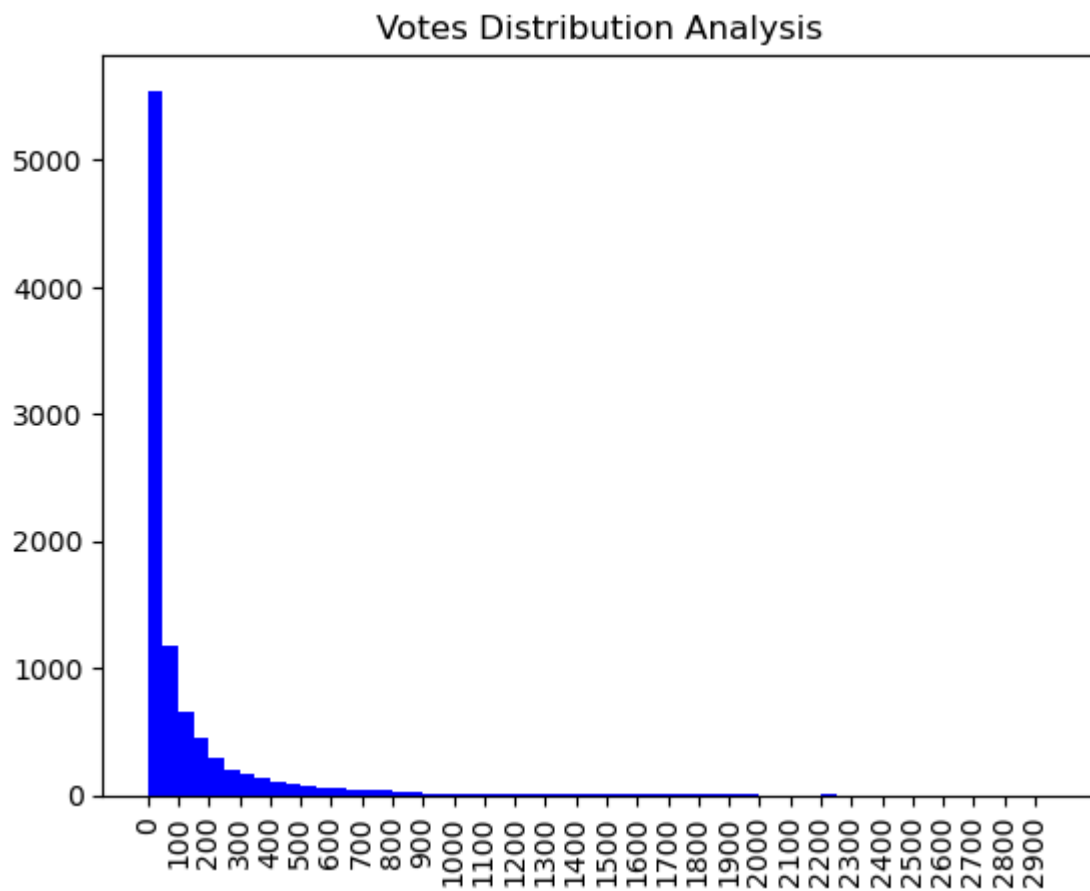


```
In [52]: plt.title('Votes Distribution Analysis')

plt.hist(df2['VOTES'], color='b', bins=range(0, 3000, 50))
```

```
plt.xticks(range(0, 3000, 100), rotation=90, fontsize=10)

plt.show()
```



```
In [53]: len(num_cols)
```

```
Out[53]: 8
```

```
In [54]: for i,j in enumerate(cat_cols):
          print(f'value counts by {j}'.title())
          print(df2[j.upper()].value_counts().head(10))

          print('-----',end = '\n'*2)
```

Value Counts By Restaurantname

RESTAURANTNAME

CAFE COFFEE DAY	83
DOMINO'S PIZZA	79
SUBWAY	63
GREEN CHICK CHOP	51
MCDONALD'S	48
KEVENTERS	34
PIZZA HUT	30
GIANI	29
BASKIN ROBBINS	28
BARBEQUE NATION	26

Name: count, dtype: int64

Value Counts By City

CITY

NEW DELHI	5473
GURGAON	1118
NOIDA	1080
FARIDABAD	251
GHAZIABAD	25
GUWAHATI	21
AMRITSAR	21
AHMEDABAD	21
LUCKNOW	21
BHUBANESHWAR	21

Name: count, dtype: int64

Value Counts By Address

ADDRESS

DILLI HAAT, INA, NEW DELHI	11
SECTOR 41, NOIDA	11
GREATER KAILASH (GK) 1, NEW DELHI	10
HUDA MARKET, SECTOR 56, GURGAON	9
THE IMPERIAL, JANPATH, NEW DELHI	9
CYBER HUB, DLF CYBER CITY, GURGAON	8
FOOD COURT, 3RD FLOOR, LOGIX CITY CENTRE, SECTOR 32, NEAR SECTOR 34, NOIDA	8
3RD FLOOR, DLF MALL OF INDIA, SECTOR 18, NOIDA	8
PALATE OF DELHI, DHAULA KUAN METRO STATION, CHANAKYAPURI, NEW DELHI	8
THE LALIT, BARAKHAMBHA AVENUE, BARAKHAMBHA ROAD, NEW DELHI	8

Name: count, dtype: int64

Value Counts By Locality

LOCALITY

CONNAUGHT PLACE	122
RAJOURI GARDEN	99
SHAHDARA	87
DEFENCE COLONY	86
MALVIYA NAGAR	85
PITAMPURA	85
MAYUR VIHAR PHASE 1	84
RAJINDER NAGAR	81
SAFDARJUNG	80
SATYANIKETAN	79

Name: count, dtype: int64

Value Counts By Localityverbose

LOCALITYVERBOSE

CONNAUGHT PLACE, NEW DELHI	122
RAJOURI GARDEN, NEW DELHI	99
SHAHDARA, NEW DELHI	87
DEFENCE COLONY, NEW DELHI	86
PITAMPURA, NEW DELHI	85
MALVIYA NAGAR, NEW DELHI	84

```

MAYUR VIHAR PHASE 1, NEW DELHI      84
RAJINDER NAGAR, NEW DELHI           81
SAFDARJUNG, NEW DELHI               80
SATYANIKETAN, NEW DELHI             79
Name: count, dtype: int64
Value Counts By Cuisines
CUISINES
NORTH INDIAN                        936
NORTH INDIAN, CHINESE               511
FAST FOOD                           354
CHINESE                             354
NORTH INDIAN, MUGHLAI               334
CAFE                                 299
BAKERY                               218
NORTH INDIAN, MUGHLAI, CHINESE      197
BAKERY, DESSERTS                    170
STREET FOOD                          149
Name: count, dtype: int64
Value Counts By Currency
CURRENCY
INDIAN RUPEES(RS.)                  8652
DOLLAR($)                           482
POUNDS(£)                           80
EMIRATI DIRAM(AED)                  60
BRAZILIAN REAL(R$)                  60
RAND(R)                             60
NEWZEALAND($)                       40
TURKISH LIRA(TL)                    34
BOTSWANA PULA(P)                    22
INDONESIAN RUPIAH(IDR)              21
Name: count, dtype: int64
Value Counts By Has_Table_Booking
HAS_TABLE_BOOKING
NO      8393
YES     1158
Name: count, dtype: int64
Value Counts By Has_Online_Delivery
HAS_ONLINE_DELIVERY
NO      7100
YES     2451
Name: count, dtype: int64
Value Counts By Is_Delivering_Now
IS_DELIVERING_NOW
NO      9517
YES      34
Name: count, dtype: int64
Value Counts By Switch_To_Order_Menu
SWITCH_TO_ORDER_MENU
NO      9551
Name: count, dtype: int64
Value Counts By Datekey_Opening
DATEKEY_OPENING
2010_7_14    11
2011_7_16    10
2011_4_6      10
2015_10_5     10
2011_7_11      9
2018_1_23      9
2013_5_11      9
2018_6_8       9

```

```

2012_8_7      9
2011_8_2      9
Name: count, dtype: int64
Value Counts By Cuisines 1
CUISINES 1
NORTH INDIAN    2992
CHINESE         855
FAST FOOD       672
BAKERY          621
CAFE            617
AMERICAN        278
SOUTH INDIAN    262
MITHAI          246
STREET FOOD     236
CONTINENTAL     235
Name: count, dtype: int64
-----

```

```

In [55]: for i,j in enumerate(cat_cols):
          temp_df = df2[j.upper()].value_counts().head(10)
          print(temp_df.shape)

```

```

(10,)
(10,)
(10,)
(10,)
(10,)
(10,)
(10,)
(2,)
(2,)
(2,)
(1,)
(10,)
(10,)

```

```

In [56]: plt.figure(figsize=(50, 30))

temp_cat_cols = list(cat_cols)
temp_cat_cols.remove('Switch_to_order_menu')

for i, j in enumerate(temp_cat_cols):
    plt.subplot(4,3,i+1)
    temp_df = df2[j.upper()].value_counts().head(10).sort_values()
    row = temp_df.shape[0]

    x = temp_df.index
    y = temp_df.values

    if row <= 1:
        # Skip this chart
        pass

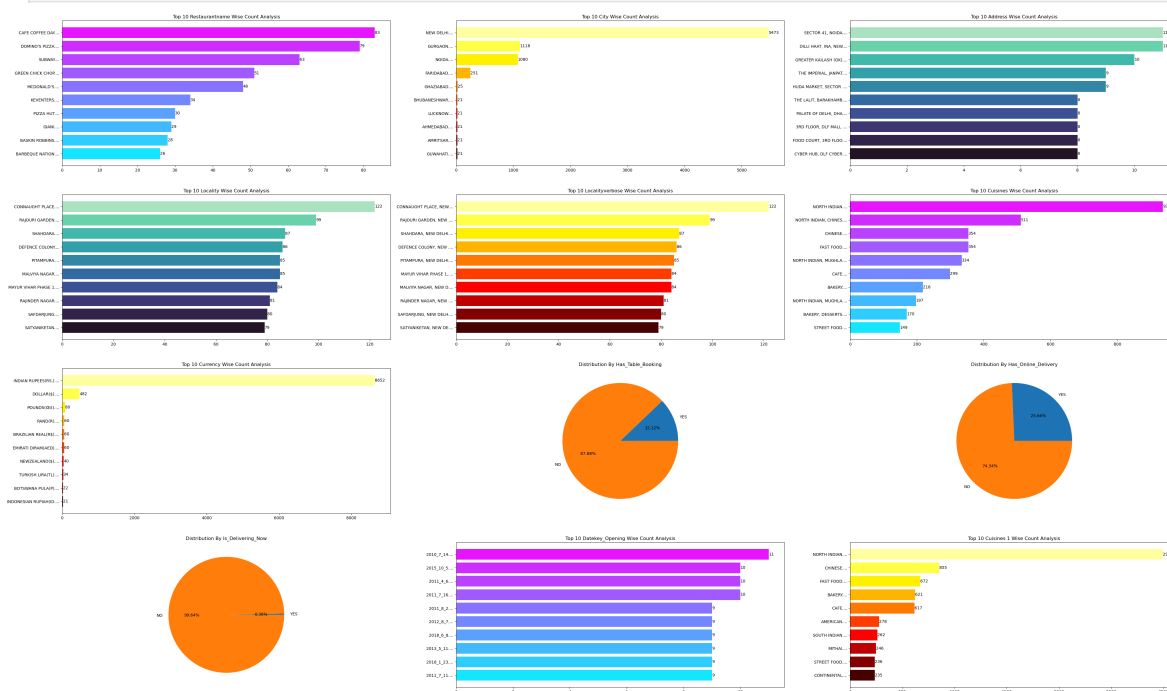
    elif row <= 5:
        # For pie chart
        plt.title(f'Distribution by {j}'.title())
        plt.pie(y, labels=x, autopct='%.2f%%')

    else:

```

```
# For bar chart
```

```
plt.title(f'Top 10 {j} Wise Count Analysis'.title())
r_color = ['cool', 'magma', 'mako', 'hot', 'rainbow']
ax = plt.barh(x, y, color=sns.color_palette(random.choice(r_color),10))
plt.bar_label(ax)
y_new_name = [i[:20] + '....' for i in list(x)]
plt.yticks(range(10),y_new_name)
```



```
In [57]: len(cat_cols)
```

```
Out[57]: 13
```

```
In [58]: len(temp_cat_cols)
```

```
Out[58]: 12
```

```
In [59]: temp_cat_cols
```

```
Out[59]: ['RestaurantName',
          'City',
          'Address',
          'Locality',
          'LocalityVerbose',
          'Cuisines',
          'Currency',
          'Has_Table_booking',
          'Has_Online_delivery',
          'Is_delivering_now',
          'Datekey_Opening',
          'Cuisines 1']
```

Bivariate Analysis

```
In [60]: df.columns
```



```
Out[60]: Index(['RestaurantID', 'RestaurantName', 'CountryCode', 'City', 'Address',
              'Locality', 'LocalityVerbose', 'Longitude', 'Latitude', 'Cuisines',
              'Currency', 'Has_Table_booking', 'Has_Online_delivery',
              'Is_delivering_now', 'Switch_to_order_menu', 'Price_range', 'Votes',
              'Average_Cost_for_two', 'Rating', 'Datekey_Opening', 'Unnamed: 20',
              'Cuisines 1', 'Cuisines 2', 'Cuisines 3', 'Cuisines 4', 'Cuisines 5',
              'Cuisines 6', 'Cuisines 7', 'Cuisines 8'],
              dtype='object')
```

```
In [61]: df2.columns
```

```
Out[61]: Index(['RESTAURANTID', 'RESTAURANTNAME', 'COUNTRYCODE', 'CITY', 'ADDRESS',
              'LOCALITY', 'LOCALITYVERBOSE', 'LONGITUDE', 'LATITUDE', 'CUISINES',
              'CURRENCY', 'HAS_TABLE_BOOKING', 'HAS_ONLINE_DELIVERY',
              'IS_DELIVERING_NOW', 'SWITCH_TO_ORDER_MENU', 'PRICE_RANGE', 'VOTES',
              'AVERAGE_COST_FOR_TWO', 'RATING', 'DATEKEY_OPENING', 'CUISINES 1'],
              dtype='object')
```

```
In [62]: len(df2.columns)
```

```
Out[62]: 21
```

```
In [63]: df2['PRICE_RANGE'].value_counts()
```

```
Out[63]: PRICE_RANGE
1      4444
2      3113
3      1408
4       586
Name: count, dtype: int64
```

```
In [64]: df2['AVERAGE_COST_FOR_TWO']
```

```
Out[64]: 0      300
1      200
2      400
3      100
4      150
...
9546    50
9547    50
9548    45
9549   2500
9550    70
Name: AVERAGE_COST_FOR_TWO, Length: 9551, dtype: int64
```

```
In [65]: df2[df2['PRICE_RANGE'] == 1]['AVERAGE_COST_FOR_TWO'].agg(['min', 'max']).values
```

```
Out[65]: array([ 0, 450], dtype=int64)
```

```
In [66]: p1_min, p1_max = df2[df2['PRICE_RANGE'] == 1]['AVERAGE_COST_FOR_TWO'].agg(['min',
p2_min, p2_max = df2[df2['PRICE_RANGE'] == 2]['AVERAGE_COST_FOR_TWO'].agg(['min',
p3_min, p3_max = df2[df2['PRICE_RANGE'] == 3]['AVERAGE_COST_FOR_TWO'].agg(['min',
p4_min, p4_max = df2[df2['PRICE_RANGE'] == 4]['AVERAGE_COST_FOR_TWO'].agg(['min',

temp_dict = {1: f'{p1_min} - {p1_max}',
              2: f'{p2_min} - {p2_max}',
              3: f'{p3_min} - {p3_max}',
              4: f'{p4_min} - {p4_max}'}
```

```
temp_dict_df = pd.DataFrame(temp_dict,index = [0])
temp_dict_df
```

```
Out[66]:
```

	1	2	3	4
0	0 - 450	15 - 70000	30 - 800000	50 - 8000

```
In [67]: pip install currencyconverter
```

Requirement already satisfied: currencyconverter in c:\users\ankit\anaconda3\lib\site-packages (0.18.12)

Note: you may need to restart the kernel to use updated packages.

```
In [68]: from currency_converter import CurrencyConverter
```

```
c = CurrencyConverter()

print(c.convert(100000, 'IDR', 'INR'))
```

531.2537681442643

```
In [69]: from currency_converter import CurrencyConverter
```

```
c = CurrencyConverter()
amount_idr = 800000
inr_value = c.convert(amount_idr,'IDR','INR')

print(f"{amount_idr} IDR = {inr_value:.2f} INR")
```

800000 IDR = 4250.03 INR

```
In [70]: currency_codes = {
    "INDIAN RUPEES": "INR",
    "DOLLAR": "USD",
    "POUNDS": "GBP",
    "EMIRATI DIRHAM": "AED",
    "BRAZILIAN REAL": "BRL",
    "RAND": "ZAR",
    "NEWZEALAND": "NZD",
    "TURKISH LIRA": "TRY",
    "BOTSWANA PULA": "BWP",
    "INDONESIAN RUPIAH": "IDR"
}

print(currency_codes)
```

```
{'INDIAN RUPEES': 'INR', 'DOLLAR': 'USD', 'POUNDS': 'GBP', 'EMIRATI DIRHAM': 'AED', 'BRAZILIAN REAL': 'BRL', 'RAND': 'ZAR', 'NEWZEALAND': 'NZD', 'TURKISH LIRA': 'TRY', 'BOTSWANA PULA': 'BWP', 'INDONESIAN RUPIAH': 'IDR'}
```

```
In [71]: temp_curr_symbol_index = df2['CURRENCY'].value_counts().head(10).index
temp_curr_symbol = [i.split('(')[0] for i in temp_curr_symbol_index]

final_curr_symbol_dict = dict(zip(temp_curr_symbol_index,temp_curr_symbol))
print(temp_curr_symbol_index)
```

```
Index(['INDIAN RUPEES(RS.)', 'DOLLAR($)', 'POUNDS(£)', 'EMIRATI DIRAM(AED)',
      'BRAZILIAN REAL(R$)', 'RAND(R)', 'NEWZEALAND($)', 'TURKISH LIRA(TL)',
      'BOTSWANA PULA(P)', 'INDONESIAN RUPIAH(IDR)'],
      dtype='object', name='CURRENCY')
```

```
In [72]: df2['CURRENCY'].value_counts()
```

```
Out[72]: CURRENCY
INDIAN RUPEES(RS.)      8652
DOLLAR($)               482
POUNDS(£)               80
EMIRATI DIRAM(AED)      60
BRAZILIAN REAL(R$)      60
RAND(R)                 60
NEWZEALAND($)           40
TURKISH LIRA(TL)        34
BOTSWANA PULA(P)        22
INDONESIAN RUPIAH(IDR)  21
QATARI RIAL(QR)         20
SRI LANKAN RUPEE(LKR)   20
Name: count, dtype: int64
```

```
In [73]: df2['CURRENCY_TEMP'] = df2['CURRENCY'].apply(lambda row: final_curr_symbol_dict[
```

```
In [74]: df2['CURRENCY_CODE'] = df2['CURRENCY_TEMP'].apply(lambda row : currency_codes[ro
```

```
In [75]: df2['CURRENCY_CODE'].value_counts().head(10)
```

```
Out[75]: CURRENCY_CODE
INR      8652
USD      482
GBP       80
EMIRATI DIRAM  60
BRL       60
ZAR       60
NZD       40
TRY       34
BWP       22
IDR       21
Name: count, dtype: int64
```

```
In [76]: from currency_converter import CurrencyConverter
```

```
def currency_convert_to_inr(code, amount):
    c = CurrencyConverter()
    inr_value = c.convert(amount, code, 'INR')
    return inr_value
```

```
In [77]: currency_convert_to_inr('USD',1)
```

```
Out[77]: 88.77410109431996
```

```
In [78]: cache = {}
valid_codes = set(currency_codes.values())

final_amount = []
append = final_amount.append

for _, curr_code, amount in df2[['CURRENCY_CODE', 'AVERAGE_COST_FOR_TWO']].itertu
```

```
if curr_code in valid_codes:

    if curr_code not in cache:
        try:
            cache[curr_code] = currency_convert_to_inr(curr_code, 1) # per
        except:
            cache[curr_code] = None

    if cache[curr_code] is not None:
        append(cache[curr_code] * amount)
    else:
        append(amount)

else:
    append(amount)

print(final_amount)
```

37/65

38/65

650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0
50.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.
0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	6
50.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.
0	1100.0	1500.0	1800.0	1350.0	1550.0	1700.0	1600.0	1200.0	1200.0	1400.		
0	1900.0	1300.0	1500.0	1500.0	1600.0	1200.0	1500.0	1500.0	1400.0	1300.		
0	1300.0	1500.0	1300.0	1800.0	1700.0	1200.0	1100.0	1500.0	1800.0	1200.		
0	1800.0	1200.0	1100.0	1500.0	1900.0	1300.0	1550.0	1750.0	1100.0	1800.		
0	1200.0	1500.0	1800.0	1400.0	1300.0	1200.0	1250.0	1200.0	1500.0	1600.		
0	1200.0	1800.0	1850.0	1900.0	1400.0	1600.0	1200.0	1100.0	1050.0	1800.		
0	1200.0	1200.0	1400.0	1500.0	1200.0	1200.0	1650.0	1700.0	1250.0	1400.		
0	1600.0	1500.0	1200.0	1500.0	1500.0	1200.0	1100.0	1200.0	1500.0	1600.		
0	1200.0	1500.0	1100.0	1700.0	1400.0	1800.0	1200.0	1800.0	1700.0	1200.		
0	1500.0	1400.0	1100.0	1200.0	1900.0	1300.0	1200.0	1900.0	1100.0	1500.		
0	1600.0	1400.0	1100.0	1200.0	1500.0	1300.0	1100.0	1800.0	1600.0	1700.		
0	1300.0	1500.0	1100.0	1800.0	1800.0	1500.0	1500.0	1100.0	1500.0	1500.		
0	1800.0	1200.0	1200.0	1250.0	1900.0	1400.0	1100.0	1100.0	1800.0	1800.		
0	1700.0	1300.0	1100.0	1200.0	1300.0	1950.0	1100.0	1300.0	1600.0	1650.		
0	1200.0	1500.0	1500.0	1500.0	1650.0	1500.0	1800.0	1400.0	1900.0	1250.		
0	1200.0	1200.0	1100.0	1600.0	1100.0	1600.0	1200.0	1600.0	1600.0	1100.		
0	1100.0	1250.0	1500.0	1350.0	1300.0	1400.0	1600.0	1500.0	1700.0	1300.		
0	1200.0	1600.0	1200.0	1800.0	1500.0	1250.0	1200.0	1200.0	1800.0	1600.		
0	1350.0	1550.0	1200.0	1500.0	1700.0	1650.0	1500.0	1200.0	1400.0	1100.		
0	1100.0	1100.0	1400.0	1300.0	1500.0	1250.0	1200.0	1700.0	1400.0	1200.		
0	1700.0	1500.0	1600.0	1200.0	1900.0	1150.0	1200.0	1200.0	1700.0	1500.		
0	1650.0	1500.0	1700.0	1100.0	1200.0	1500.0	1600.0	1100.0	1500.0	1200.		
0	1500.0	1500.0	1300.0	1600.0	1600.0	1800.0	1250.0	1500.0	1800.0	1600.		
0	1800.0	1800.0	1100.0	1800.0	1200.0	1100.0	1200.0	1550.0	1200.0	1800.		
0	1200.0	1200.0	1700.0	1500.0	1400.0	1500.0	1500.0	1500.0	1600.0	1800.		
0	1500.0	1600.0	1750.0	1600.0	1900.0	1200.0	1500.0	1200.0	1500.0	1500.		
0	1700.0	1550.0	1500.0	1400.0	1100.0	1100.0	1200.0	1500.0	1600.0	1400.		
0	1500.0	1300.0	1800.0	1300.0	1200.0	1450.0	1400.0	1300.0	1800.0	1650.		
0	1500.0	1300.0	1700.0	1								

40/65

41/65

42/65

43/65

[illegible]

0, 200.0, 400.0, 200.0, 150.0, 350.0, 300.0, 200.0, 400.0, 250.0, 400.0, 400.0, 100.0, 250.0, 300.0, 200.0, 250.0, 450.0, 200.0, 200.0, 350.0, 150.0, 150.0, 200.0, 250.0, 200.0, 500.0, 950.0, 800.0, 750.0, 550.0, 500.0, 550.0, 600.0, 500.0, 600.0, 750.0, 600.0, 650.0, 600.0, 600.0, 500.0, 650.0, 500.0, 600.0, 650.0, 800.0, 700.0, 650.0, 800.0, 700.0, 700.0, 600.0, 500.0, 500.0, 600.0, 600.0, 700.0, 600.0, 600.0, 600.0, 500.0, 500.0, 550.0, 500.0, 650.0, 500.0, 600.0, 500.0, 500.0, 650.0, 550.0, 600.0, 600.0, 550.0, 500.0, 500.0, 600.0, 500.0, 650.0, 800.0, 600.0, 500.0, 600.0, 500.0, 600.0, 600.0, 600.0, 500.0, 500.0, 750.0, 600.0, 550.0, 500.0, 500.0, 550.0, 500.0, 700.0, 500.0, 700.0, 500.0, 550.0, 600.0, 500.0, 550.0, 550.0, 1300.0, 1200.0, 1800.0, 1500.0, 1500.0, 1100.0, 1500.0, 1200.0, 1500.0, 1400.0, 1500.0, 1300.0, 1100.0, 1500.0, 1300.0, 1400.0, 1800.0, 1500.0, 1500.0, 1000.0, 1500.0, 1300.0, 1400.0, 1600.0, 1500.0, 1500.0, 1600.0, 1500.0, 1500.0, 100.0, 300.0, 500.0, 350.0, 250.0, 150.0, 400.0, 100.0, 200.0, 400.0, 500.0, 200.0, 500.0, 400.0, 500.0, 500.0, 200.0, 100.0, 500.0, 200.0, 200.0, 500.0, 300.0, 600.0, 200.0, 200.0, 200.0, 500.0, 250.0, 250.0, 350.0, 400.0, 600.0, 200.0, 200.0, 100.0, 200.0, 350.0, 400.0, 150.0, 200.0, 300.0, 300.0, 250.0, 400.0, 450.0, 100.0, 200.0, 200.0, 200.0, 200.0, 350.0, 500.0, 250.0, 150.0, 500.0, 450.0, 150.0, 350.0, 350.0, 250.0, 200.0, 250.0, 350.0, 450.0, 200.0, 200.0, 200.0, 200.0, 200.0, 250.0, 200.0, 200.0, 300.0, 100.0, 100.0, 350.0, 200.0, 150.0, 550.0, 400.0, 300.0, 250.0, 500.0, 500.0, 350.0, 500.0, 250.0, 150.0, 200.0, 400.0, 150.0, 200.0, 500.0, 300.0, 350.0, 350.0, 150.0, 200.0, 250.0, 300.0, 400.0, 300.0, 450.0, 200.0, 260.0, 450.0, 100.0, 200.0, 350.0, 300.0, 450.0, 350.0, 200.0, 150.0, 200.0, 300.0, 500.0, 100.0, 250.0, 300.0, 300.0, 600.0, 200.0, 200.0, 200.0, 500.0, 300.0, 500.0, 300.0, 100.0, 200.0, 200.0, 200.0, 250.0, 100.0, 400.0, 350.0, 100.0, 300.0, 300.0, 400.0, 600.0, 550.0, 1000.0, 500.0, 300.0, 1200.0, 750.0, 1000.0, 600.0, 1000.0, 450.0, 400.0, 350.0, 1500.0, 1500.0, 1400.0, 1000.0, 1500.0, 1500.0, 700.0, 1200.0, 1300.0, 800.0, 700.0, 1500.0, 150.0, 1000.0, 550.0, 550.0, 1000.0, 1500.0, 800.0, 1000.0, 700.0, 1500.0, 450.0, 750.0, 900.0, 100.0, 150.0, 1500.0, 550.0, 700.0, 800.0, 650.0, 350.0, 150.0, 1500.0, 1250.0, 1850.0, 150.0, 800.0, 150.0, 450.0, 450.0, 1100.0, 1500.0, 1500.0, 1500.0, 700.0, 700.0, 800.0, 200.0, 200.0, 1500.0, 1000.0, 1500.0, 1200.0, 1800.0, 1100.0, 1200.0, 1700.0, 800.0, 1500.0, 750.0, 700.0, 550.0, 200.0, 750.0, 550.0, 800.0, 1500.0, 1500.0, 450.0, 1500.0, 1500.0, 1900.0, 1200.0, 800.0, 450.0, 700.0, 200.0, 250.0, 250.0, 700.0, 450.0, 250.0, 800.0, 900.0, 1500.0, 1500.0, 1500.0, 700.0, 1400.0, 850.0, 800.0, 200.0, 100.0, 800.0, 350.0, 700.0, 350.0, 1000.0, 800.0, 750.0, 950.0, 800.0, 700.0, 650.0, 1300.0, 200.0, 1300.0, 1300.0, 1600.0, 1500, 1700.0, 700.0, 1500, 800.0, 650.0, 200.0, 350.0, 150.0, 1000.0, 850.0, 650.0, 1500, 1500.0, 800.0, 7898.437476479355, 450.0, 1800.0, 1600.0, 350.0, 1800.0, 450.0, 1000.0, 550.0, 350.0, 900.0, 800.0, 350.0, 200.0, 450.0, 650.0, 450.0, 900.0, 550.0, 1600.0, 1200.0, 650.0, 450.0, 350.0, 150.0, 250.0, 150.0, 350.0, 650.0, 200.0, 200.0, 1000.0, 250.0, 650.0, 1600.0, 1600.0, 200.0, 1000.0, 450.0, 200.0, 250.0, 150.0, 250.0, 350.0, 1300.0, 800.0, 200.0, 350.0, 700.0, 800.0, 900.0, 1750.0, 650.0, 800.0, 200.0, 350.0, 1600.0, 650.0, 550.0, 1000.0, 1500.0, 1300.0, 1850.0, 1600.0, 1600.0, 350.0, 1500.0, 200.0, 350.0, 1800.0, 700.0, 1500.0, 1800.0, 200.0, 900.0, 200.0, 1000.0, 350.0, 850.0, 800.0, 1400.0, 1500.0, 1500.0, 1500.0, 1600.0, 1200.0, 200.0, 250.0, 250.0, 650.0, 1600.0, 250.0, 100.0, 200.0, 1600.0, 1800.0, 150.0, 250.0, 200.0, 350.0, 1000.0, 1500.0, 1400.0, 1600.0, 1500.0, 700.0, 550.0, 700.0, 250.0, 800.0, 1000.0, 700.0, 750.0, 750.0, 1000.0, 1000.0, 1700.0, 1600.0, 1000.0, 1600.0, 1400.0, 1650.0, 550.0, 1500.0, 550.0, 1700.0, 700.0, 250.0, 150.0, 200.0, 1400.0, 1800.0, 350.0, 700.0, 100.0, 200.0, 700.0, 150.0, 800.0, 250.0, 250.0, 700.0, 1500.0, 750.0, 1700.0, 1400.0, 1500.0, 1700.0, 1800.0, 1800.0, 1800.0, 1500.0, 1200.0, 700.0, 200.0, 200.0, 550.0, 250.0, 800.0, 650.0, 1100.0, 1000.0, 1100.0, 1800, 800.0, 550.0, 800.0, 1000.0, 1000.0, 550.0, 200.0, 700.0, 200.0, 200.0, 450.0, 200.0, 1200.0, 200.0, 350.0, 150.0, 650.0, 800.0, 800.0, 750.0, 850.0, 1950.0, 2000.0, 550.0, 800.0, 1200.0, 2000.0, 2000.0, 350.0, 350.0, 450.0, 350.0, 200.0, 200.0, 200.0, 700.0, 800.0, 1250.0, 1500.0, 850.0, 700.0, 1400.0, 1400.0, 1200.0, 1200.0, 650.0, 900.0, 250.0, 200.0, 800.0, 800.0, 650.0, 1700.0, 250.0, 100.0, 350.0, 450.0, 700.0, 650.0, 200.0, 700.0, 800.0, 700.0, 1800.0, 1600.0, 450.0, 200.0, 750.0, 200.0, 350.0, 950.0, 2000.0, 700.0, 250.0, 550.0, 200.0, 250.0, 2000.0, 200.0, 100.0, 2000.0, 700.0, 650.0, 800.0, 650.0, 700.0, 350.0, 800.0, 80

46/65

0, 250.0, 350.0, 100.0, 450.0, 100.0, 200.0, 150.0, 400.0, 300.0, 450.0, 200.0, 250.0, 200.0, 400.0, 300.0, 300.0, 200.0, 150.0, 400.0, 350.0, 400.0, 200.0, 300.0, 150.0, 400.0, 400.0, 250.0, 350.0, 200.0, 150.0, 350.0, 350.0, 300.0, 200.0, 200.0, 250.0, 150.0, 150.0, 400.0, 300.0, 300.0, 200.0, 250.0, 450.0, 250.0, 200.0, 200.0, 150.0, 600.0, 550.0, 600.0, 500.0, 350.0, 200.0, 500.0, 200.0, 250.0, 100.0, 300.0, 400.0, 400.0, 450.0, 500.0, 300.0, 500.0, 300.0, 450.0, 200.0, 250.0, 700.0, 300.0, 300.0, 300.0, 200.0, 300.0, 350.0, 100.0, 200.0, 300.0, 400.0, 400.0, 400.0, 200.0, 500.0, 250.0, 200.0, 100.0, 800.0, 600.0, 300.0, 150.0, 200.0, 1200.0, 300.0, 250.0, 300.0, 200.0, 250.0, 500.0, 400.0, 400.0, 100.0, 550.0, 300.0, 500.0, 300.0, 500.0, 400.0, 200.0, 200.0, 350.0, 600.0, 250.0, 500.0, 800.0, 450.0, 300.0, 50.0, 700.0, 400.0, 500.0, 400.0, 500.0, 300.0, 100.0, 400.0, 250.0, 500.0, 200.0, 100.0, 350.0, 250.0, 450.0, 1000.0, 300.0, 400.0, 200.0, 100.0, 300.0, 200.0, 500.0, 300.0, 400.0, 300.0, 200.0, 300.0, 100.0, 200.0, 300.0, 400.0, 500.0, 500.0, 200.0, 250.0, 2000.0, 200.0, 200.0, 200.0, 450.0, 200.0, 300.0, 300.0, 150.0, 300.0, 200.0, 100.0, 100.0, 150.0, 200.0, 350.0, 500.0, 400.0, 500.0, 300.0, 300.0, 250.0, 500.0, 450.0, 150.0, 450.0, 300.0, 450.0, 1000.0, 200.0, 500.0, 500.0, 250.0, 200.0, 250.0, 300.0, 500.0, 900.0, 100.0, 300.0, 450.0, 400.0, 500.0, 150.0, 200.0, 200.0, 400.0, 250.0, 250.0, 500.0, 250.0, 350.0, 2000.0, 400.0, 300.0, 200.0, 250.0, 500.0, 500.0, 300.0, 500.0, 300.0, 450.0, 200.0, 200.0, 500.0, 300.0, 300.0, 300.0, 200.0, 300.0, 350.0, 300.0, 500.0, 300.0, 150.0, 400.0, 500.0, 200.0, 300.0, 500.0, 200.0, 300.0, 500.0, 600.0, 700.0, 500.0, 500.0, 500.0, 800.0, 600.0, 550.0, 700.0, 500.0, 500.0, 850.0, 2000.0, 600.0, 600.0, 600.0, 800.0, 600.0, 500.0, 600.0, 500.0, 250.0, 300.0, 250.0, 250.0, 150.0, 100.0, 400.0, 150.0, 100.0, 200.0, 100.0, 150.0, 300.0, 250.0, 300.0, 300.0, 300.0, 350.0, 400.0, 250.0, 200.0, 300.0, 450.0, 150.0, 250.0, 400.0, 100.0, 400.0, 200.0, 400.0, 150.0, 200.0, 450.0, 300.0, 200.0, 300.0, 300.0, 200.0, 400.0, 300.0, 150.0, 200.0, 250.0, 250.0, 400.0, 100.0, 150.0, 200.0, 300.0, 150.0, 300.0, 200.0, 400.0, 200.0, 350.0, 200.0, 150.0, 100.0, 350.0, 400.0, 200.0, 350.0, 150.0, 350.0, 450.0, 350.0, 350.0, 1100.0, 2000.0, 1600.0, 650.0, 350.0, 250.0, 1500.0, 950.0, 1000.0, 2000.0, 300.0, 1400.0, 1000.0, 800.0, 1100.0, 300.0, 250.0, 550.0, 200.0, 300.0, 600.0, 450.0, 300.0, 450.0, 300.0, 450.0, 1300.0, 700.0, 800.0, 300.0, 800.0, 800.0, 750.0, 650.0, 300.0, 650.0, 650.0, 450.0, 800.0, 600.0, 150.0, 350.0, 150.0, 450.0, 800.0, 700.0, 750.0, 1200.0, 100.0, 600.0, 600.0, 450.0, 700.0, 600.0, 350.0, 1300.0, 600.0, 2000.0, 550.0, 1500.0, 600.0, 800.0, 350.0, 700.0, 200.0, 600.0, 750.0, 300.0, 800.0, 550.0, 650.0, 650.0, 300.0, 200.0, 200.0, 250.0, 100.0, 450.0, 700.0, 850.0, 550.0, 350.0, 800.0, 1100.0, 150.0, 1200.0, 650.0, 1900.0, 1200.0, 600.0, 1300.0, 650.0, 350.0, 800.0, 800.0, 200.0, 700.0, 600.0, 600.0, 700.0, 300.0, 200.0, 300.0, 600.0, 450.0, 100.0, 350.0, 600.0, 350.0, 700.0, 300.0, 600.0, 600.0, 300.0, 600.0, 600.0, 150.0, 300.0, 1500.0, 650.0, 900.0, 550.0, 300.0, 800.0, 1000.0, 200.0, 1000.0, 1200.0, 200.0, 350.0, 900.0, 1200.0, 700.0, 600.0, 900.0, 1000.0, 350.0, 350.0, 2000.0, 450.0, 450.0, 1200.0, 350.0, 250.0, 100.0, 250.0, 1100.0, 300.0, 200.0, 650.0, 600.0, 300.0, 350.0, 300.0, 600.0, 200.0, 300.0, 900.0, 650.0, 800.0, 650.0, 250.0, 450.0, 1300.0, 800.0, 600.0, 350.0, 200.0, 350.0, 750.0, 250.0, 100.0, 600.0, 550.0, 550.0, 2000.0, 450.0, 300.0, 300.0, 1500.0, 450.0, 300.0, 1500.0, 200.0, 600.0, 800.0, 600.0, 600.0, 450.0, 900.0, 150.0, 1400.0, 700.0, 1300.0, 450.0, 300.0, 800.0, 300.0, 700.0, 350.0, 600.0, 600.0, 600.0, 300.0, 550.0, 600.0, 650.0, 150.0, 350.0, 1100.0, 250.0, 750.0, 200.0, 650.0, 800.0, 250.0, 600.0, 700.0, 200.0, 300.0, 1000.0, 200.0, 1100.0, 150.0, 800.0, 550.0, 550.0, 150.0, 700.0, 350.0, 600.0, 1300.0, 600.0, 350.0, 650.0, 650.0, 700.0, 600.0, 350.0, 1000.0, 300.0, 300.0, 300.0, 600.0, 450.0, 1300.0, 2000.0, 1500.0, 1500.0, 600.0, 450.0, 250.0, 100.0, 300.0, 1000.0, 550.0, 300.0, 200.0, 150.0, 200.0, 150.0, 200.0, 200.0, 300.0, 350.0, 100.0, 350.0, 300.0, 600.0, 700.0, 200.0, 450.0, 1300.0, 700.0, 350.0, 700.0, 350.0, 650.0, 300.0, 300.0, 200.0, 150.0, 700.0, 300.0, 850.0, 700.0, 300.0, 1400.0, 300.0, 600.0, 1000.0, 650.0, 250.0, 600.0, 300.0, 600.0, 1500.0, 1600.0, 350.0, 650.0, 450.0, 600.0, 2000.0, 600.0, 700.0, 600.0, 600.0, 800.0, 800.0, 350.0, 1100.0, 150.0, 800.0, 350.0, 1250.0, 100.0, 200.0, 650.0, 600.0, 150.0, 350.0, 350.0, 600.0, 150.0, 600.0, 250.0, 150.0, 300.0, 700.0, 650.0, 300.0, 350.0, 300.0, 300.0, 700.0, 600.0, 1200.0, 700.0, 2000.0, 800.0, 1000.0, 600.0, 600.0, 100.0, 300.0, 300.0, 300.0, 300.0, 600.0, 600.0, 700.0, 700.0, 800.0, 350.0, 350.0,

48/65

0.0, 750.0, 400.0, 400.0, 100.0, 300.0, 200.0, 200.0, 400.0, 400.0, 550.0, 100.0, 300.0, 300.0, 850.0, 1500.0, 1000.0, 850.0, 1000.0, 1500.0, 1500.0, 1000.0, 1500.0, 800.0, 1200.0, 1500.0, 1800.0, 1000.0, 1900.0, 1200.0, 1000.0, 800.0, 800.0, 750.0, 1700.0, 1500.0, 1000.0, 1500.0, 1400.0, 850.0, 1800.0, 700.0, 2000.0, 900.0, 1300.0, 850.0, 1100.0, 800.0, 1000.0, 1000.0, 850.0, 2000.0, 850.0, 1500.0, 1200.0, 1000.0, 1000.0, 1000.0, 1200.0, 1200.0, 1200.0, 1500.0, 1000.0, 1400.0, 1500.0, 1800.0, 1200.0, 1400.0, 1600.0, 1800.0, 800.0, 1300.0, 2000.0, 900.0, 2000.0, 1500.0, 1000.0, 1200.0, 1400.0, 1000.0, 1000.0, 950.0, 300.0, 1800.0, 2000.0, 2000.0, 600.0, 1500.0, 800.0, 600.0, 500.0, 800.0, 1050.0, 900.0, 600.0, 750.0, 500.0, 700.0, 1800.0, 1300.0, 800.0, 400.0, 700.0, 300.0, 650.0, 1400.0, 1800.0, 600.0, 600.0, 650.0, 550.0, 850.0, 300.0, 550.0, 1000.0, 600.0, 1200.0, 600.0, 1100.0, 1000.0, 800.0, 500.0, 1300.0, 1000.0, 450.0, 350.0, 500.0, 1200.0, 700.0, 1400.0, 500.0, 750.0, 700.0, 400.0, 750.0, 1200.0, 900.0, 600.0, 900.0, 600.0, 550.0, 1400.0, 1000.0, 550.0, 550.0, 400.0, 250.0, 800.0, 700.0, 650.0, 550.0, 450.0, 500.0, 1200.0, 400.0, 1000.0, 700.0, 1500.0, 0.0, 1400.0, 500.0, 500.0, 500.0, 1400.0, 1800.0, 2000.0, 2000.0, 2000.0, 2000.0, 1000.0, 500.0, 400.0, 800.0, 550.0, 800.0, 0.0, 800.0, 300.0, 450.0, 900.0, 700.0, 1500.0, 900.0, 2000.0, 700.0, 550.0, 500.0, 600.0, 300.0, 2000.0, 550.0, 1500.0, 500.0, 800.0, 800.0, 550.0, 700.0, 350.0, 500.0, 150.0, 400.0, 500.0, 700.0, 1000.0, 500.0, 500.0, 400.0, 900.0, 1400.0, 2000.0, 2000.0, 250.0, 2000.0, 800.0, 2000.0, 600.0, 700.0, 600.0, 1400.0, 1600.0, 1500.0, 400.0, 650.0, 1000.0, 500.0, 2000.0, 800.0, 1000.0, 500.0, 650.0, 900.0, 400.0, 600.0, 800.0, 1400.0, 350.0, 500.0, 1000.0, 350.0, 1500.0, 700.0, 850.0, 450.0, 1500.0, 600.0, 700.0, 550.0, 500.0, 2000.0, 800.0, 2000.0, 2000.0, 2000.0, 500.0, 1200.0, 1000.0, 1500.0, 1000.0, 350.0, 350.0, 400.0, 1000.0, 2000.0, 1400.0, 1200.0, 1200.0, 800.0, 200.0, 1000.0, 500.0, 1000.0, 650.0, 2000.0, 2500.0, 800.0, 2000.0, 1000.0, 2000.0, 1000.0, 2000.0, 200.0, 1200.0, 200.0, 500.0, 600.0, 200.0, 900.0, 700.0, 2000.0, 2000.0, 1600.0, 600.0, 2000.0, 1400.0, 500.0, 2000.0, 400.0, 2000.0, 2000.0, 600.0, 1000.0, 700.0, 800.0, 650.0, 2000.0, 2000.0, 900.0, 300.0, 1000.0, 700.0, 1000.0, 1000.0, 400.0, 2000.0, 650.0, 2000.0, 800.0, 500.0, 1300.0, 800.0, 1300.0, 600.0, 500.0, 2000.0, 2000.0, 700.0, 1200.0, 500.0, 400.0, 500.0, 450.0, 600.0, 1200.0, 350.0, 450.0, 500.0, 400.0, 300.0, 300.0, 1000.0, 850.0, 2000.0, 850.0, 650.0, 500.0, 1600.0, 2000.0, 2000.0, 1500.0, 400.0, 800.0, 1000.0, 400.0, 700.0, 1100.0, 500.0, 1200.0, 350.0, 2000.0, 600.0, 600.0, 450.0, 1000.0, 600.0, 500.0, 1400.0, 900.0, 2000.0, 350.0, 1200.0, 500.0, 2000.0, 500.0, 800.0, 800.0, 500.0, 1400.0, 550.0, 400.0, 2000.0, 600.0, 650.0, 1200.0, 2000.0, 150.0, 2000.0, 800.0, 1500.0, 1000.0, 1200.0, 1900.0, 450.0, 400.0, 1500.0, 1100.0, 1000.0, 600.0, 500.0, 700.0, 800.0, 700.0, 400.0, 1000.0, 800.0, 900.0, 1400.0, 1000.0, 800.0, 600.0, 2000.0, 0.0, 800.0, 700.0, 600.0, 2000.0, 2000.0, 1500.0, 600.0, 500.0, 400.0, 1100.0, 450.0, 1200.0, 400.0, 1400.0, 2000.0, 2100.0, 350.0, 2100.0, 1000.0, 800.0, 900.0, 700.0, 800.0, 1600.0, 1100.0, 800.0, 700.0, 800.0, 1200.0, 290.0, 1000.0, 1000.0, 100.0, 2100.0, 600.0, 150.0, 2100.0, 650.0, 450.0, 1000.0, 500.0, 200.0, 450.0, 700.0, 400.0, 600.0, 500.0, 1100.0, 900.0, 900.0, 500.0, 400.0, 500.0, 450.0, 450.0, 800.0, 950.0, 250.0, 600.0, 500.0, 1000.0, 750.0, 100.0, 600.0, 400.0, 800.0, 2100.0, 600.0, 750.0, 500.0, 2100.0, 700.0, 1000.0, 1000.0, 1000.0, 250.0, 600.0, 1000.0, 600.0, 500.0, 1000.0, 900.0, 2100.0, 1700.0, 900.0, 800.0, 1000.0, 600.0, 450.0, 400.0, 700.0, 450.0, 500.0, 500.0, 450.0, 300.0, 650.0, 350.0, 350.0, 500.0, 350.0, 450.0, 600.0, 2100.0, 800.0, 500.0, 1300.0, 700.0, 400.0, 300.0, 500.0, 700.0, 400.0, 500.0, 600.0, 2100.0, 800.0, 2100.0, 250.0, 500.0, 600.0, 0.0, 600.0, 2200.0, 850.0, 550.0, 600.0, 700.0, 300.0, 400.0, 1000.0, 1600.0, 800.0, 0.0, 600.0, 150.0, 600.0, 500.0, 2200.0, 800.0, 1100.0, 1650.0, 1500.0, 500.0, 850.0, 650.0, 1600.0, 800.0, 800.0, 800.0, 500.0, 1100.0, 400.0, 400.0, 2200.0, 800.0, 500.0, 450.0, 1500.0, 450.0, 150.0, 600.0, 300.0, 600.0, 1200.0, 0.0, 600.0, 700.0, 600.0, 650.0, 400.0, 1600.0, 1200.0, 1300.0, 1400.0, 1400.0, 700.0, 600.0, 500.0, 650.0, 1800.0, 1000.0, 1000.0, 1200.0, 1100.0, 1500.0, 500.0, 1500.0, 700.0, 2200.0, 700.0, 2200.0, 400.0, 400.0, 250.0, 600.0, 600.0, 400.0, 300.0, 500.0, 650.0, 550.0, 1400.0, 450.0, 650.0, 600.0, 700.0, 1500.0, 250.0, 800.0, 700.0, 0.0, 1000.0, 600.0, 550.0, 400.0, 2200.0, 1400.0, 2200.0, 800.0, 1400.0, 2200.0, 1000.0, 1000.0, 600.0, 1000.0, 1200.0, 2200.0, 150.0, 2200.0, 800.0, 300.0, 450.0, 2200.0, 500.0, 400.0, 1400.0, 1300.0, 800.0, 800.0, 700.0, 600.0, 150.0, 500.0, 700.0, 500.0, 1000.0, 1500.0, 60

50/65

51/65

52/65

50.0301451541145, 2535.5849374875966, 2535.5849374875966, 2535.5849374875966, 5247.749572162009, 2500, 147.78722402265447]

```
In [79]: print(len(final_amount))
```

9551

```
In [80]: df2['New_AVERAGE_COST_FOR_TWO'] = final_amount
```

```
In [81]: p1_min,p1_max = df2[df2['PRICE_RANGE'] == 1]['New_AVERAGE_COST_FOR_TWO'].agg(['min', 'max'])
p2_min,p2_max = df2[df2['PRICE_RANGE'] == 2]['New_AVERAGE_COST_FOR_TWO'].agg(['min', 'max'])
p3_min,p3_max = df2[df2['PRICE_RANGE'] == 3]['New_AVERAGE_COST_FOR_TWO'].agg(['min', 'max'])
p4_min,p4_max = df2[df2['PRICE_RANGE'] == 4]['New_AVERAGE_COST_FOR_TWO'].agg(['min', 'max'])

temp_dict = {1:f'{round(p1_min)} - {round(p1_max)}',
              2:f'{round(p2_min)} - {round(p2_max)}',
              3:f'{round(p3_min)} - {round(p3_max)}',
              4:f'{round(p4_min)} - {round(p4_max)}'}

temp_dict_df = pd.DataFrame(temp_dict,index = [0])
temp_dict_df
```

```
Out[81]:
```

	1	2	3	4
0	0 - 1749	40 - 4082	60 - 6997	100 - 44387

```
In [82]: print(len(final_amount))
```

9551

```
In [ ]: # ----- Created By Ankit Kumar -----
```

```
In [83]: p1_min,p1_max = df2[df2['PRICE_RANGE'] == 1]['New_AVERAGE_COST_FOR_TWO'].agg(['min', 'max'])
p2_min,p2_max = df2[df2['PRICE_RANGE'] == 2]['New_AVERAGE_COST_FOR_TWO'].agg(['min', 'max'])
p3_min,p3_max = df2[df2['PRICE_RANGE'] == 3]['New_AVERAGE_COST_FOR_TWO'].agg(['min', 'max'])
p4_min,p4_max = df2[df2['PRICE_RANGE'] == 4]['New_AVERAGE_COST_FOR_TWO'].agg(['min', 'max'])

temp_dict = {'AVG_PRICE_RANGE': {1:f'{round(p1_min)} - {round(p1_max)}',
                                  2:f'{round(p2_min)} - {round(p2_max)}',
                                  3:f'{round(p3_min)} - {round(p3_max)}',
                                  4:f'{round(p4_min)} - {round(p4_max)}'}}

temp_dict_df = pd.DataFrame(temp_dict)
temp_dict_df
```

```
Out[83]:
```

	AVG_PRICE_RANGE
1	0 - 1749
2	40 - 4082
3	60 - 6997
4	100 - 44387

```
In [84]: p1_min,p1_max = df2[df2['PRICE_RANGE'] == 1]['New_AVERAGE_COST_FOR_TWO'].agg(['min', 'max'])
p2_min,p2_max = df2[df2['PRICE_RANGE'] == 2]['New_AVERAGE_COST_FOR_TWO'].agg(['min', 'max'])
```

```

p3_min,p3_max = df2[df2['PRICE_RANGE'] == 3]['New_AVERAGE_COST_FOR_TWO'].agg(['min','max'])
p4_min,p4_max = df2[df2['PRICE_RANGE'] == 4]['New_AVERAGE_COST_FOR_TWO'].agg(['min','max'])

temp_dict = {'AVG_PRICE_RANGE': {1:f'{round(p1_min)} - {round(p1_max)}',
                                2:f'{round(p2_min)} - {round(p2_max)}',
                                3:f'{round(p3_min)} - {round(p3_max)}',
                                4:f'{round(p4_min)} - {round(p4_max)}'}}

temp_dict_df = pd.DataFrame(temp_dict)
temp_dict_df.loc[2,:].values[0]

```

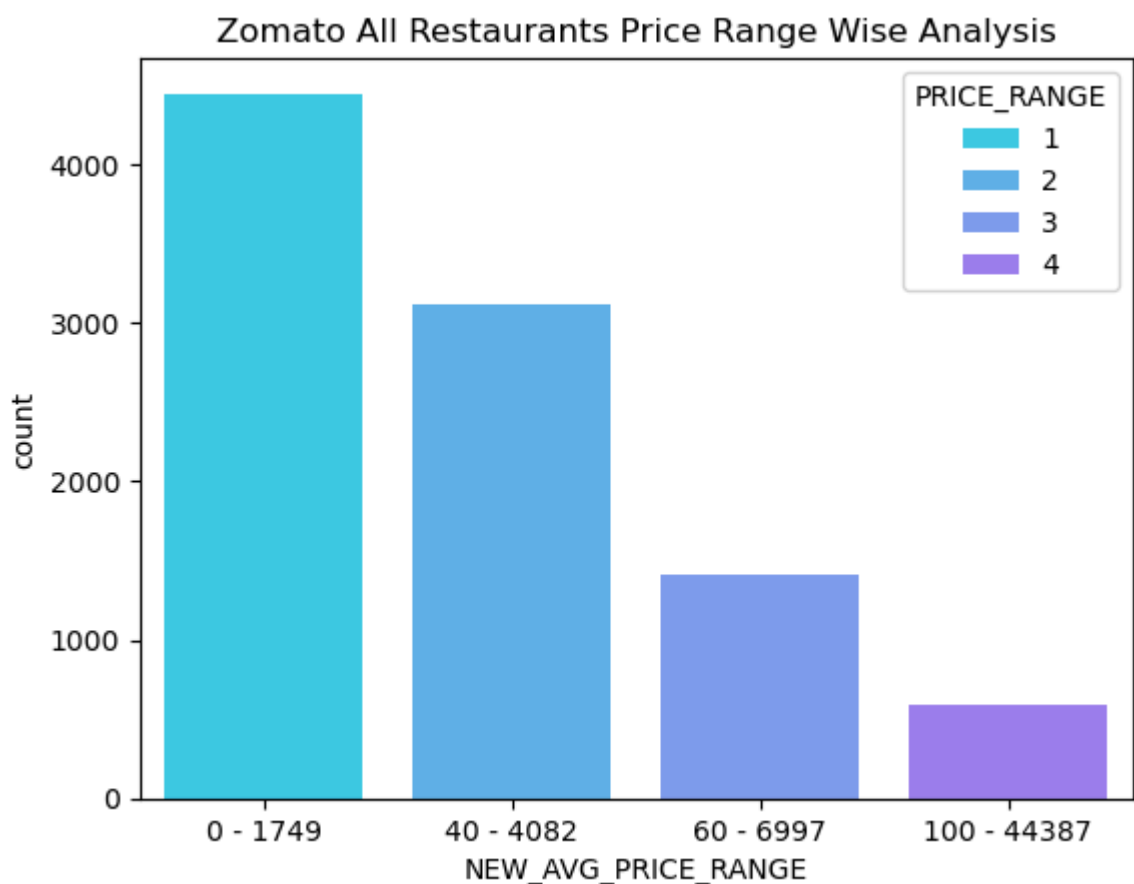
Out[84]: '40 - 4082'

In [85]: df2['NEW_AVG_PRICE_RANGE'] = df2['PRICE_RANGE'].apply(lambda row:temp_dict_df.loc

In [86]: df2['NEW_AVG_PRICE_RANGE'].value_counts()

Out[86]: NEW_AVG_PRICE_RANGE
0 - 1749 4444
40 - 4082 3113
60 - 6997 1408
100 - 44387 586
Name: count, dtype: int64

In [87]: plt.title('Zomato All Restaurants Price Range Wise Analysis')
sns.countplot(data = df2, x = 'NEW_AVG_PRICE_RANGE',hue = 'PRICE_RANGE',palette=
plt.show())



In [88]: df2.columns

```
Out[88]: Index(['RESTAURANTID', 'RESTAURANTNAME', 'COUNTRYCODE', 'CITY', 'ADDRESS',  
              'LOCALITY', 'LOCALITYVERBOSE', 'LONGITUDE', 'LATITUDE', 'CUISINES',  
              'CURRENCY', 'HAS_TABLE_BOOKING', 'HAS_ONLINE_DELIVERY',  
              'IS_DELIVERING_NOW', 'SWITCH_TO_ORDER_MENU', 'PRICE_RANGE', 'VOTES',  
              'AVERAGE_COST_FOR_TWO', 'RATING', 'DATEKEY_OPENING', 'CUISINES 1',  
              'CURRENCY_TEMP', 'CURRENCY_CODE', 'New_AVERAGE_COST_FOR_TWO',  
              'NEW_AVG_PRICE_RANGE'],  
             dtype='object')
```

```
In [ ]:
```

Multivariate Analysis

```
In [89]: df2['NEW_AVG_PRICE_RANGE']
```

```
Out[89]: 0      0 - 1749  
         1      0 - 1749  
         2      0 - 1749  
         3      0 - 1749  
         4      0 - 1749  
         ...  
        9546    60 - 6997  
        9547    60 - 6997  
        9548    60 - 6997  
        9549    60 - 6997  
        9550    60 - 6997  
        Name: NEW_AVG_PRICE_RANGE, Length: 9551, dtype: object
```

```
In [90]: df2['New_AVERAGE_COST_FOR_TWO']
```

```
Out[90]: 0      300.000000  
         1      200.000000  
         2      400.000000  
         3      100.000000  
         4      150.000000  
         ...  
        9546    2535.584937  
        9547    2535.584937  
        9548    5247.749572  
        9549    2500.000000  
        9550     147.787224  
        Name: New_AVERAGE_COST_FOR_TWO, Length: 9551, dtype: float64
```

```
In [91]: df2['RESTAURANTNAME'].value_counts().head(50)
```

```
Out[91]:
```

RESTAURANTNAME	
CAFE COFFEE DAY	83
DOMINO'S PIZZA	79
SUBWAY	63
GREEN CHICK CHOP	51
MCDONALD'S	48
KEVENTERS	34
PIZZA HUT	30
GIANI	29
BASKIN ROBBINS	28
BARBEQUE NATION	26
DUNKIN' DONUTS	22
BARISTA	22
GIANI'S	22
COSTA COFFEE	20
PIND BALLUCHI	20
TWENTY FOUR SEVEN	19
PIZZA HUT DELIVERY	19
SAGAR RATNA	19
WAH JI WAH	19
CHAAYOS	18
KFC	18
REPUBLIC OF CHICKEN	18
STARBUCKS	18
BURGER KING	16
HALDIRAM'S	16
SHREE RATHNAM	15
BIKANERVALA	14
FRONTIER	14
MOTI MAHAL DELUX	14
AGGARWAL SWEETS	14
KARIM'S	13
BIKANER SWEETS	13
BEHROUZ BIRYANI	13
34, CHOWRINGHEE LANE	12
APNI RASOI	12
CHICAGO PIZZA	12
MADRAS CAFE	11
BURGER POINT	11
WOW! MOMO	11
GOPALA	10
YO! CHINA	9
BERCO'S	9
SHAMA CHICKEN CORNER	9
NIRULA'S ICE CREAM	9
COCOBERRY	9
SARDAR A PURE MEAT SHOP	9
PUNJABI TADKA	8
PUNJABI CHAAP CORNER	8
FAASOS	8
OVENSTORY PIZZA	8

Name: count, dtype: int64

```
In [92]: df2['RESTAURANTNAME'].value_counts().head(50).index
```



```
Out[92]: Index(['CAFE COFFEE DAY', 'DOMINO'S PIZZA', 'SUBWAY', 'GREEN CHICK CHOP',
               'MCDONALD'S', 'KEVENTERS', 'PIZZA HUT', 'GIANI', 'BASKIN ROBBINS',
               'BARBEQUE NATION', 'DUNKIN' DONUTS', 'BARISTA', 'GIANI'S',
               'COSTA COFFEE', 'PIND BALLUCHI', 'TWENTY FOUR SEVEN',
               'PIZZA HUT DELIVERY', 'SAGAR RATNA', 'WAH JI WAH', 'CHAAYOS', 'KFC',
               'REPUBLIC OF CHICKEN', 'STARBUCKS', 'BURGER KING', 'HALDIRAM'S',
               'SHREE RATHNAM', 'BIKANERVALA', 'FRONTIER', 'MOTI MAHAL DELUX',
               'AGGARWAL SWEETS', 'KARIM'S', 'BIKANER SWEETS', 'BEHROUZ BIRYANI',
               '34, CHOWRINGHEE LANE', 'APNI RASOI', 'CHICAGO PIZZA', 'MADRAS CAFE',
               'BURGER POINT', 'WOW! MOMO', 'GOPALA', 'YO! CHINA', 'BERCO'S',
               'SHAMA CHICKEN CORNER', 'NIRULA'S ICE CREAM', 'COCOBERRY',
               'SARDAR A PURE MEAT SHOP', 'PUNJABI TADKA', 'PUNJABI CHAAP CORNER',
               'FAASOS', 'OVENSTORY PIZZA'],
              dtype='object', name='RESTAURANTNAME')
```

```
In [93]: df2.groupby(['RESTAURANTNAME'])['New_AVERAGE_COST_FOR_TWO'].mean().sort_values(a
```

```
Out[93]: RESTAURANTNAME
RESTAURANT ANDRE      44387.050547
JAAN                  38172.863471
RHUBARB LE RESTAURANT 27963.841845
RESTAURANT GORDON RAMSAY 26821.831147
SUMMER PAVILION       26632.230328
...
URBANCRAVE           0.000000
SHEROES HANGOUT      0.000000
SENROR IGUANAS        0.000000
DEENA CHAT BHANDAR    0.000000
HI LITE BAR & LOUNGE  0.000000
Name: New_AVERAGE_COST_FOR_TWO, Length: 7433, dtype: float64
```

```
In [94]: df2.groupby(['RESTAURANTNAME', 'HAS_TABLE_BOOKING'])['New_AVERAGE_COST_FOR_TWO'].
```

```

Out[94]:
  RESTAURANTNAME  HAS_TABLE_BOOKING  44387.050547
  RESTAURANT ANDRE  NO  38172.863471
  JAAN  NO  27963.841845
  RHUBARB LE RESTAURANT  NO  26821.831147
  RESTAURANT GORDON RAMSAY  NO  26632.230328
  SKY ON 57  NO  26632.230328
  SUMMER PAVILION  NO  23969.007295
  CUT BY WOLFGANG PUCK  NO  19530.302241
  COLONY  NO  18658.665145
  THE FRENCH BY SIMON ROGAN - THE MIDLAND  NO  16463.626169
  RESTAURANT MOSAIC @ THE ORIENT  NO  13993.998859
  PURNELL'S  NO  13993.998859
  HAKKASAN  YES  11661.665716
  THE WITCHERY & THE SECRET GARDEN  YES  11661.665716
  SKETCH GALLERY  NO  11661.665716
  NOBU  YES  10652.892131
  PIER 70  NO  10495.499144
  THE KITCHIN  NO  10495.499144
  YAUATCHA  YES  10142.339750
  HIPPOPOTAMUS - MUSEUM HOTEL  NO  9912.415859
  MANCHESTER HOUSE  NO  9635.222762
  EIGHT - THE LANGHAM HOTEL  NO  9329.332573
  LASAN RESTAURANT  NO  9329.332573
  GAUCHO  NO  8877.410109
  TEXAS DE BRAZIL  NO  8877.410109
  FRATINI LA TRATTORIA  NO  8433.539604
  SUPER LOCO  NO  8000.000000
  ORIENT EXPRESS - TAJ PALACE HOTEL  YES  7898.437476
  CUBE - TASTING KITCHEN  NO  7101.928088
  POTATO HEAD FOLK  NO  7101.928088
  THE REFINERY SINGAPORE  NO  7000.000000
  TIAN - ASIAN CUISINE STUDIO - ITC MAURYA  NO  6996.999430
  BANK  YES  6996.999430
  ROKA  NO  6658.057582
  ARTICHOKE CAFE  NO  6626.118242
  TERRAÍ_O ITÍÇLIA  NO  6500.000000
  BUKHARA - ITC MAURYA  NO  6413.916144
  STEAK  NO  6413.916144
  MR COOPER'S HOUSE & GARDEN - THE MIDLAND  YES  6413.916144
  THE GRILL ON THE ALLEY  NO  6413.916144
  DUCK & WAFFLE  NO  6214.187077
  CHANDLERS STEAKHOUSE  NO  6214.187077
  HENRY CAMPBELL'S STEAKHOUSE  NO  6214.187077
  BERN'S STEAK HOUSE  NO  6214.187077
  DUKE'S WAIKIKI  NO  6214.187077
  BARBACOA RESTAURANT  NO  6214.187077
  NATALIA'S  NO  6214.187077
  VIC'S ON THE RIVER  NO  6214.187077
  ROCKS ON THE RIVER  NO  6214.187077
  DUCK CITY BISTRO  NO  6214.187077
  KAHILL'S STEAK-FISH CHOPHOUSE  NO  6214.187077
  Name: New_AVERAGE_COST_FOR_TWO, dtype: float64

```

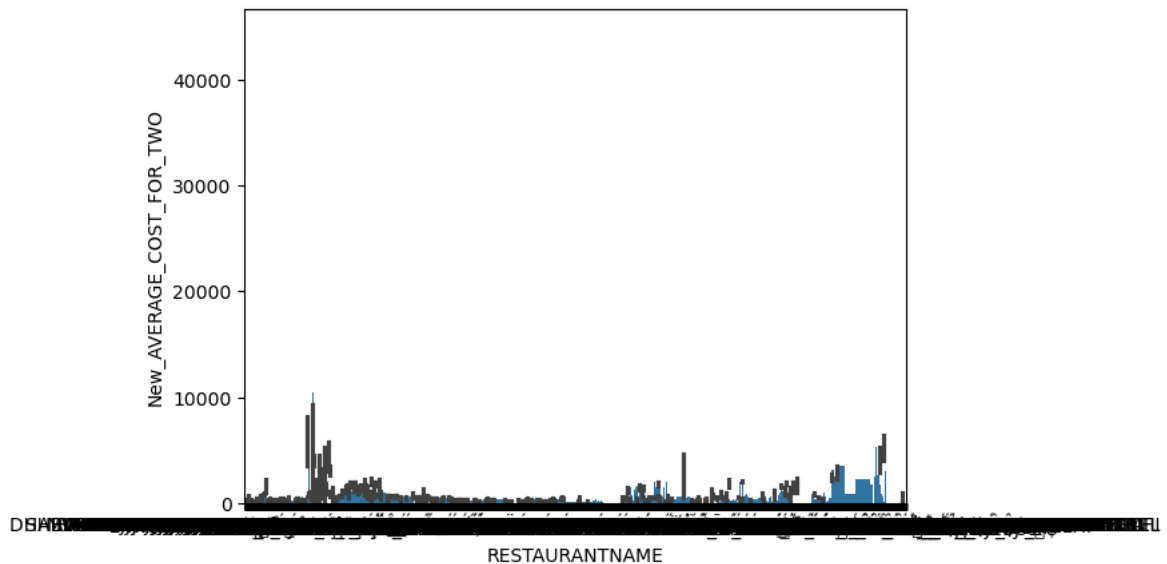
```

In [95]: df2.columns

```

```
Out[95]: Index(['RESTAURANTID', 'RESTAURANTNAME', 'COUNTRYCODE', 'CITY', 'ADDRESS',
              'LOCALITY', 'LOCALITYVERBOSE', 'LONGITUDE', 'LATITUDE', 'CUISINES',
              'CURRENCY', 'HAS_TABLE_BOOKING', 'HAS_ONLINE_DELIVERY',
              'IS_DELIVERING_NOW', 'SWITCH_TO_ORDER_MENU', 'PRICE_RANGE', 'VOTES',
              'AVERAGE_COST_FOR_TWO', 'RATING', 'DATEKEY_OPENING', 'CUISINES 1',
              'CURRENCY_TEMP', 'CURRENCY_CODE', 'New_AVERAGE_COST_FOR_TWO',
              'NEW_AVG_PRICE_RANGE'],
              dtype='object')
```

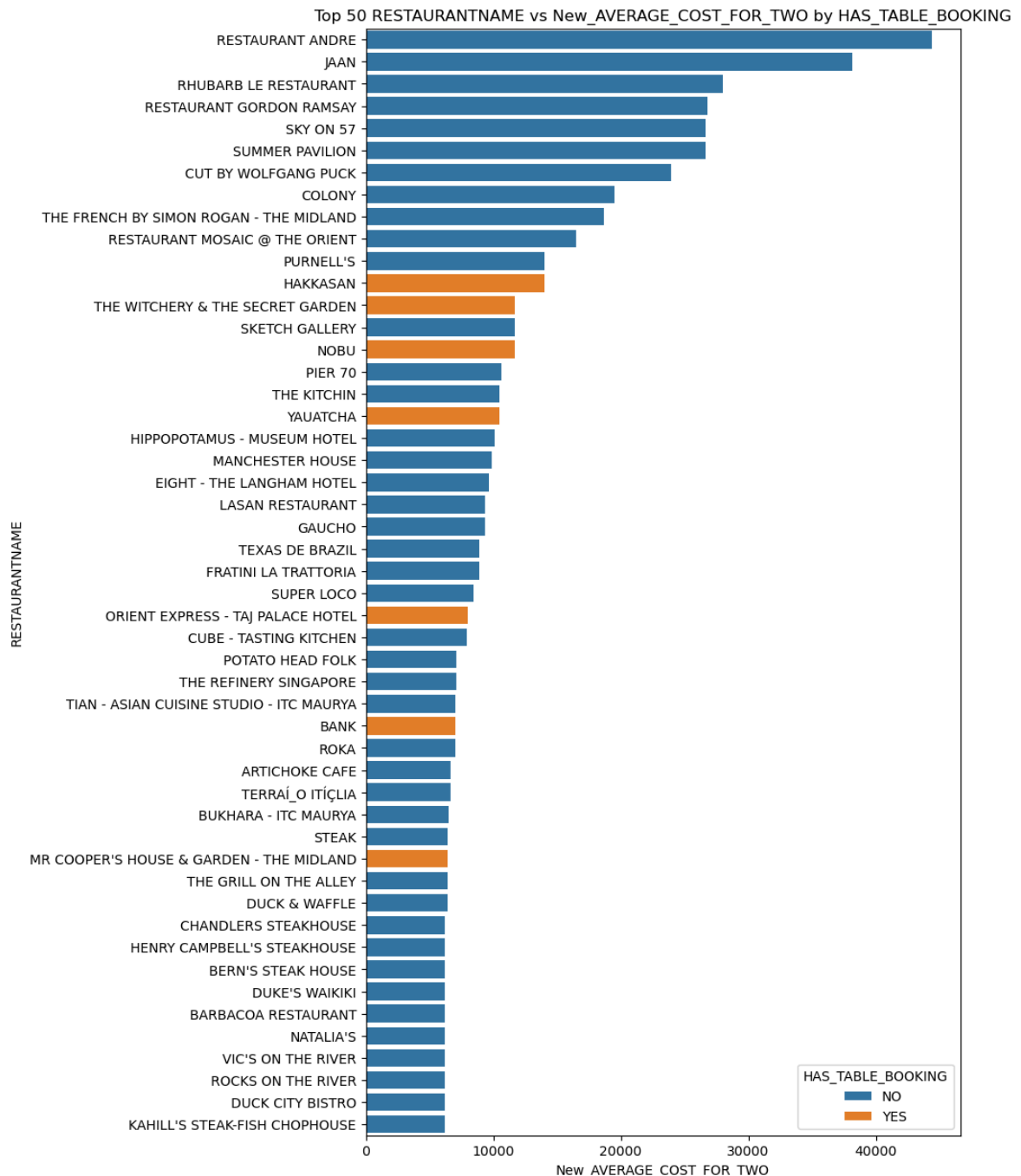
```
In [96]: sns.barplot(data = df2, x= 'RESTAURANTNAME', y = 'New_AVERAGE_COST_FOR_TWO')
plt.show()
```



```
In [97]: temp_df1 = df2.groupby(['RESTAURANTNAME', 'HAS_TABLE_BOOKING'])['New_AVERAGE_COST']
```

```
In [98]: temp_df1 = temp_df1.reset_index()
```

```
In [99]: plt.figure(figsize = (8,15))
plt.title('Top 50 RESTAURANTNAME vs New_AVERAGE_COST_FOR_TWO by HAS_TABLE_BOOKING')
sns.barplot(data = temp_df1, y = 'RESTAURANTNAME', x = 'New_AVERAGE_COST_FOR_TWO')
plt.show()
```



```
In [100... df2.columns
```

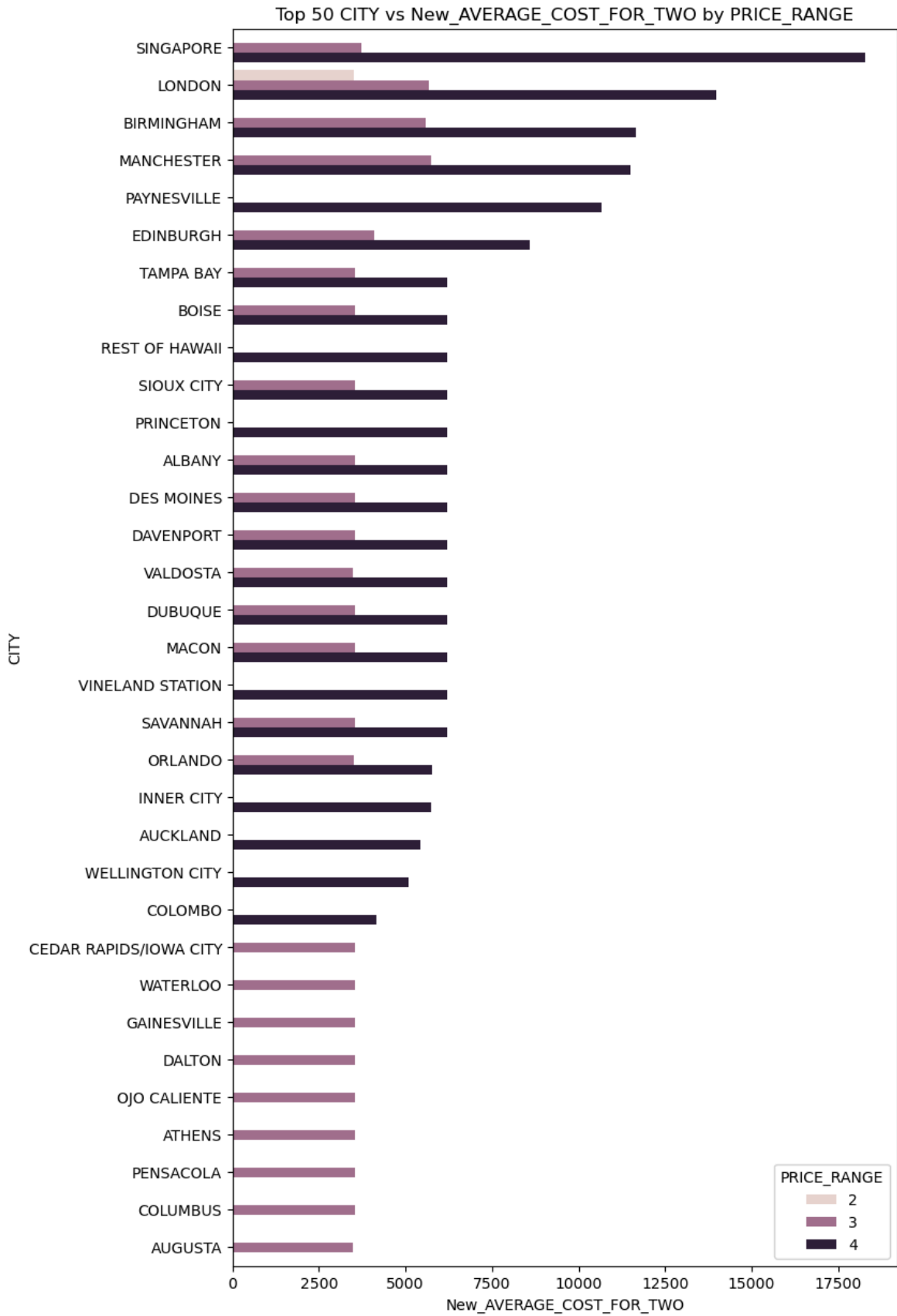
```
Out[100... Index(['RESTAURANTID', 'RESTAURANTNAME', 'COUNTRYCODE', 'CITY', 'ADDRESS',
      'LOCALITY', 'LOCALITYVERBOSE', 'LONGITUDE', 'LATITUDE', 'CUISINES',
      'CURRENCY', 'HAS_TABLE_BOOKING', 'HAS_ONLINE_DELIVERY',
      'IS_DELIVERING_NOW', 'SWITCH_TO_ORDER_MENU', 'PRICE_RANGE', 'VOTES',
      'AVERAGE_COST_FOR_TWO', 'RATING', 'DATEKEY_OPENING', 'CUISINES 1',
      'CURRENCY_TEMP', 'CURRENCY_CODE', 'New_AVERAGE_COST_FOR_TWO',
      'NEW_AVG_PRICE_RANGE'],
      dtype='object')
```

```
In [101... temp_df2 = df2.groupby(['CITY', 'PRICE_RANGE'])['New_AVERAGE_COST_FOR_TWO'].mean()
```

```
In [102... temp_df2 = temp_df2.reset_index()
```

```
In [103... plt.figure(figsize = (8,15))
plt.title('Top 50 CITY vs New_AVERAGE_COST_FOR_TWO by PRICE_RANGE')
```

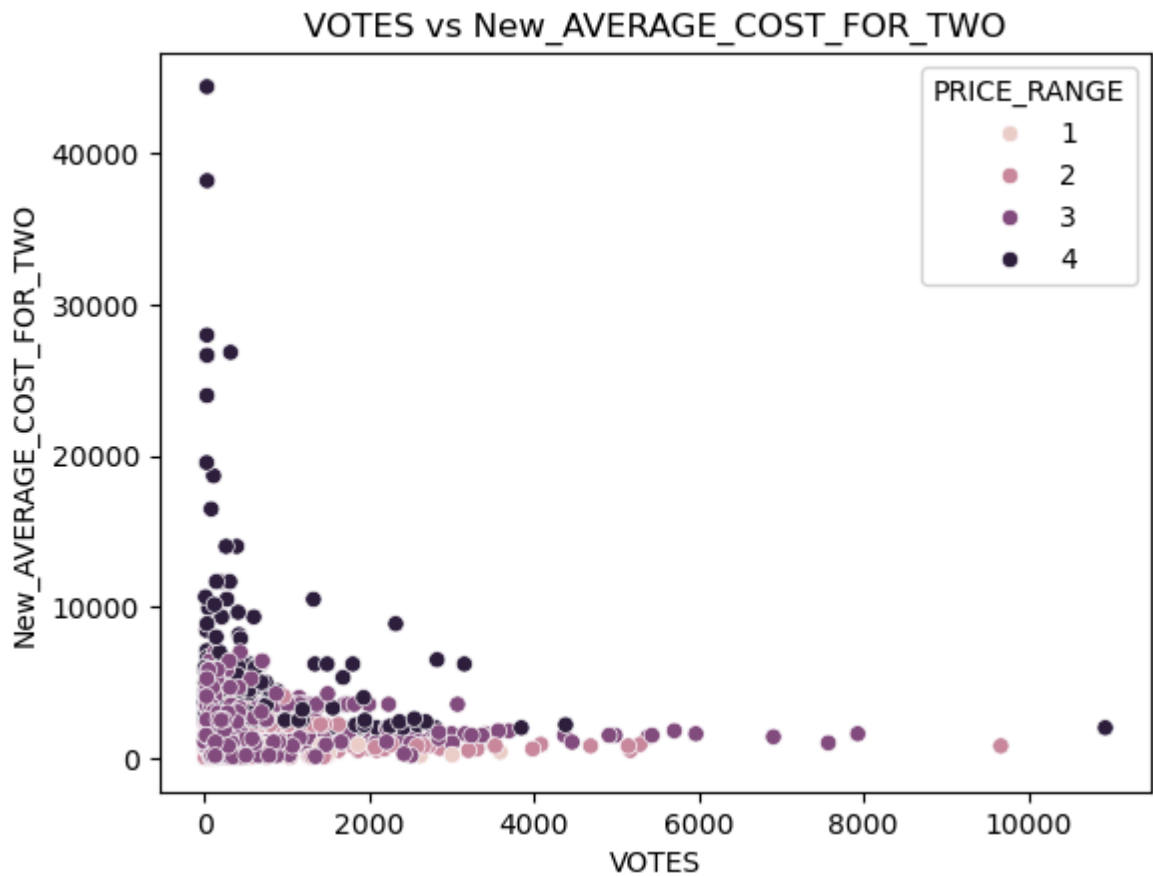
```
sns.barplot(data = temp_df2 , y = 'CITY', x = 'New_AVERAGE_COST_FOR_TWO', hue =  
plt.show()
```



In [104... df2.columns

```
Out[104... Index(['RESTAURANTID', 'RESTAURANTNAME', 'COUNTRYCODE', 'CITY', 'ADDRESS',  
      'LOCALITY', 'LOCALITYVERBOSE', 'LONGITUDE', 'LATITUDE', 'CUISINES',  
      'CURRENCY', 'HAS_TABLE_BOOKING', 'HAS_ONLINE_DELIVERY',  
      'IS_DELIVERING_NOW', 'SWITCH_TO_ORDER_MENU', 'PRICE_RANGE', 'VOTES',  
      'AVERAGE_COST_FOR_TWO', 'RATING', 'DATEKEY_OPENING', 'CUISINES 1',  
      'CURRENCY_TEMP', 'CURRENCY_CODE', 'New_AVERAGE_COST_FOR_TWO',  
      'NEW_AVG_PRICE_RANGE'],  
      dtype='object')
```

```
In [105... plt.title('VOTES vs New_AVERAGE_COST_FOR_TWO')  
sns.scatterplot(data = df2,x = 'VOTES',y = 'New_AVERAGE_COST_FOR_TWO',hue = 'PRI  
plt.show()
```



```
In [106... sns.pairplot(df2, hue = 'PRICE_RANGE')  
plt.show()
```



Findings and Outcomes & Insights

Zomato Restaurant Data – EDA Summary

Key Insights

- **Ratings & Price Range**
 - Moderate positive correlation (**0.46**) → higher-priced restaurants tend to have slightly better ratings.
 - **Votes correlate with ratings (0.35)** → restaurants with more popularity generally receive higher ratings.
 - **Votes vs Price Range correlation (0.31)** → premium restaurants attract more customer votes.
- **Country Distribution**
 - Majority of the records belong to **country code 0 (India)**.
- **Rating Trends Across Price Ranges**

- Most ratings are observed for **price range 1**, followed by **2 and 3**.
 - **Votes Distribution**
 - Most restaurants have **less than 2000 votes**, and a large portion under **100 votes**, indicating **low customer engagement**.
 - **Cuisine & Restaurant Type**
 - Zomato has more tie-ups with **Cafe, Pizza, Burger, and Coffee** restaurants.
 - Top cuisines include **North Indian, Chinese, Fast Food, and Café**.
 - **Geographical Concentration**
 - Over **80%** of restaurants listed are from **Delhi NCR**.
 - Strong clusters in **Connaught Place (CP), Rajouri Garden, and Shahdara**.
 - **Delivery & Dining Options**
 - **90%+** restaurants have **no table booking option**.
 - **80%+** do **not offer online delivery**, showing Zomato lists many **local/offline restaurants**.
 - **High-Cost Restaurants**
 - Premium restaurants like **Restaurant André, Le Restaurant, Gordon Ramsay, Summer Pavilion** show significantly **higher average cost for two**.
 - Cities such as **Singapore, London, Birmingham, Manchester** show **higher average cost for two** compared to Indian cities.
-

Conclusion

Zomato's dataset shows a strong **India-centric focus**, especially dominated by **Delhi NCR-based restaurants**. Most restaurants fall under the **affordable price range**, but these also see **lower votes and customer interaction**. Higher-priced restaurants generally perform better in terms of **ratings and votes**, indicating customers associate premium pricing with better quality or service. Additionally, limited online delivery and table booking options highlight untapped opportunities for Zomato to expand its services.

Suggestions / Recommendations

- Expand **online delivery partnerships** with offline restaurants.
 - Increase **customer engagement** through rating/voting incentives.
 - Improve geographical presence beyond **Delhi NCR**.
 - Highlight **low-cost but high-rated restaurants** to attract more customers.
 - Encourage restaurants to adopt **table booking** to improve dining experience.
 - Use correlation insights to help restaurants optimize **pricing + quality strategy**.
-

End of EDA Summary Report created by Ankit Kumar

