Zomato Sales Analysis

Big-Data Analysis with Python

Importing Libraries with Data Set

```
In [ ]:   1  Pandas is used for data manipulation and analysis.
          2  NumPy is used for numerical operators.
          3  Matplotlib.pyplot and Seaborn is used for data visualization.
```

```
In [5]:   1  import pandas as pd
          2  import numpy as np
          3  import matplotlib.pyplot as plt
          4  import seaborn as sns
```

```
In [6]:   1  data = pd.read_csv("Zomato data .csv")
```

```
In [7]:   1  data.head(10)
```

Out[7]:

	name	online_order	book_table	rate	votes	approx_cost(for two people)	listed_in(type)
0	Jalsa	Yes	Yes	4.1/5	775	800	Buffet
1	Spice Elephant	Yes	No	4.1/5	787	800	Buffet
2	San Churro Cafe	Yes	No	3.8/5	918	800	Buffet
3	Addhuri Udupi Bhojana	No	No	3.7/5	88	300	Buffet
4	Grand Village	No	No	3.8/5	166	600	Buffet
5	Timepass Dinner	Yes	No	3.8/5	286	600	Buffet
6	Rosewood International Hotel - Bar & Restaurant	No	No	3.6/5	8	800	Buffet
7	Onesta	Yes	Yes	4.6/5	2556	600	Cafes
8	Penthouse Cafe	Yes	No	4.0/5	324	700	other
9	Smacznego	Yes	No	4.2/5	504	550	Cafes

Let's convert the data type of the "rate" column to float and remove the denominator.

In [8]:
```python
def HandleRate(value):
    value = str(value).split('/')
    value = value[0];
    return float(value)
data['rate']=data['rate'].apply(HandleRate)
print(data.head())
```

```
                   name online_order book_table  rate  votes  \
0                 Jalsa          Yes        Yes   4.1    775
1        Spice Elephant          Yes         No   4.1    787
2        San Churro Cafe         Yes         No   3.8    918
3  Addhuri Udupi Bhojana          No         No   3.7     88
4          Grand Village          No         No   3.8    166

   approx_cost(for two people) listed_in(type)
0                          800          Buffet
1                          800          Buffet
2                          800          Buffet
3                          300          Buffet
4                          600          Buffet
```

Information of Data Set

```
In [9]:  1  data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148 entries, 0 to 147
Data columns (total 7 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   name                     148 non-null    object
 1   online_order             148 non-null    object
 2   book_table               148 non-null    object
 3   rate                     148 non-null    float64
 4   votes                    148 non-null    int64
 5   approx_cost(for two people)  148 non-null    int64
 6   listed_in(type)          148 non-null    object
dtypes: float64(1), int64(2), object(4)
memory usage: 8.2+ KB
```

Task.1. What type of restaurant do the majority of customers order from?

```
In [10]:    1  sns.countplot(x = data ['listed_in(type)'])
            2  plt.xlabel('Types of Restaurant')
```
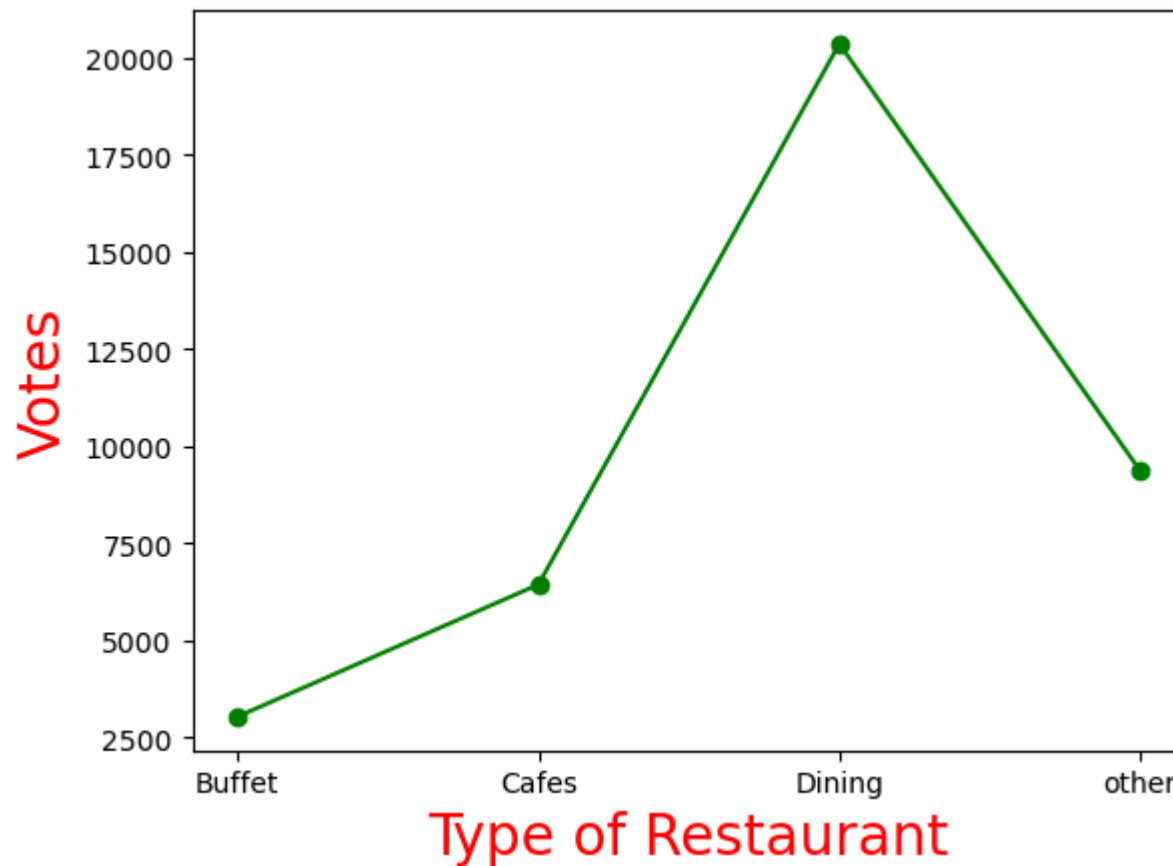
Out[10]: Text(0.5, 0, 'Types of Restaurant')
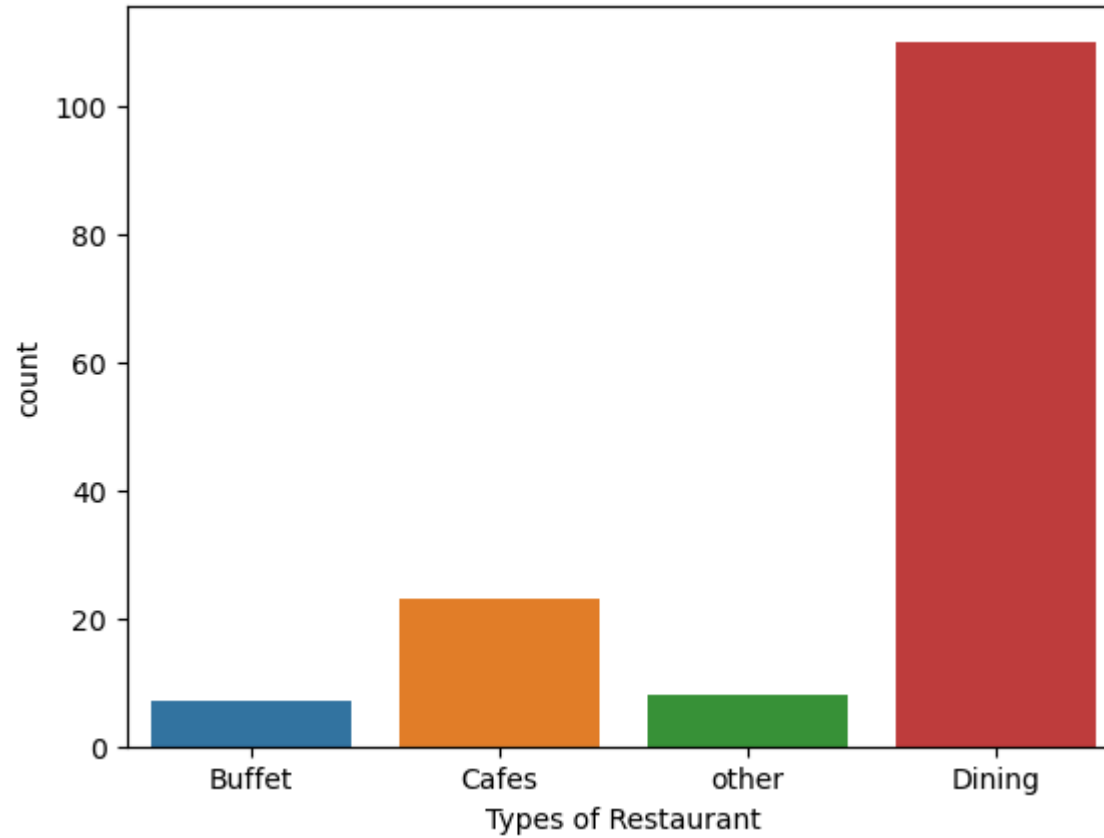
Task.2. How many votes has each type of restaurant received from customers?

```
In [15]:   1  grouped = data.groupby('listed_in(type)')['votes'].sum()
           2  result = pd.DataFrame({'votes': grouped})
           3  plt.plot(result, c="green", marker="o")
           4  plt.xlabel("Type of Restaurant", c="red", size=20)
           5  plt.ylabel("Votes", c="red", size=20)
```
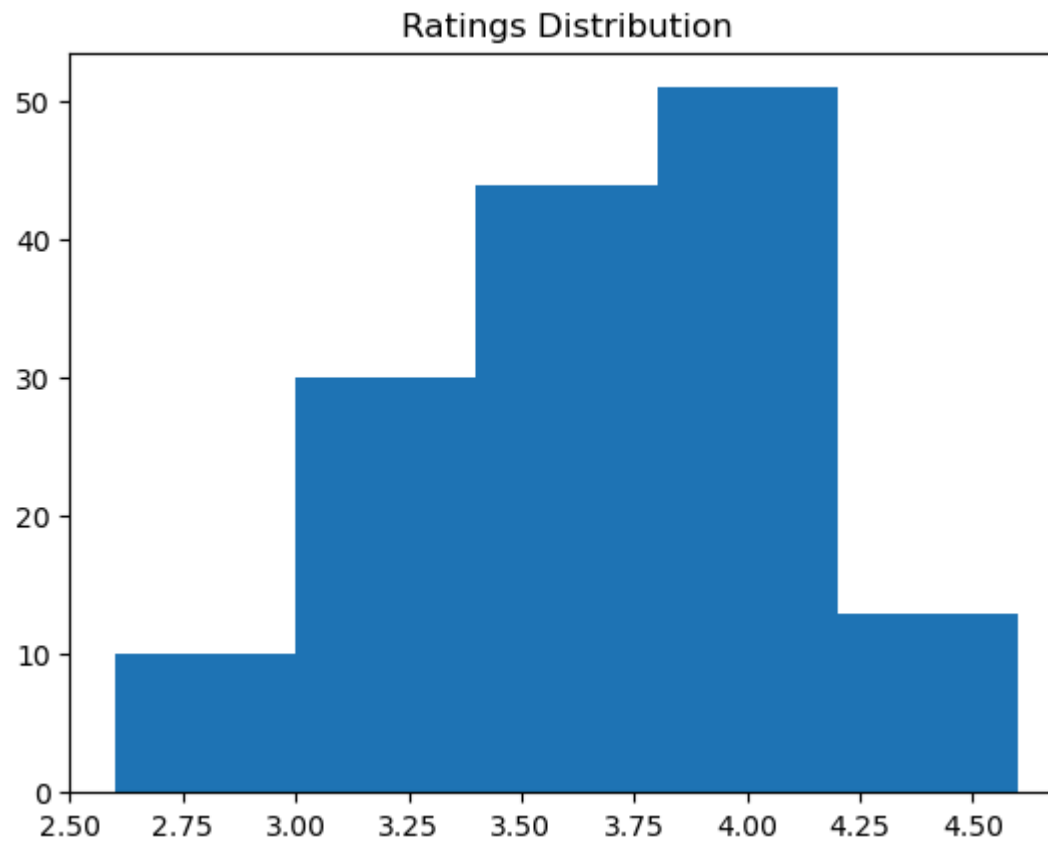
Out[15]: Text(0, 0.5, 'Votes')

## Task.3. What are the ratings that the majority of restaurants have received?

In [19]:
```
1 plt.hist(data['rate'],bins=5)
2 plt.title("Ratings Distribution")
3 plt.show()
```


Ratings Distribution

Task.4. Zomato has observed that most couples order most of their food online. What is their average spending on each order?

In [30]:
```
1  couple_data=data['approx_cost(for two people)']
2  sns.countplot(x=couple_data)
```

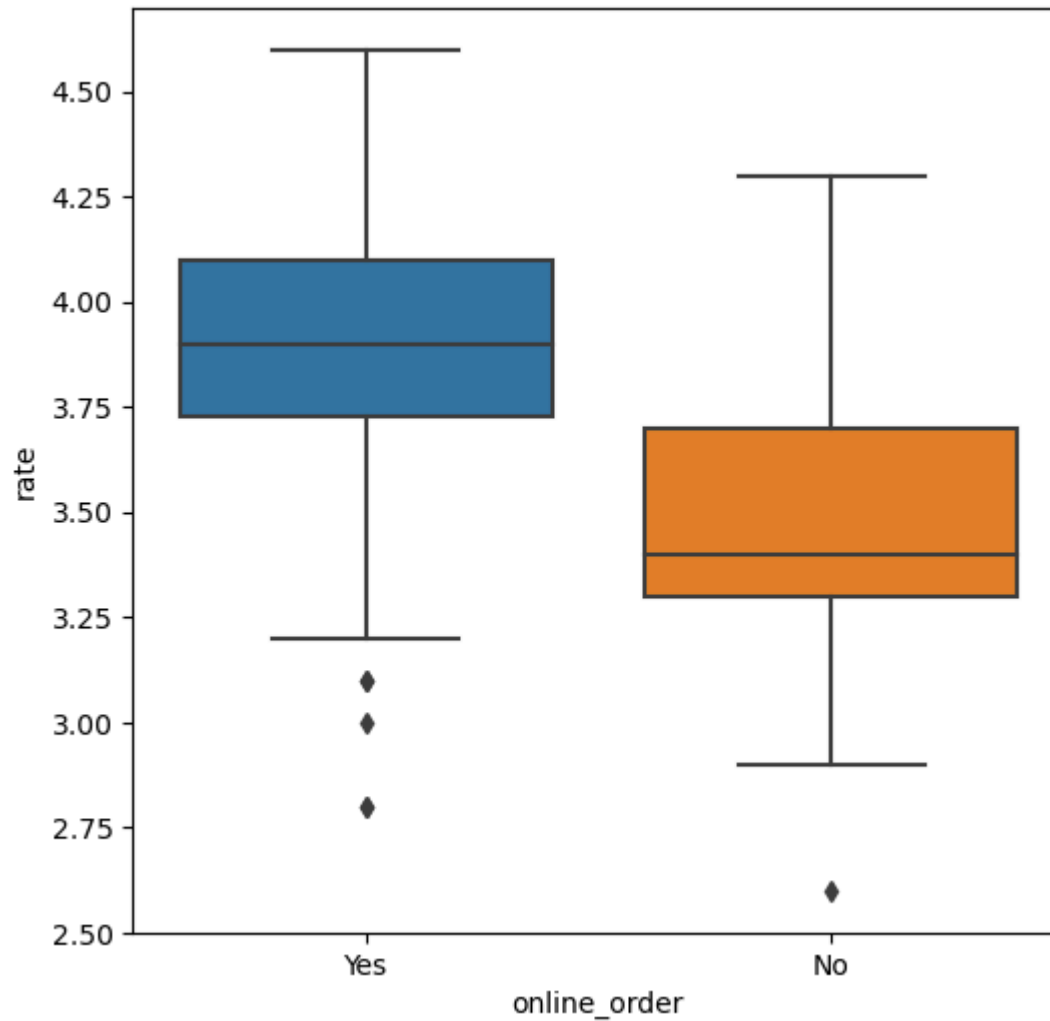Out[30]: <Axes: xlabel='approx_cost(for two people)', ylabel='count'>
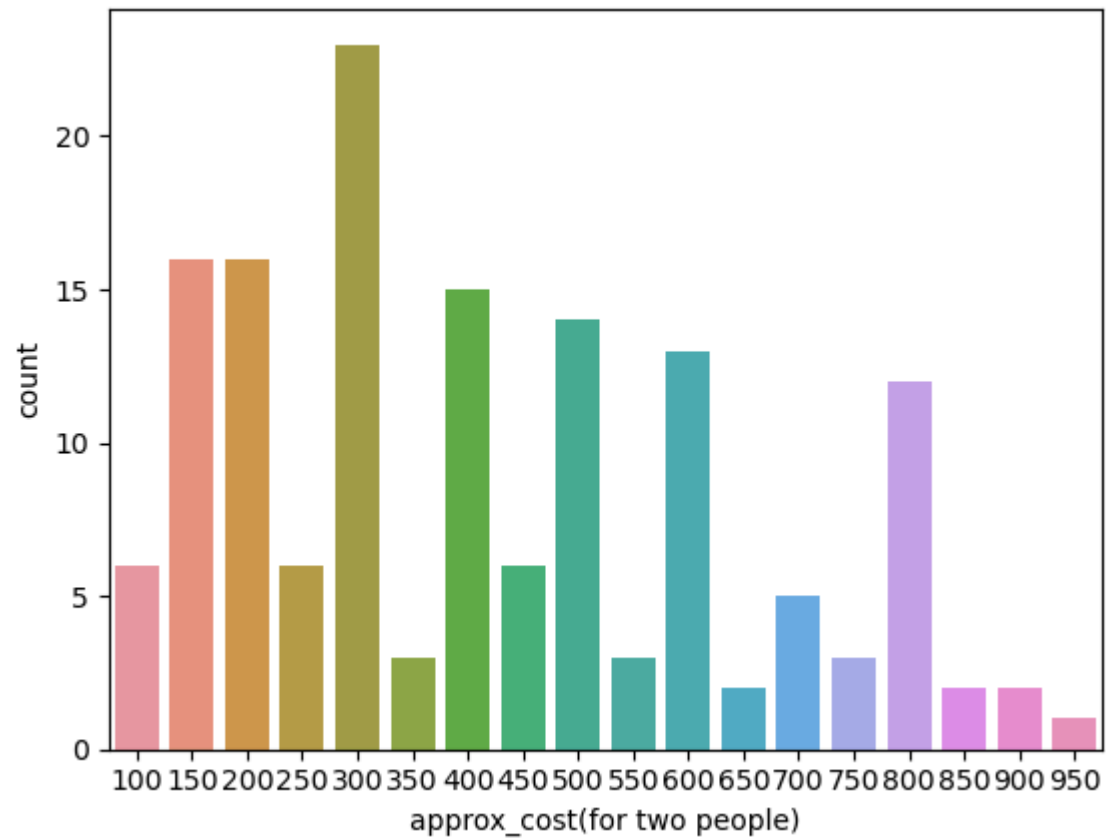
## Task.5. Which mode (online or offline) has received the maximum rating?

```
In [31]:    1  plt.figure(figsize = (6,6))
            2  sns.boxplot(x = 'online_order', y = 'rate', data = data)
```

Out[31]: <Axes: xlabel='online_order', ylabel='rate'>

Task.6. Which type of restaurant received more offline orders, so that Zomato can provide those customers with some good offers?

In [27]:
```
1  pivot_table=data.pivot_table(index='listed_in(type)', columns='online_order', aggfunc='size', fill_value=0)
2  sns.heatmap(pivot_table, annot=True, cmap="YlGnBu", fmt='d')
3  plt.title("Heatmap")
4  plt.xlabel("Online Order")
5  plt.ylabel("Listed In (Type)")
6  plt.show()
7
```
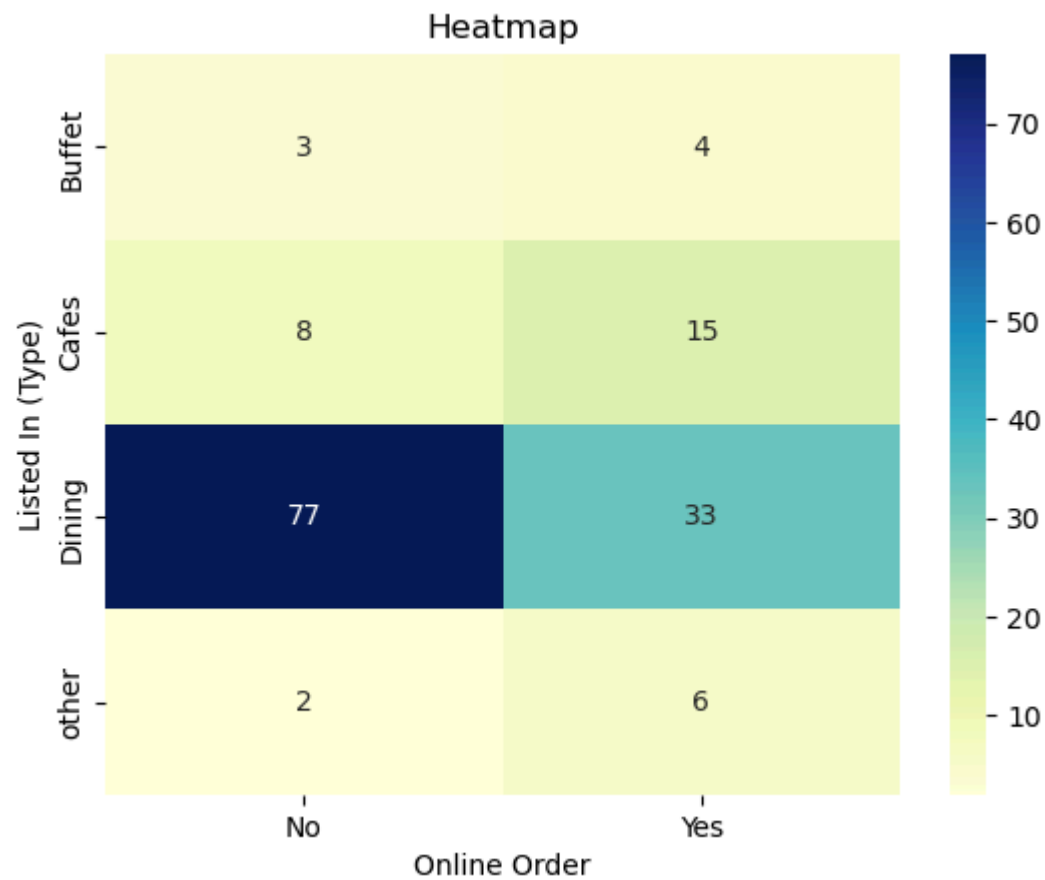
```
Conclusion - There is no Null value in Data Set.

Conclusion 1 - The Majority of the Restaurant Fall Under Dining Category.

Conclusion 2 - Dining Restaurant has Received Maximums Votes.

Conclusion 3 - The Majority of Restaurant Received Ranging from 3.5 to 4.

Conclusion 4 - The Majority of Couples Restaurant with an Approximate Cost of 300 Rupees.

Conclusion 5 - Offline Orders Received Lower Ratings in Comparison to Online Orders,
Which Obtained Excellent Ratings.

Conclusion 6 - Dining Restaurant Primarily Accept Online Orders,
Whereas Cafes Primarily Receive Online Orders.
This Suggest That Client Prefer to Place Orders in Person at Restaurant,
but Prefer Online Ordering at Cafes.
```