

Zomato Data Analysis Project

Step 1 - Importing Libraries

```
In [ ]: Pandas is used for data manipulation and analysis.  
        Numpy is used for numerical operation.  
        Matplotlib.pyplot and seaborn are used for data visualization.
```

```
In [1]: import numpy as np  
        import pandas as pd  
        import matplotlib.pyplot as plt  
        import seaborn as sns  
        import warnings  
        warnings.filterwarnings("ignore")
```

Step 2 - Create a Dataframe

```
In [9]: dataframe = pd.read_csv(r"C:\Users\Itians\downloads\Zomato data .csv")  
        dataframe
```

Out[9]:

	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
...
143	Melting Melodies	No	No	3.3/5	0	100	Dining
144	New Indraprasta	No	No	3.3/5	0	150	Dining
145	Anna Kuteera	Yes	No	4.0/5	771	450	Dining
146	Darbar	No	No	3.0/5	98	800	Dining
147	Vijayalakshmi	Yes	No	3.9/5	47	200	Dining

148 rows × 7 columns

Convert the data type of column - rate

```
In [27]: def handlerate (value):
          value = str(value).split('/')
          value = value[0];
          return float (value)
          dataframe['rate'] = dataframe['rate'].apply(handlerate)
          print(dataframe.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

```
In [29]: dataframe.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
```

Type of Restaurant

```
In [32]: dataframe.head()
```

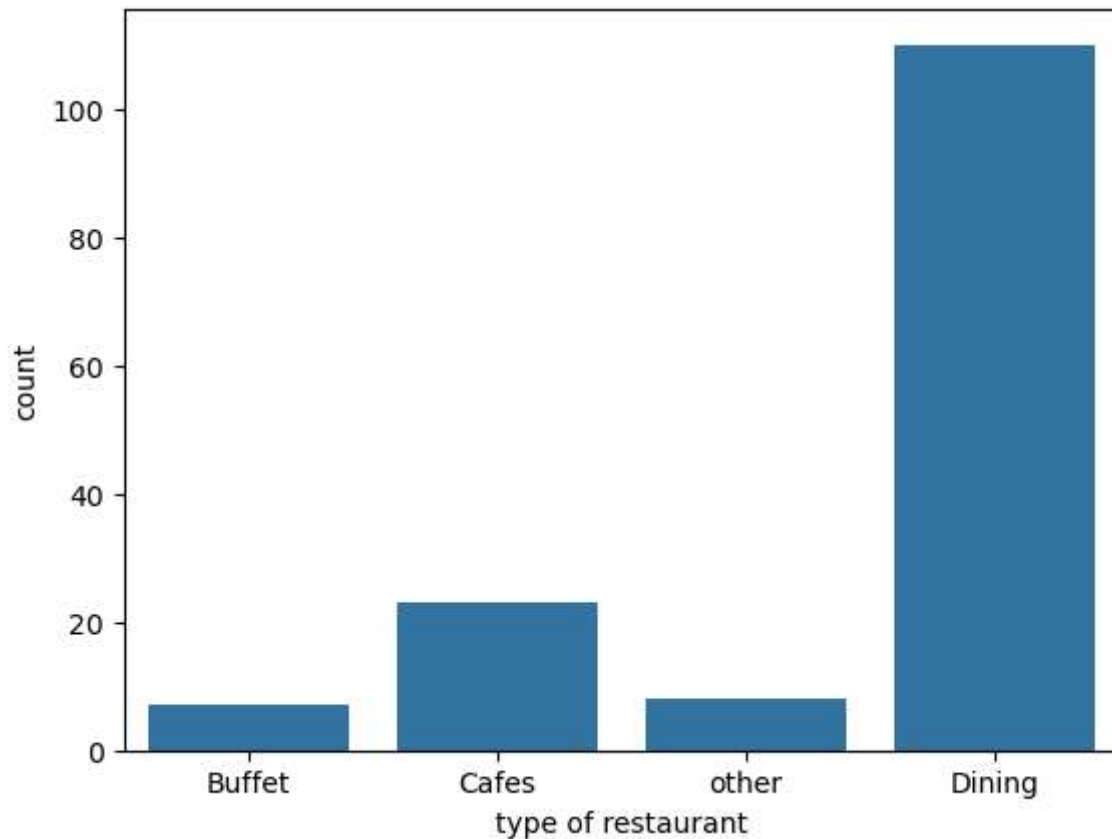
```
Out[32]:
```

	name	online_order	book_table	rate	votes	approx_cost(for two people)	listed_in(type)
0	Jalsa	Yes	Yes	4.1	775	800	Buffet
1	Spice Elephant	Yes	No	4.1	787	800	Buffet
2	San Churro Cafe	Yes	No	3.8	918	800	Buffet
3	Addhuri Udupi Bhojana	No	No	3.7	88	300	Buffet
4	Grand Village	No	No	3.8	166	600	Buffet

Q.1) What type of restaurant do the majority of customers orders from ?

```
In [104... sns.countplot(x = dataframe['listed_in(type)'])
plt.xlabel("type of restaurant")
```

```
Out[104... Text(0.5, 0, 'type of restaurant')
```



conclusion - majority of the restaurant falls in dining category

In []:

Q.2)How many votes has each type of restaurant received from customers?

In [61]: `dataframe.head()`

Out[61]:

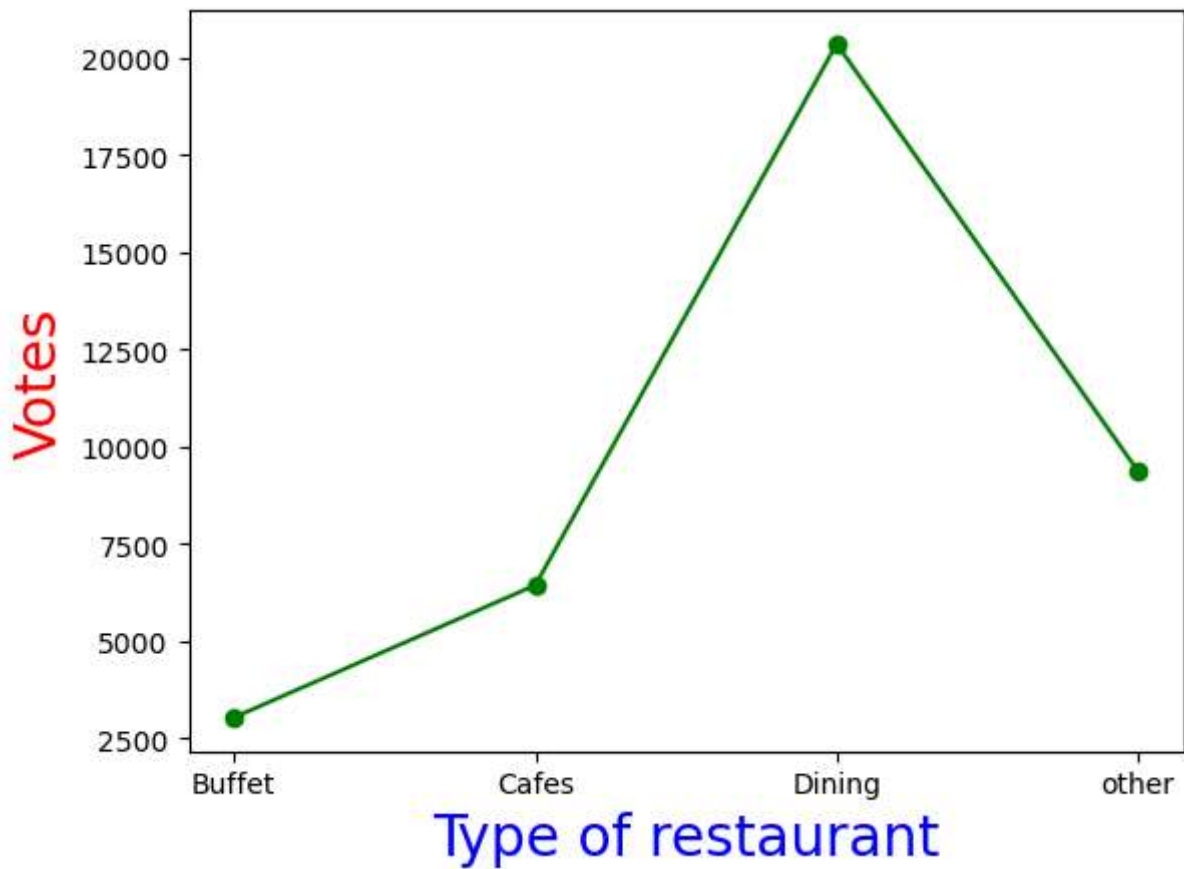
	name	online_order	book_table	rate	votes	approx_cost(for two people)	listed_in(type)
0	Jalsa	Yes	Yes	4.1	775	800	Buffet
1	Spice Elephant	Yes	No	4.1	787	800	Buffet
2	San Churro Cafe	Yes	No	3.8	918	800	Buffet
3	Addhuri Udupi Bhojana	No	No	3.7	88	300	Buffet
4	Grand Village	No	No	3.8	166	600	Buffet

In [109...]

```
grouped_data = dataframe.groupby ('listed_in(type)')['votes'].sum()
result= pd.DataFrame({'votes':grouped_data})
plt.plot(result,c= "green",marker = "o")
plt.xlabel("Type of restaurant",c= "blue",size =20 )
plt.ylabel("Votes",c="red",size=20)
```

Out[109...]

Text(0, 0.5, 'Votes')



Conclusion - Dinning restaurant has received maximum votes

In []:

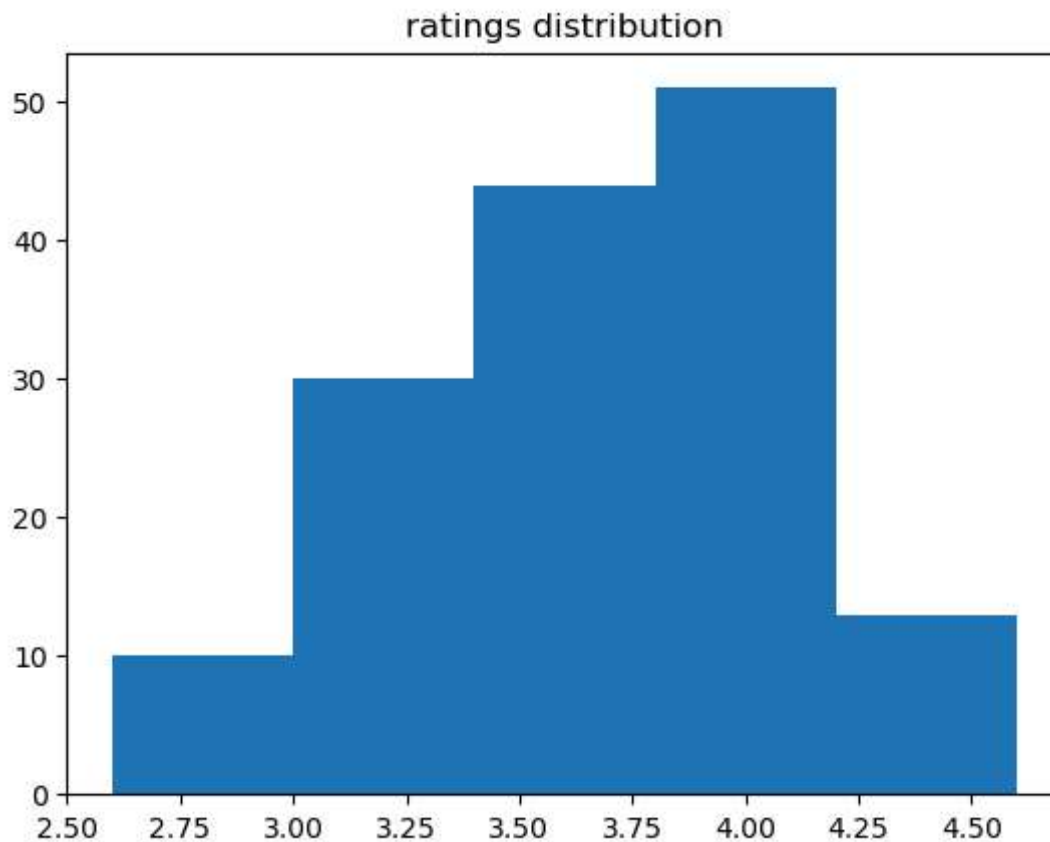
Q.3) What are the ratings that the majority of restaurants have received ?

In [63]: `dataframe.head()`

Out[63]:

	name	online_order	book_table	rate	votes	approx_cost(for two people)	listed_in(type)
0	Jalsa	Yes	Yes	4.1	775	800	Buffet
1	Spice Elephant	Yes	No	4.1	787	800	Buffet
2	San Churro Cafe	Yes	No	3.8	918	800	Buffet
3	Addhuri Udupi Bhojana	No	No	3.7	88	300	Buffet
4	Grand Village	No	No	3.8	166	600	Buffet

In [69]: `plt.hist(dataframe['rate'],bins= 5)
plt.title("ratings distribution")
plt.show()`



Conclusion - the majority restaurants received ratings from 3.5 to 4

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

```
In [72]: dataframe.head()
```

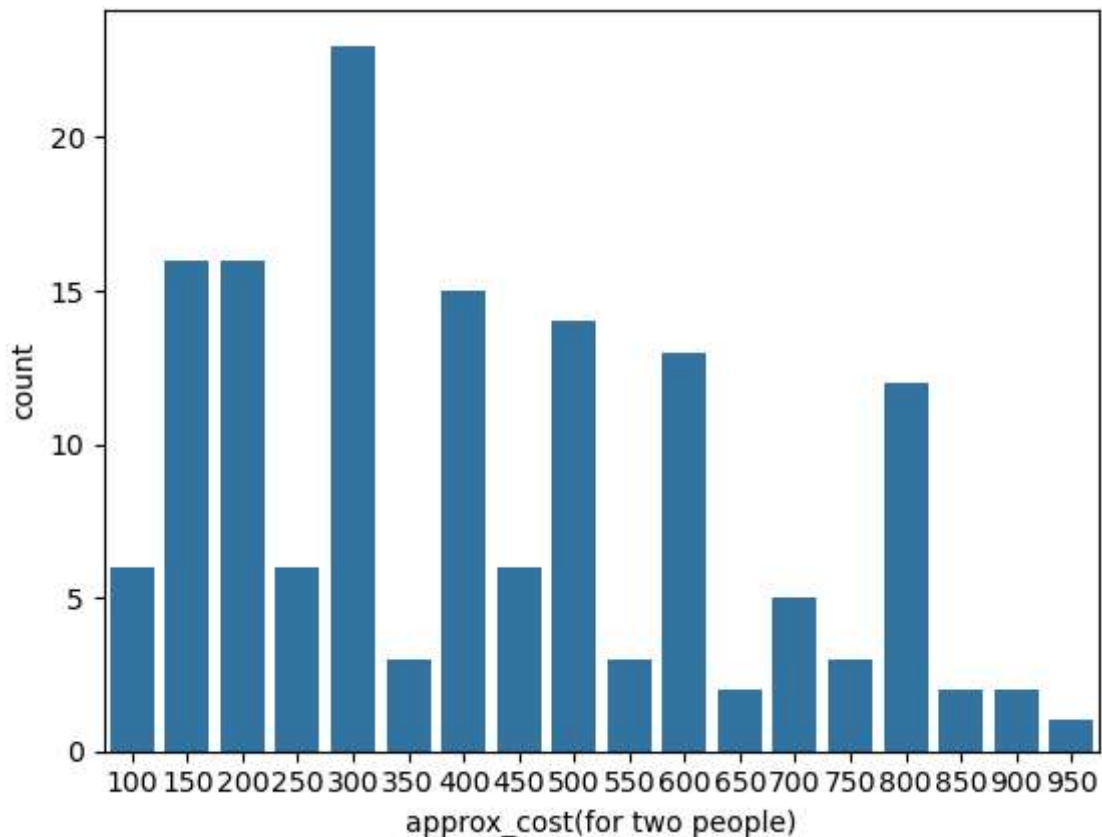
Out[72]:

	name	online_order	book_table	rate	votes	approx_cost(for two people)	listed_in(type)
0	Jalsa	Yes	Yes	4.1	775	800	Buffet
1	Spice Elephant	Yes	No	4.1	787	800	Buffet
2	San Churro Cafe	Yes	No	3.8	918	800	Buffet
3	Addhuri Udupi Bhojana	No	No	3.7	88	300	Buffet
4	Grand Village	No	No	3.8	166	600	Buffet

Average order spending by couples

```
In [79]: couple_data = dataframe['approx_cost(for two people)']  
sns.countplot(x=couple_data)
```

Out[79]: <Axes: xlabel='approx_cost(for two people)', ylabel='count'>



Conclusion - the majority of couples
preferr restaurant with an approximate
cost of 300 rupees

In []:

Q.5) Which mode (offline or online) has
received the maximum rating?

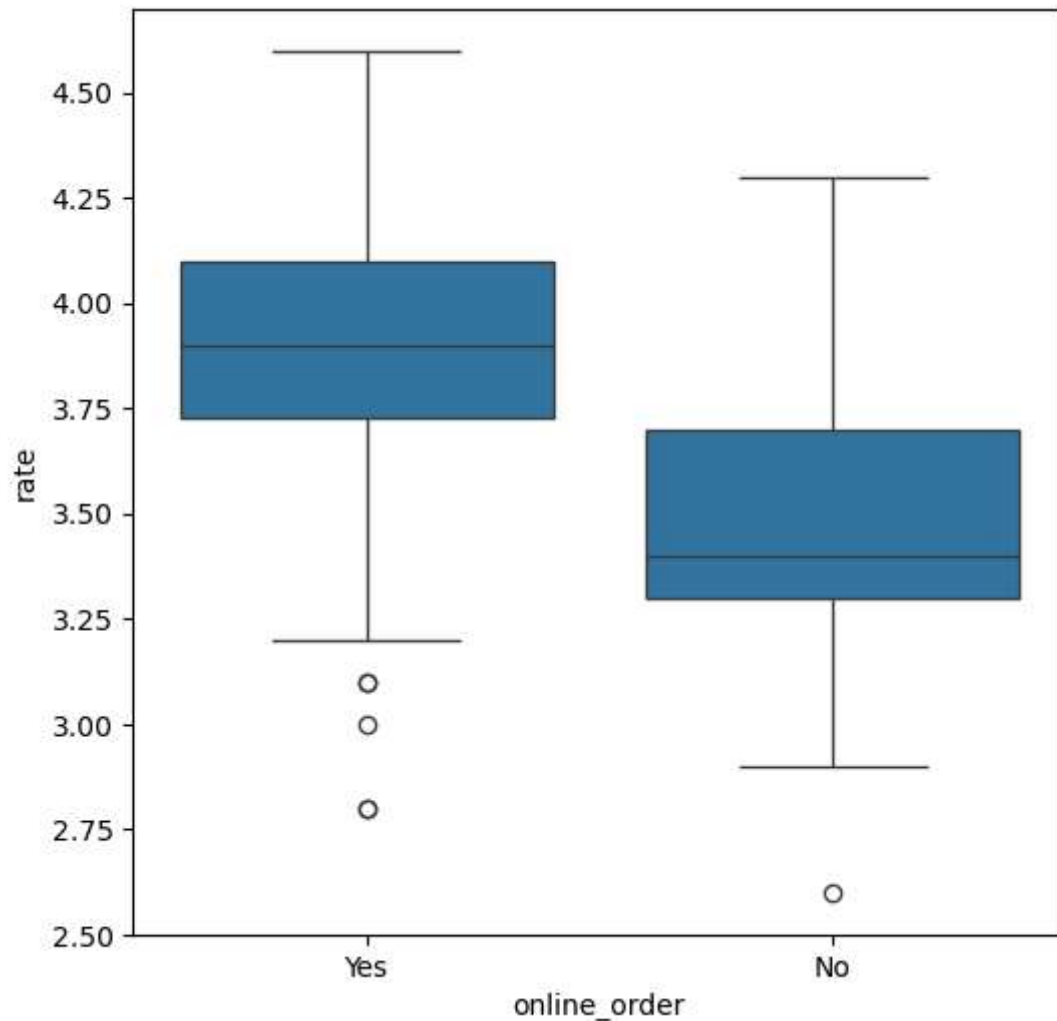
In [82]: dataframe .head()

Out[82]:

	name	online_order	book_table	rate	votes	approx_cost(for two people)	listed_in(type)
0	Jalsa	Yes	Yes	4.1	775	800	Buffet
1	Spice Elephant	Yes	No	4.1	787	800	Buffet
2	San Churro Cafe	Yes	No	3.8	918	800	Buffet
3	Addhuri Udupi Bhojana	No	No	3.7	88	300	Buffet
4	Grand Village	No	No	3.8	166	600	Buffet

In [85]: plt.figure(figsize= (6,6))
sns.boxplot(x="online_order",y="rate", data= dataframe)

Out[85]: <Axes: xlabel='online_order', ylabel='rate'>



Conclusion - offline order received lower rating in comparison to online order

In []:

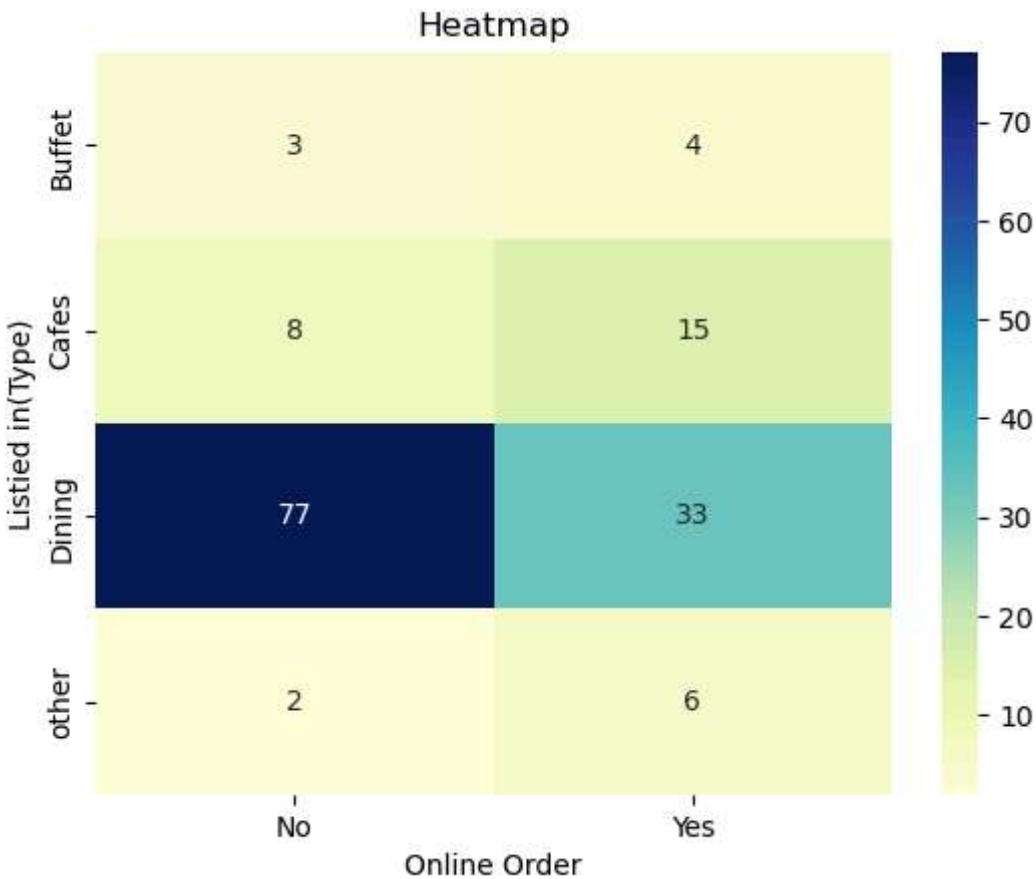
Q.6) Which type of restaurant received more offline orders, so that Zomato can provide customer with some good offers?

In [94]: `dataframe.head()`

Out[94]:

	name	online_order	book_table	rate	votes	approx_cost(for two people)	listed_in(type)
0	Jalsa	Yes	Yes	4.1	775	800	Buffet
1	Spice Elephant	Yes	No	4.1	787	800	Buffet
2	San Churro Cafe	Yes	No	3.8	918	800	Buffet
3	Addhuri Udupi Bhojana	No	No	3.7	88	300	Buffet
4	Grand Village	No	No	3.8	166	600	Buffet

```
In [92]: pivot_table = dataframe.pivot_table(index='listed_in(type)',columns = 'online_order',
sns.heatmap(pivot_table,annot=True,cmap="YlGnBu",fmt = 'd')
plt.title("Heatmap")
plt.xlabel("Online Order")
plt.ylabel("Listied in(Type)")
plt.show()
```



CONCLUSION - Dining Restaurants primarily accept offline orders, whereas Cafes primarily receive online orders.

This suggests that clients prefer placing orders in person at restaurants, but prefer online ordering at Cafes.

In []: