Zomato Data Analysis project

Step 1 ;- Importing Libraries

pandas is used for data manipulation and analysis numpy is used for numerical problems matplotlib.pyplot and seaborn are used for data visualization

```python
In [2]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
```

Step 2 ;- Create the data frame

```python
In [3]: df= pd.read_csv("Zomato data .csv")
        print(df)
```

```
                       name online_order book_table   rate  votes  \
0                     Jalsa          Yes        Yes  4.1/5    775
1             Spice Elephant          Yes         No  4.1/5    787
2            San Churro Cafe          Yes         No  3.8/5    918
3        Addhuri Udupi Bhojana          No         No  3.7/5     88
4               Grand Village          No         No  3.8/5    166
..                      ...          ...        ...    ...    ...
143           Melting Melodies          No         No  3.3/5      0
144            New Indraprasta          No         No  3.3/5      0
145              Anna Kuteera          Yes         No  4.0/5    771
146                    Darbar          No         No  3.0/5     98
147              Vijayalakshmi          Yes         No  3.9/5     47

     approx_cost(for two people) listed_in(type)
0                            800          Buffet
1                            800          Buffet
2                            800          Buffet
3                            300          Buffet
4                            600          Buffet
..                           ...             ...
143                          100          Dining
144                          150          Dining
145                          450          Dining
146                          800          Dining
147                          200          Dining

[148 rows x 7 columns]
```

```python
In [4]: df
```

	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 [5]:
```python
def handleRate(value):
    value= str(value).split('/')
    value = value[0];
    return float(value)
df['rate'] = df['rate'].apply(handleRate)
print(df.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 [6]: df.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 Restraunt

```
In [7]: df.head()
```

Out[7]:

	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 [9]: sns.countplot(x=df['listed_in(type)'])
        plt.xlabel("type of retraunt")
```

```
Out[9]: Text(0.5, 0, 'type of retraunt')
```
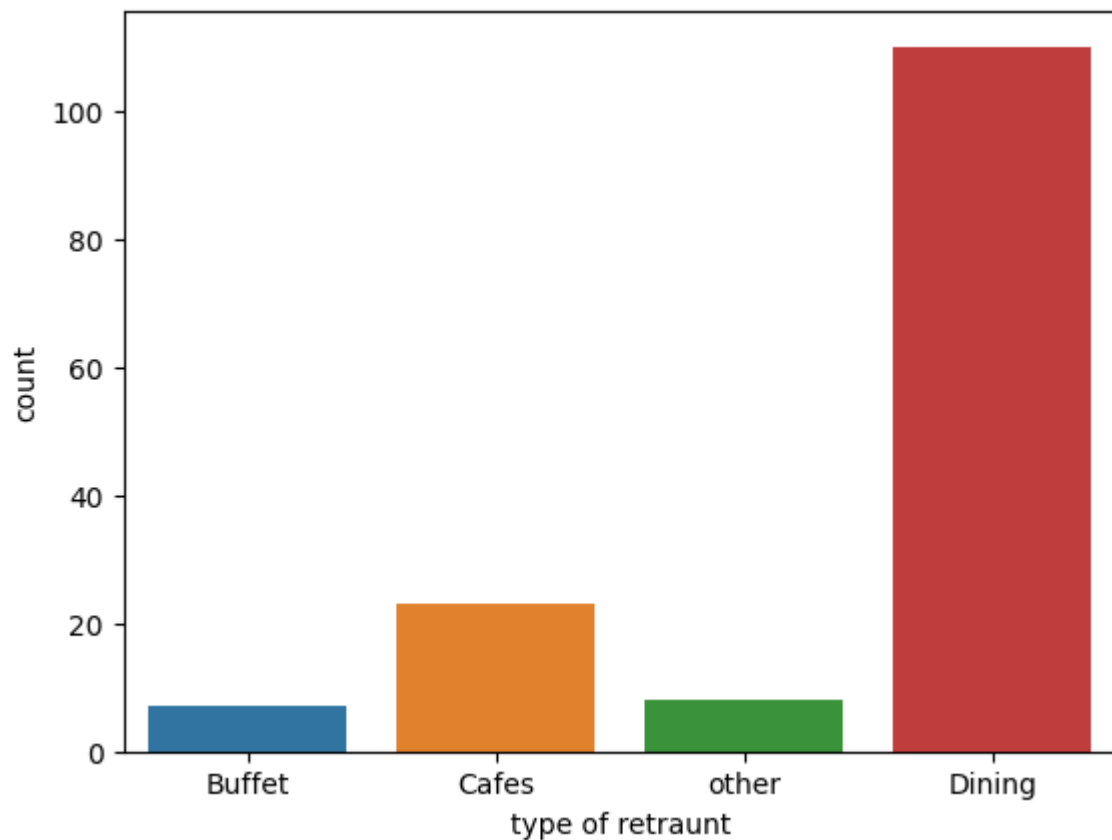
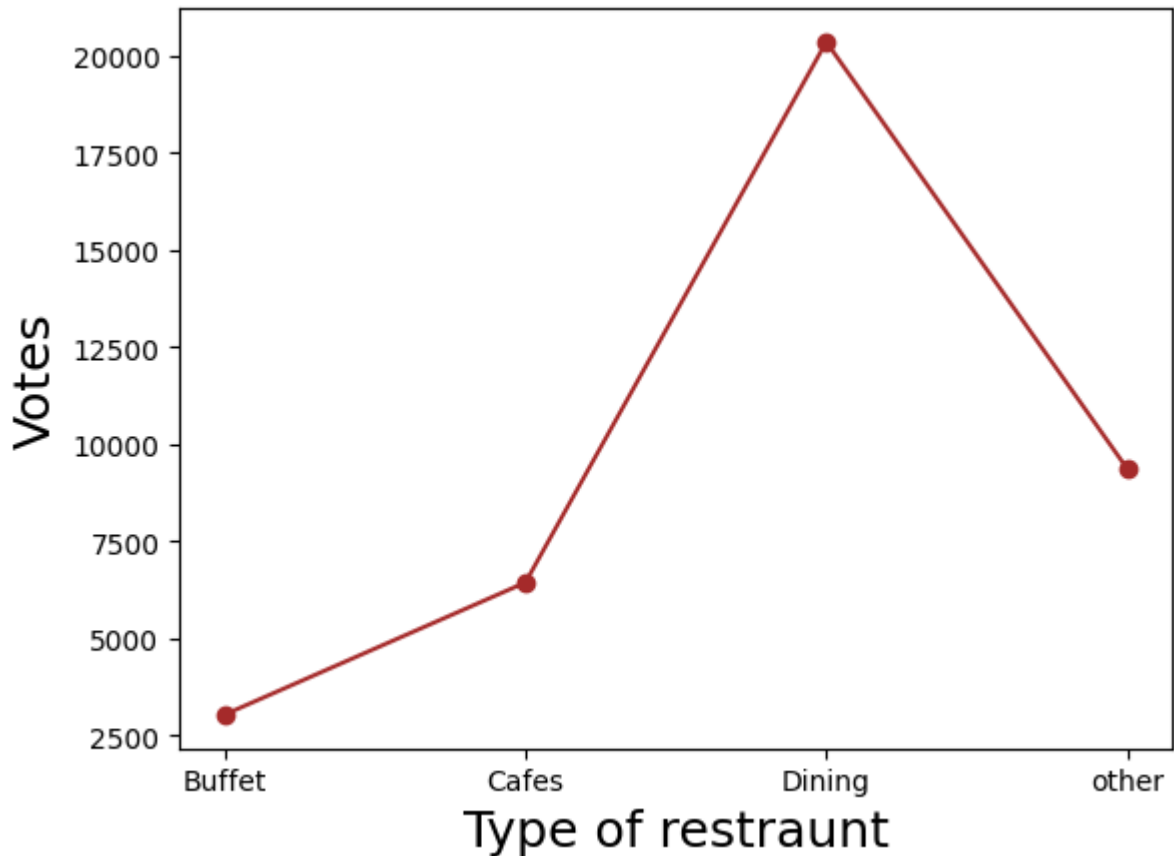# Conclusion - Majority of the restraunt falls in dining category

In [10]: `df.head()`

Out[10]:

	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 [19]:
```
grouped_data =df.groupby('listed_in(type)')['votes'].sum()
result = pd.DataFrame({'votes': grouped_data})
plt.plot(result, c="Brown", marker = "o")
```

```
plt.xlabel("Type of restraunt", c = "Black", size = 18)
plt.ylabel("Votes", c= "Black", size = 18)
```

Out[19]: Text(0, 0.5, 'Votes')

Conclusion - Dinning restraunts has recieved maximum votes

In [21]: `df.head()`

Out[21]:

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

ratings distribution
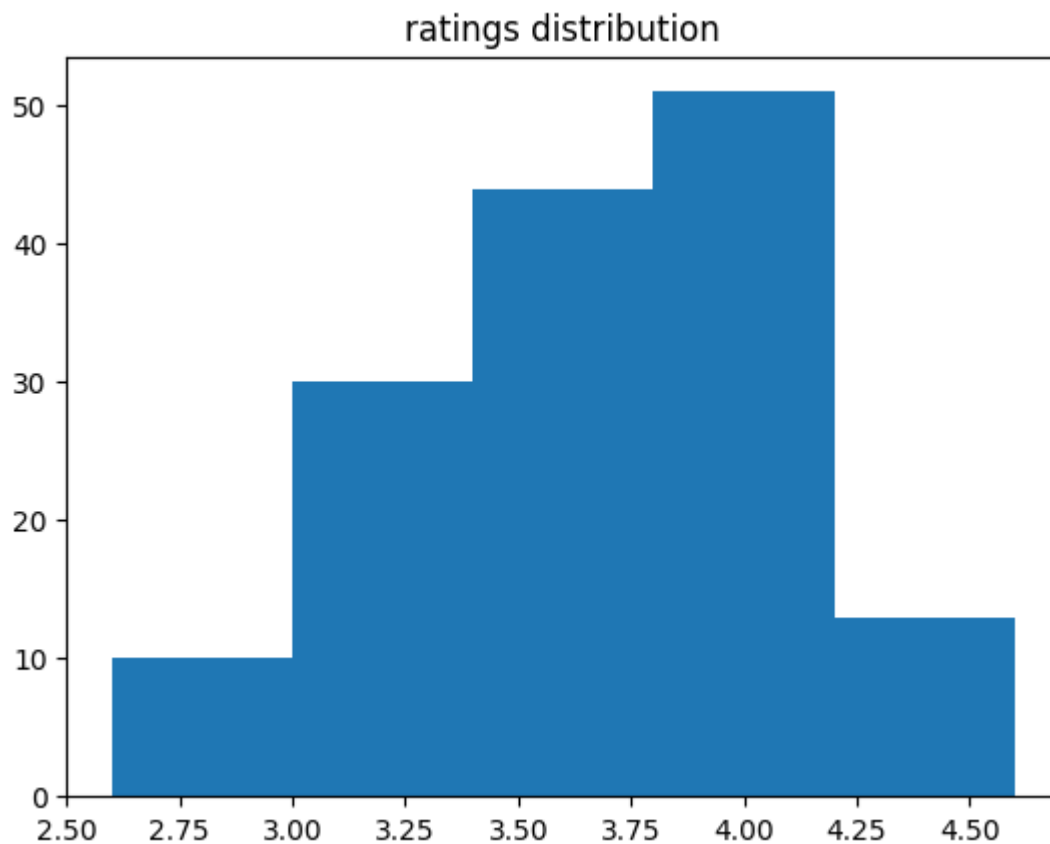
Conclusion - The Majority restraunts Received ratings from 3.5 to 4
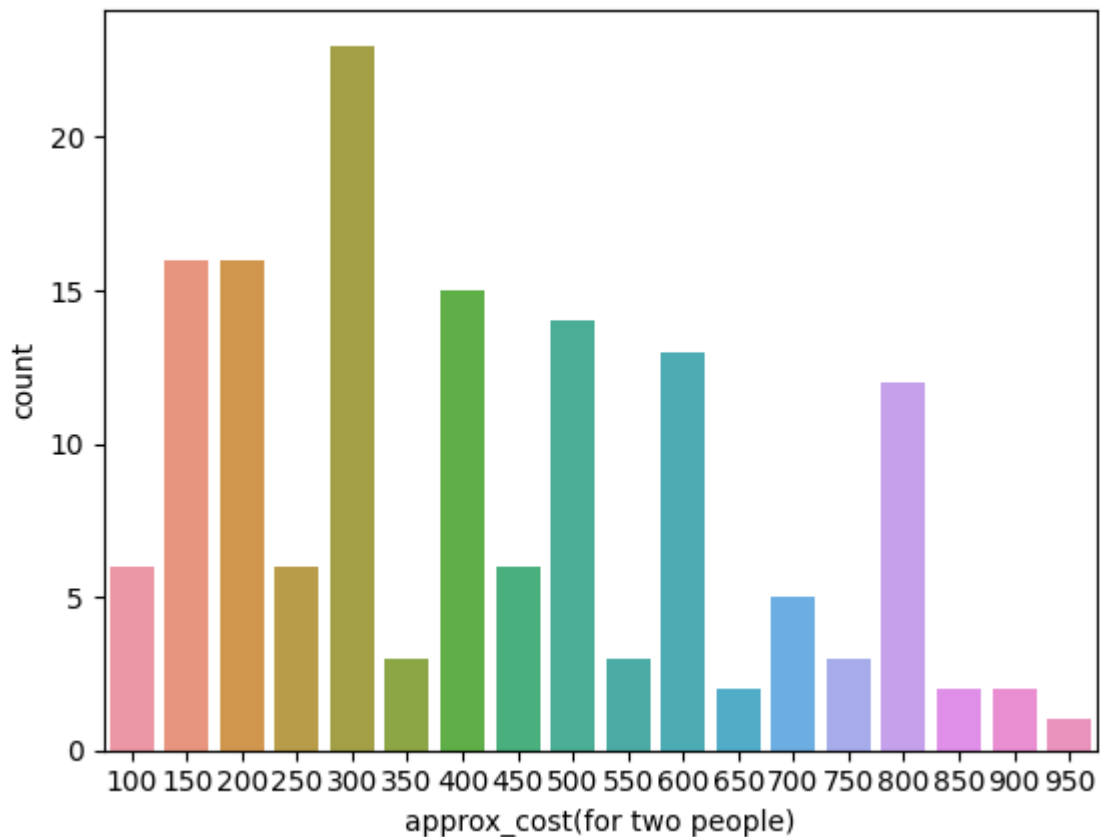
Average order spending by couples

```
In [24]: df.head()
```

	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 [26]:
```python
couple_data = df['approx_cost(for two people)']
sns.countplot(x= couple_data)
```

Out[26]: <Axes: xlabel='approx_cost(for two people)', ylabel='count'>

Conclusion - Average cost for two people Lies in 150Rs to 200Rs
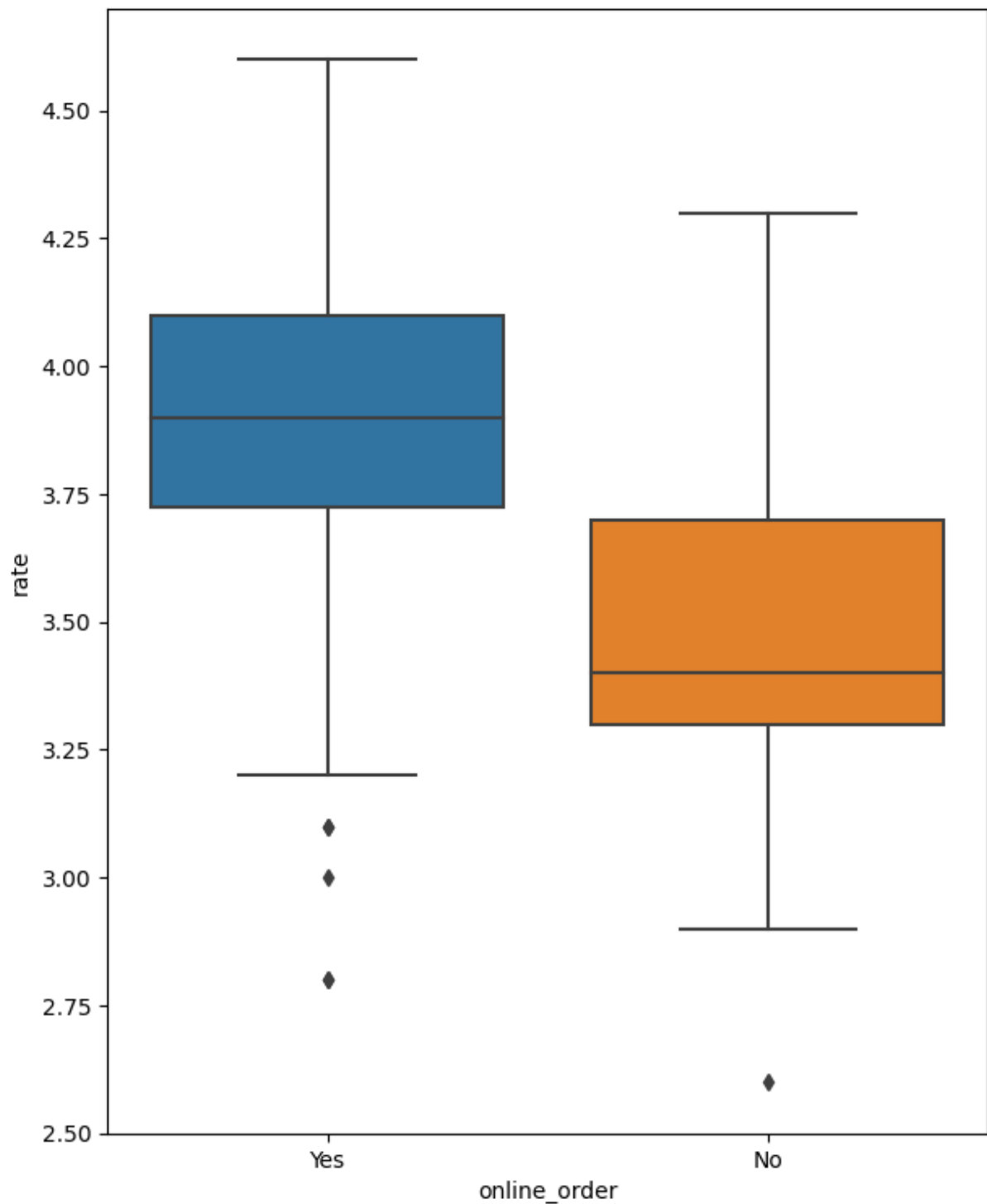
Which mode receives maximum rating

```
In [28]: df.head()
```

Out[28]:

	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 [32]: plt.figure(figsize = (7, 9))
         sns.boxplot(x= 'online_order', y = 'rate', data = df)
```

Out[32]: <Axes: xlabel='online_order', ylabel='rate'>

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

```
In [34]: df.head()
```

	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 [39]:
```python
pivot_table = df.pivot_table(index = 'listed_in(type)', columns = 'online_order', a
sns.heatmap(pivot_table, annot = True , cmap = "YlGnBu", fmt = 'd')
plt.title("Heatmap")
plt.xlabel("Online Order")
plt.ylabel("Listed in (Type)")
plt.show()
```
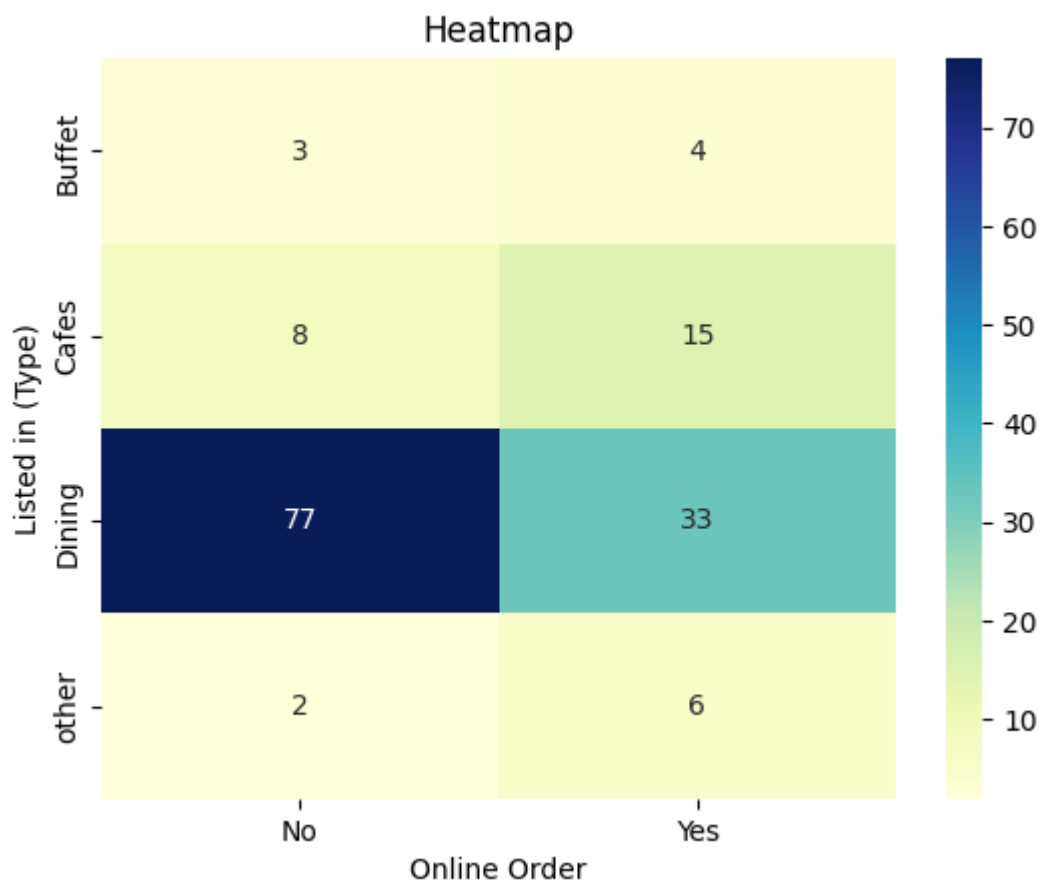

Conclusion - Dining restraunts primarily

accept offline orders, whereas cafe's primarily receive online orders. This suggests that client prefers orders in person at restraunts, but prefer online ordering at cafe's

In [ ]: