

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

```
In [7]: df = pd.read_csv("Zomato data .csv")
```

```
In [9]: df
```

```
Out[9]:
```

	name	online_order	book_table	rate	votes	approx_cost(for two people)
0	Jalsa	Yes	Yes	4.1/5	775	800
1	Spice Elephant	Yes	No	4.1/5	787	800
2	San Churro Cafe	Yes	No	3.8/5	918	800
3	Addhuri Udupi Bhojana	No	No	3.7/5	88	300
4	Grand Village	No	No	3.8/5	166	600
...	...	...	...	...	...	...
143	Melting Melodies	No	No	3.3/5	0	100
144	New Indraprasta	No	No	3.3/5	0	150
145	Anna Kuteera	Yes	No	4.0/5	771	450
146	Darbar	No	No	3.0/5	98	800
147	Vijayalakshmi	Yes	No	3.9/5	47	200

148 rows x 7 columns

```
In [10]: df.describe
```

```
Out[10]: <bound method NDFrame.describe of
e_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]>

```
In [11]: 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 [12]: df.info()
```

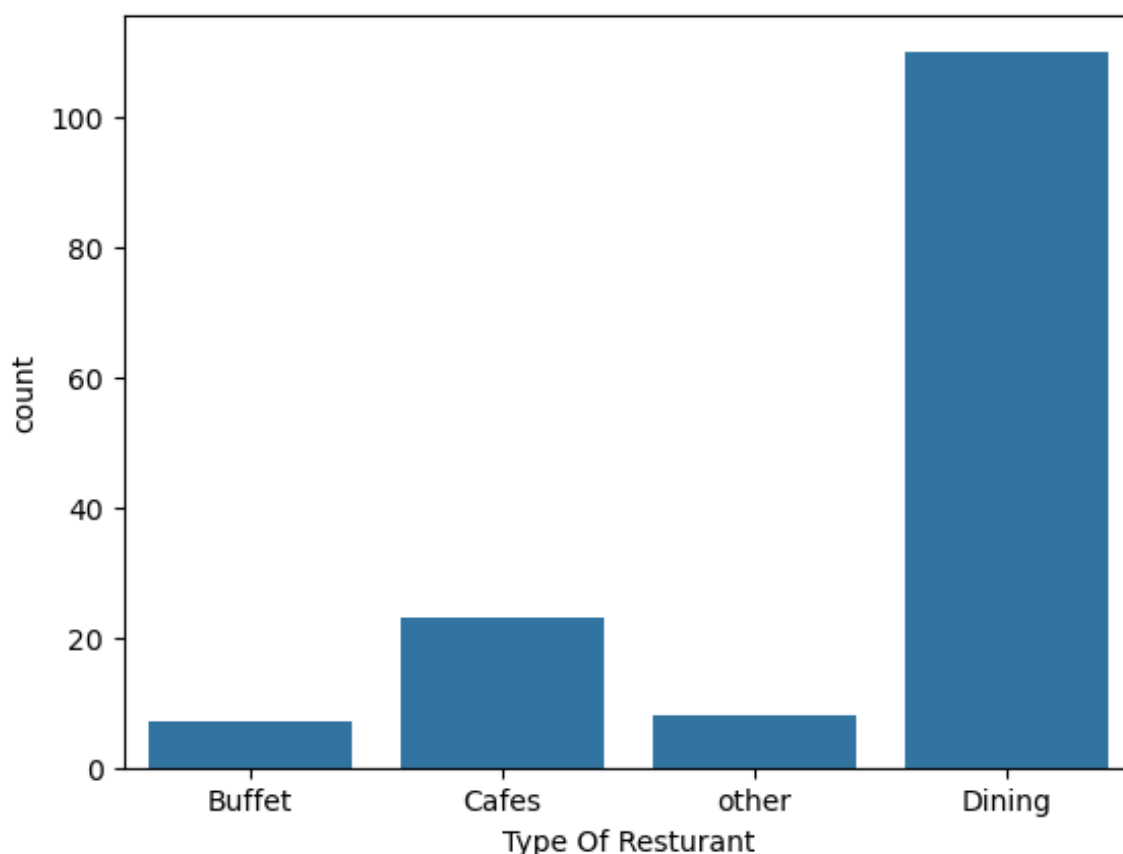
```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 148 entries, 0 to 147  
Data columns (total 7 columns):  
#   Column                                Non-Null Count  Dtype  
---  -  
0   name                                  148 non-null    object  
1   online_order                         148 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
```

There is no null value in data frame

WHAT TYPE OF REATURANT DO MAJORITY OF CUSTOMERS ORDER FROM?

```
In [14]: sns.countplot(x=df['listed_in(type)'])  
plt.xlabel("Type Of Resturant")
```

```
Out[14]: Text(0.5, 0, 'Type Of Resturant')
```



So, the majority of resturant fall into the dining category

```
In [15]: df.head()
```

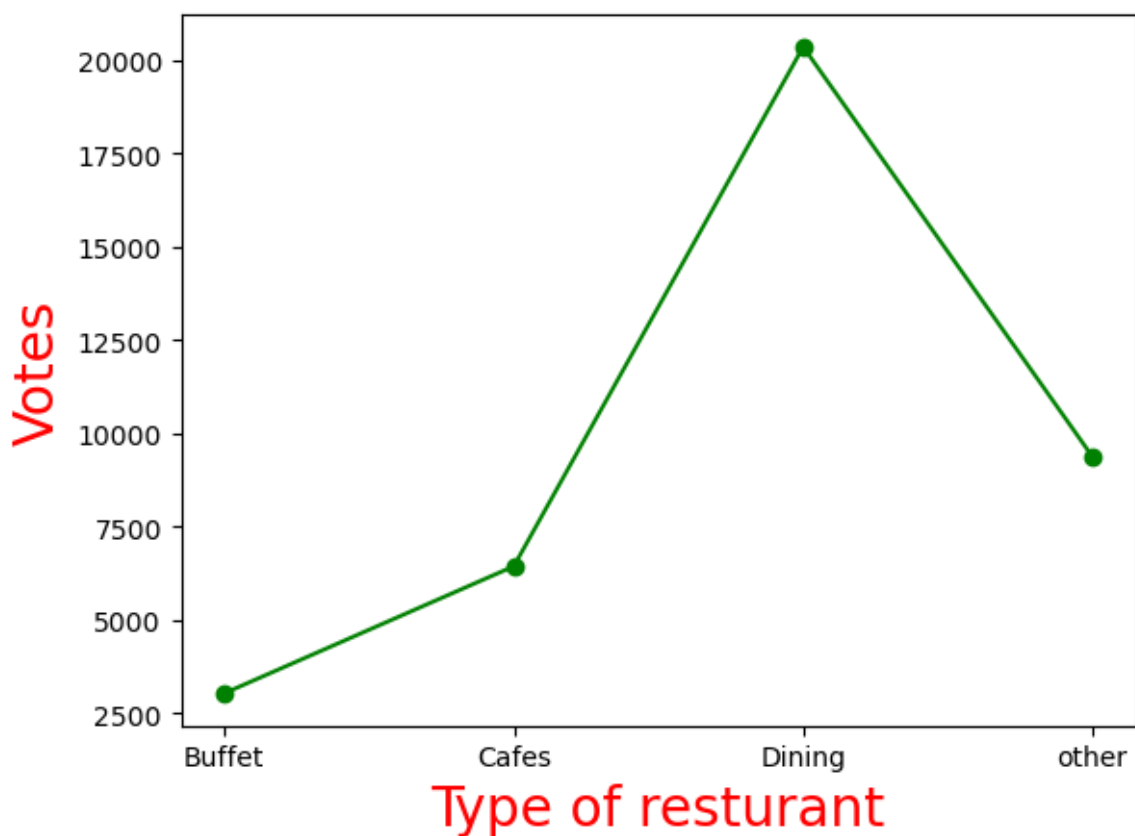
Out[15]:

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

How many votes has each type of restaurant received from customers?

```
In [17]: grouped_data=df.groupby('listed_in(type)')['votes'].sum()
result=pd.DataFrame({'votes': grouped_data})
plt.plot(result, c="green", marker="o")
plt.xlabel("Type of resturant", c="red", size=20)
plt.ylabel("Votes", c="red", size=20)
```

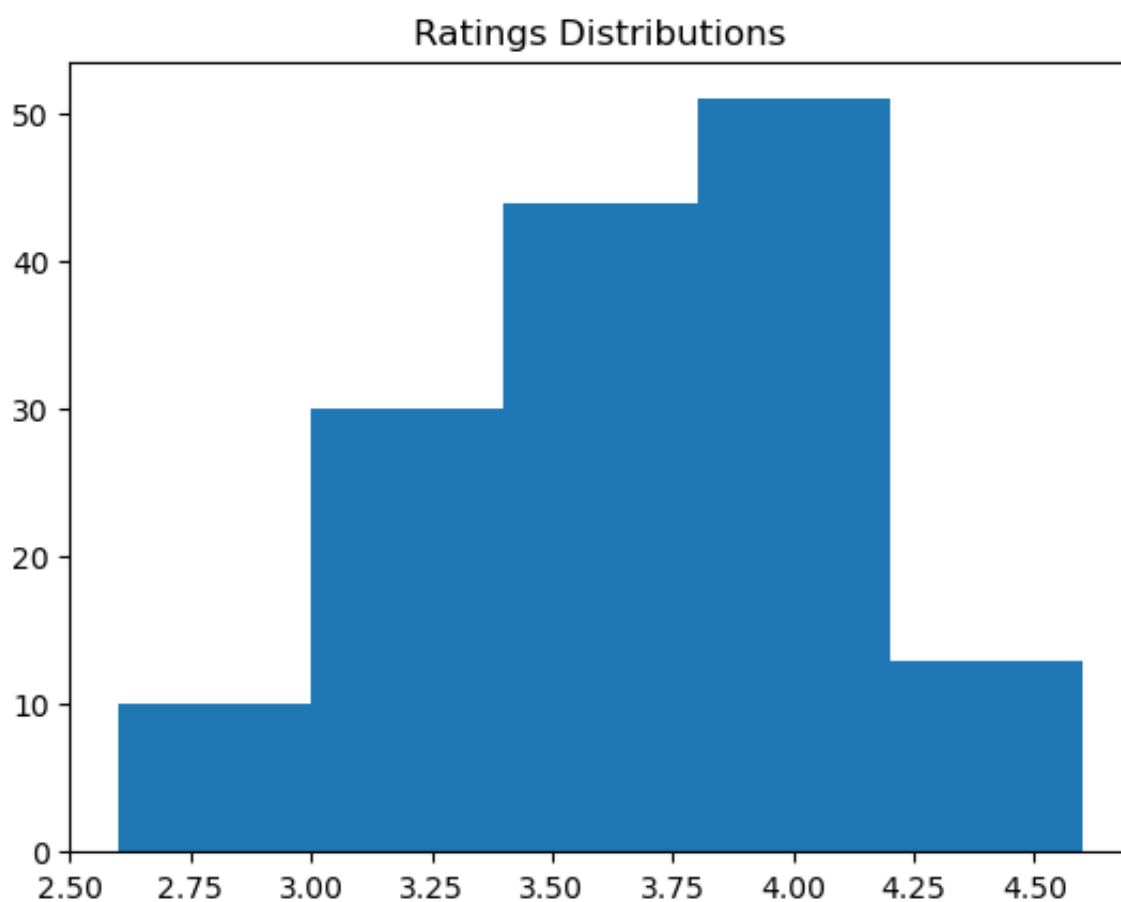
Out[17]: Text(0, 0.5, 'Votes')



Dining resturants have recieved maximum votes

What are the ratings that the majority of restaurants have received?

```
In [18]: plt.hist(df['rate'], bins=5)
plt.title("Ratings Distributions")
plt.show()
```

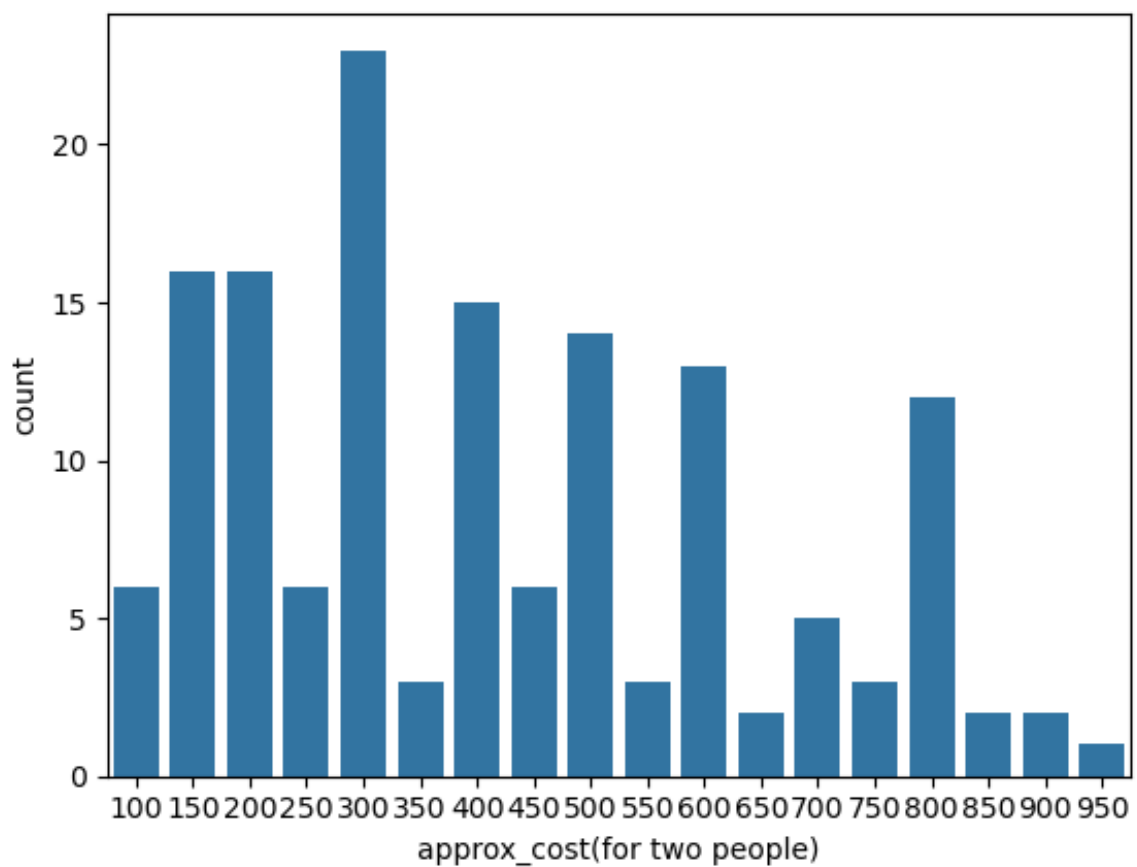


So, The majority of resturants recieved the ratings ranging from 3.5 to 4

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

```
In [20]: couple_data=df["approx_cost(for two people)"]
sns.countplot(x=couple_data)
```

```
Out[20]: <Axes: xlabel='approx_cost(for two people)', ylabel='count'>
```

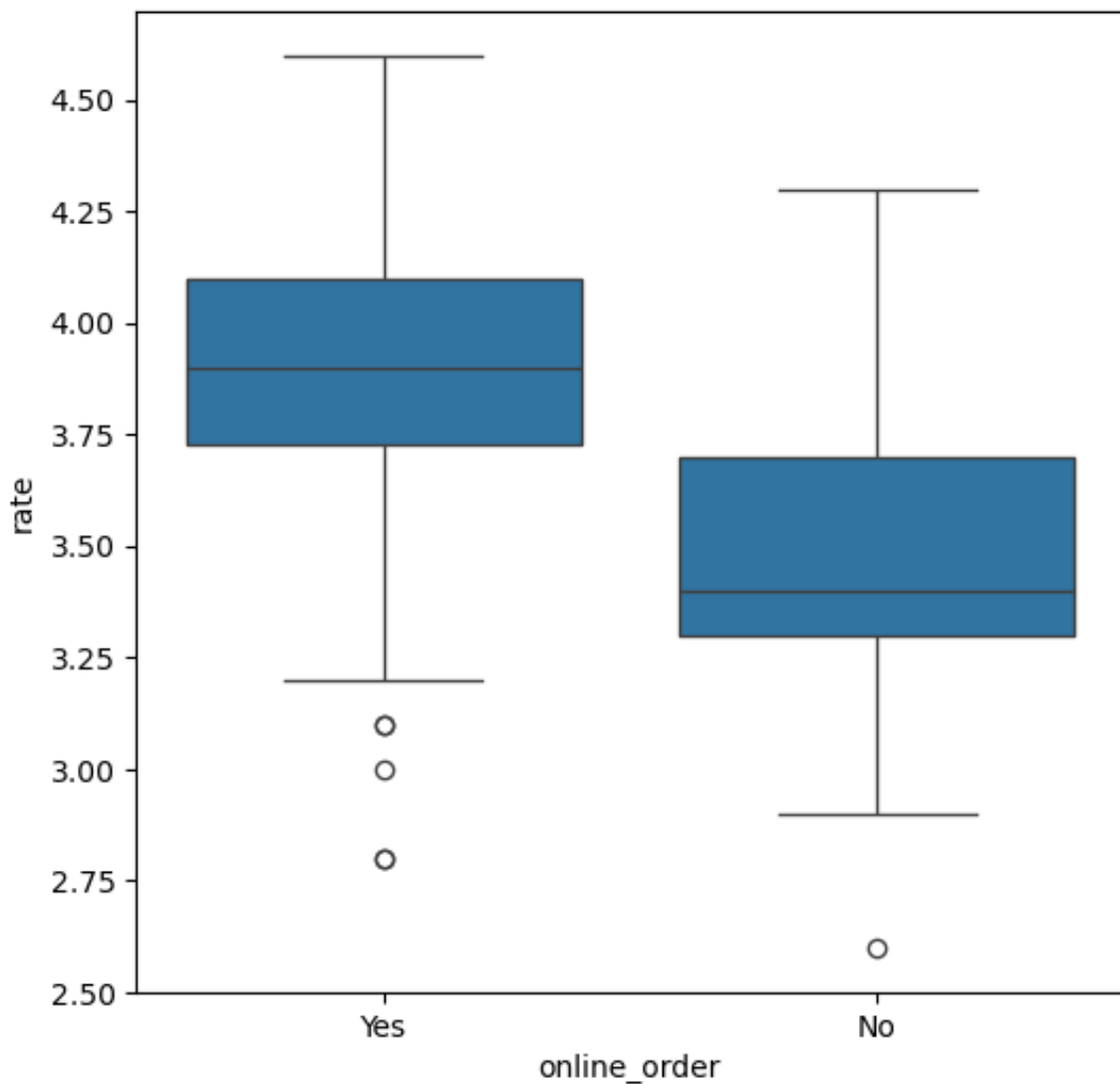


The majority of couples prefer restaurants with an approximate cost of 300 rupees.

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

```
In [21]: plt.figure(figsize=(6,6))  
sns.boxplot(x='online_order', y='rate', data=df)
```

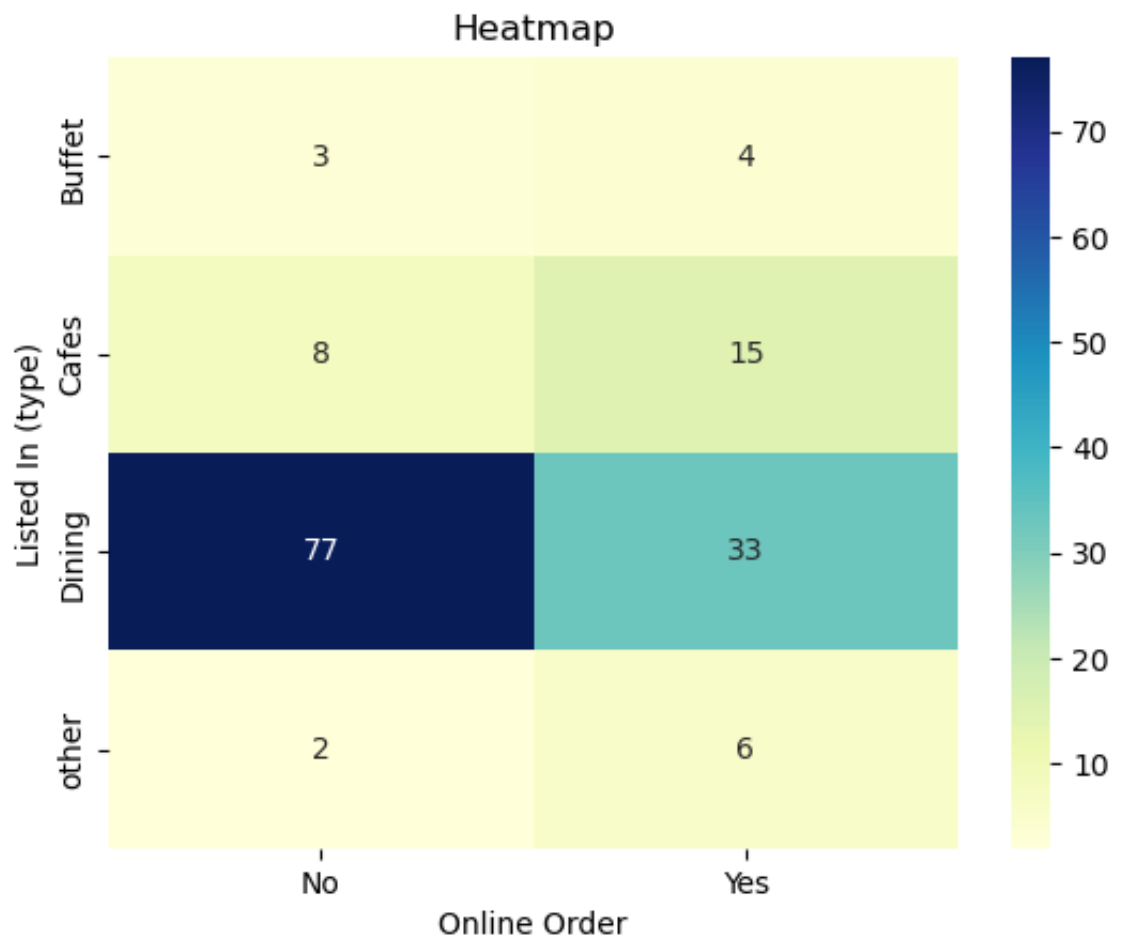
```
Out[21]: <Axes: xlabel='online_order', ylabel='rate'>
```



Offline orders recieved lower ratings compared to online orders,which obtained excellent ratings.

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

```
In [28]: pivot_table=df.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("Listed In (type)")
plt.show()
```



So, in conclusion, dining restaurants primarily receive offline orders, whereas cafes primarily receive online orders. This suggests that customers prefer to place orders in person at restaurants, but prefer online ordering at cafes.

In [ ]: