

Zomato Data Analysis

Questions:

1. What type of restaurent do the majority of customers order from?
2. How many votes has each type of restraurants received from customers?
3. What are the ratings that the majority of restaurants have received?
4. Zomato has observed that most couples order most of their food online. What is their average spending on each other?
5. Which mode (online or offline) has received the maximum rating?
6. Which type of restaurants received more offline orders, so that Zomato can provide those customers with good offers?

step-1 Importing Libraries

```
In [58]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Step-2 Create the data frame

```
In [59]: dataframe=pd.read_csv("Zomato data .csv")
```

```
In [60]: print(dataframe)
```

```
   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 Indraprastha      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 [5]: dataframe
```

```
Out[5]:
```

	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 Indraprastha	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 x 7 columns

Convert the data type of column - rate

```
In [6]: def handleRate(value):
value=str(value).split('/')
value=value[0]/int(value[1])
return float(value)

dataframe['rate']=dataframe['rate'].apply(handleRate)
```

```
In [7]: dataframe
```

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

148 rows x 7 columns

```
In [9]: 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 restaurent

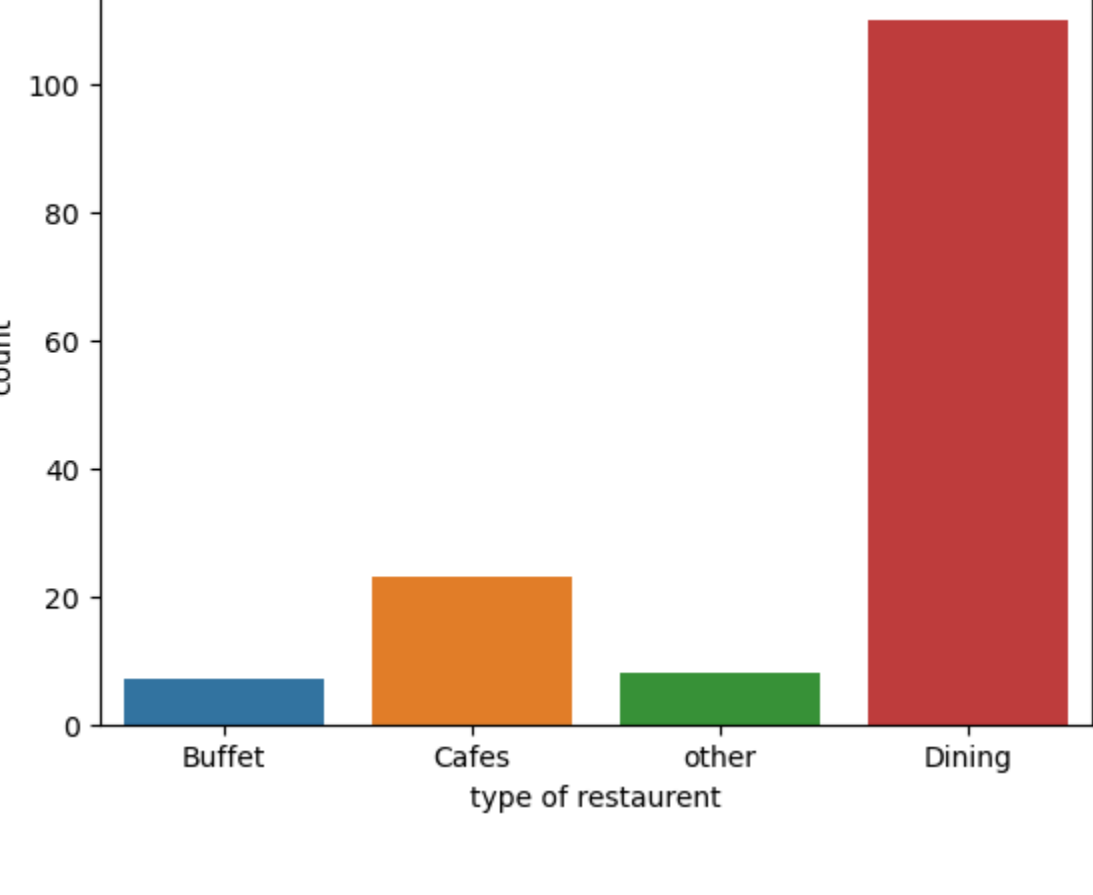
```
In [10]: dataframe.head()
```

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

```
Out[19]: Text(0.5, 0, 'type of restaurant')
```



1. Conclusion - Majority of the restaurent falls in dining category

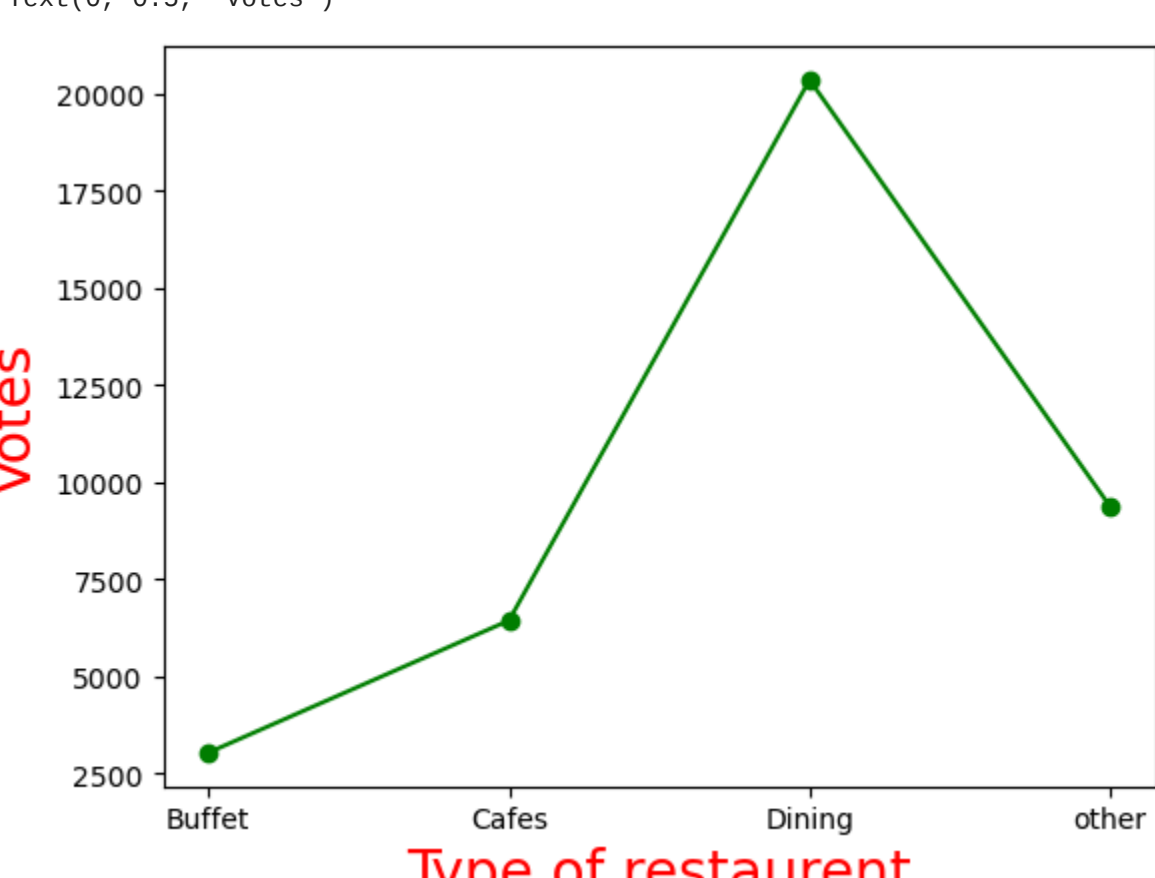
```
In [20]: dataframe.head()
```

```
Out[20]:
```

	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

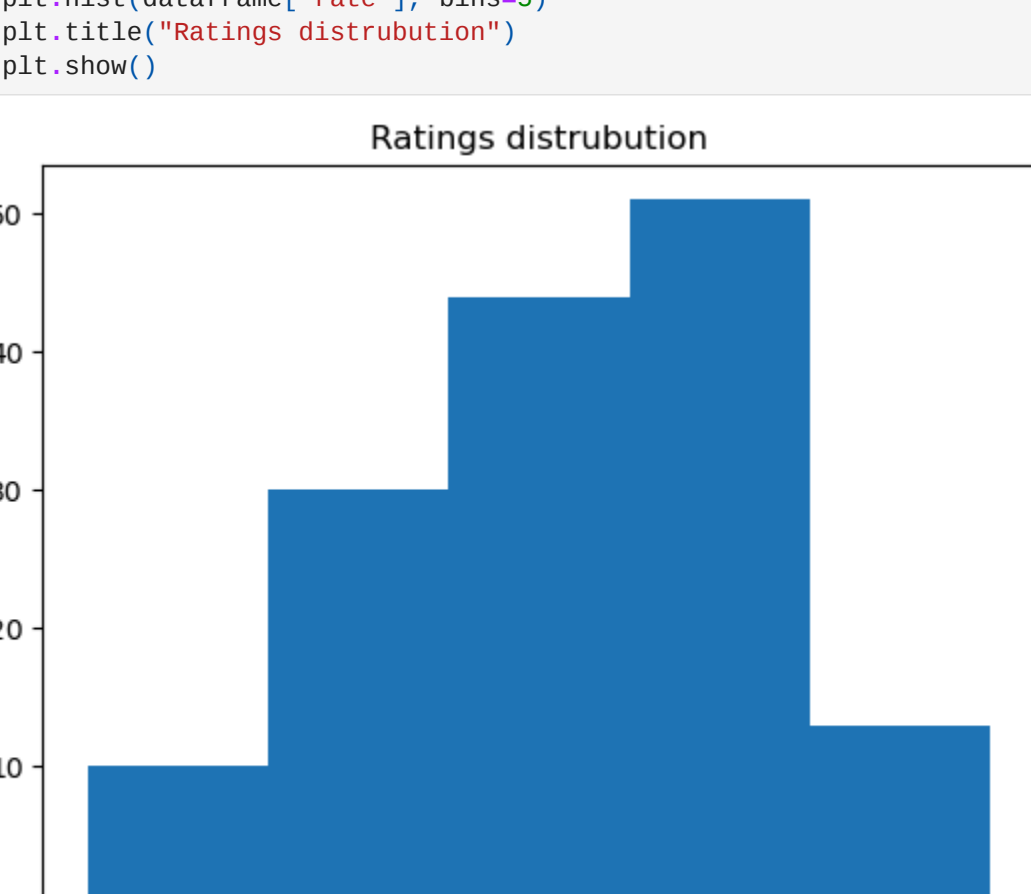
```
In [27]: grouped_data=dataframe.groupby(['listed_in(type)'])['votes'].sum()
result=pd.DataFrame({'votes': grouped_data})
plt.plot(result, c='green', marker='o')
plt.xlabel("type of restaurant", c='red', size=28)
plt.ylabel("votes", c='red', size=20)
```

```
Out[27]: Text(0, 0.5, 'votes')
```



2. Conclusion - dining type has received maximum votes

```
In [28]: plt.hist(dataframe['rate'], bins=5)
plt.title("Ratings distribution")
plt.show()
```



3. Conclusion - the majority restaurants received ratings from 3.5 to 4

Average orders spending by couples

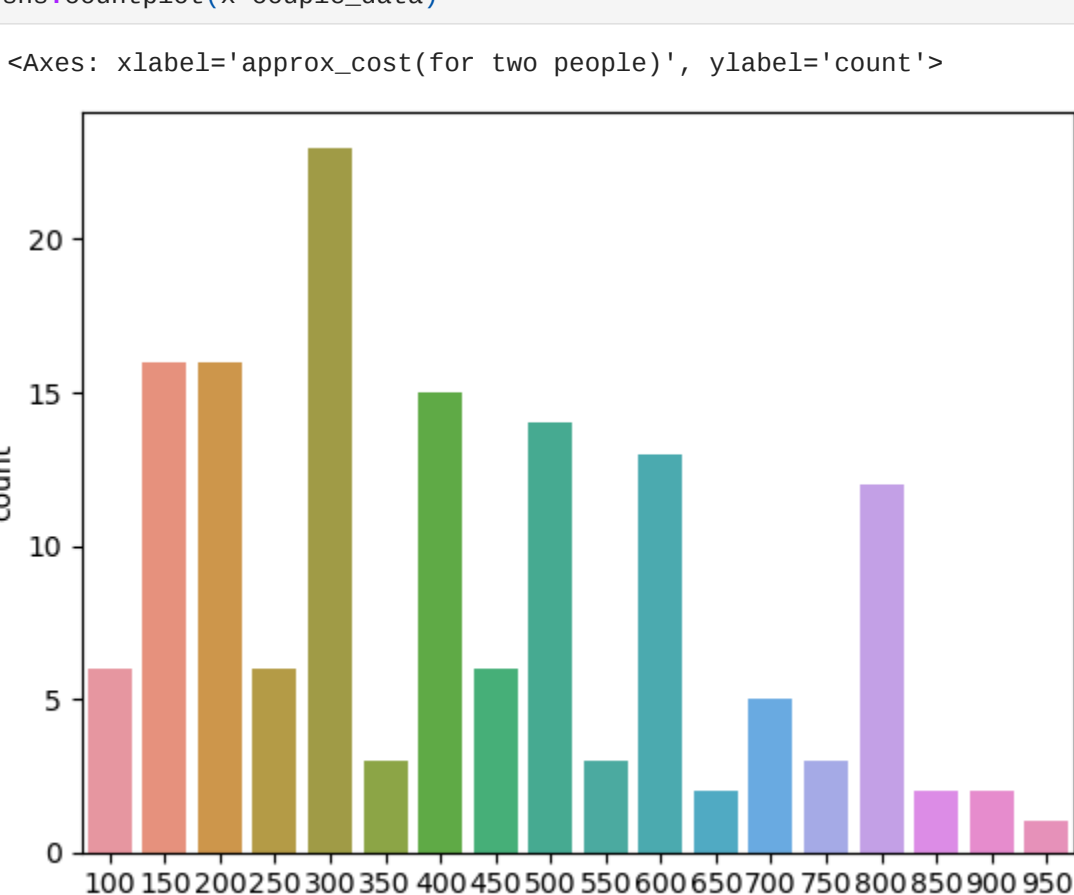
```
In [31]: dataframe.head()
```

```
Out[31]:
```

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

```
Out[36]: <Axes: xlabel='approx_cost(for two people)', ylabel='count'>
```



4. Conclusion - the majority of couples prefer restaurants with an approx cost of Rs. 300

```
In [37]: dataframe.head()
```

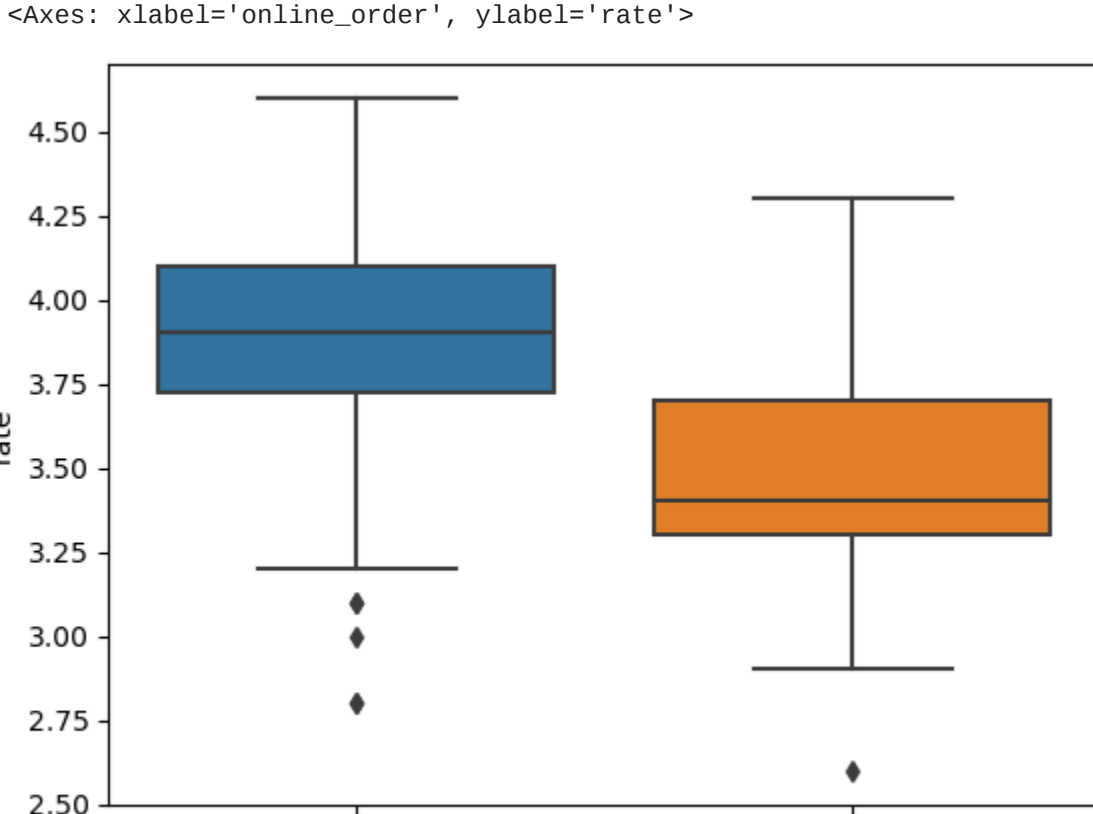
```
Out[37]:
```

	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

Which mode reveives maximum rating

```
In [39]: sns.boxplot(x='online_order', y='rate', data=dataframe)
```

```
Out[39]: <Axes: xlabel='online_order', ylabel='rate'>
```



5. Conclusion - online order receives higher rating

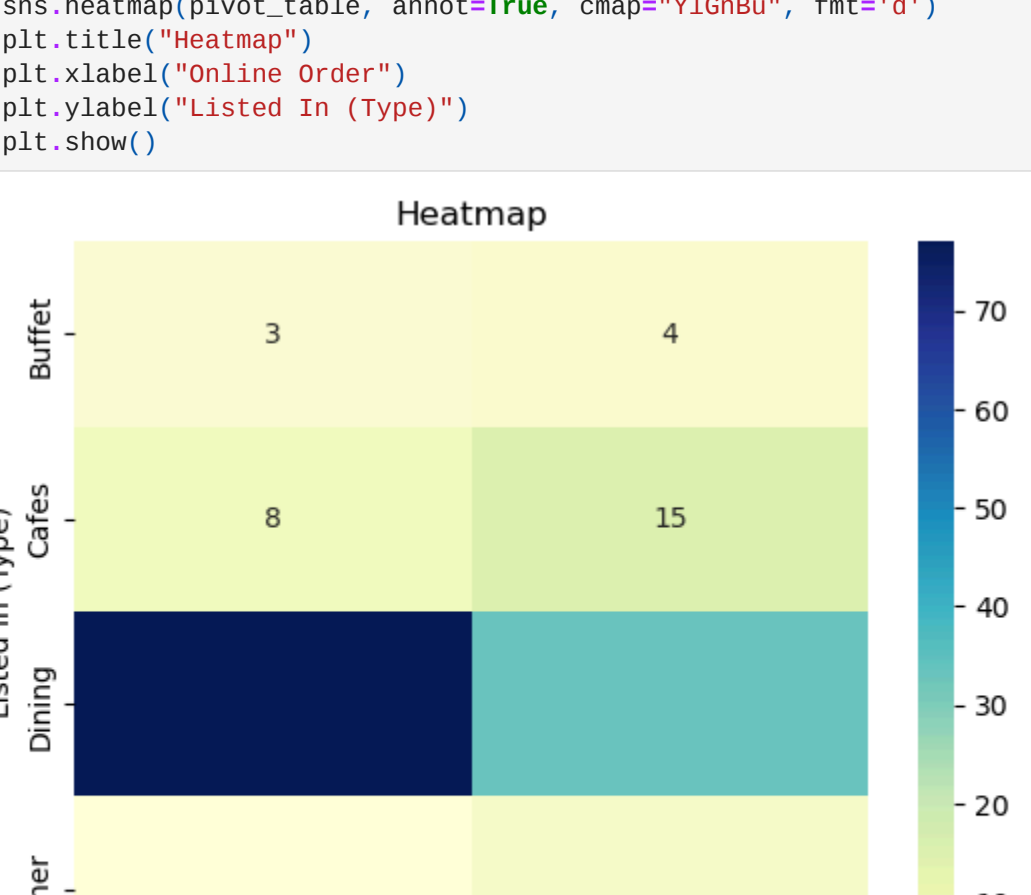
```
In [41]: dataframe.head()
```

```
Out[41]:
```

	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 [63]: pivot_table(dataframe.pivot_table(index='listed_in(type)', columns='online_order', aggfunc='size', fill_value=0))
sns.heatmap(pivot_table, annot=True, cmap='YlGnBu', fmt='d')
```

```
plt.title("Heatmap")
plt.xlabel("Online Order")
plt.ylabel("Listed In (Type)")
plt.show()
```



6. Conclusion - Dining restaurants primarily accept offline orders, whereas cafes primarily receive online orders. This suggests that clients prefers orders in person at restaurants, but prefer online ordering at cafes

