

Zomato data analysis project

Step 1 - Importing libraries

```
In [24]: import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns  
import os
```

Step 2 - Create the data frame

```
In [25]: df=pd.read_csv("Zomato data .csv")  
In [26]: df  
Out[26]:  


|     | 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 × 7 columns


```

convert the data type of column- rate

```
In [27]: 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 | 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          |


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


```

```
In [28]: 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    int64     
 4   votes           148 non-null    int64     
 5   approx_cost     148 non-null    int64     
 6   listed_in       148 non-null    object    
dtypes: float64(1), int64(2), object(4)  
memory usage: 8.2+ KB
```

Type of restaurant

```
In [29]: df.head()  
Out[29]:  


|   | 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 [30]: sns.countplot(x=df['listed_in(type)'])  
plt.xlabel('type of restaurant')
```



Conclusion- majority of the restaurant falls in dining category

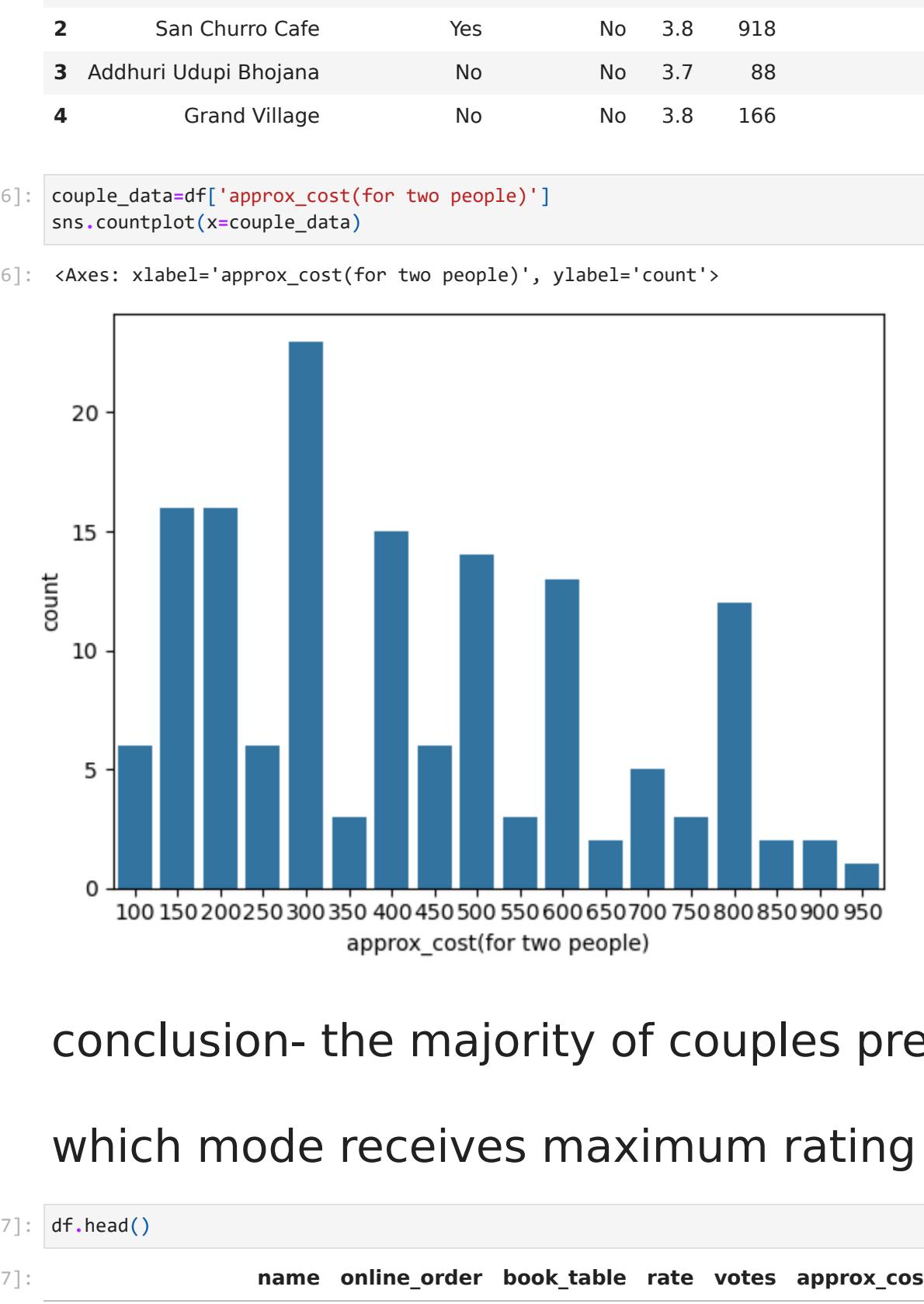
```
In [31]: df.head()  
Out[31]:  


|   | 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          |


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


```



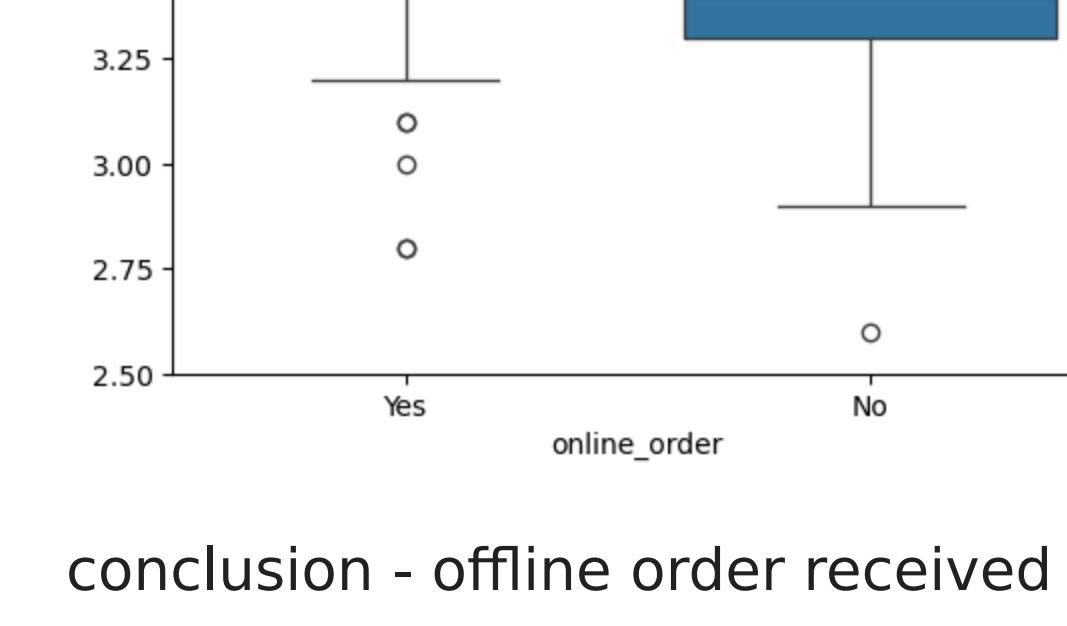
conclusion-dinning restaurants has received maximum votes

```
In [32]: df.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          |


```

```
In [33]: plt.hist(df['rate'],bins=10)  
plt.title("ratings distribution")  
plt.show()
```



conclusion - the majority restaurants received rating from 3.5 to 4

Average order spending by couples

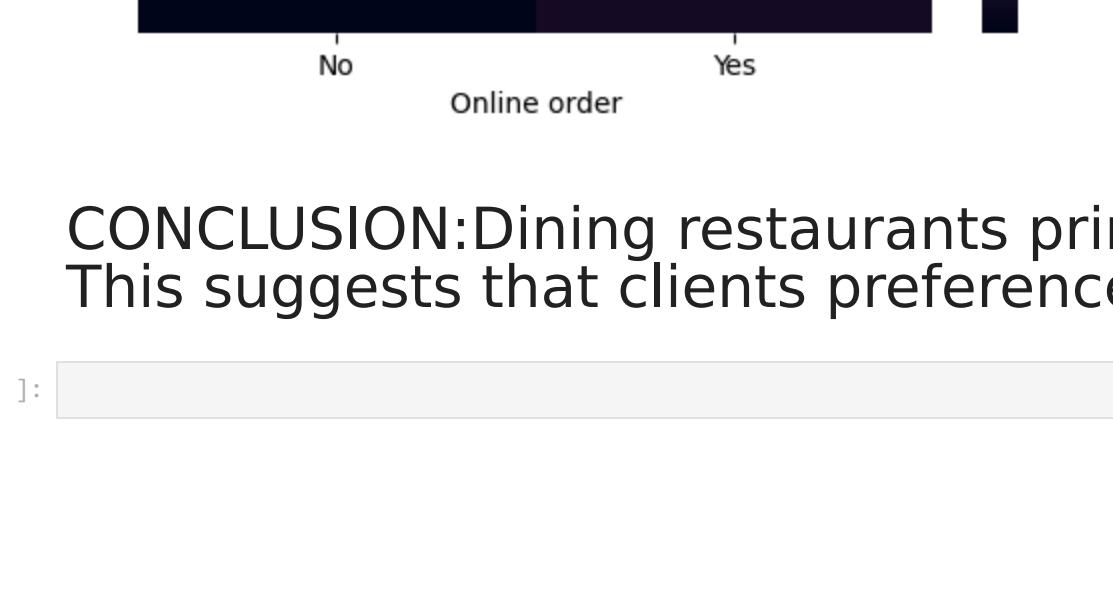
```
In [34]: df.head()  
Out[34]:  


|   | 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          |


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


```



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

```
In [35]: plt.figure(figsize=(6,6))  
sns.boxplot(x=df['online_order'],y=df['rate'])  
plt.xlabel("online_order", y=450)  
plt.ylabel("rate", y=300)
```



CONCLUSION:Dining restaurants primarily accept offline order , where as cafes primarily receive online orders. This suggests that clients preference orders in person at restaurants , but prefer online ordering at cafes