

Zomato data analysis project

Step 1 - Importing Libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

step 2 - Create the data frame

```
dataframe = pd.read_csv("Zomato data.csv")
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 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]

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 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]

convert the data type of column - rate

```
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

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

```

1) What type of restaurant do the majority of customers order from?

```

dataframe.head()

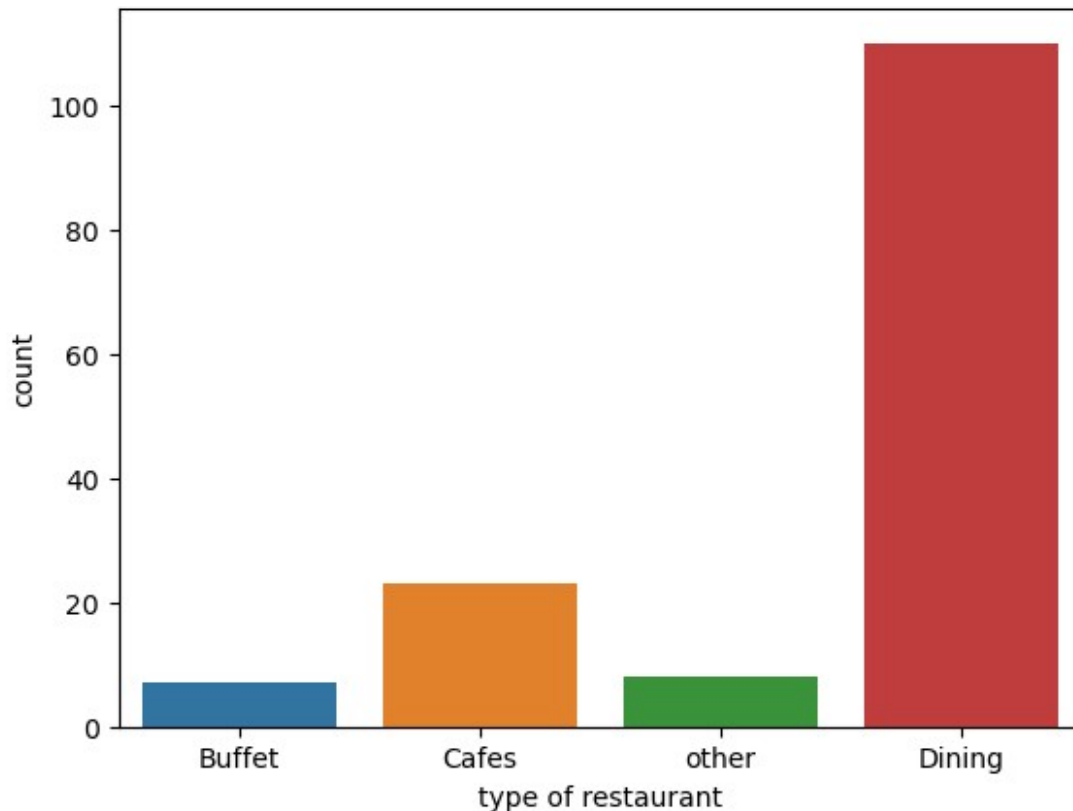
   name online_order book_table  rate  votes  \
0   Jalsa          Yes        Yes   4.1    775
1  Spice Elephant          Yes         No   4.1    787
2  San Churro Cafe          Yes         No   3.8    918
3  Addhuri Udupi Bhojana          No         No   3.7     88
4   Grand Village          No         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

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

Text(0.5, 0, 'type of restaurant')

```



conclusion - majority of the restaurant falls in dining category

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

```
dataframe.head()
```

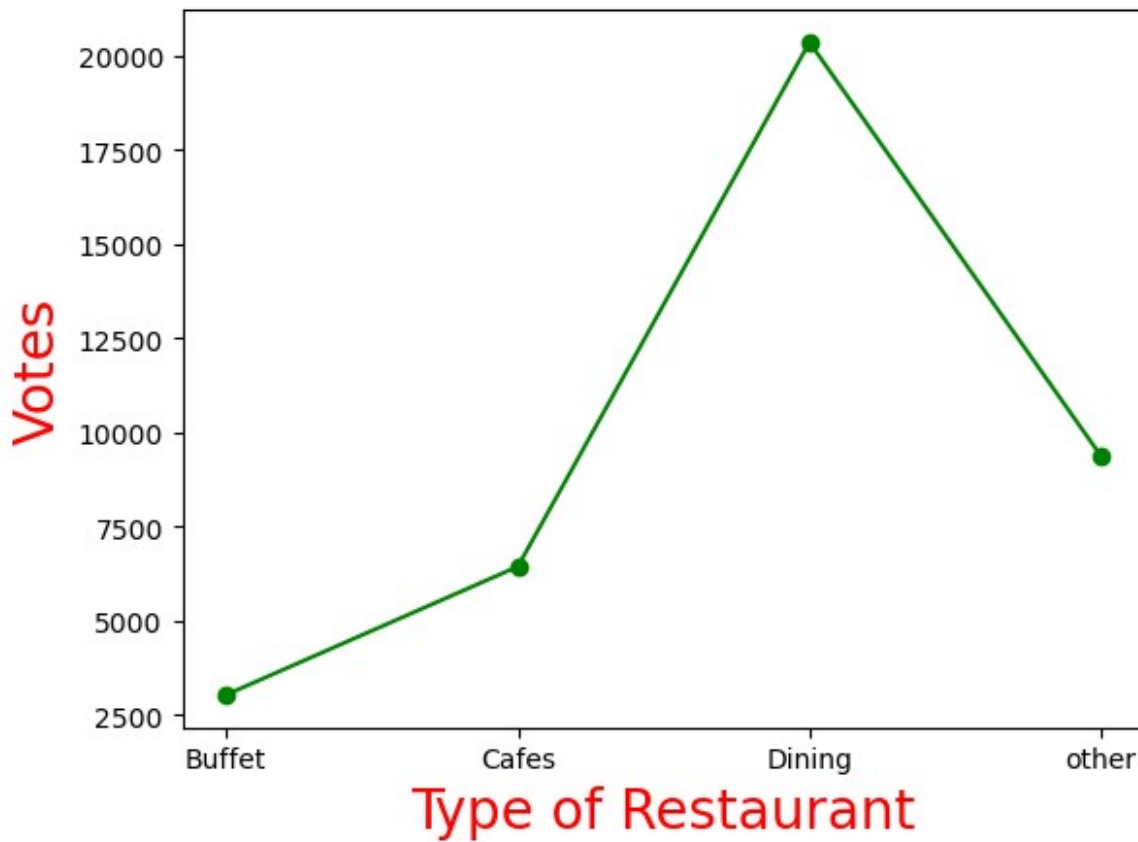
| | name | online_order | book_table | rate | votes | \ |
|---|-----------------------|--------------|------------|------|-------|---|
| 0 | Jalsa | Yes | Yes | 4.1 | 775 | |
| 1 | Spice Elephant | Yes | No | 4.1 | 787 | |
| 2 | San Churro Cafe | Yes | No | 3.8 | 918 | |
| 3 | Addhuri Udupi Bhojana | 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 |

```
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=20)  
plt.ylabel('Votes',c='red',size=20)
```

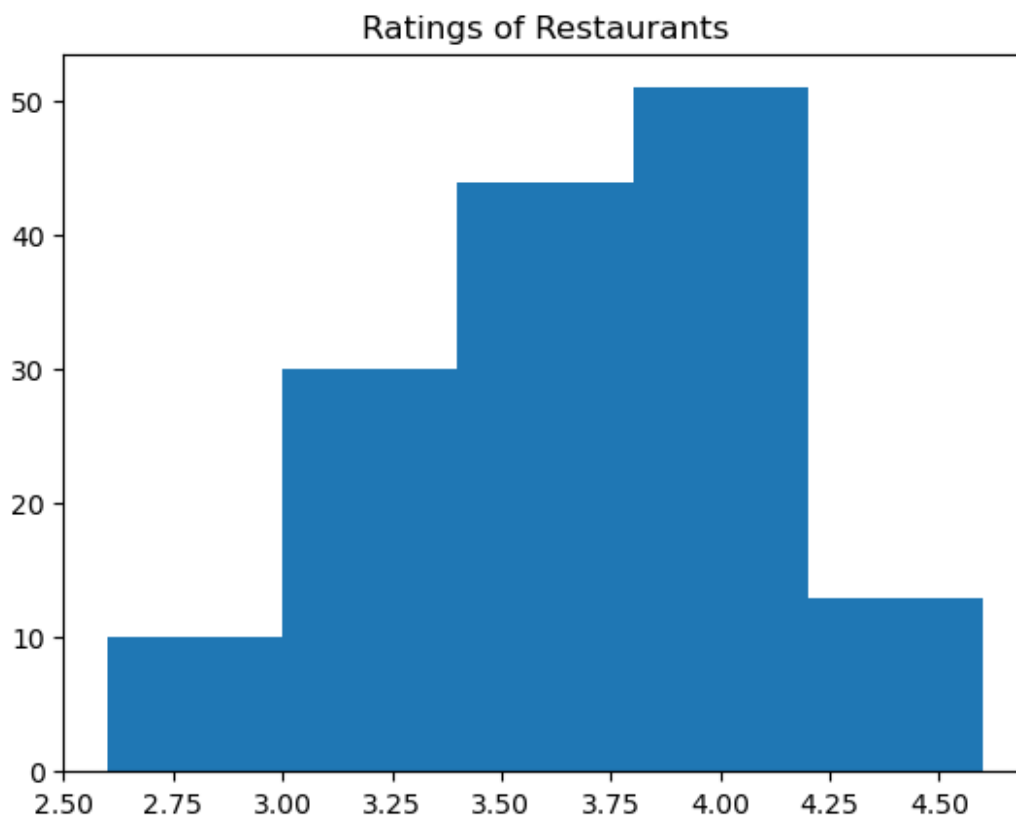
```
Text(0, 0.5, 'Votes')
```



Conclusion - dinning restaurant has recieved maximum votes

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

```
plt.hist(dataframe['rate'],bins=5)  
plt.title('Ratings of Restaurants')  
plt.show()
```



Conclusion - the majority of restaurants received ratings from 3.5 to 4

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

```
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 Udipi Bhojana | No | No | 3.7 | 88 | |
| 4 | Grand Village | No | No | 3.8 | 166 | |

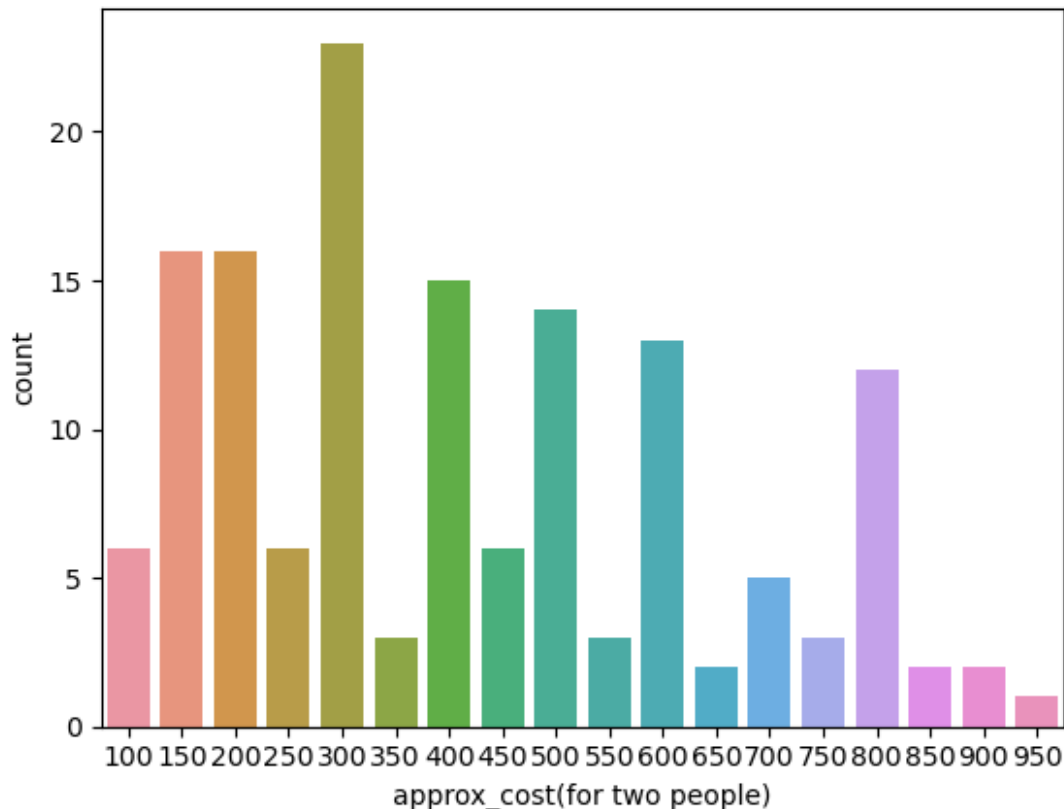
```

    approx_cost(for two people) listed_in(type)
0                800          Buffet
1                800          Buffet
2                800          Buffet
3                300          Buffet
4                600          Buffet

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

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

```



conclusion - the mahority of couples preferr
resturants with an approxiamate cost of 300
rupee

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

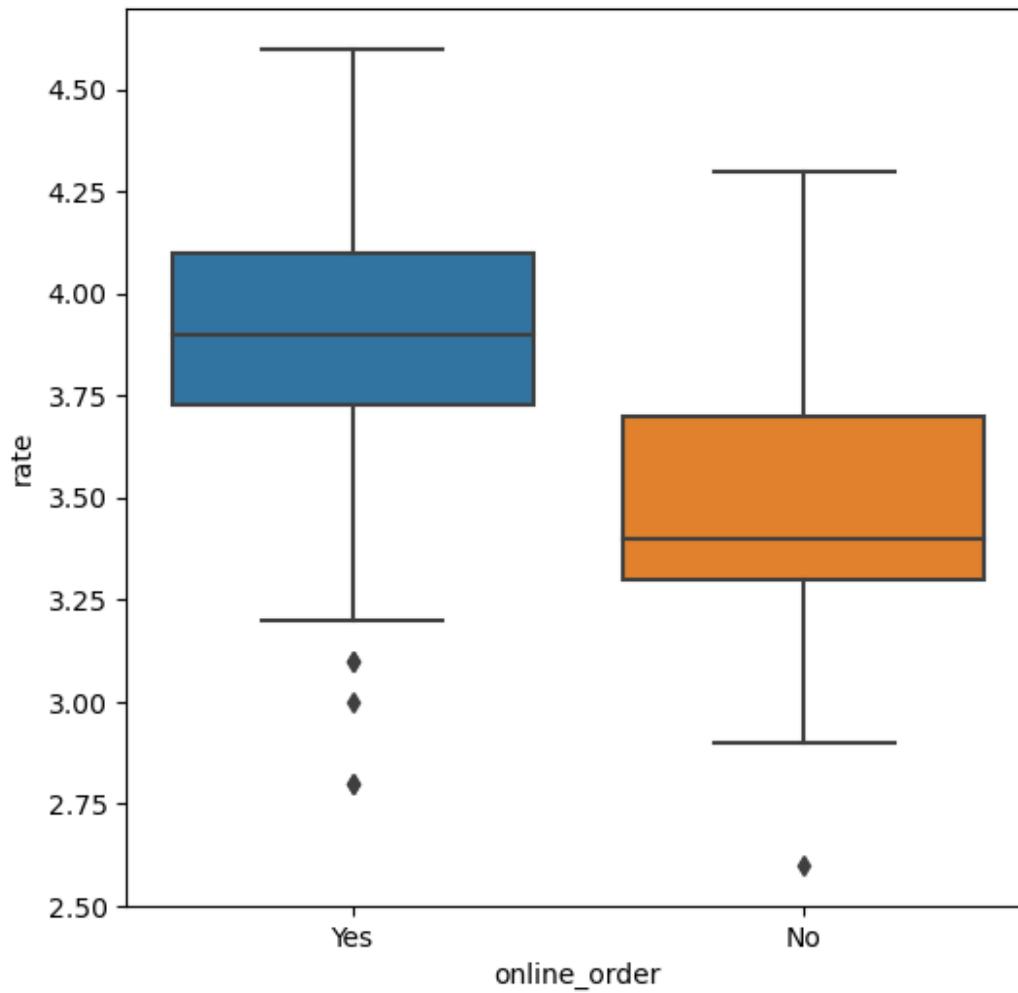
```
dataframe.head()
```

| | 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 Udipi 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 |

```
plt.figure(figsize=(6,6))
sns.boxplot(x='online_order',y='rate',data=dataframe)
<Axes: xlabel='online_order', ylabel='rate'>
```



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

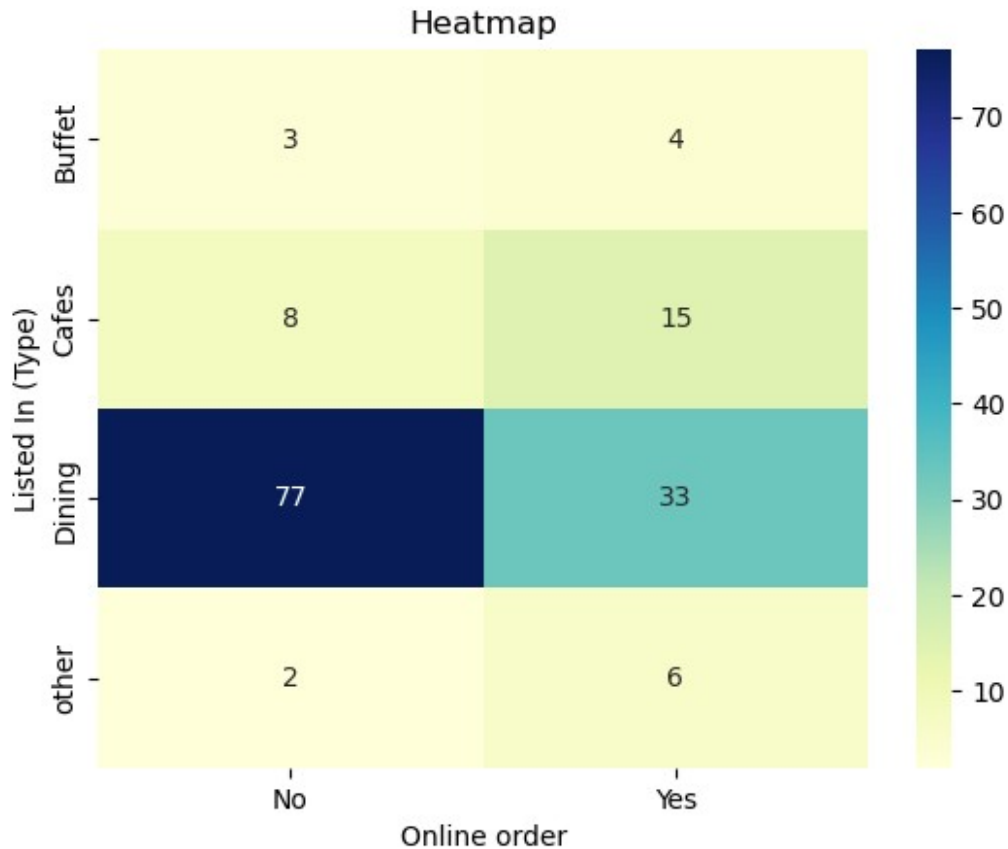
6) Which type of restaurant received more offline orders so that zomato can provide customers with some good offers?

```
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

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()
```



conclusion - Dining restaurants primarily accept offline orders m whereas cafes primarily receive online orders. This suggests that clients prefer orders in person at restaurants but prefer online ordering at cafes.