ZOMATO DATA ANALYSIS

IMPORTING LIBRARIES

```
In [1]:  import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import seaborn as sns
```

Load Dataset

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

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

Out[4]:

	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]:
```python
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
```

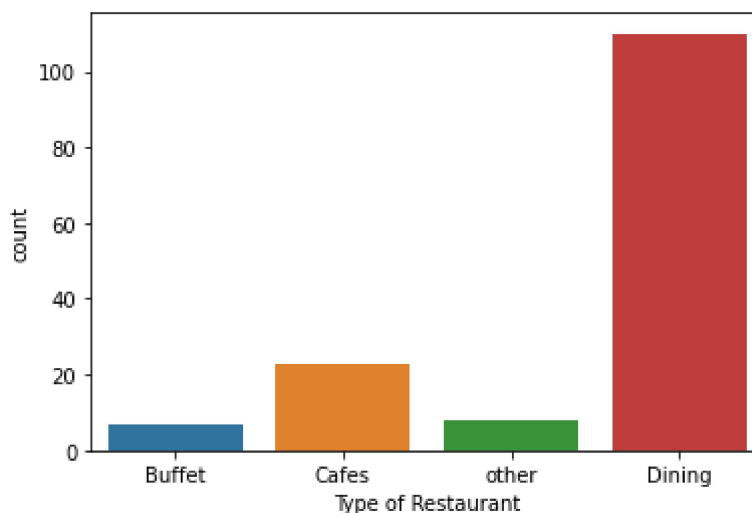# What type of Restaurant do majority customer order from ?

In [7]:
```python
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 [8]:
```python
sns.countplot(x=df['listed_in(type)'])
plt.xlabel("Type of Restaurant")
```

Out[8]: Text(0.5, 0, 'Type of Restaurant')

Conclusion : majority of the restaurant falls in Dinning Category so majority people order from dining restaurants

In [ ]:

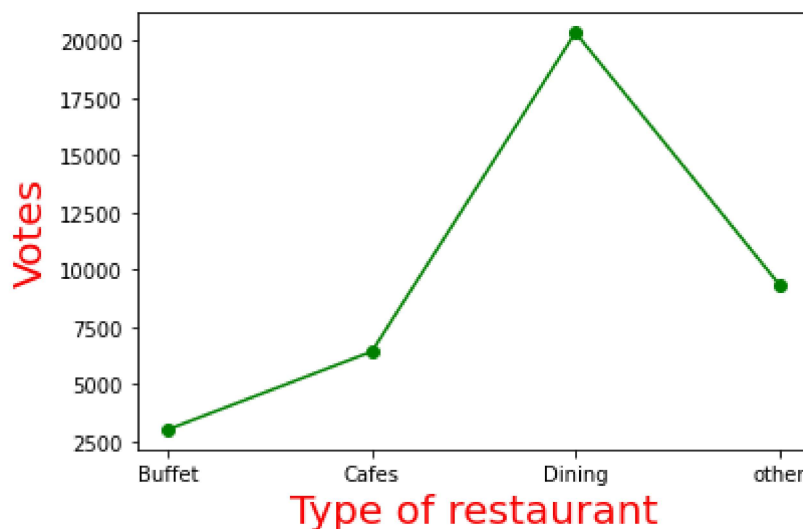# How many votes has each type of restaurants recieved from customer

In [9]: `df.head()`

Out[9]:

	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 [13]:
```python
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 restaurant", c="red", size=20)
plt.ylabel("Votes", c="red", size=20)
```

Out[13]: `Text(0, 0.5, 'Votes')`

Conclusion: Dinning Restaurants has recieved maximum votes

In [ ]:

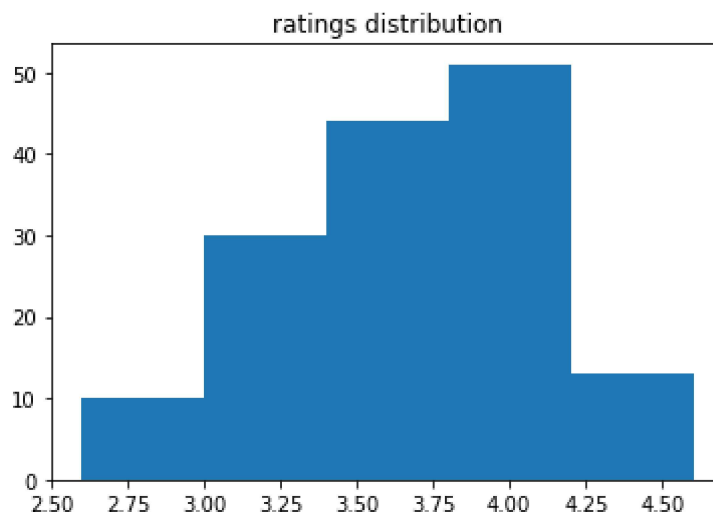# What are the ratings that the majority of restaurants have recieved

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

Out[15]:

	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 [16]:
```
plt.hist(df['rate'],bins=5)
plt.title("ratings distribution")
plt.show
```

Out[16]: `<function matplotlib.pyplot.show(close=None, block=None)>`

Conclusion: The majority Restaurants recieved ratings from 3.5 to 4
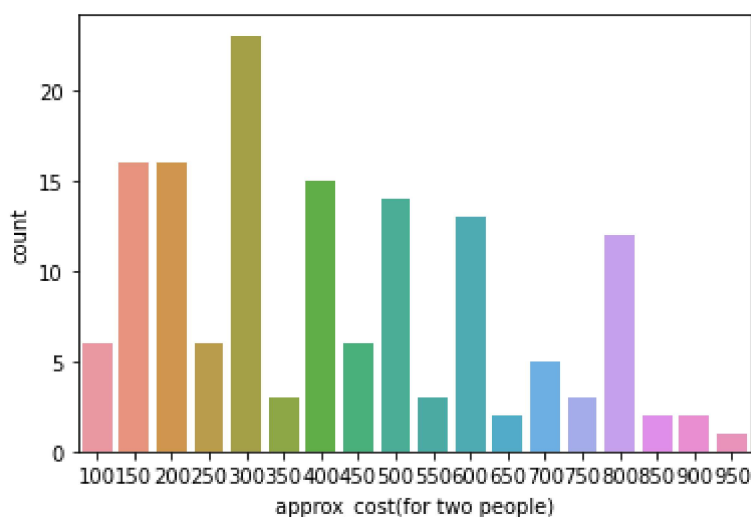
In [ ]:

In [ ]:

Average order spending by Couples

In [18]:
```python
df.head()
```

Out[18]:

	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 [21]:
```python
couple_data=df['approx_cost(for two people)']
sns.countplot(x=couple_data)
```

Out[21]: <AxesSubplot:xlabel='approx_cost(for two people)', ylabel='count'>

Conclusion: The majority of couples prefer restaurants withan approximate cost of 300 rupees for 2 people

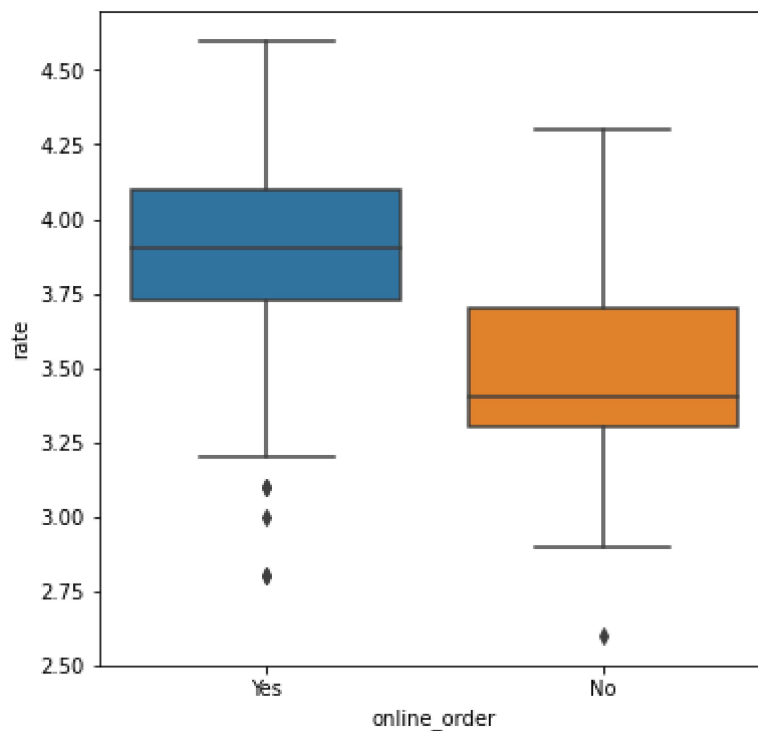In [ ]:

In [ ]:

Which mode recieves maximum rating

In [23]: `df.head()`

Out[23]:

	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 [24]:
```
plt.figure(figsize=(6,6))
sns.boxplot(x='online_order', y='rate',data=df)
```

Out[24]: `<AxesSubplot:xlabel='online_order', ylabel='rate'>`

Conclusion: Offline order recieved lower rating in comparison to online order

In [26]: `df.head()`

Out[26]:

	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 [28]:
```python
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()
```
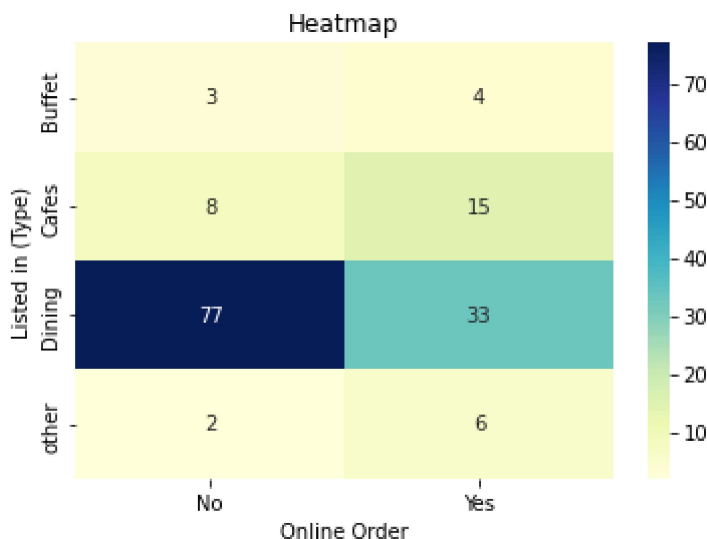

CONCLUSION: Dining Restaurants primarily accept offline orders, whereas cafes primarily recieve online order. This suggests that clients prefers oders in person at restaurants, but prefer online order in Cafes.

In [ ]: