zomato data analysis project

In [18]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [19]:
```
df = pd.read_csv("Zomato data .csv")
df
```

Out[19]:

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

148 rows × 7 columns

In [20]:
```
df.describe()
```

Out[20]:

	votes	approx_cost(for two people)
count	148.000000	148.000000
mean	264.810811	418.243243
std	653.676951	223.085098
min	0.000000	100.000000
25%	6.750000	200.000000
50%	43.500000	400.000000
75%	221.750000	600.000000
max	4884.000000	950.000000

In [21]:
```
def handleRate(value):
    value=str(value).split('/')
    value=value[0];
    return float(value)

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

In [22]:
```
df.head()
```

Out[22]:

	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 [23]:
```
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
```

type of restuarant

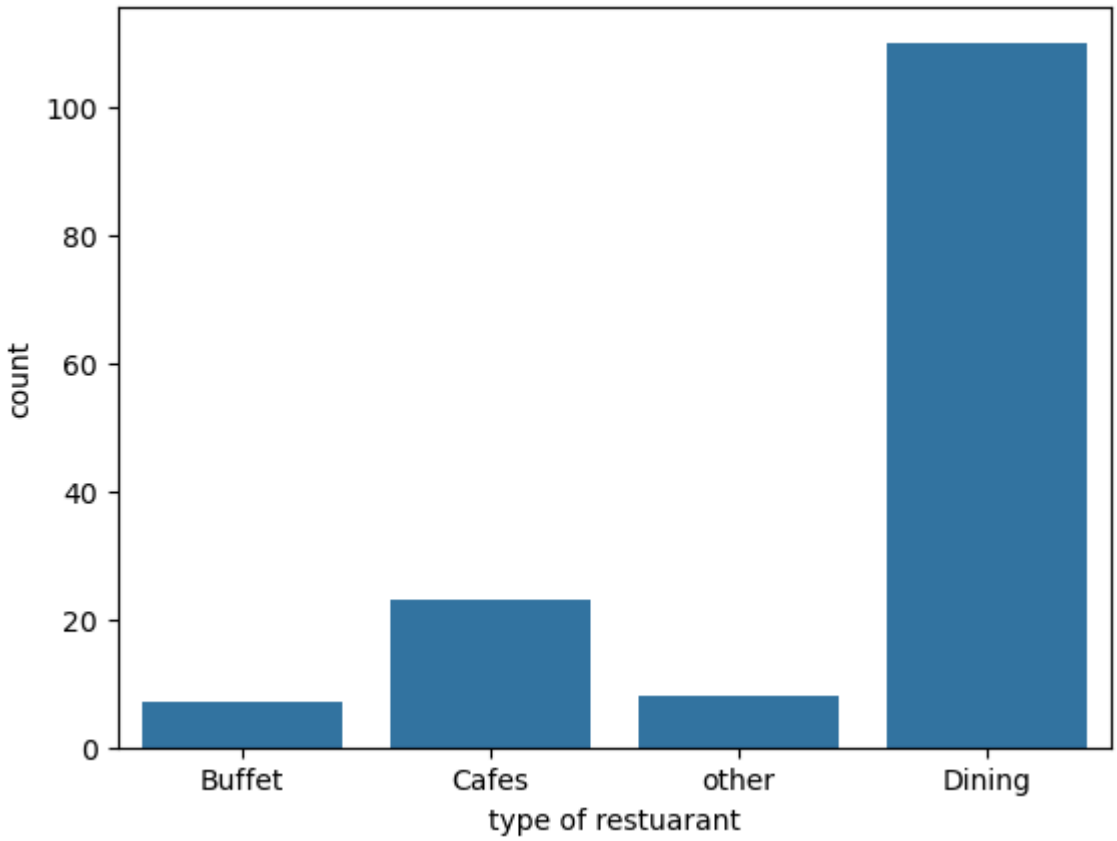
In [24]:
```
df.head()
```

Out[24]:

	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 [25]:
```
sns.countplot(x=df['listed_in(type)'])
plt.xlabel("type of restuarant")
```

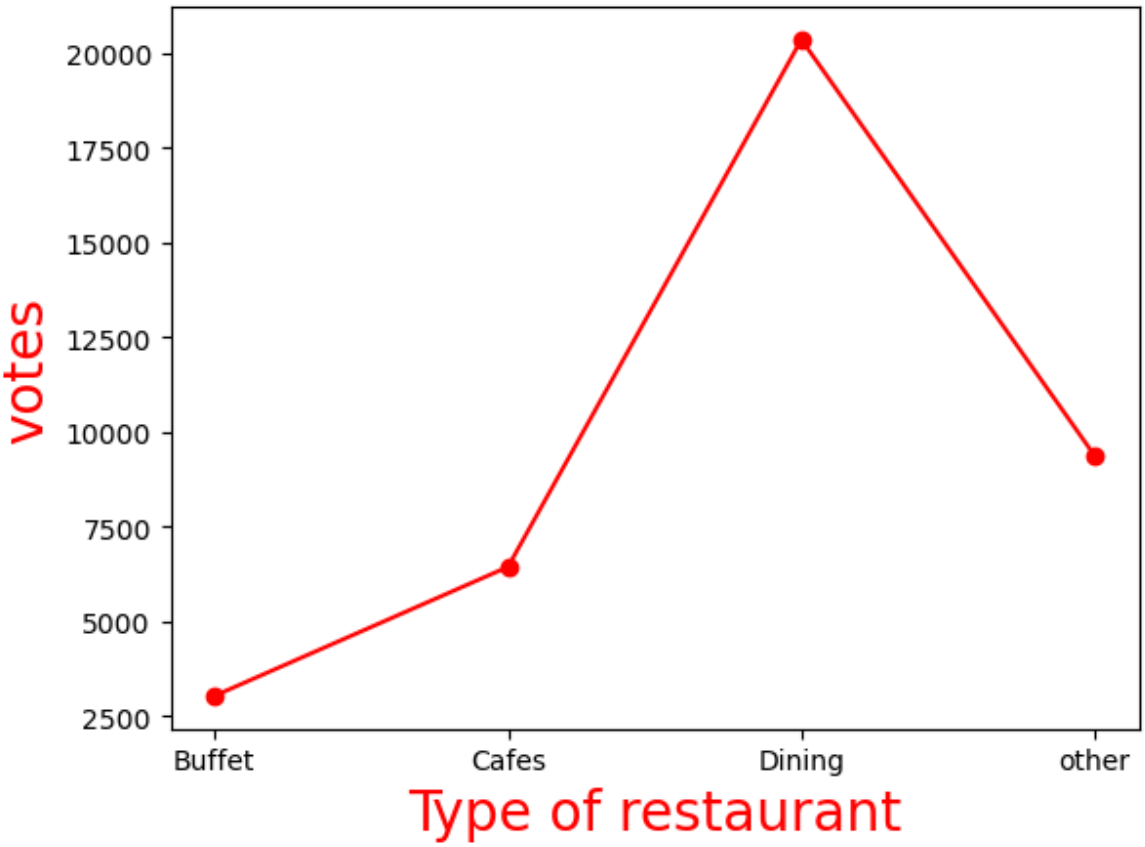
Out[25]:
```
Text(0.5, 0, 'type of restuarant')
```



conclusion: majority of the restaurant falls in dinning category

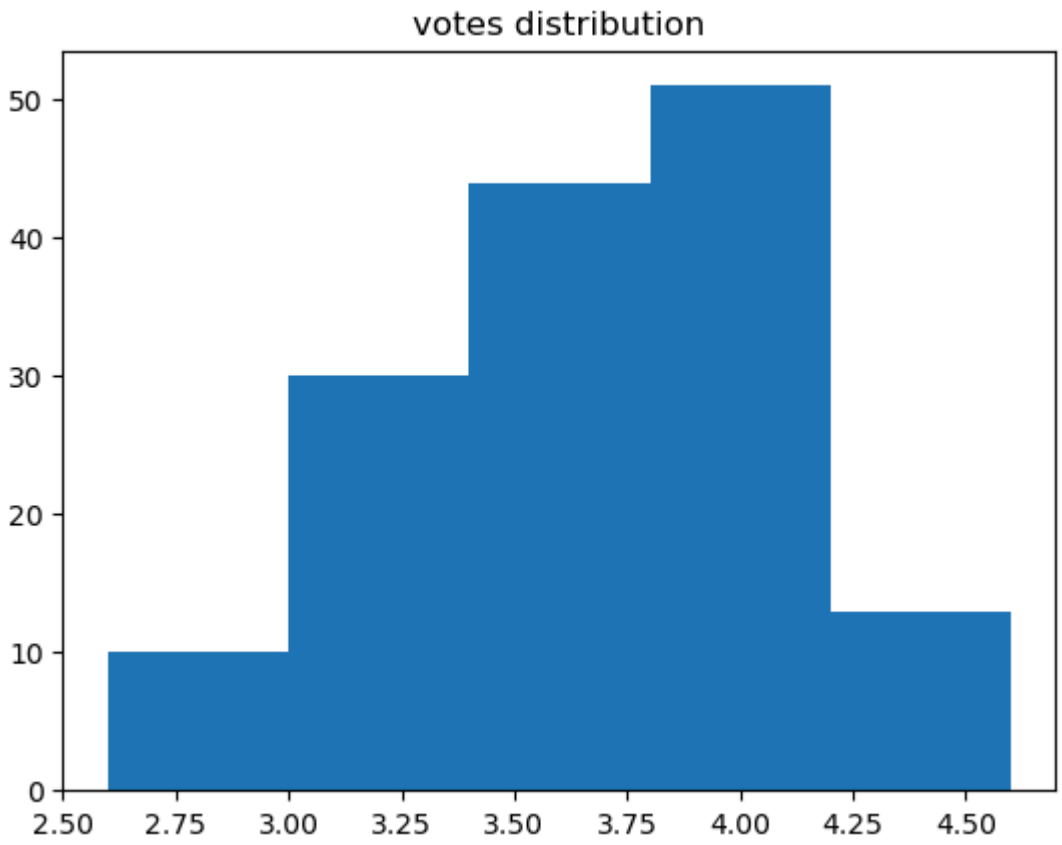
In [29]:
```
grouped_data = df.groupby('listed_in(type)')['votes'].sum()
result=pd.DataFrame({'votes': grouped_data})
plt.plot(result,c='red',marker="o")
plt.xlabel("Type of restaurant",c="red",size=20)
plt.ylabel("votes",c="red",size=20)
```

Out[29]:
```
Text(0, 0.5, 'votes')
```



conclusion: dining restaurant has recieved maximum votes

In [32]:
```
plt.hist(df['rate'],bins=5)
plt.title("votes distribution")
plt.show()
```



conclusion: majority restaurant recieved rating from 3.5to 4

Average order spending by couples

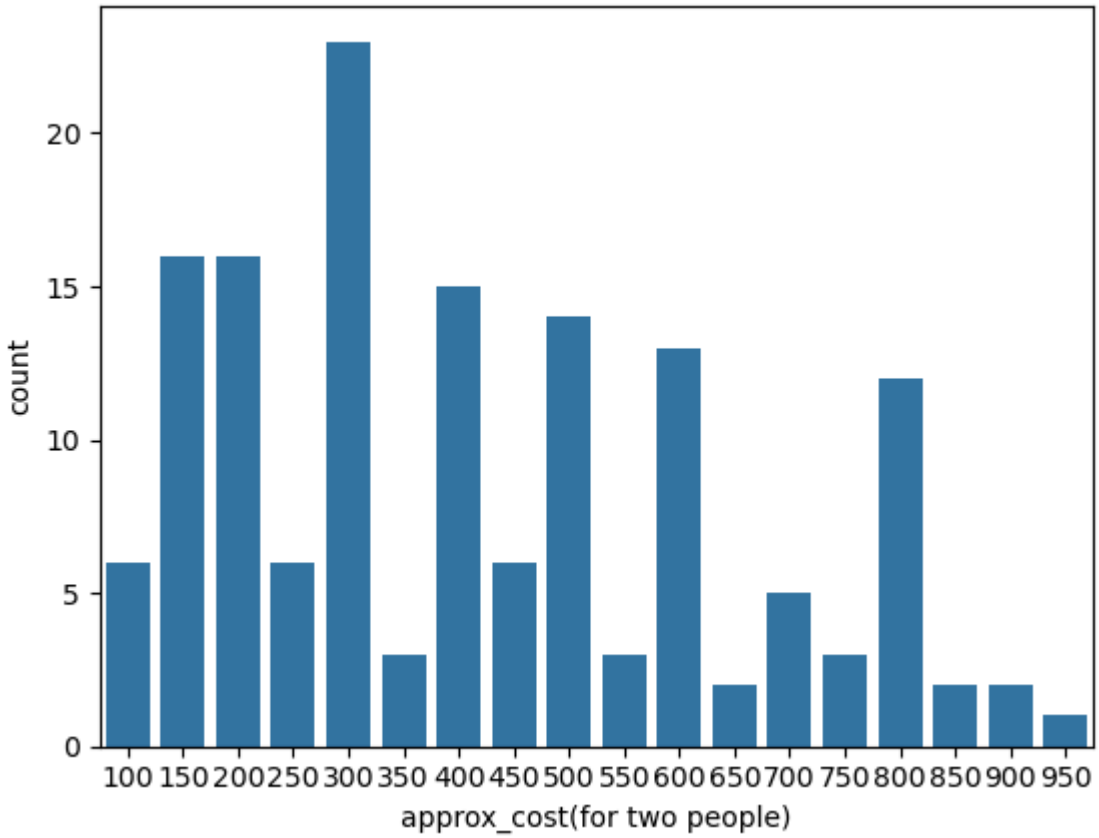
In [33]:
```
df.head()
```

Out[33]:

	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 [34]:
```
couple_data=df['approx_cost(for two people)']
sns.countplot(x=couple_data)
```

Out[34]: <Axes: xlabel='approx_cost(for two people)', ylabel='count'>

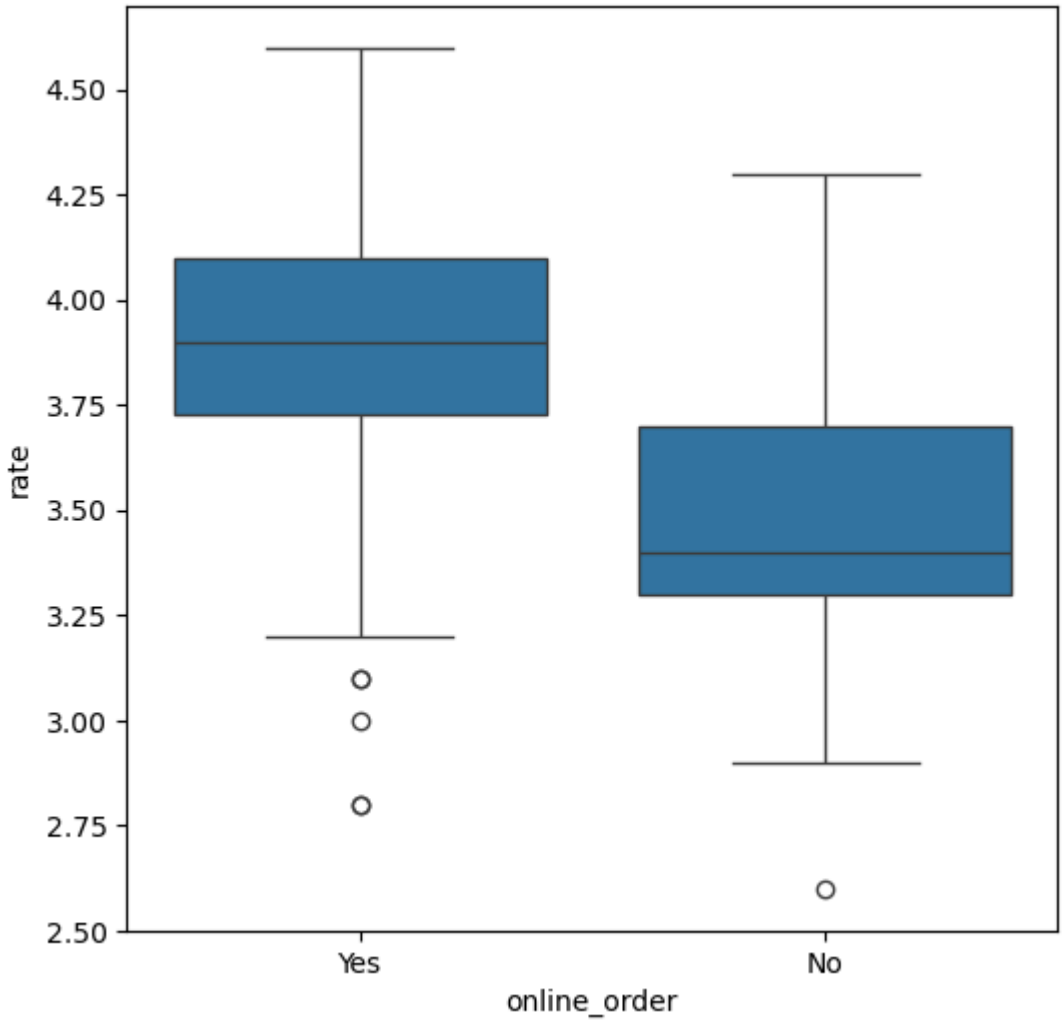


conclusion: the majority of couple preffred restuarant with an approxiamate cost of 300 rupees

which modes recieves maximum rating

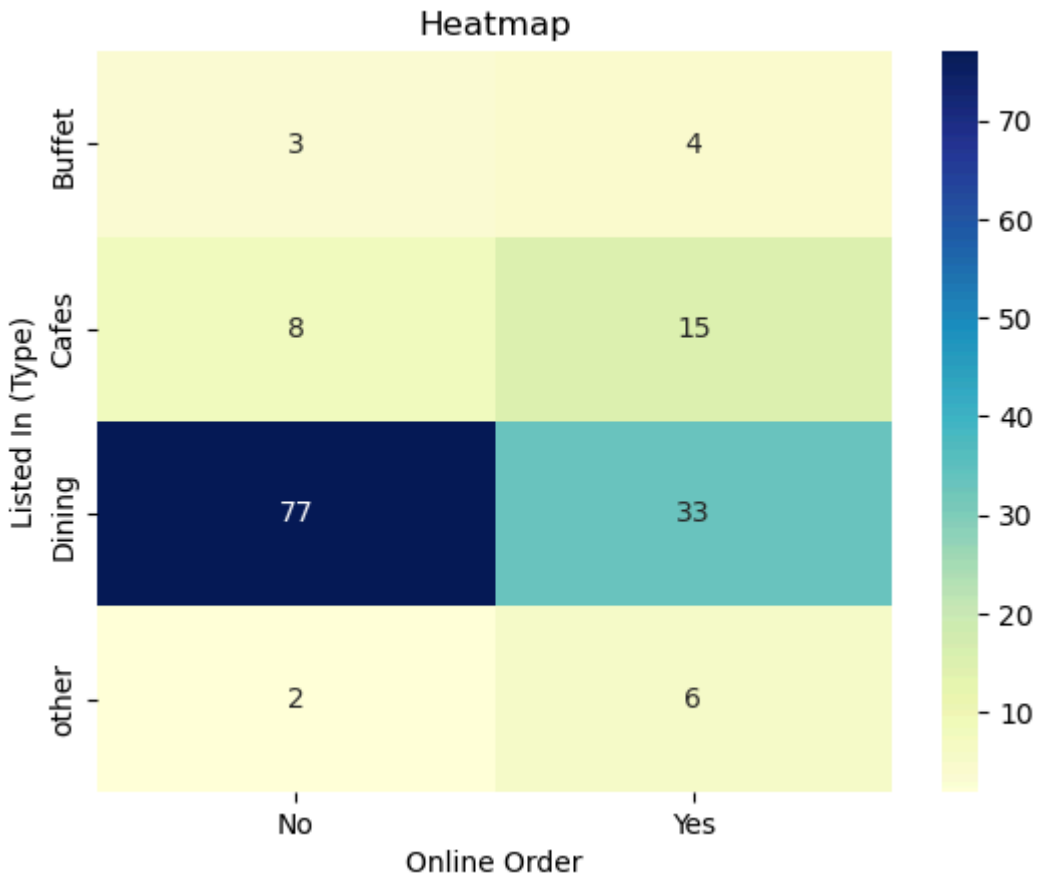
In [37]:
```
plt.figure(figsize = (6,6))
sns.boxplot( x = 'online_order', y= 'rate',data = df)
```

Out[37]: <Axes: xlabel='online_order', ylabel='rate'>



conclusion: offline order recieved lower rating in comparison to online order

In [39]:
```
pivot_table = df.pivot_table(index='listed_in(type)', columns='online_order', ag
sns.heatmap(pivot_table, annot=True, cmap="YlGnBu", fmt='d')
plt.title("Heatmap")
plt.xlabel("Online Order")
plt.ylabel("Listed In (Type)")
plt.show()
```



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

In [ ]: