In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [5]:
```
df=pd.read_csv("zomato.csv")
print(df.head())
```

```
                   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

   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]:
```
df=pd.read_csv("zomato.csv")
```

In [7]:
```
df
```

Out[7]:

	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

convert the data type of the "rate" column to float and removed the denominator

In [4]:
```
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
```

summary of the dataframe

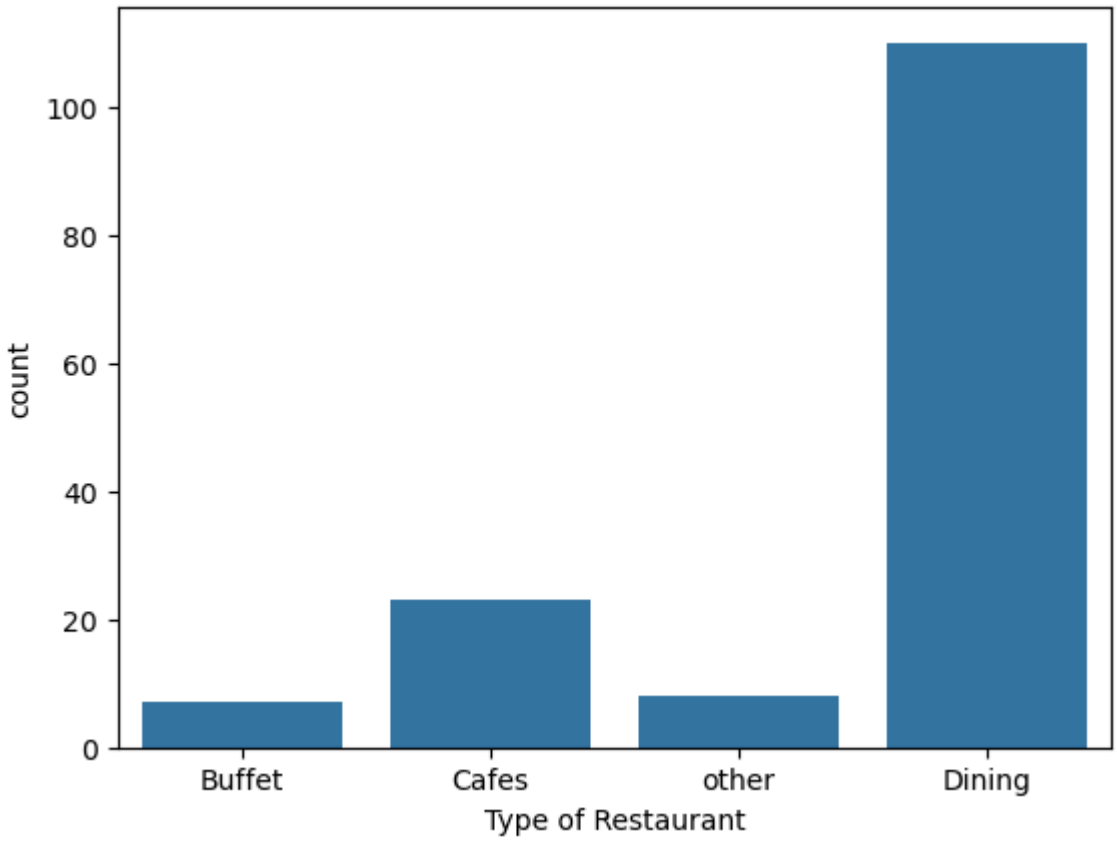
In [18]:
```
df.info()
```

```
<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
```

1>type of restaurant

In [23]:
```
df.head()
sns.countplot(x=df['listed_in(type)'])
plt.xlabel("Type of Restaurant")
```

Out[23]:
```
Text(0.5, 0, 'Type of Restaurant')
```

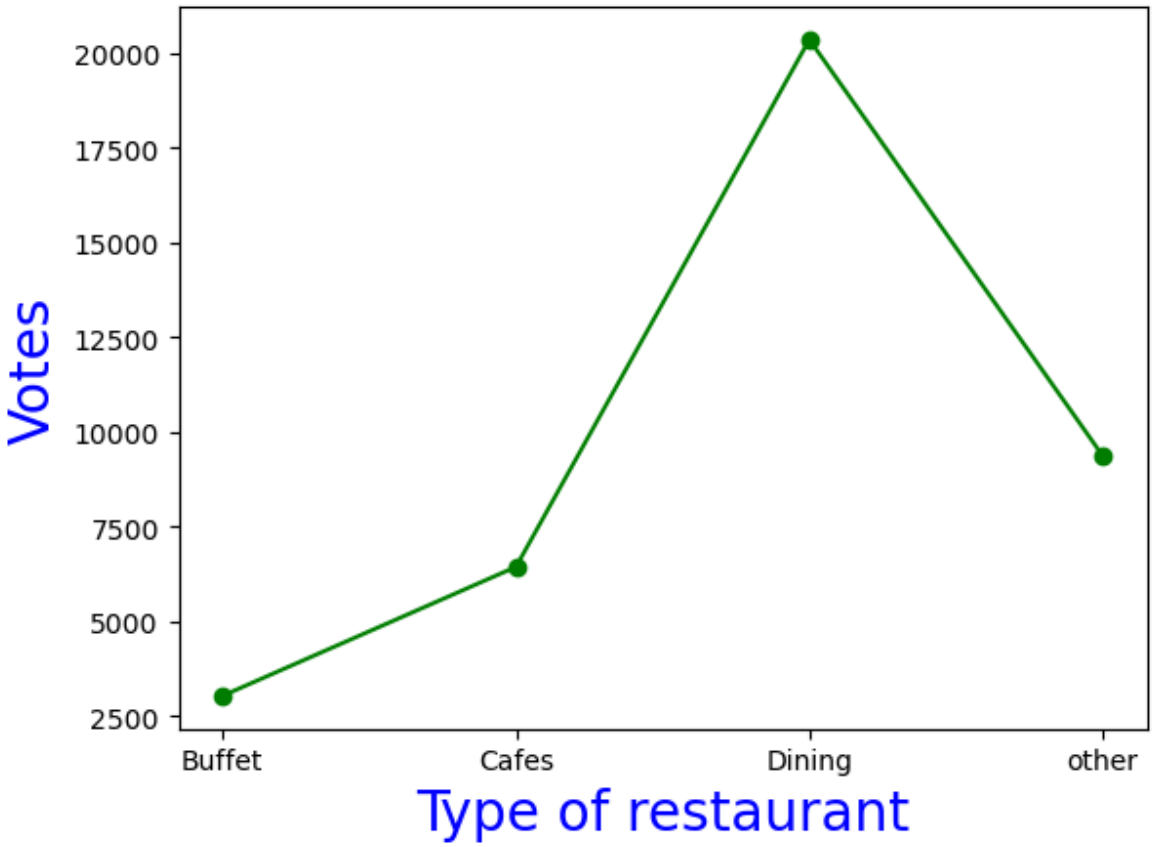


conclusion:The majority of the restaurant fall into the dining category.

2>votes gained by the restaurant

In [27]:
```
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="blue",size=20)
plt.ylabel("Votes",c="blue",size=20)
```

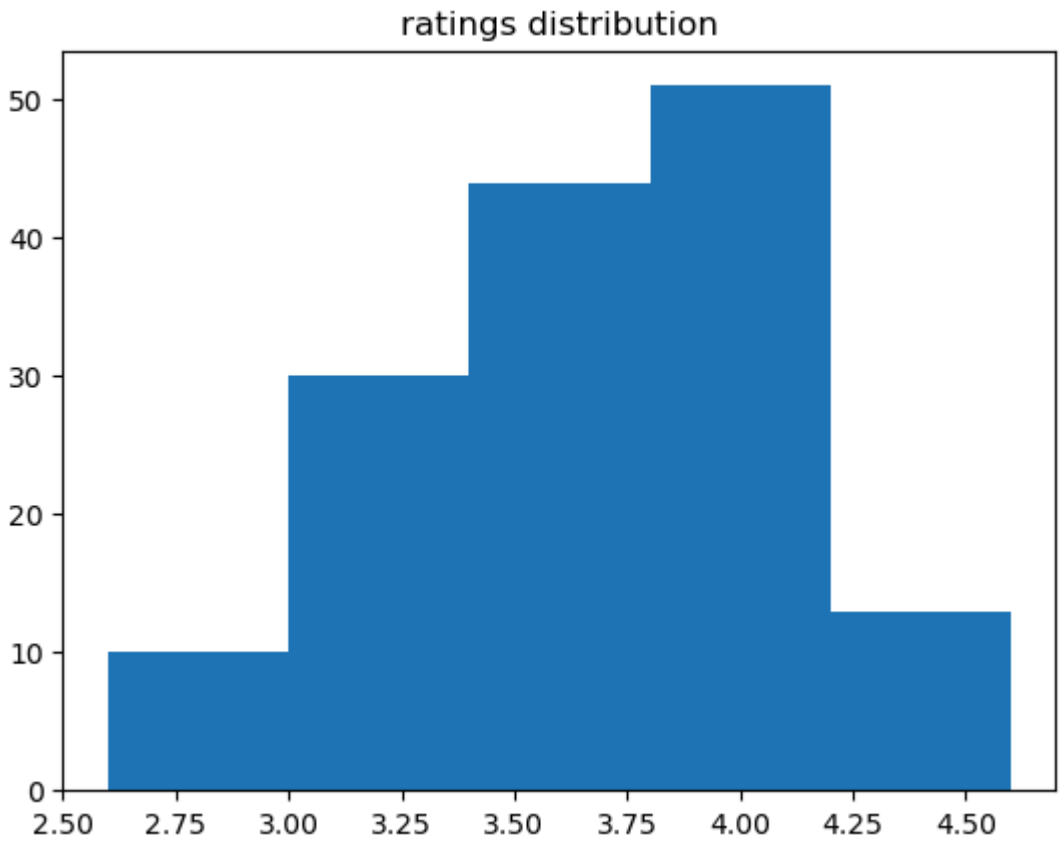
Out[27]:
```
Text(0, 0.5, 'Votes')
```



conclusion:the majority of the restaurants received ratings

3>majority of the ratings

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

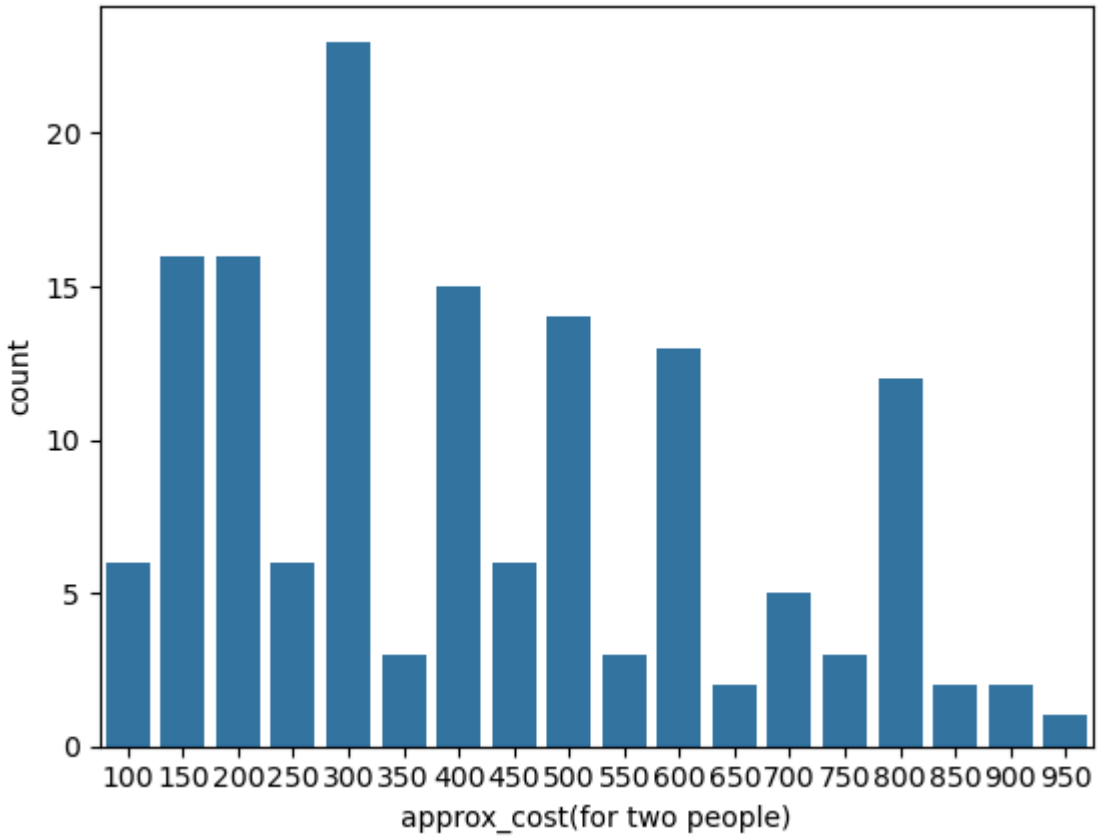


Conclusion: The majority of restaurants received ratings ranging from 3.5 to 4

4>average order spending by couples

In [29]:
```
couple_data=df['approx_cost(for two people)']
sns.countplot(x=couple_data)
```

Out[29]:
```
<Axes: xlabel='approx_cost(for two people)', ylabel='count'>
```

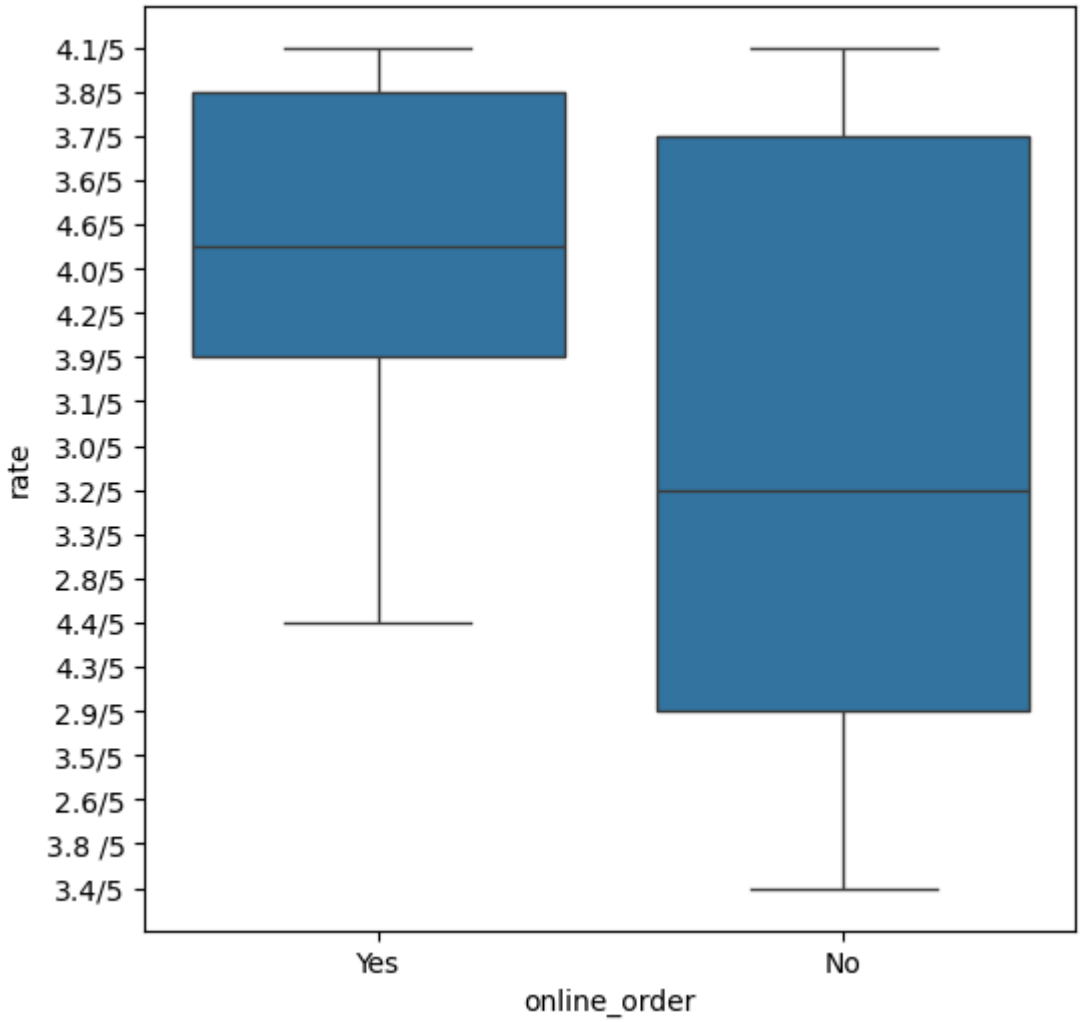


conclusion: online order receives more rating than offline orders

5>which mode(online or offline)received the maximum rating..?

In [8]:
```
plt.figure(figsize=(6,6))
sns.boxplot(x='online_order',y='rate',data=df)
```

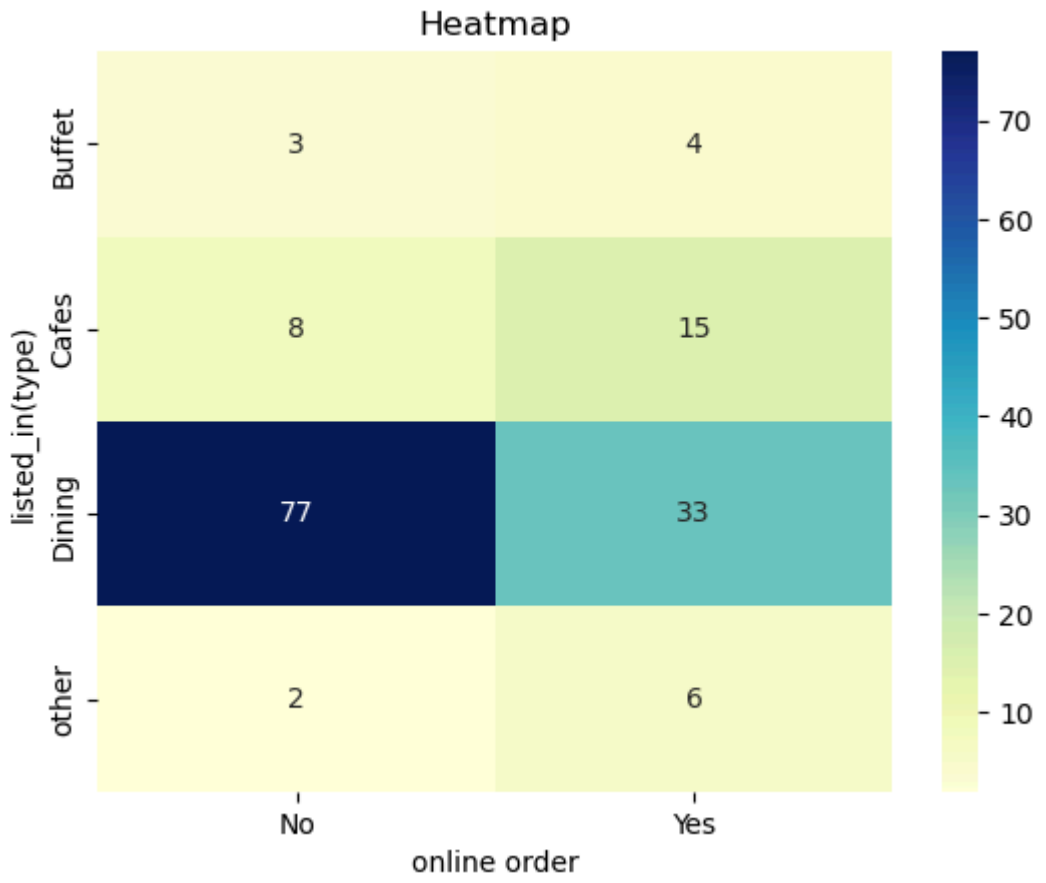
Out[8]:
```
<Axes: xlabel='online_order', ylabel='rate'>
```



conclusion-offline order receives lower rating compares to online order

Which type of restaurant received more offline orders, so that Zomato can provide those customers with some good offers..?

In [31]:
```
pivot_table=df.pivot_table(index='listed_in(type)',columns='online_order',aggfun
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 receives online order.This suggests that clients prefferd orders inperson at restaurants,but prefer online ordering at cafes.

In [ ]: