

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 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 x 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())
```

```
0      name online_order book_table rate votes \
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 [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   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
```

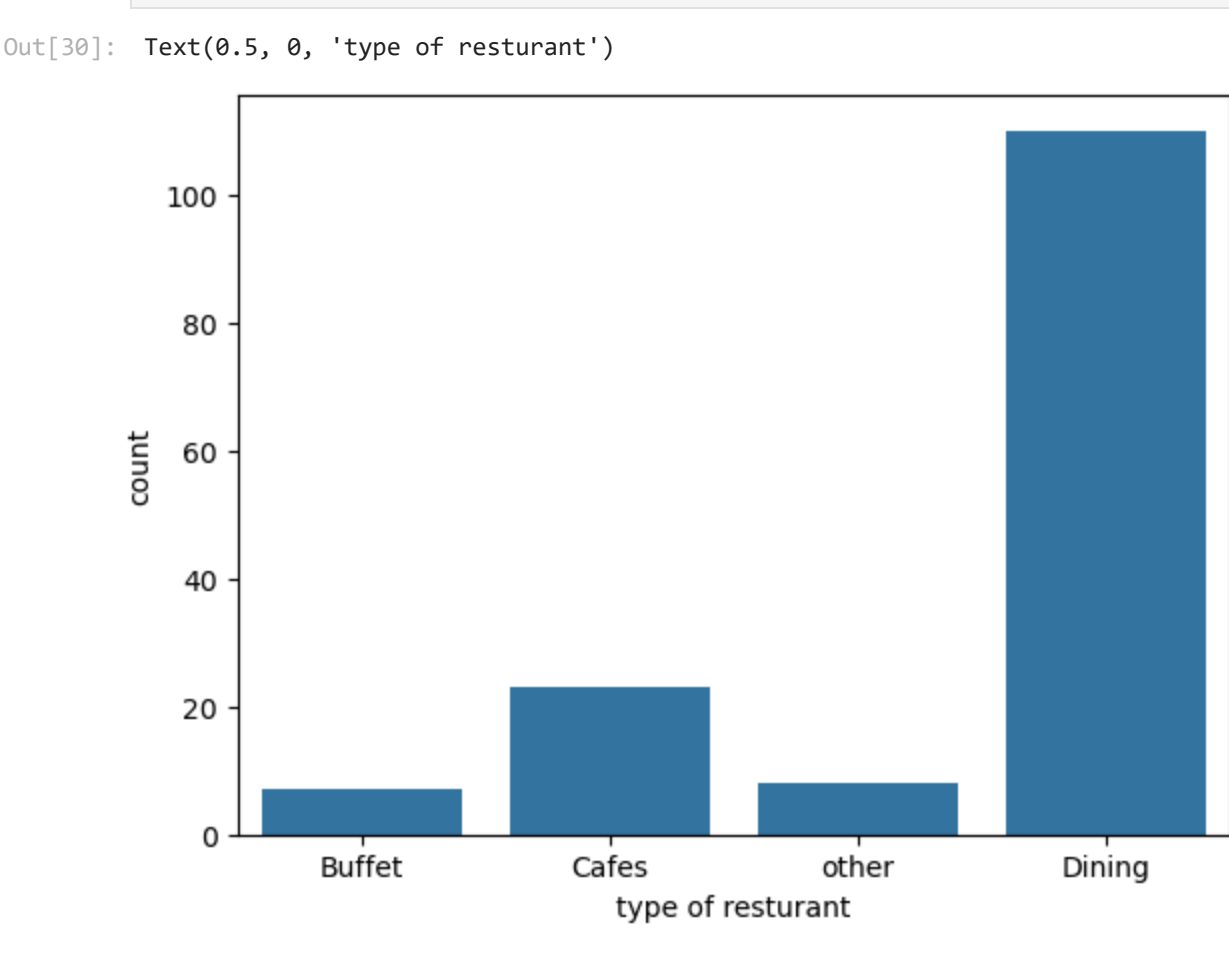
Type of restrurant

```
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 restrurant')
```



Conclusion- majority of the restrurant falls in dinning 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

```
In [32]: 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 restraurant' , c='red',size=20)
plt.ylabel('Votes', c='red', size=20)
```



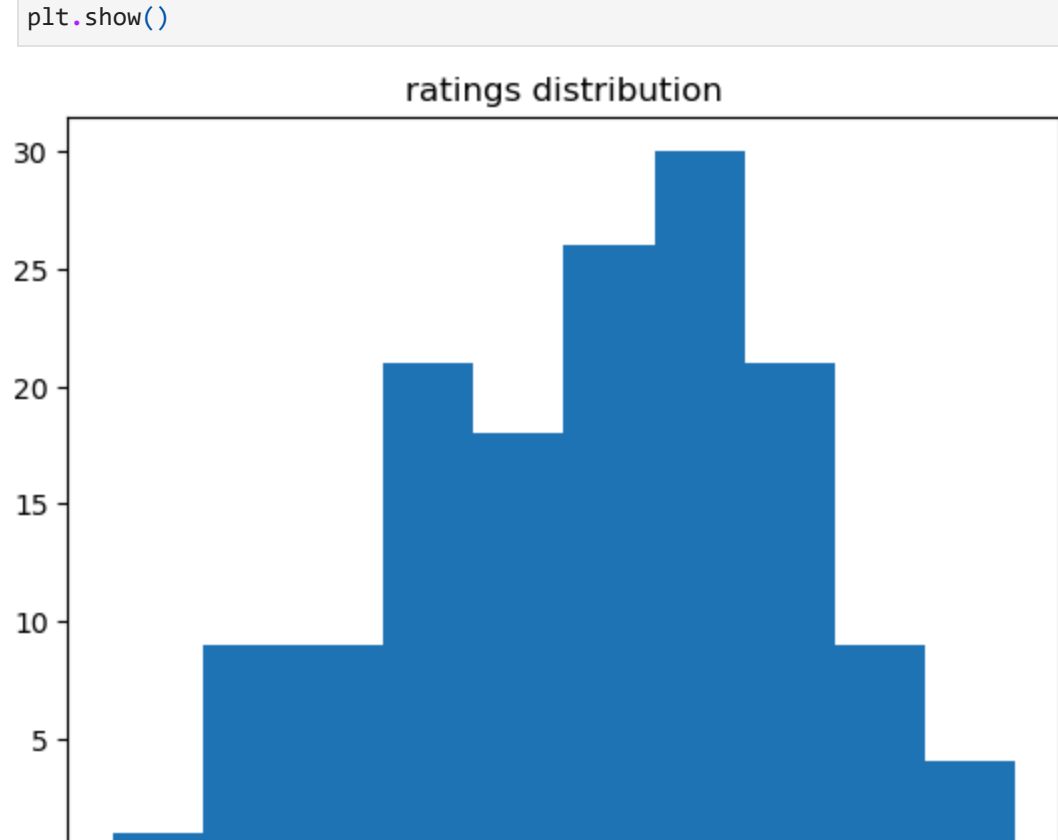
conclusion-dinning restrutants has recieved maximum votes

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



conclusion - the majority restrutants received rating from 3.5 to 4

Average order spending by couples

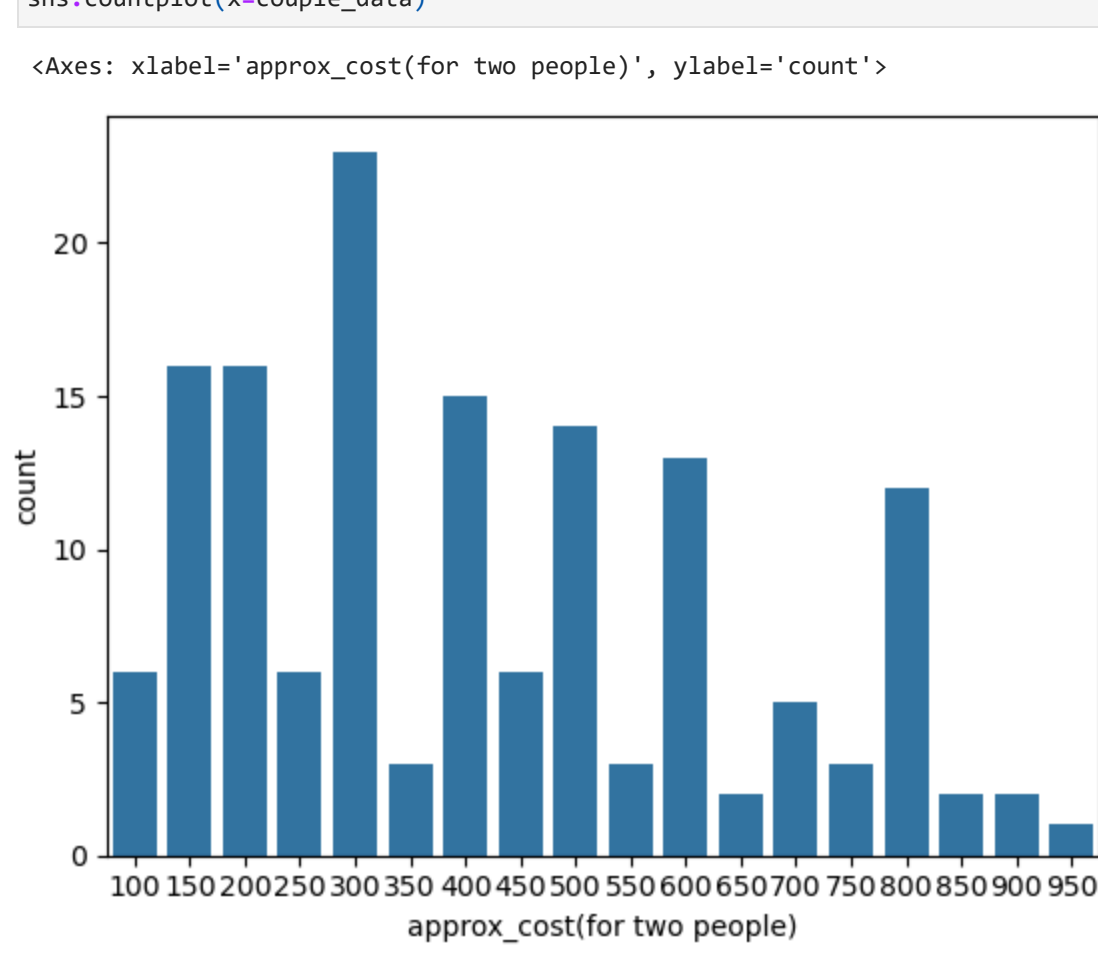
```
In [35]: df.head()
```

```
Out[35]:
```

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

Out[36]: <Axes: xlabel='approx_cost(for two people)', ylabel='count'>



conclusion- the majority of couples preferr restrutants with an approximate cost of 300 rupees

which mode receives maximum rating

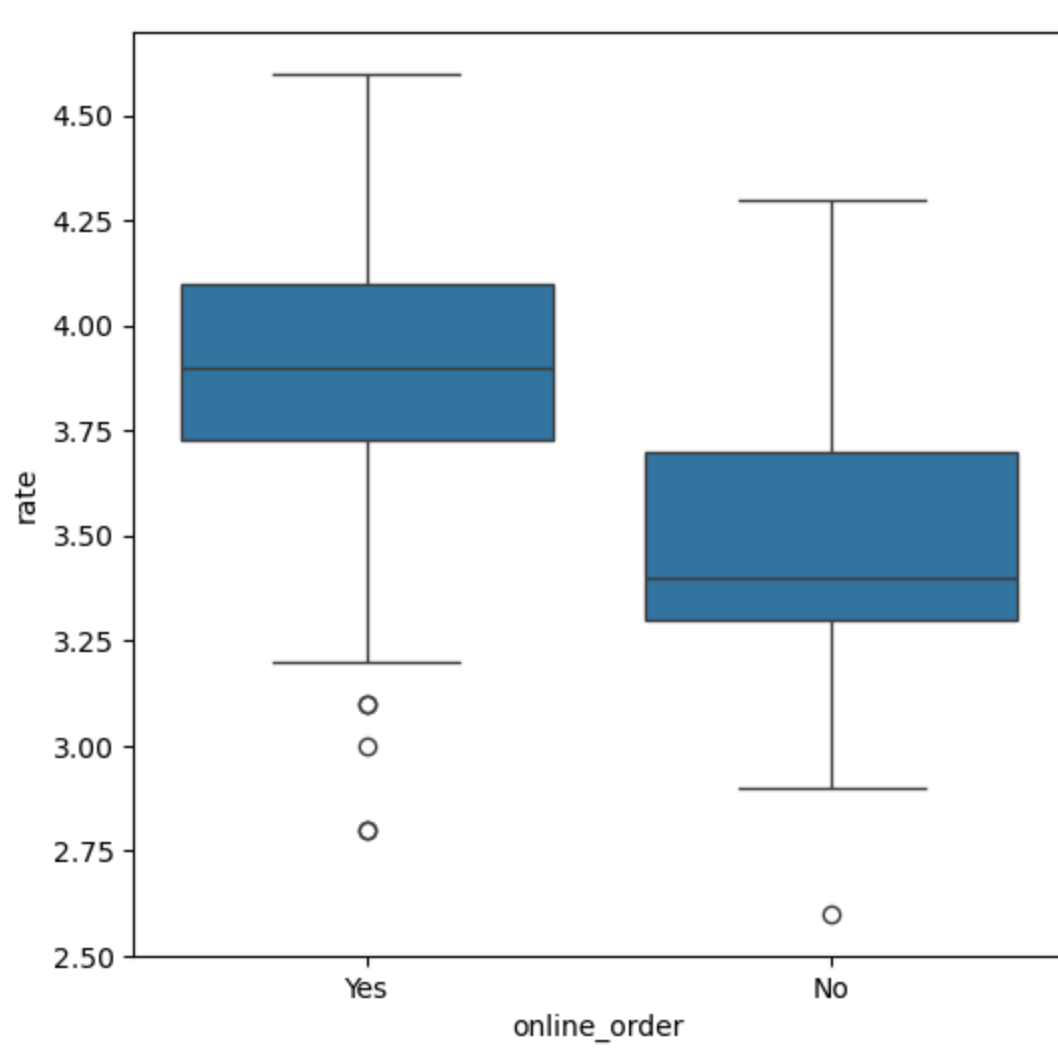
```
In [37]: df.head()
```

```
Out[37]:
```

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

Out[38]: <Axes: xlabel='online_order', ylabel='rate'>



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

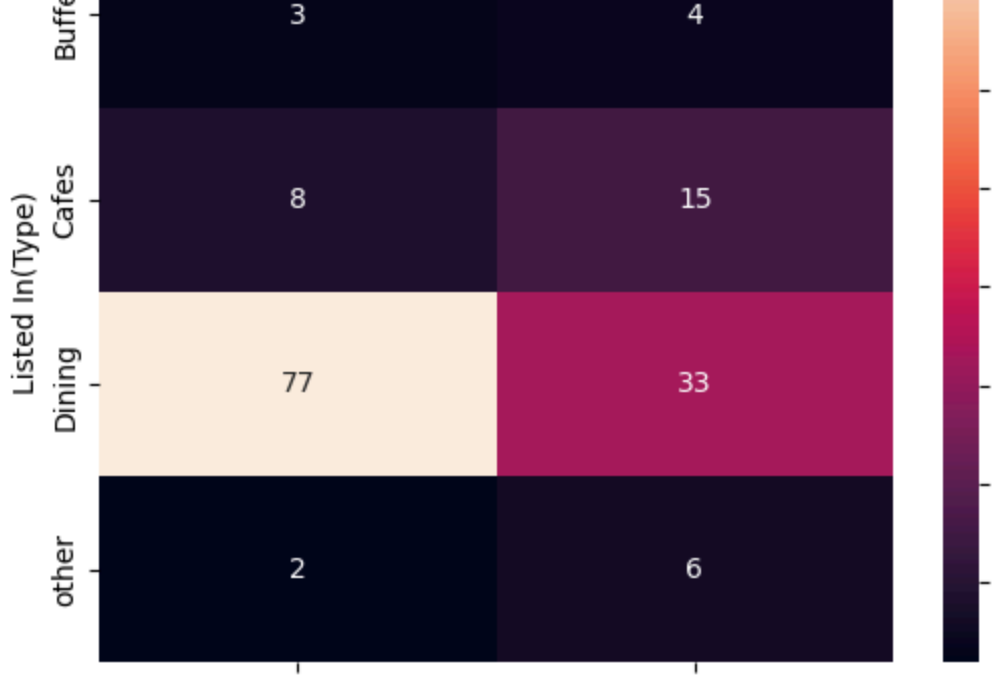
```
In [39]: pivot_table=df.pivot_table(index='listed_in(type)',columns='online_order',aggfunc='size',fill_value=0)
sns.heatmap(pivot_table, annot=True ,fat='d')
```

plt.title("Heatmap")

plt.xlabel("Online order")

plt.ylabel("listed In(type)")

plt.show()



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

```
In [ ]:
```