zomato

April 24, 2025

1 ZOMATO DATA ANALYSIS PROJECT

STEP-1:- Importing Libraries

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

Step-2:- Create the data frame

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

[4]:
```
df
```

[4]:

	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]
```

Step-3:-Convert the data type of column-rate

[5]:
```
def correctRate(value):
    value=str(value).split('/')
    value=value[0]
    return float(value)

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

[6]:
```
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
```

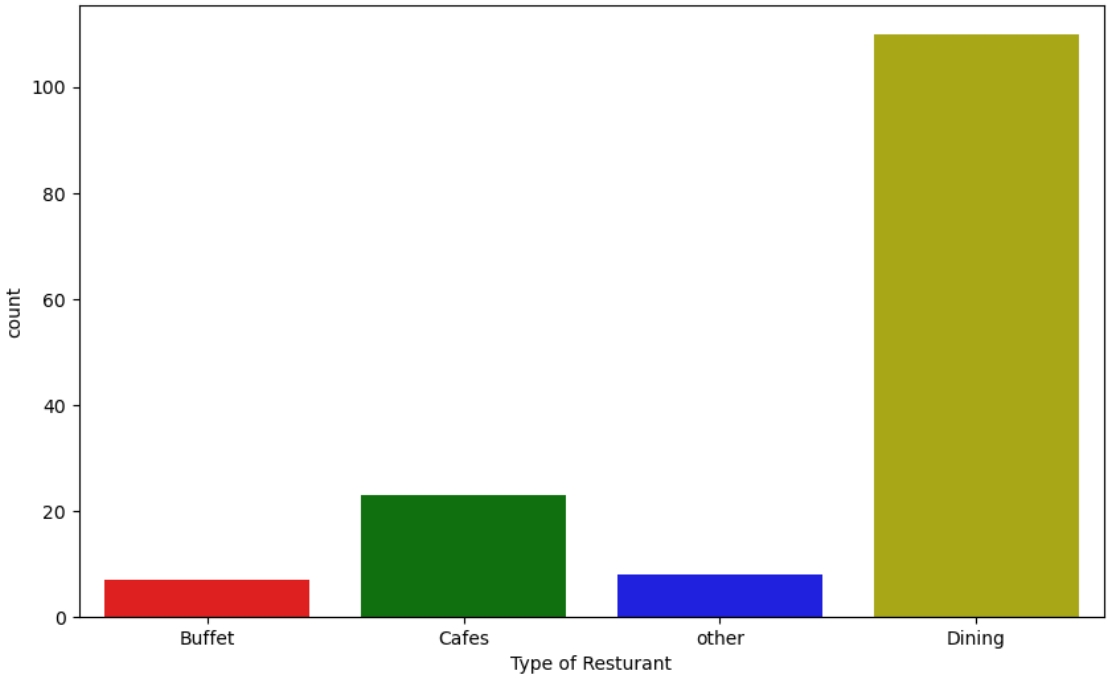
TYPE OF RESTURANT

[7]:
```
df.head()
```

[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

[40]:
```
c=['r','g','b','y']
plt.figure(figsize=(10, 6))
sns.countplot(x=df['listed_in(type)'],hue=df['listed_in(type)'],palette=c)
plt.xlabel("Type of Resturant")
plt.show()
```



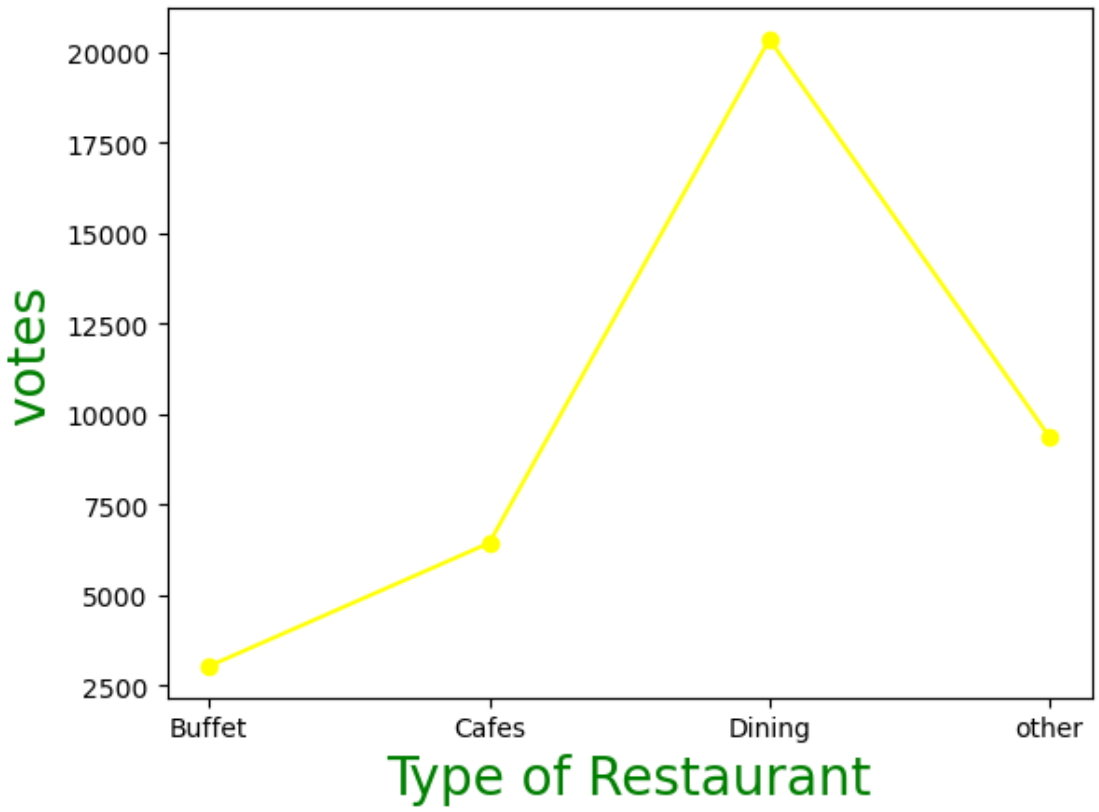
Conclusion:-majority of the resturant falls in dinning category

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

[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

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



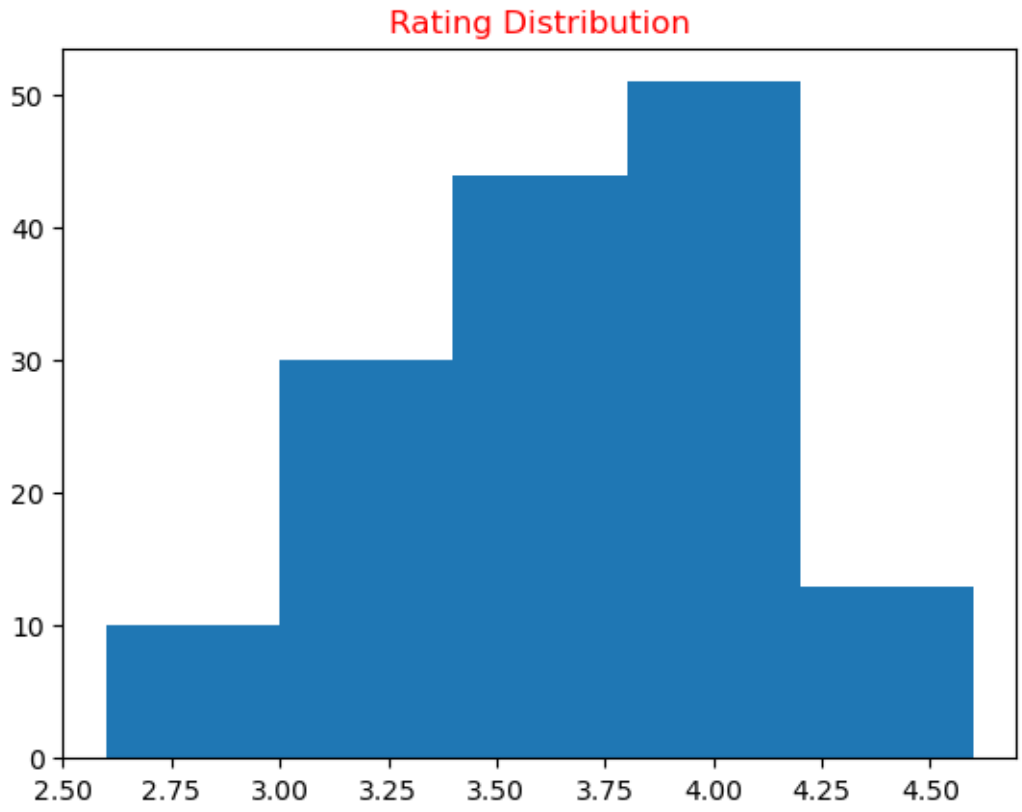
Conclusion:-Dinning resturants has recieved maximum votes

[11]:
```
df.head()
```

[11]:

	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

[17]:
```
plt.hist(df['rate'],bins=5)
plt.title("Rating Distribution",color='red')
plt.show()
```



Conclusion:- The majority resturants received ratings from 3.5 to 4

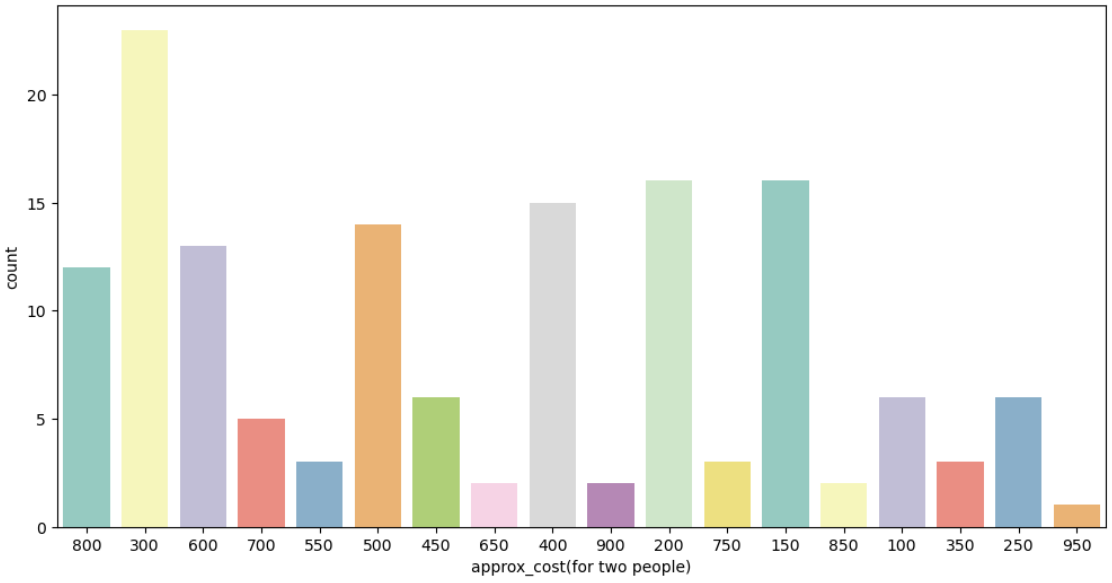
Average order spending by couples:-

[14]:
```
df.head()
```

[14]:
```
                   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
```

[44]:
```
couple_data=df['approx_cost(for two people)'].astype(str)
unique_vals = couple_data.nunique()
palette = sns.color_palette("Set3", unique_vals)
plt.figure(figsize=(12, 6))
sns.countplot(x=couple_data, hue=couple_data, palette=palette, legend=False)
plt.show()
```



Conclusion:-The majority of couples prefer resturants with an approximate cost of 300 rupees

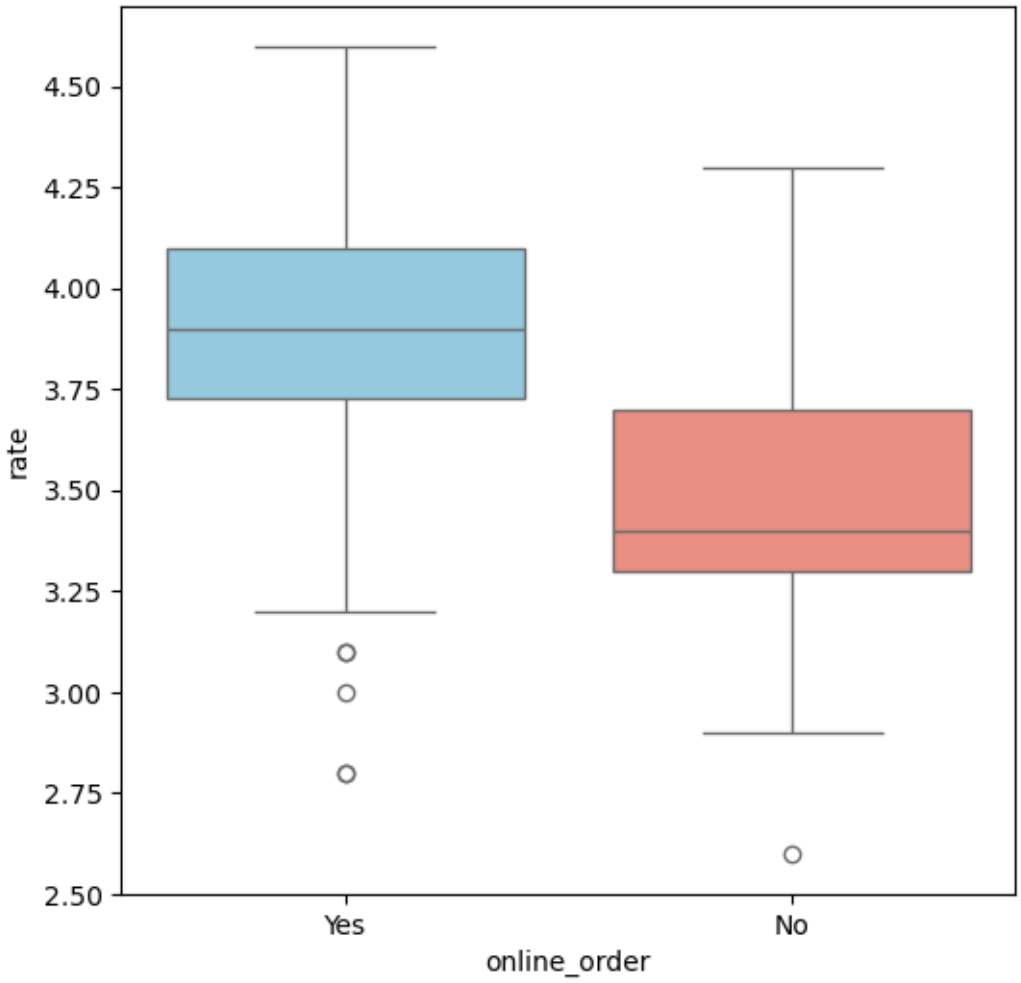
Which mode receives maximum rating

[19]:
```
df.head()
```

[19]:
```
                   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
```

[46]:
```
plt.figure(figsize=(6,6))
e = {'Yes': 'skyblue', 'No': 'salmon'}
sns.boxplot(x='online_order',y='rate',hue='online_order',data=df,palette=e,␣
 ↪legend=False)
plt.show()
```



Conclusion:-Offline order received lower rating in comparison to online order.

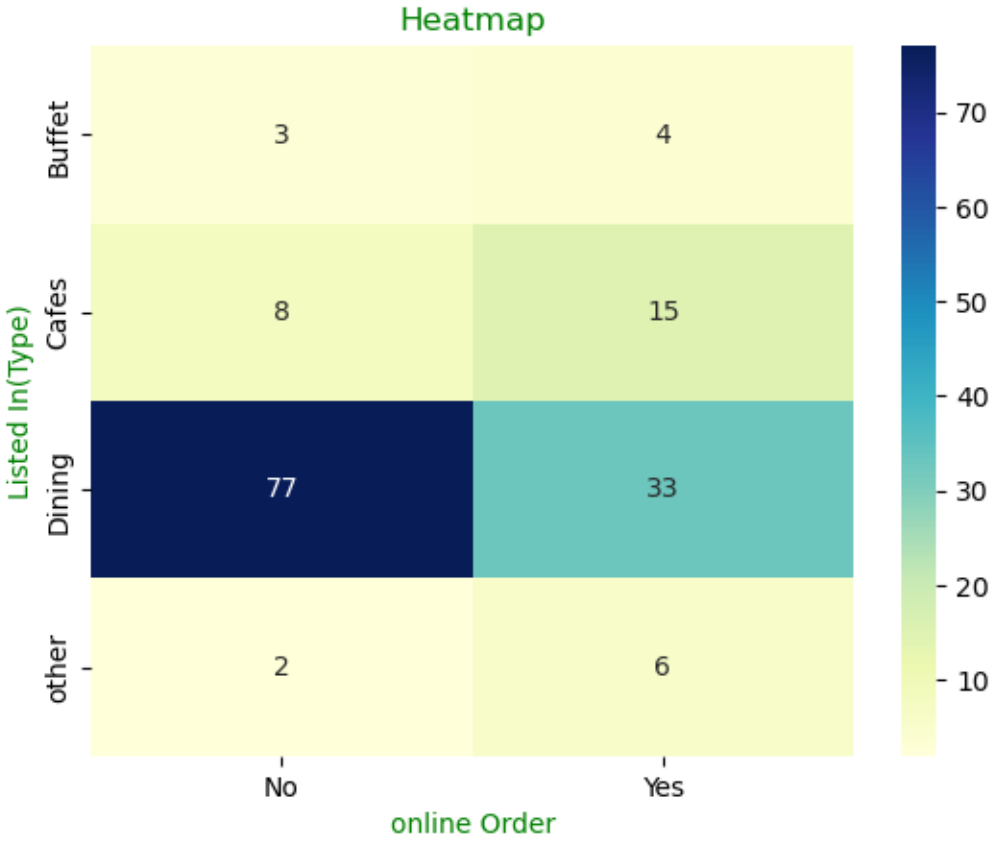
[22]:
```
df.head()
```

[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		

```
4                              600           Buffet
```

[27]:
```
pivot_table=df.
  ↪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",color="green")
plt.xlabel("online Order",color="green")
plt.ylabel("Listed In(Type)",color="green")
plt.show()
```



Conclusion:-Restaurants mostly get offline orders, while cafes mostly get online ones. This means people like to order in person at restaurants but prefer online orders at cafes.

[ ]: