ZOMATO DATA ANALYSIS

Import Python Libraries

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

In [ ]:
```
#pandas-for data manipulation and analysis
#numpy- for numerical operations
#matplotlib & seaborn- for data visualisation
```

Create dataframe

In [ ]:
```
dataframe=pd.read_csv(r"C:\Users\Shashank Shahi\Desktop\Coding\Python\Zomato Analysis Project\Zor
print(dataframe.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 [ ]:
```
dataframe
```

Out[ ]:

	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 Data type of column- rate to pure number

In [ ]:
```
def rating(val):
    val=str(val).split('/')
    val=val[0];
    return float(val)

dataframe['rate']=dataframe['rate'].apply(rating)
print(dataframe.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
```

Check if all entries in the dataset are NOT NULL

In [ ]:
```
dataframe.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
```

Orders from each Type of Restaurant

In [ ]:

```
dataframe.head()
```

Out[ ]:

	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

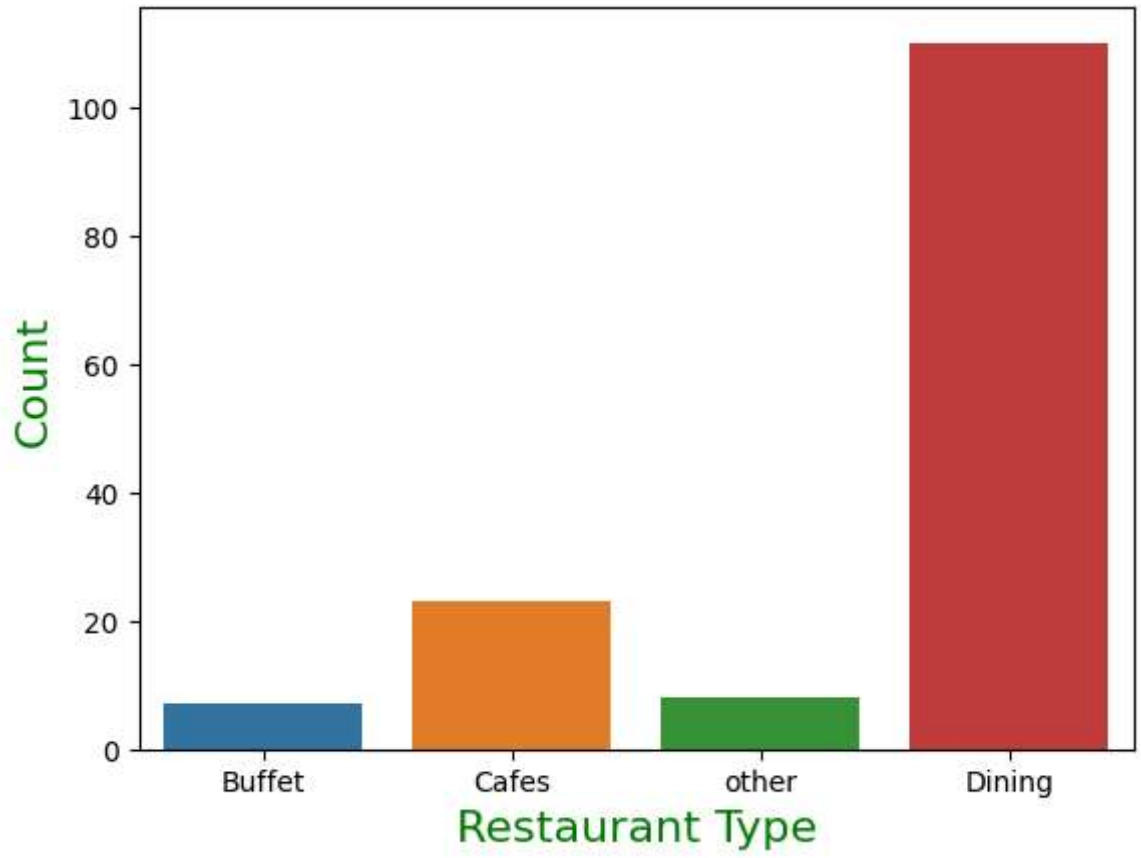
In [ ]:

```
sns.countplot(x=dataframe['listed_in(type)'])
plt.xlabel("Restaurant Type",size="16",c="Green")
plt.ylabel("Count",size="16",c="Green")
```

```
C:\Users\Shashank Shahi\AppData\Roaming\Python\Python311\site-packages\seaborn\_oldcore.py:1498:
FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use is
instance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
C:\Users\Shashank Shahi\AppData\Roaming\Python\Python311\site-packages\seaborn\_oldcore.py:1498:
FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use is
instance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
C:\Users\Shashank Shahi\AppData\Roaming\Python\Python311\site-packages\seaborn\_oldcore.py:1498:
FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use is
instance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
```

Out[ ]:

```
Text(0, 0.5, 'Count')
```



Result- Majority Restaurants fall under Dining Category

Votes for each type of Restaurant

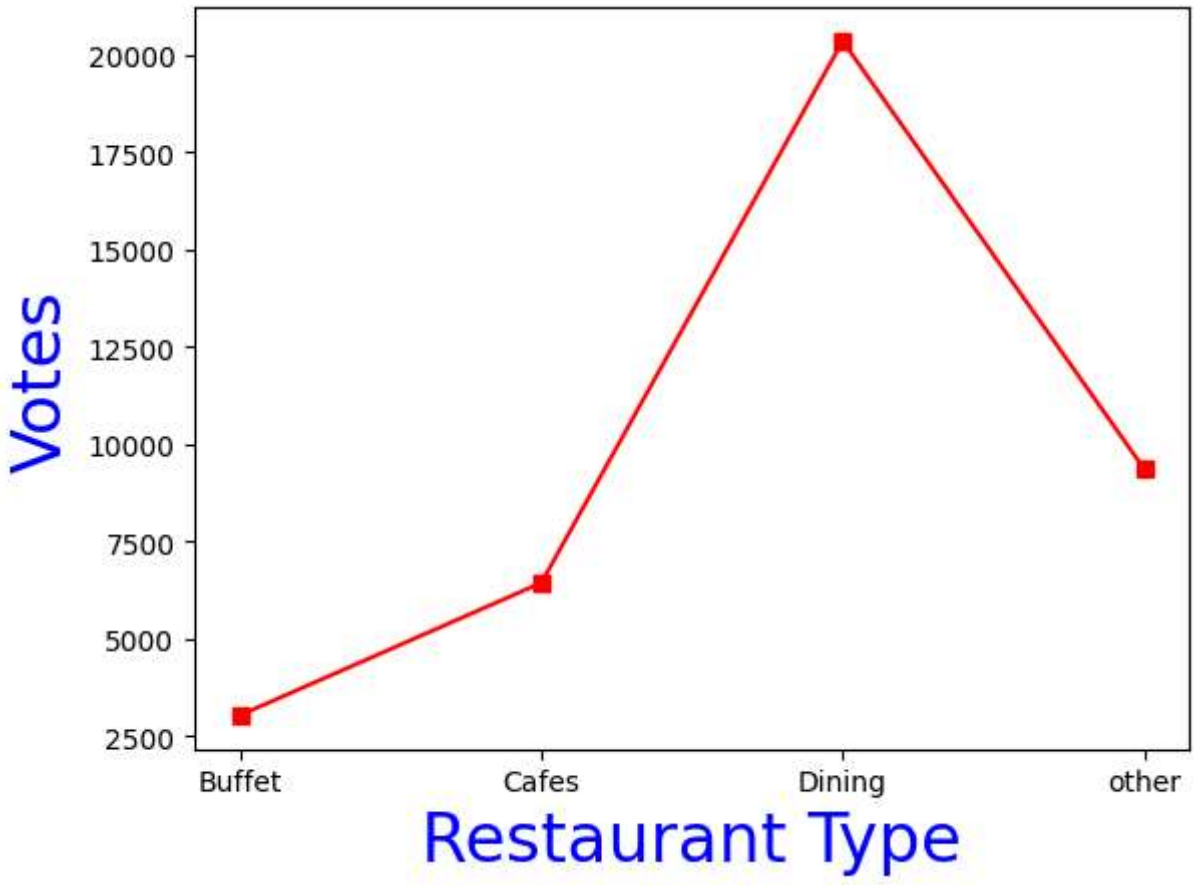
In [ ]:
```
dataframe.head()
```

Out[ ]:

	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 [ ]:
```
GroupedData=dataframe.groupby('listed_in(type)')['votes'].sum()
result= pd.DataFrame({'votes': GroupedData})
plt.plot(result,c="red",marker="s")
plt.xlabel("Restaurant Type",c="blue",size=24)
plt.ylabel("Votes",c="blue",size=24)
```

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



Result- Dining Restaurant have received Maximum votes

Rating Histogram

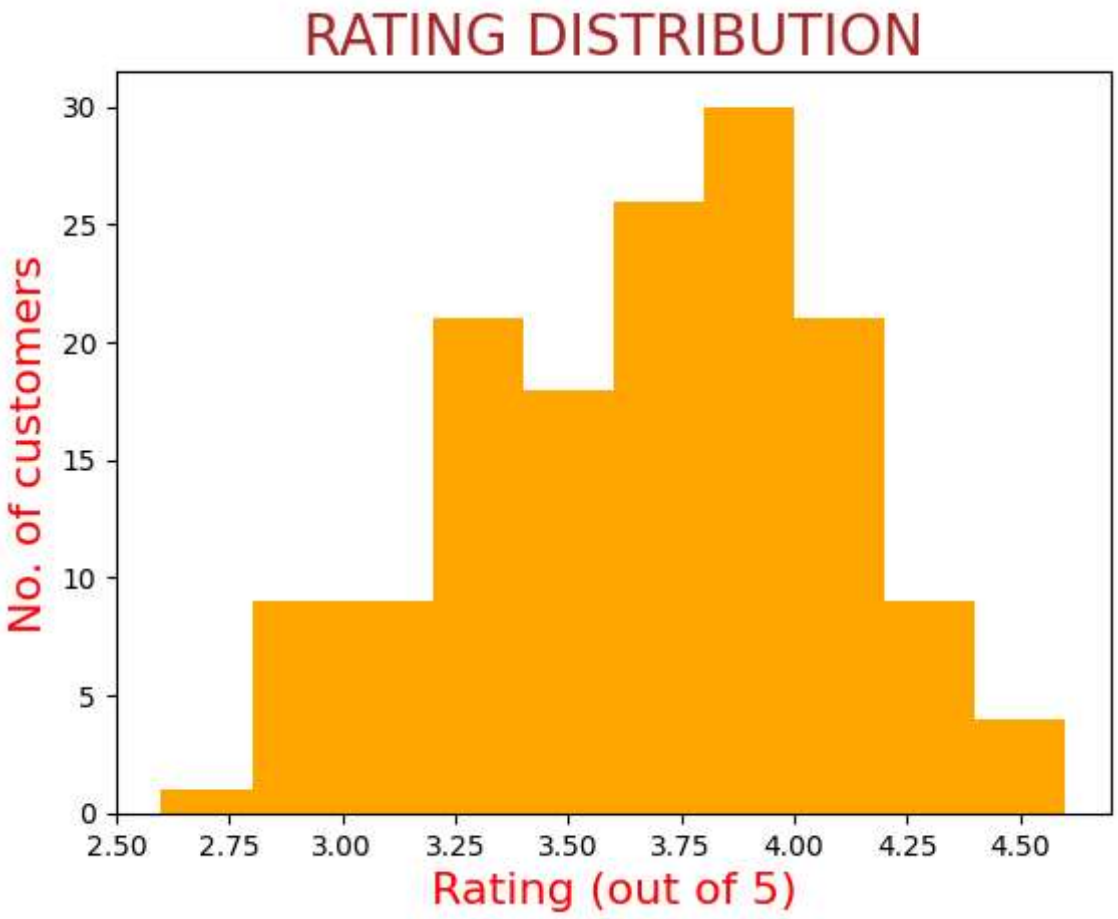
In [ ]:
```
dataframe.head()
```

Out[ ]:

	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 [ ]:
```
plt.hist(dataframe['rate'],bins=10,color="orange")
plt.title("RATING DISTRIBUTION",c="brown",size=20)
plt.ylabel("No. of customers",c="red",size=16)
plt.xlabel("Rating (out of 5)",c="red",size=16)
```

Out[ ]:
```
Text(0.5, 0, 'Rating (out of 5)')
```



Result- Majority Restaurants received rating from 3.5 to 4

Avg Spending per Order by Couples

In [ ]:
```
dataframe.head()
```

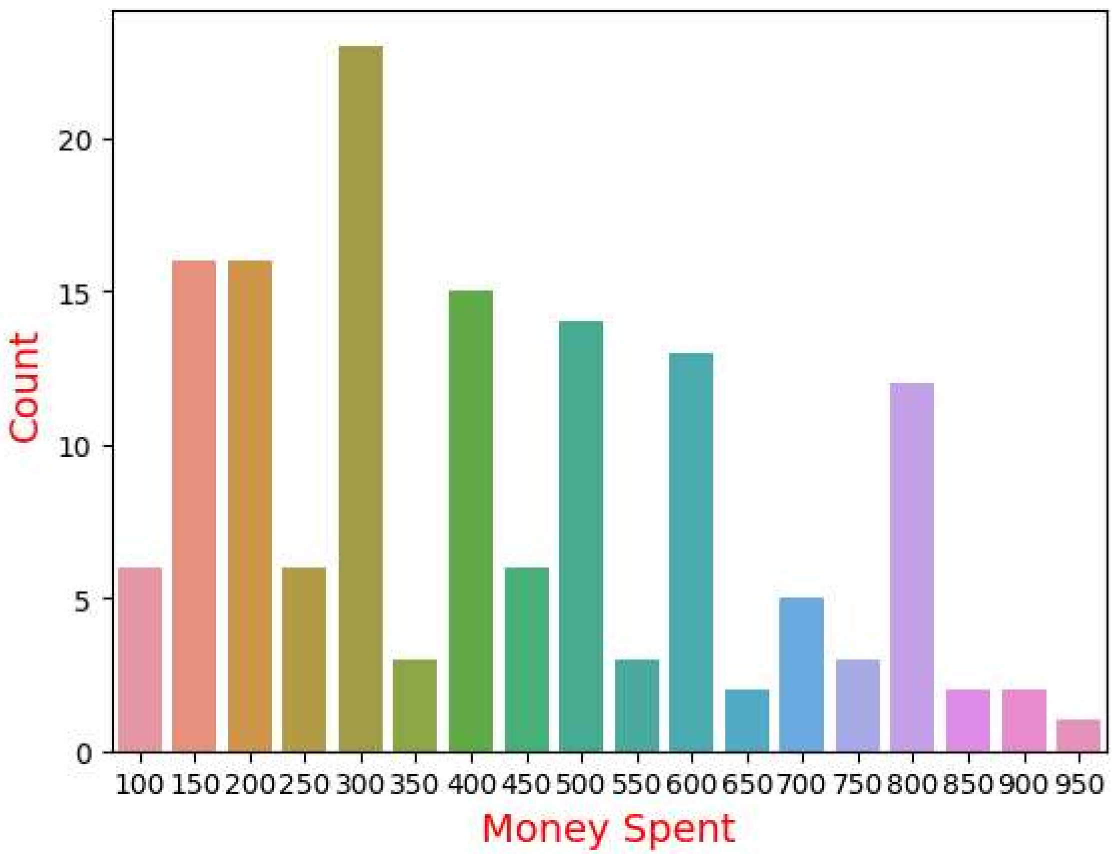
Out[ ]:

	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 [ ]:
```
CoupleData=dataframe['approx_cost(for two people)']
sns.countplot(x=CoupleData)
plt.xlabel("Money Spent",c="red",size=14)
plt.ylabel("Count",c="red",size=14)
```

```
C:\Users\Shashank Shahi\AppData\Roaming\Python\Python311\site-packages\seaborn\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
C:\Users\Shashank Shahi\AppData\Roaming\Python\Python311\site-packages\seaborn\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
C:\Users\Shashank Shahi\AppData\Roaming\Python\Python311\site-packages\seaborn\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
```

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



Result- Avg order Value for 2 people id 300Rs.

Mode (Online/Offline) having Maximum Rating

In [ ]:
```
dataframe.head()
```

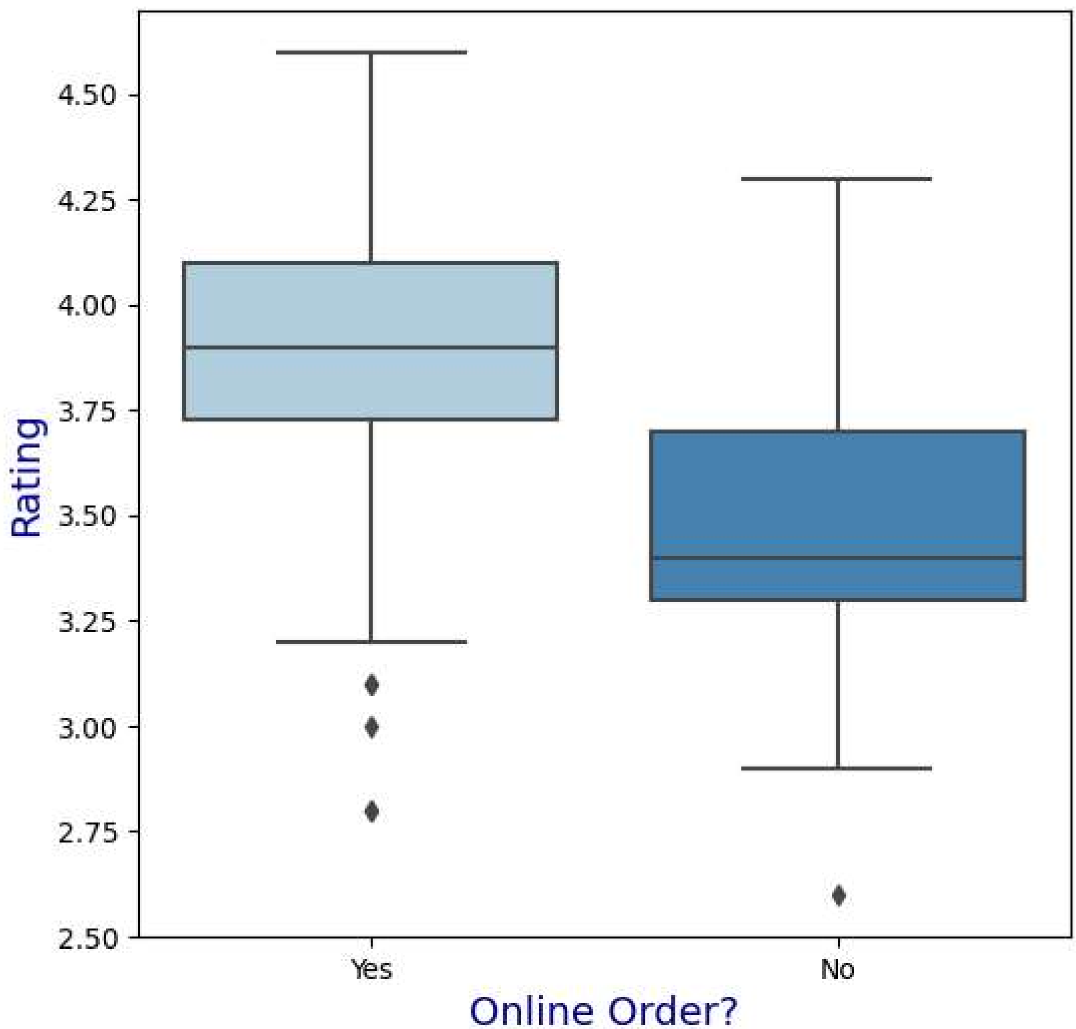
Out[ ]:

	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 [ ]:
```
plt.figure(figsize=(6,6))
sns.boxplot(x='online_order', y='rate', data=dataframe, palette = 'Blues')
plt.xlabel("Online Order?",c="darkblue",size=14)
plt.ylabel("Rating",c="darkblue",size=14)
```

```
C:\Users\Shashank Shahi\AppData\Roaming\Python\Python311\site-packages\seaborn\_oldcore.py:1498:
FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use is
instance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
C:\Users\Shashank Shahi\AppData\Roaming\Python\Python311\site-packages\seaborn\_oldcore.py:1498:
FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use is
instance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
C:\Users\Shashank Shahi\AppData\Roaming\Python\Python311\site-packages\seaborn\_oldcore.py:1498:
FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use is
instance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
```

Out[ ]: Text(0, 0.5, 'Rating')



Result- Online Orders receive higher ratings

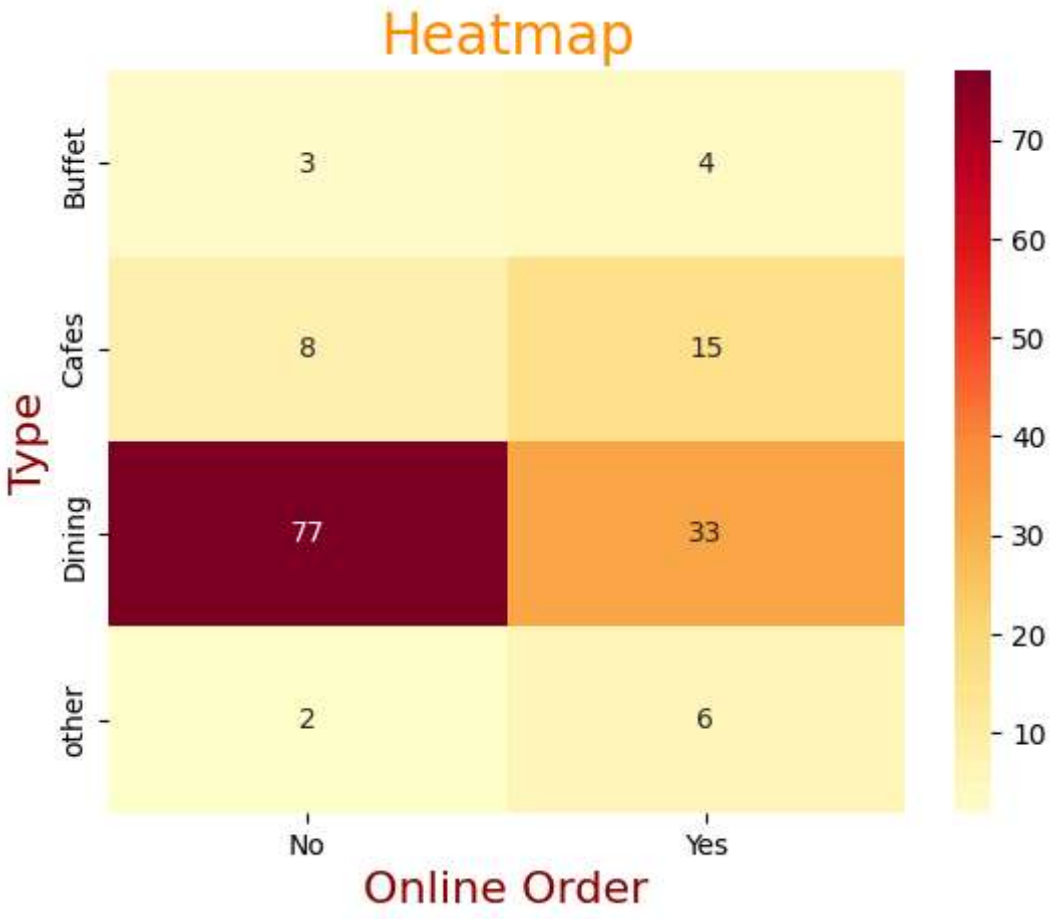
Type of Restraunt Vs No. of Orders

In [ ]:
```
df=dataframe
df.head()
```

Out[ ]:

	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 [ ]:
```
pivot_table= df.pivot_table(index='listed_in(type)',columns='online_order',aggfunc='size',fill_v
sns.heatmap(pivot_table,annot=True,cmap="YlOrRd",fmt='d')
plt.title("Heatmap", color="darkorange",size="20")
plt.xlabel("Online Order",c="darkred",size=16)
plt.ylabel("Type",c="darkred",size=16)
plt.show()
```



Result- This Heatmap shows that Dining restraunts mostly receive offline orders whereas Cafes, Buffets and other type of restaurants receive manly online orders