Zomato Data Analysis Project

Objectives of the Project-

1. what type of Resturant do the majority of customers order from?
2. How many votes has each type of resturant received from customer?
3. What are the ratings that the majority of resturants have received?
4. Zomato has observed that most couples order most of their food online. What's their average spending on each order?
5. Which mode(online or offline) has received the maximum rating?
6. which type of resturant received more offline orders, sothat Zomato can give customers with some good offers.

Importing Libraries-

In [1]:
```
import numpy as np    #use for numerical operation
import pandas as pd   #use for data manipulation and analysis
import matplotlib.pyplot as plt    #use for Data visualization
import seaborn as sns

import plotly.express as px
import plotly.graph_objects as go
```

Loading dataset-

In [2]:
```
data = pd.read_csv("Zomato data .csv")
data.head()
```

Out[2]:

	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

Manipluation Data in rate column-

In [3]:

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

data["rate"]=data["rate"].apply(HandleRate)
data.head()
```

Out[3]:

	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 [4]:

```
data.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
```

Checking Null Values-

In [5]:
```
data.isnull().sum()
```

Out[5]:
```
name                           0
online_order                   0
book_table                     0
rate                           0
votes                          0
approx_cost(for two people)    0
listed_in(type)                0
dtype: int64
```

Descriptive Analysis-

In [6]:
```
data.describe()
```

Out[6]:

	rate	votes	approx_cost(for two people)
count	148.000000	148.000000	148.000000
mean	3.633108	264.810811	418.243243
std	0.402271	653.676951	223.085098
min	2.600000	0.000000	100.000000
25%	3.300000	6.750000	200.000000
50%	3.700000	43.500000	400.000000
75%	3.900000	221.750000	600.000000
max	4.600000	4884.000000	950.000000

Type of Resturant-

In [8]:

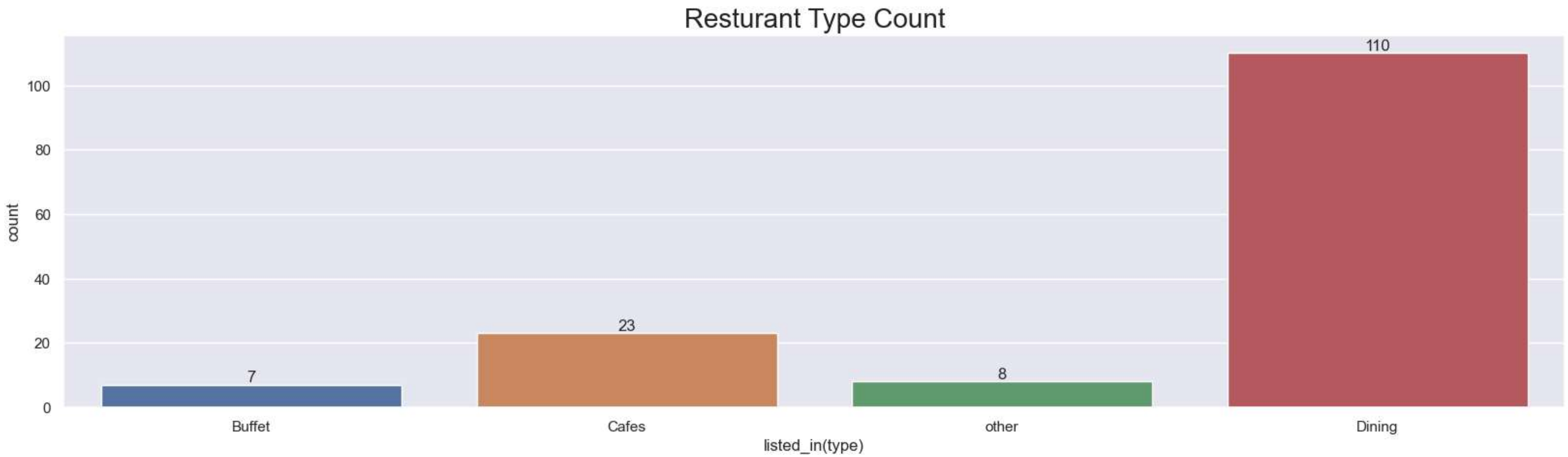
```
data["listed_in(type)"].value_counts()
```

Out[8]:

```
listed_in(type)
Dining    110
Cafes      23
other       8
Buffet      7
Name: count, dtype: int64
```

In [63]:

```
fig = sns.countplot(data, x="listed_in(type)")
for bars in fig.containers:
    fig.bar_label(bars)
plt.title("Resturant Type Count", fontsize=20)
plt.show()
```



Conclusion:Majority of the Resturant fall into Dinning Category.

Votes gained by resturants-

In [25]:
```
data.head()
```

Out[25]:

	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 [27]:
```
votes_gained = data.groupby("listed_in(type)")["votes"].sum().reset_index()
votes_gained
```

Out[27]:

	listed_in(type)	votes
0	Buffet	3028
1	Cafes	6434
2	Dining	20363
3	other	9367

In [34]:

```
fig = px.pie(votes_gained, names="listed_in(type)", values="votes", hole=0.5)
fig.update_traces(textinfo="percent+label")
fig.update_layout(title="Voted Gained by Resturant from Customers")
fig.show()
```

Conclusion: Dinning Resturant has received maximum Votes.

Ratings Distributions-

In [36]:
```
data.head()
```

Out[36]:

	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]:

```
fig = px.histogram(data, x="rate", nbins=10, title="Ratings Distribution")
fig.show()
```

Conclusion: Ratings received by majority of Resturant is from 3.5 to 4.0

Couples Average Cost Spending Analysis-

In [39]:
```
data.head()
```

Out[39]:

	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 [41]:

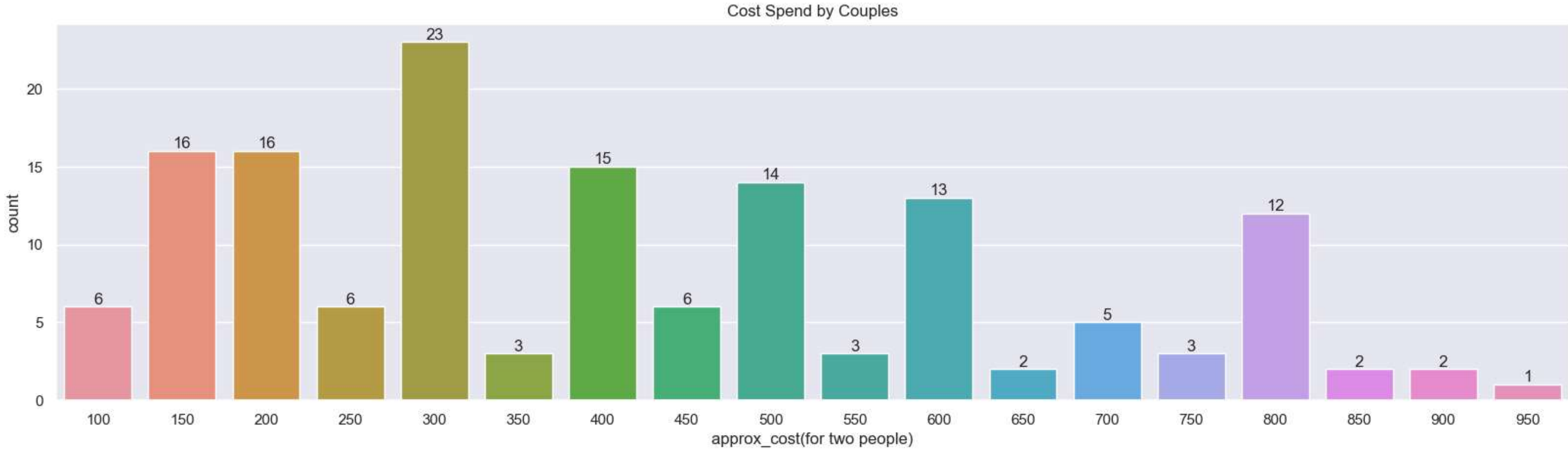
```
data["approx_cost(for two people)"].value_counts().head()
```

Out[41]:

```
approx_cost(for two people)
300    23
200    16
150    16
400    15
500    14
Name: count, dtype: int64
```

In [57]:

```
fig = sns.countplot(data, x="approx_cost(for two people)")
for bars in fig.containers:
    fig.bar_label(bars)
plt.title("Cost Spend by Couples")
plt.show()
```



In [47]:
```
ques4 = data.groupby("listed_in(type)")["approx_cost(for two people)"].mean().reset_index()
ques4
```

Out[47]:

	listed_in(type)	approx_cost(for two people)
0	Buffet	671.428571
1	Cafes	545.652174
2	Dining	357.272727
3	other	668.750000

In [64]:
```
fig = px.bar(ques4, x="listed_in(type)", y="approx_cost(for two people)", title="Resturant Type v/s Cost spend by Couples")
fig.show()
```

Conclusion: The majority of couples prefered Resturants with an approximated average cost of 300 rupees and They prefered Buffet Resturant type

Online or Offline Mode Analysis-

In [65]:
```
data.head()
```

Out[65]:

	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 [66]:

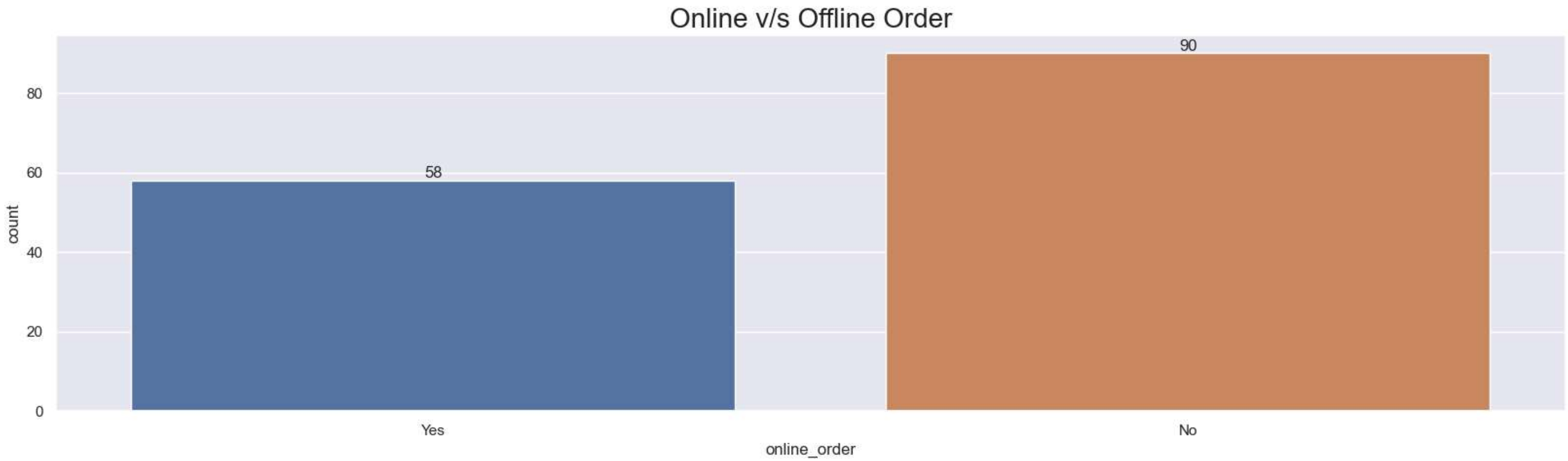
```
data["online_order"].value_counts()
```

Out[66]:

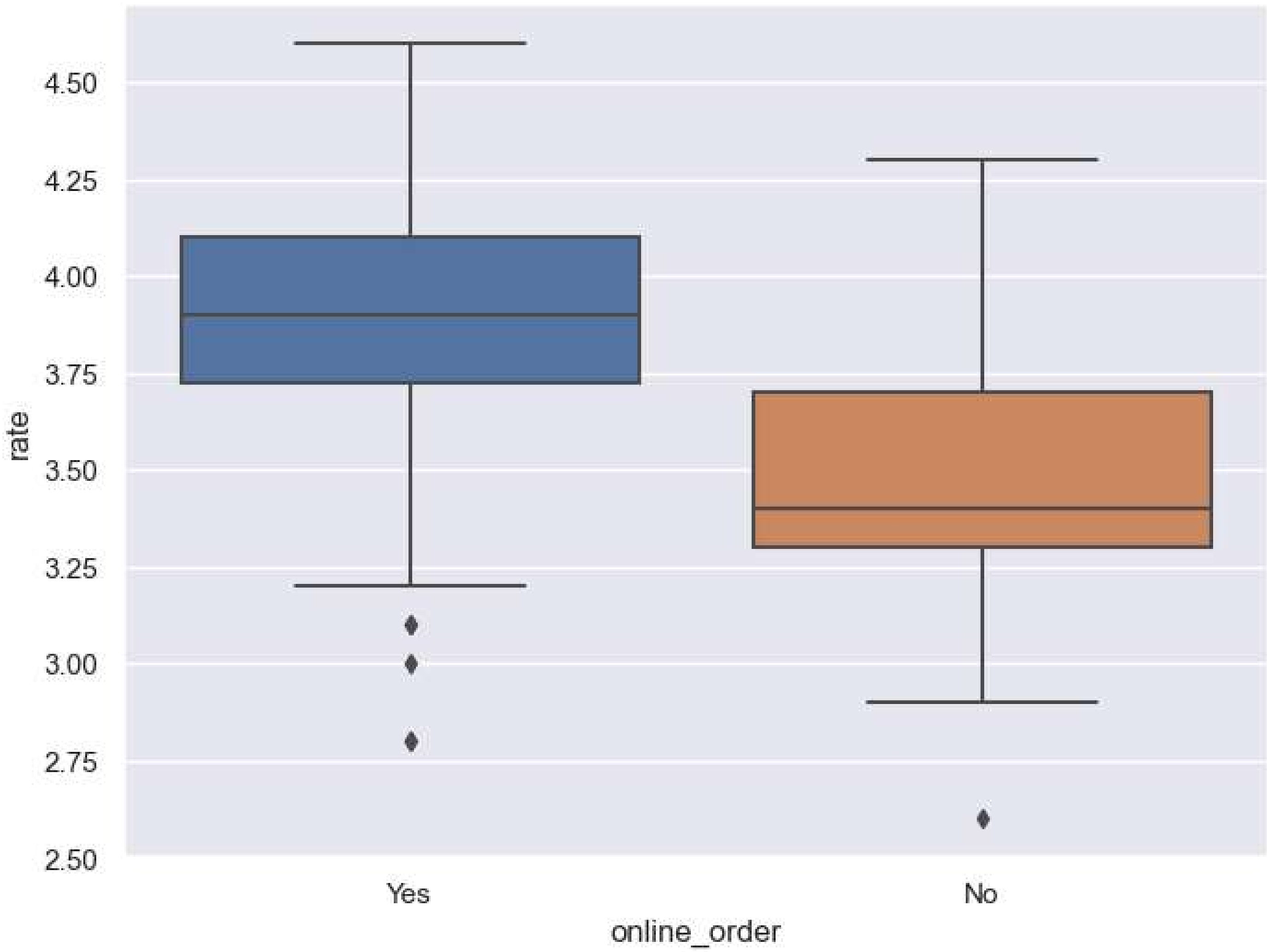
```
online_order
No     90
Yes    58
Name: count, dtype: int64
```

In [71]:

```
fig = sns.countplot(data, x="online_order")
for bars in fig.containers:
    fig.bar_label(bars)
plt.title("Online v/s Offline Order", fontsize=20)
plt.show()
```



In [74]:
```
plt.figure(figsize=[8,6])
fig = sns.boxplot(data, x="online_order", y="rate")
```



Conclusion: Offline Order received Lower Rating in comparison to online orders.

Heatmap Analysis-

In [75]:
```
data.head()
```

Out[75]:

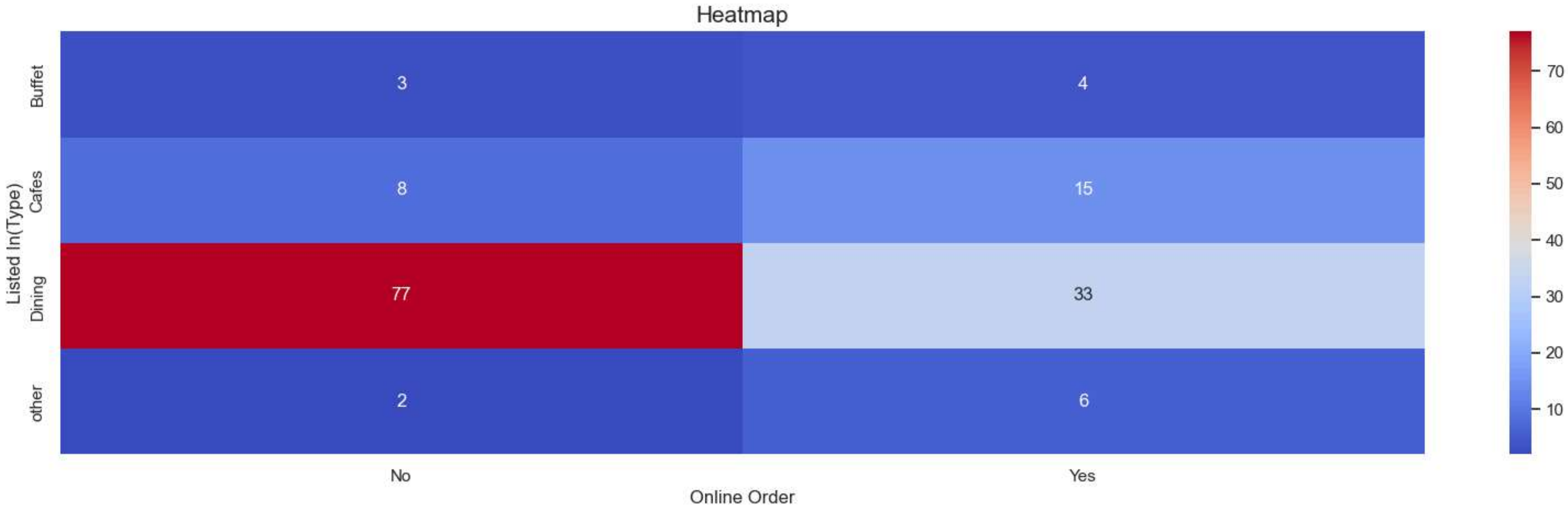
	name	online_order	book_table	rate	votes	approx_cost(for two people)	listed_in(type)
0	Jalsa	Yes	Yes	4.1	775	800	Buffet
1	Spice Elephant	Yes	No	4.1	787	800	Buffet
2	San Churro Cafe	Yes	No	3.8	918	800	Buffet
3	Addhuri Udupi Bhojana	No	No	3.7	88	300	Buffet
4	Grand Village	No	No	3.8	166	600	Buffet

In [81]:
```
# Assuming 'data' is your DataFrame
pivot_table = data.pivot_table(index="listed_in(type)", columns="online_order", aggfunc="size")

# Plotting the heatmap
sns.heatmap(pivot_table, annot=True, cmap="coolwarm", fmt="d")

# Adding title and axis labels
plt.title("Heatmap", fontsize=15)
plt.xlabel("Online Order")
plt.ylabel("Listed In(Type)")

# Display the heatmap
plt.show()
```



Conclusion: Dinning Resturants primarily accept offline orders, whereas cafes primarily receive online orders. This suggests that clients prefered orders in person at resturants, but prefer ordering at cafes.