Zomato data analysis project

Step 1 - Importing Libraries

pandas is used for data manipulation and analysis. numpy is used for numerical operations. matplotlib.pyplot and seaborn are used for data visualization.

In [5]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
```

Step 2 - Create the data frame

In [5]:
```
import pandas as pd

dataframe = pd.read_csv("Zomato data .csv")
print(dataframe)
```

```
                    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]
```

In [6]:
```
dataframe
```

Out[6]:

	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 the data type of column - rate

In [7]:
```
def handleRate(value):
    value = str(value).split('/')
    value = value[0];
    return float(value)

dataframe['rate']=dataframe['rate'].apply(handleRate)
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  
```

In [8]:
```
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
```

Type of resturant

In [9]:
```
dataframe.head()
```

Out[9]:

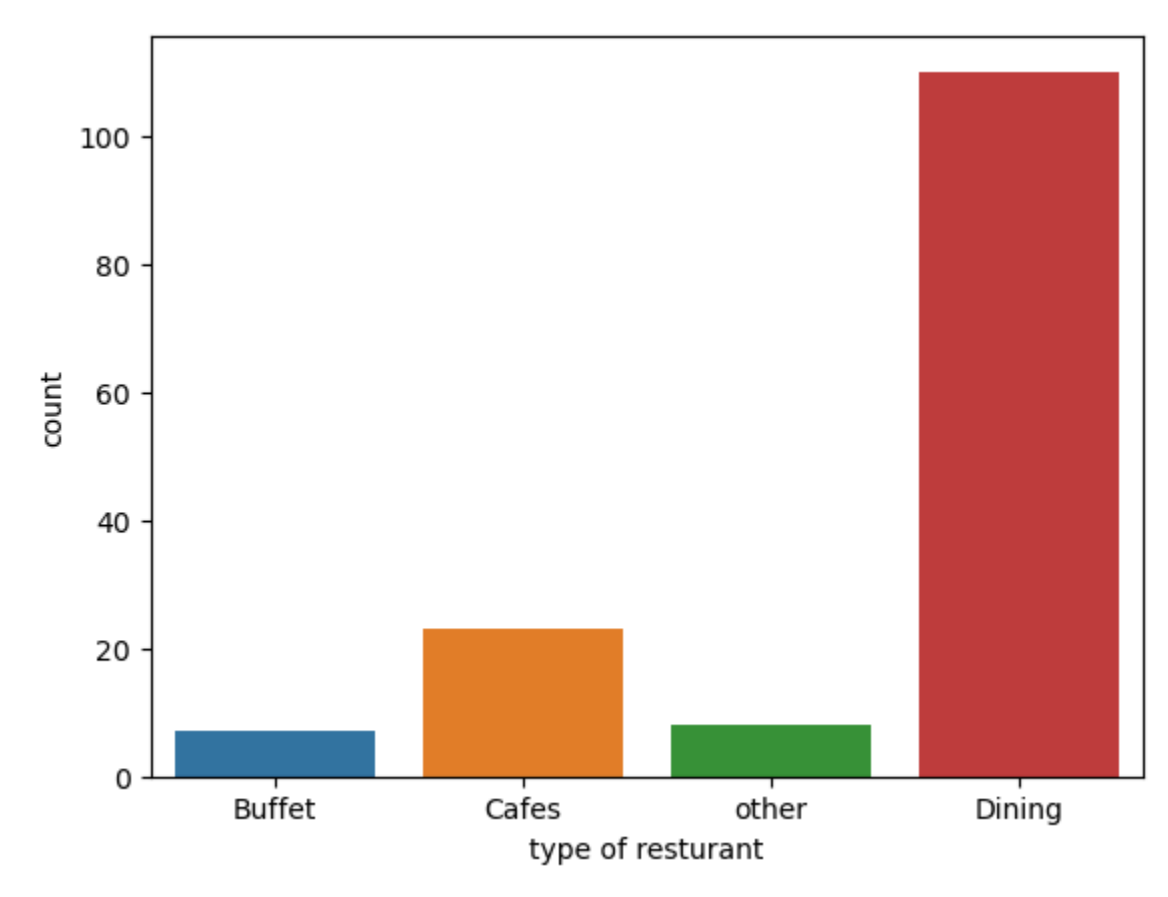
	name	online_order	book_table	rate	votes	approx_cost(for two people)	listed_in(type)
0	Jalsa	Yes	Yes	4.1	775	800	Buffet
1	Spice Elephant	Yes	No	4.1	787	800	Buffet
2	San Churro Cafe	Yes	No	3.8	918	800	Buffet
3	Addhuri Udupi Bhojana	No	No	3.7	88	300	Buffet
4	Grand Village	No	No	3.8	166	600	Buffet

Question - What type of restaurant do majority of customers order from?

In [10]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.countplot(x=dataframe['listed_in(type)'])
plt.xlabel("type of resturant")
```

Out[10]:
```
Text(0.5, 0, 'type of resturant')
```



conclusion - majority of the restaurant falls in dinning category

Question - How many votes has each type of restaurant received from customers?

In [11]:
```
dataframe.head()
```

Out[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

In [18]:
```
grouped_data = dataframe.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="red", size=20)
```

Out[18]:
```
Text(0, 0.5, 'Votes')
```

conclusion - dinning restaurant has received maximum votes

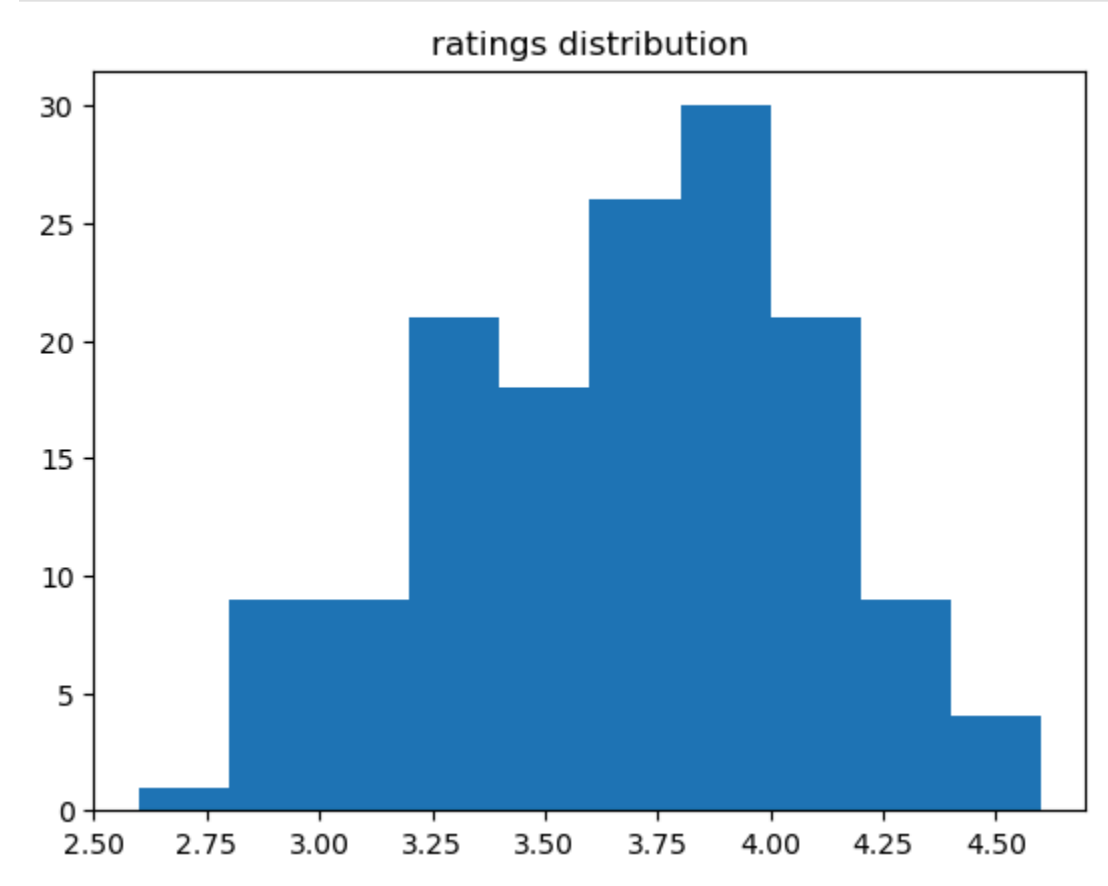
Question - What are the ratings that the majority restaurants have received?

In [19]:
```
dataframe.head()
```

Out[19]:

	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 [21]:
```
plt.hist(dataframe['rate'],bins = 10)
plt.title("ratings distribution")
plt.show()
```



conclusion - the majority restaurants received ratings from 3.5 to 4

Question - Zomato has observed that most couples order most of their food online.what is their average spending on each order?

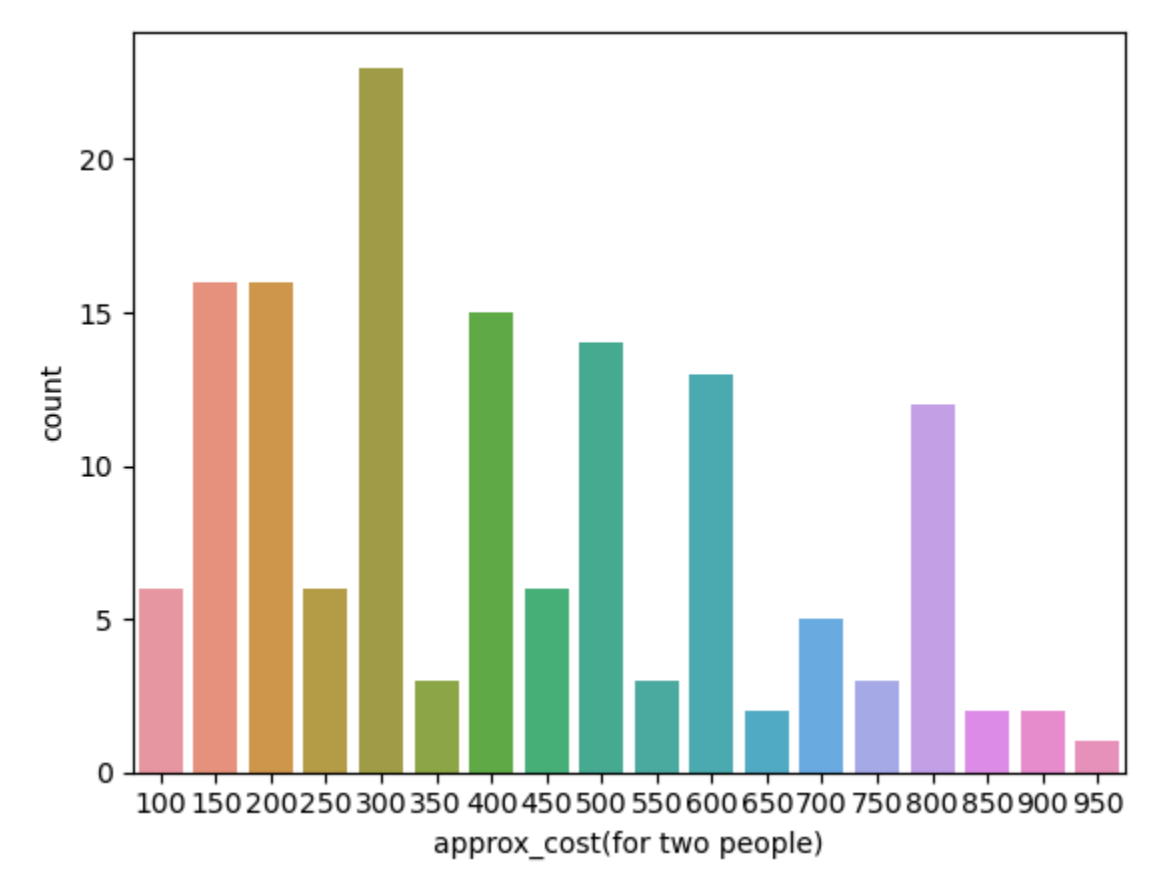
In [22]:
```
dataframe.head()
```

Out[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	600	Buffet

In [23]:
```
couple_data=dataframe['approx_cost(for two people)']
sns.countplot(x=couple_data)
```

Out[23]:
```
<Axes: xlabel='approx_cost(for two people)', ylabel='count'>
```



conclusion - the majority of couples prefer restaurants with an approximate cost of 300 rupees

Question - Which mode (online or offline) has received the maximum rating?

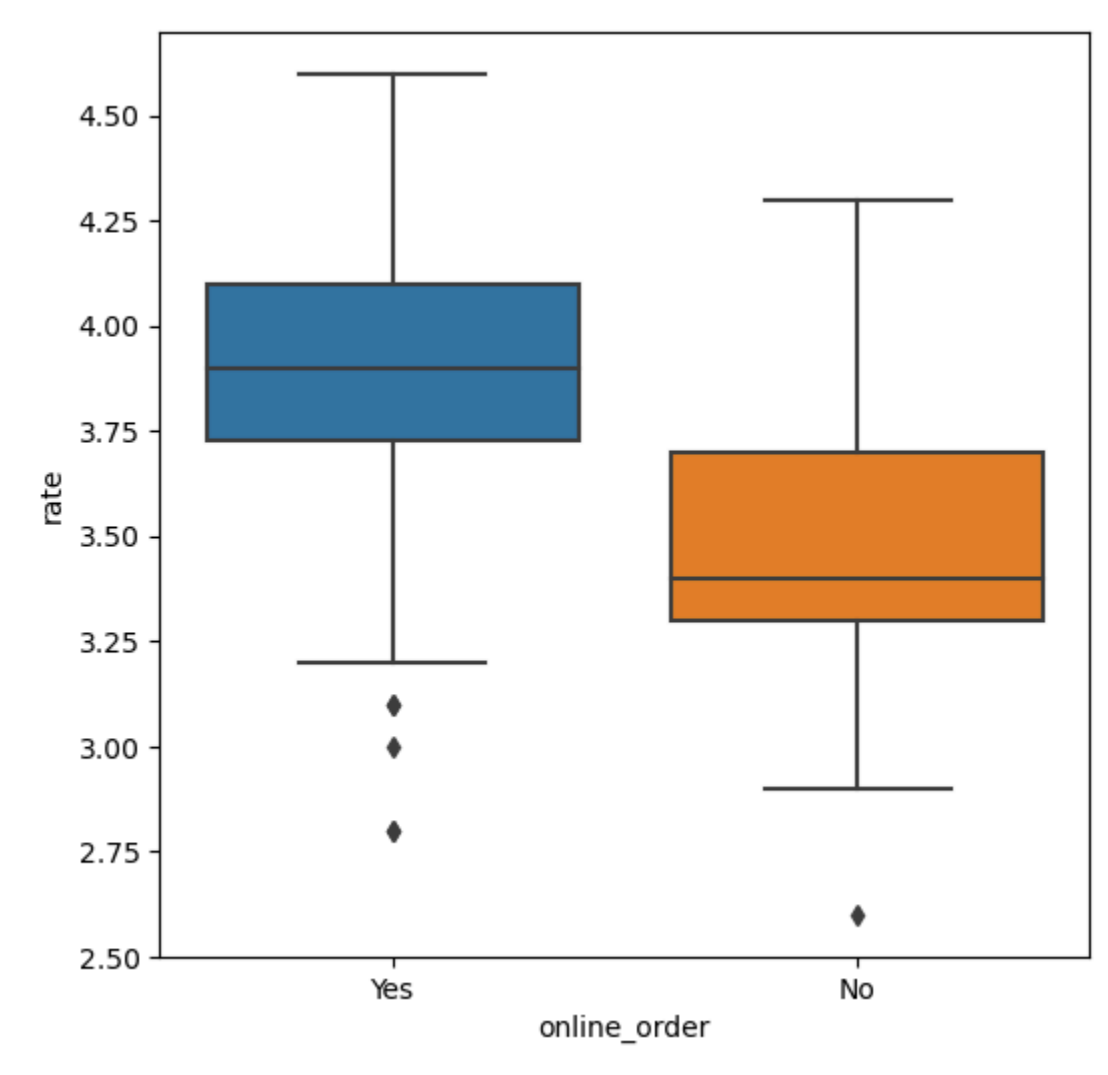
In [24]:
```
dataframe.head()
```

Out[24]:

	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 [25]:
```
plt.figure(figsize = (6,6))
sns.boxplot(x = 'online_order', y = 'rate', data = dataframe)
```

Out[25]:
```
<Axes: xlabel='online_order', ylabel='rate'>
```



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

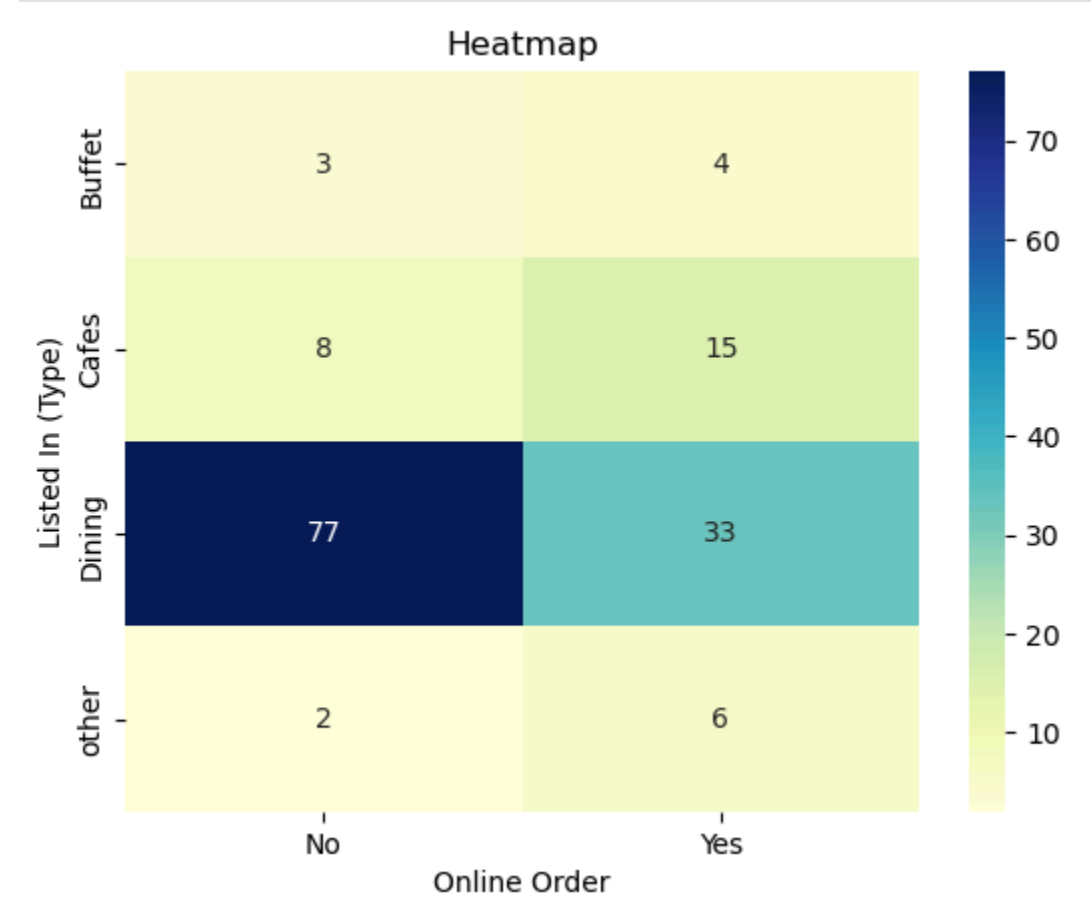
Question - which type restaurant received more offline orders, so that Zomato can prove customers with some good offers?

In [26]:
```
dataframe.head()
```

Out[26]:

	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 [28]:
```
pivot_table = dataframe.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")
plt.xlabel("Online Order")
plt.ylabel("Listed In (Type)")
plt.show()
```



conclusion - Dinning restaurants primarily accept offline orders, whereas cafes primarily receive online orders. This suggests that clients prefered oreders in person at restaurants, but prefer online ordering at cafes.

In [ ]: