zomato-eda

August 22, 2025

Zomato Data Analysis Using Python

[52]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[53]:
```
import os

print(os.listdir("C:/Users/Tanya Raj/OneDrive/Desktop"))
```

```
['02_numpy(linear algebra).ipynb', '1.xlsx', 'ADHAAR CARD.pdf', 'aditya',
'Arduino IDE.lnk', 'Canva.lnk', 'desktop.ini', 'FITA', 'major_project doc',
'myself info', 'python basics', 'python function-2']
```

[54]:
```
pip install openpyxl
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: openpyxl in c:\users\tanya
raj\appdata\roaming\python\python313\site-packages (3.1.5)
Requirement already satisfied: et-xmlfile in c:\users\tanya
raj\appdata\roaming\python\python313\site-packages (from openpyxl) (2.0.0)
Note: you may need to restart the kernel to use updated packages.


[notice] A new release of pip is available: 25.1.1 -> 25.2
[notice] To update, run: python.exe -m pip install --upgrade pip
```

[55]:
```
import pandas as pd

dataframe = pd.read_excel(r"C:\Users\Tanya Raj\OneDrive\Desktop\1.xlsx")
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
```

[56]:
```
#Data cleaning and preparation

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
```

[57]:
```
#getting summary of the dataframe use df.info().
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)
```

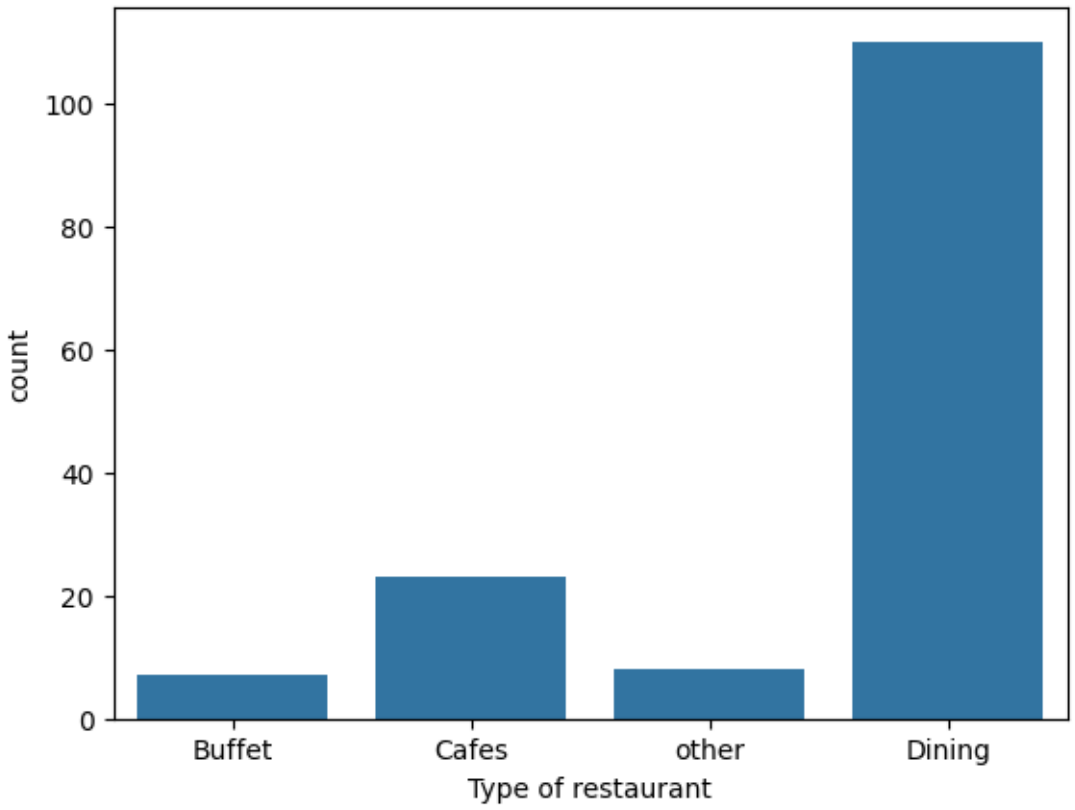
```
memory usage: 8.2+ KB
```

[58]:
```
#checking for missing or null values to identitfy any data gaps

print(dataframe.isnull().sum()) #There is no NULL in dataframe.
```

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

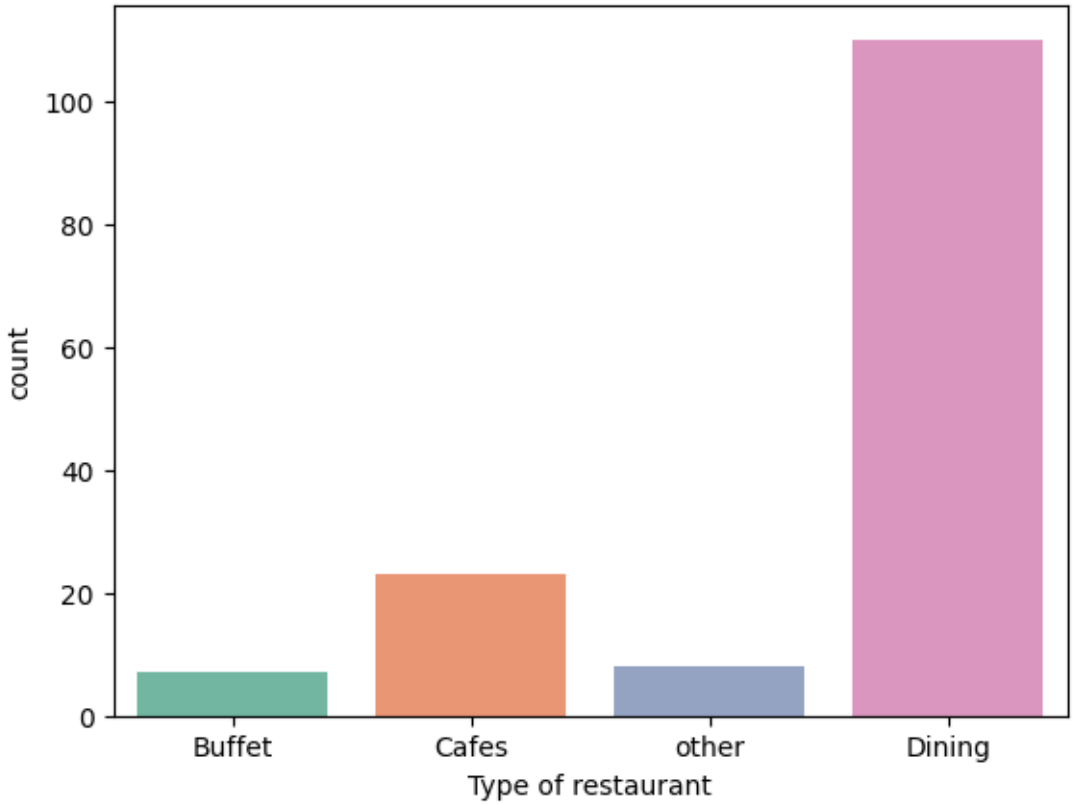
[59]:
```
#step 4:exploring Restaurant types

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

[59]:
```
Text(0.5, 0, 'Type of restaurant')
```



[60]:
```
# Step 4: Exploring Restaurant types
sns.countplot(x=dataframe['listed_in(type)'], palette="Set2")
plt.xlabel("Type of restaurant")
plt.show()
```

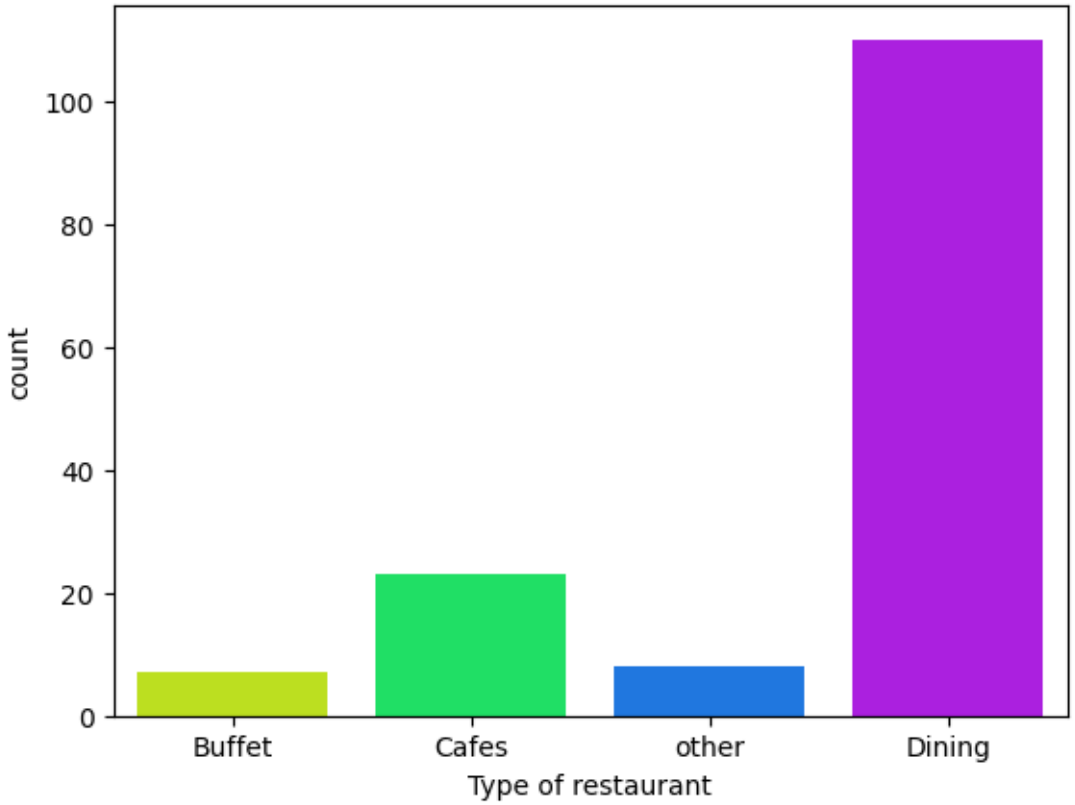
```
C:\Users\Tanya Raj\AppData\Local\Temp\ipykernel_23064\123284339.py:2:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.countplot(x=dataframe['listed_in(type)'], palette="Set2")
```



[51]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Step 4: Exploring Restaurant types
sns.countplot(x=dataframe['listed_in(type)'], palette=sns.color_palette("hsv",␣
 ↪len(dataframe['listed_in(type)'].unique())))
plt.xlabel("Type of restaurant")
```

```
plt.show()
```

```
C:\Users\Tanya Raj\AppData\Local\Temp\ipykernel_23064\2550283831.py:5:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.countplot(x=dataframe['listed_in(type)'], palette=sns.color_palette("hsv",
len(dataframe['listed_in(type)'].unique())))
```



[63]:
```
#Votes by Restaurant Type

grouped_data = dataframe.groupby('listed_in(type)')['votes'].sum()
result = pd.DataFrame({'votes': grouped_data})


#visualization

plt.plot(result, c='red', marker='o')
```

```
plt.xlabel('Type of restaurant')
plt.ylabel('Votes')
```

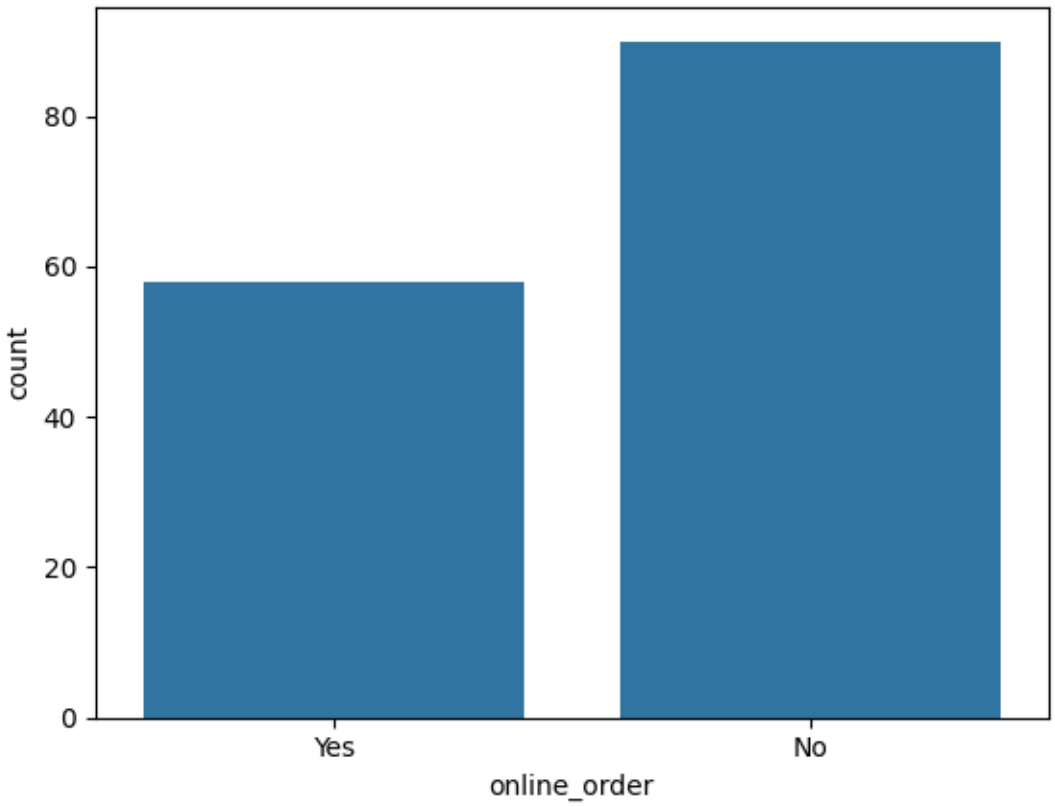
[63]:
```
Text(0, 0.5, 'Votes')
```

[64]:
```
#step 5: identify the most voted restaurant

max_votes = dataframe['votes'].max()
restaurant_with_max_votes = dataframe.loc[dataframe['votes'] == max_votes,␣
 ↪'name']

print('Restaurant(s) with the maximum votes:')
print(restaurant_with_max_votes)
```

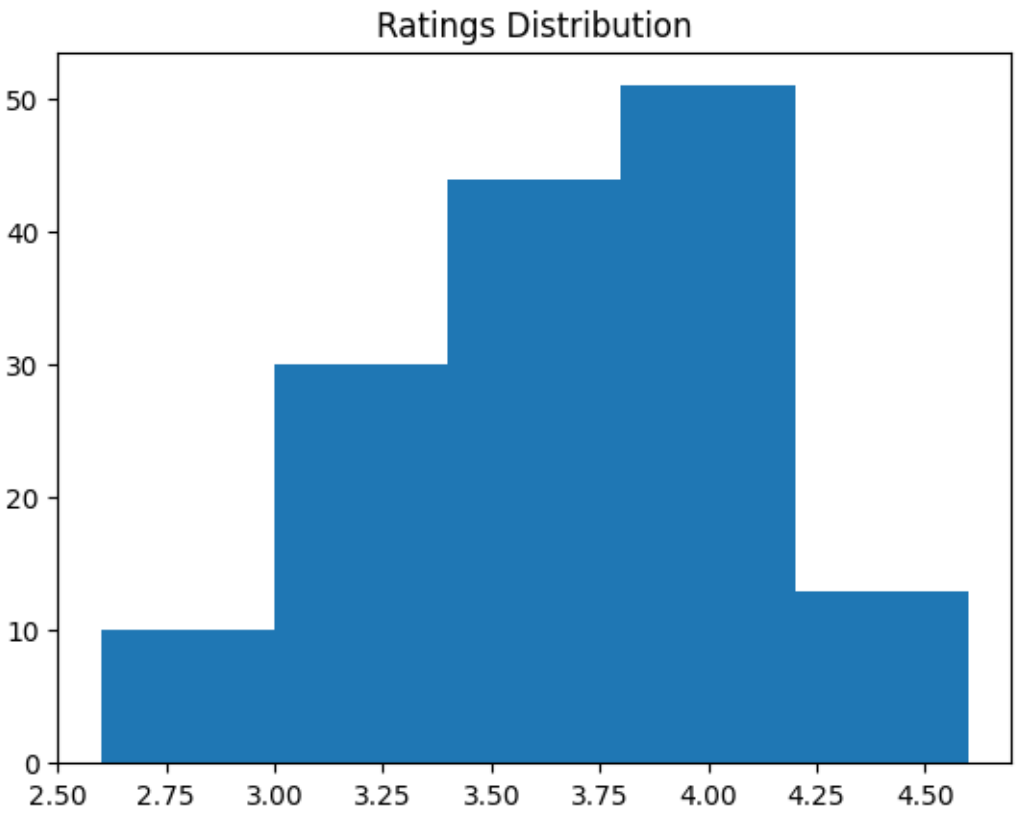
```
Restaurant(s) with the maximum votes:
38    Empire Restaurant
Name: name, dtype: object
```

[65]:
```
#Online order Availability

sns.countplot(x=dataframe['online_order'])
```

[65]: <Axes: xlabel='online_order', ylabel='count'>

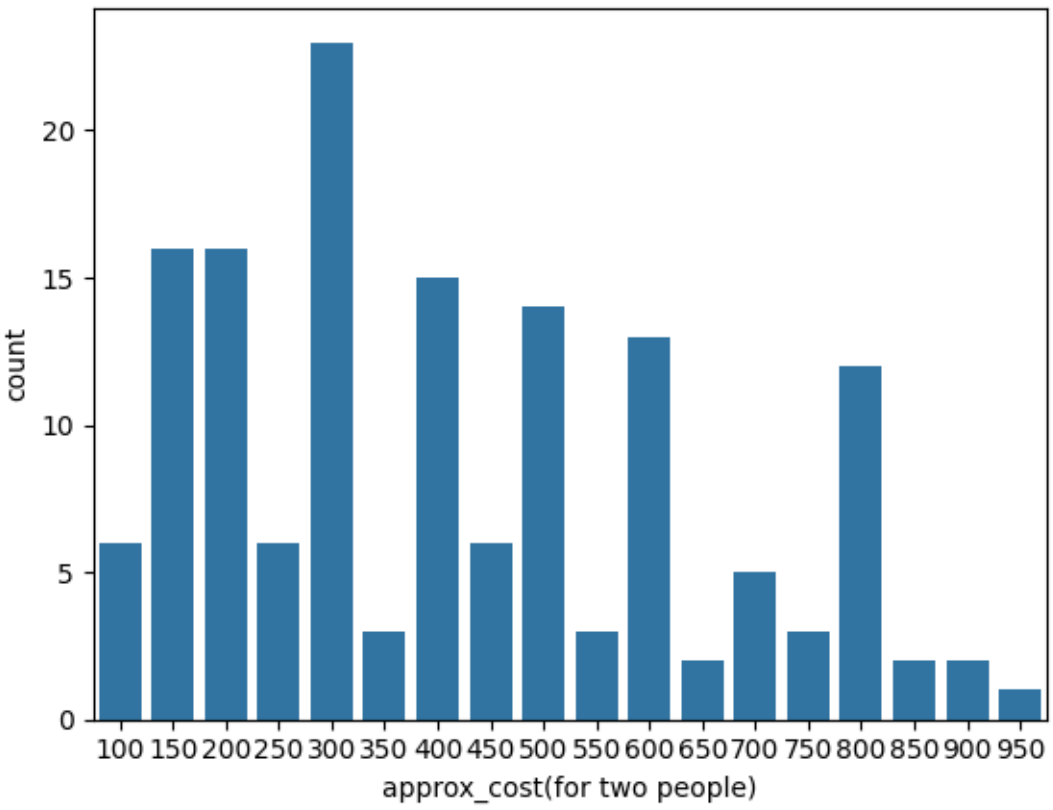


[66]:
```
#Analyze Ratings

plt.hist(dataframe['rate'],bins=5)
plt.title('Ratings Distribution')
plt.show()
```



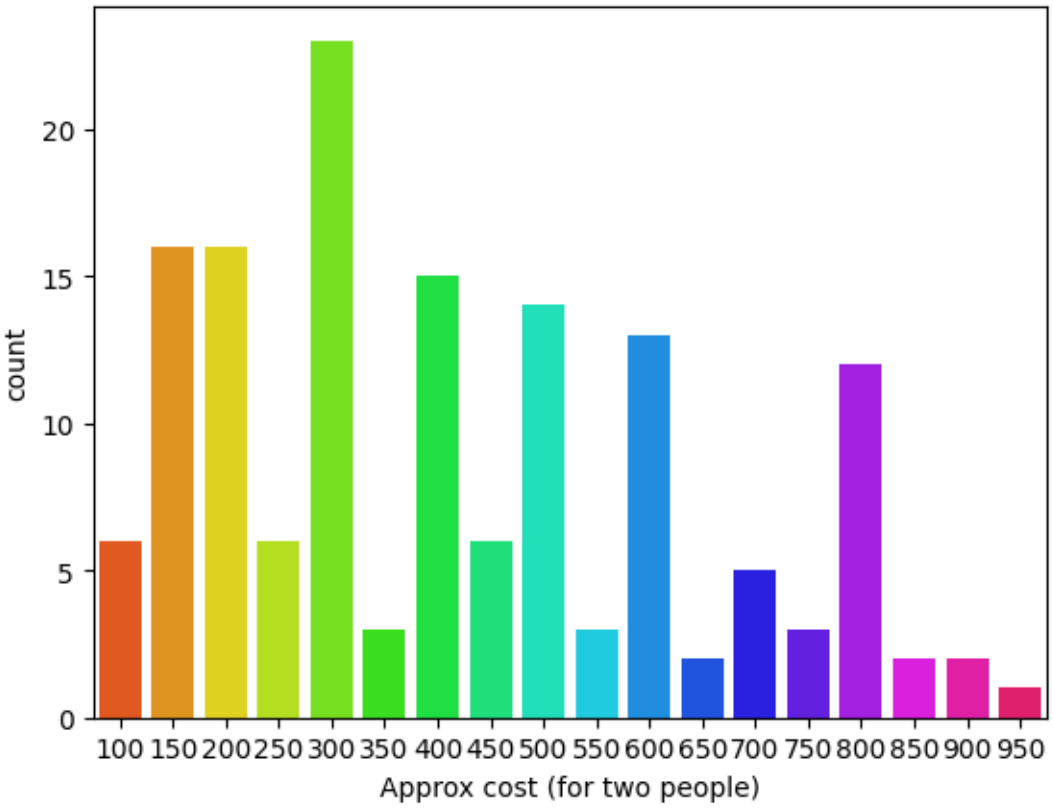
[67]:
```
c
```

[67]:
```
<Axes: xlabel='approx_cost(for two people)', ylabel='count'>
```



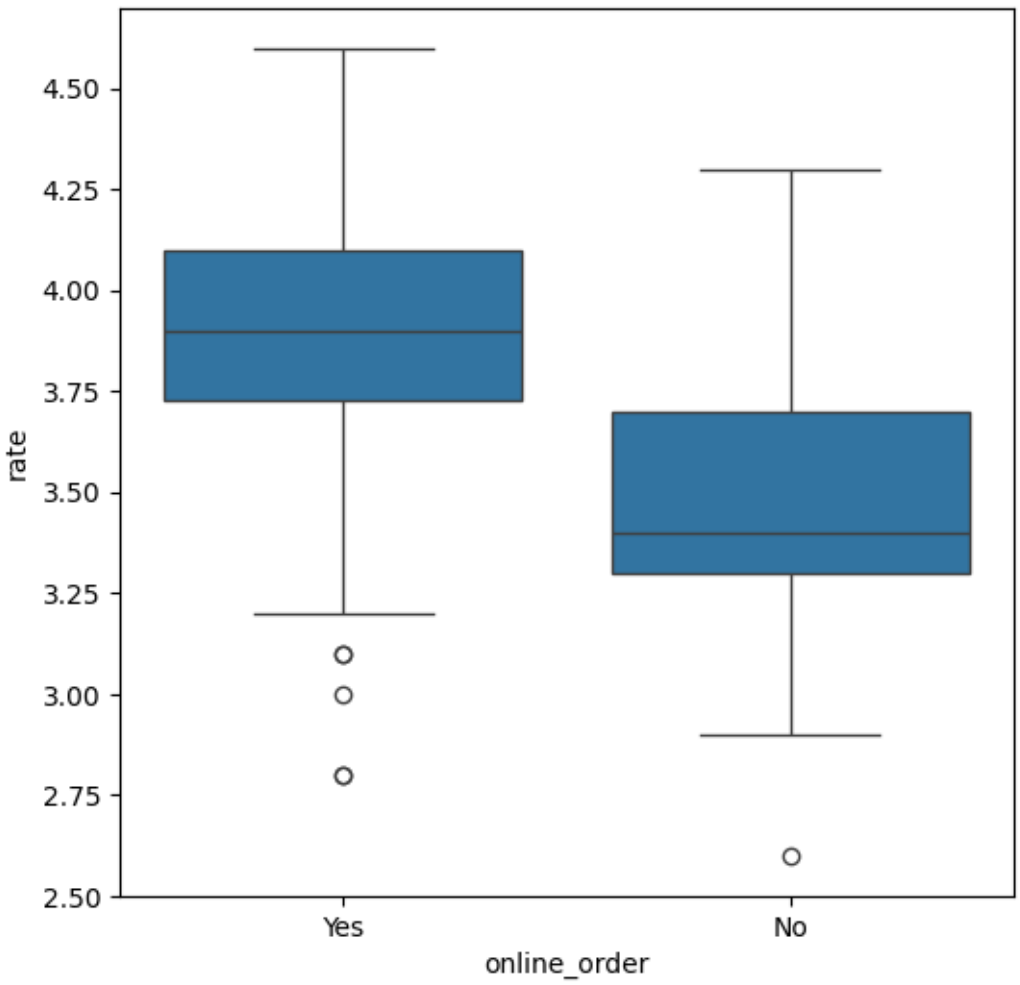
[68]:
```
import seaborn as sns
import matplotlib.pyplot as plt

couple_data = dataframe['approx_cost(for two people)']

sns.countplot(
    x=couple_data,
    palette=sns.color_palette("hsv", len(couple_data.unique()))
)

plt.xlabel("Approx cost (for two people)")
plt.show()
```

```
C:\Users\Tanya Raj\AppData\Local\Temp\ipykernel_23064\2068408218.py:6:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.countplot(
```



[69]:
```
#Step 9: Ratings Comparison - Online vs Offline Orders


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

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

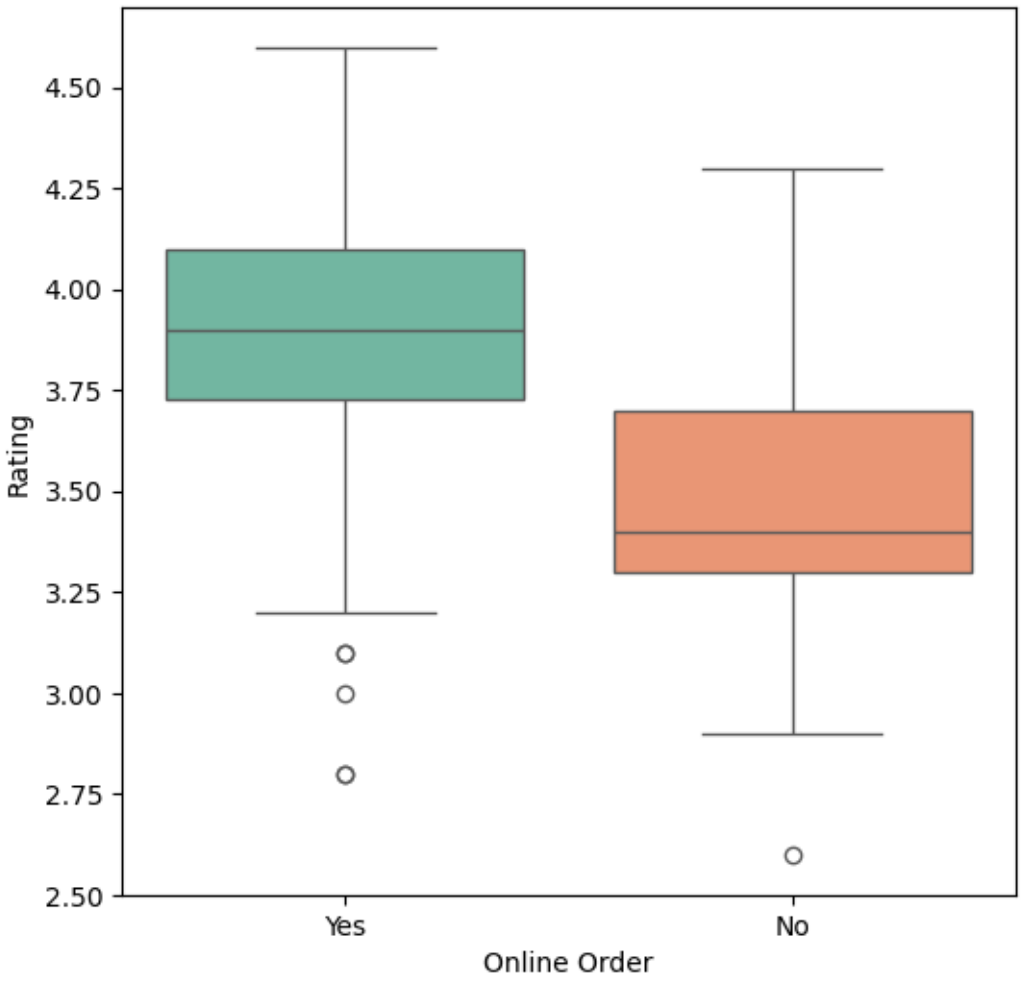


[70]:
```
plt.figure(figsize=(6,6))
sns.boxplot(
    x='online_order',
    y='rate',
    data=dataframe,
    palette="Set2"   # you can try "Paired", "coolwarm", "hsv", etc.
)
plt.xlabel("Online Order")
plt.ylabel("Rating")
plt.show()
```

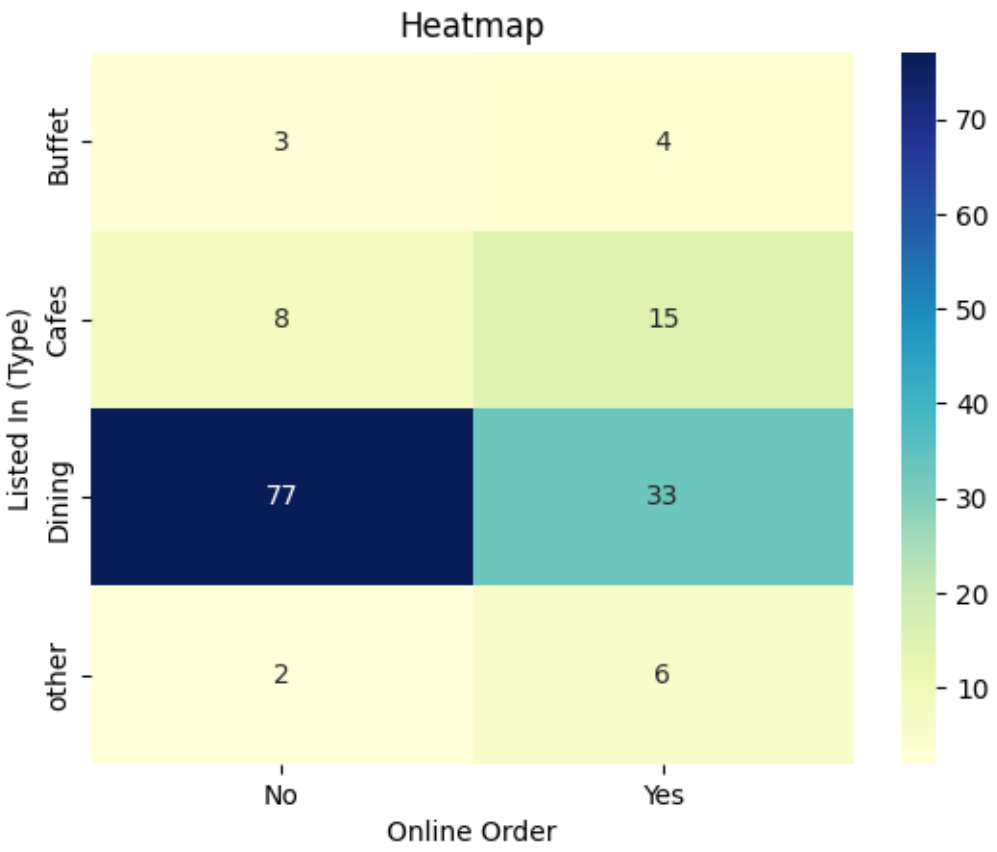
```
C:\Users\Tanya Raj\AppData\Local\Temp\ipykernel_23064\1666518352.py:2:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
```

effect.

sns.boxplot(



[71]:
```
#Step 10:Order Mode Preferences by Restaurant Type


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()
```


Heatmap
Buffet
Cafes
Dining
other
Listed In (Type)
3
4
8
15
77
33
2
6
No
Yes
Online Order
70
60
50
40
30
20
10