```
# This Python 3 environment comes with many helpful analytics
libraries installed
# It is defined by the kaggle/python Docker image:
https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/"
directory
# For example, running this (by clicking run or pressing Shift+Enter)
will list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/)
that gets preserved as output when you create a version using "Save &
Run All"
# You can also write temporary files to /kaggle/temp/, but they won't
be saved outside of the current session

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

pip install kaggle
```

```
Requirement already satisfied: kaggle in
/opt/conda/lib/python3.10/site-packages (1.6.14)
Requirement already satisfied: six>=1.10 in
/opt/conda/lib/python3.10/site-packages (from kaggle) (1.16.0)
Requirement already satisfied: certifi>=2023.7.22 in
/opt/conda/lib/python3.10/site-packages (from kaggle) (2024.7.4)
Requirement already satisfied: python-dateutil in
/opt/conda/lib/python3.10/site-packages (from kaggle) (2.9.0.post0)
Requirement already satisfied: requests in
/opt/conda/lib/python3.10/site-packages (from kaggle) (2.32.3)
Requirement already satisfied: tqdm in /opt/conda/lib/python3.10/site-
packages (from kaggle) (4.66.4)
Requirement already satisfied: python-slugify in
/opt/conda/lib/python3.10/site-packages (from kaggle) (8.0.4)
Requirement already satisfied: urllib3 in
/opt/conda/lib/python3.10/site-packages (from kaggle) (1.26.18)
Requirement already satisfied: bleach in
/opt/conda/lib/python3.10/site-packages (from kaggle) (6.1.0)
```

```
Requirement already satisfied: webencodings in
/opt/conda/lib/python3.10/site-packages (from bleach->kaggle) (0.5.1)
Requirement already satisfied: text-unidecode>=1.3 in
/opt/conda/lib/python3.10/site-packages (from python-slugify->kaggle)
(1.3)
Requirement already satisfied: charset-normalizer<4,>=2 in
/opt/conda/lib/python3.10/site-packages (from requests->kaggle)
(3.3.2)
Requirement already satisfied: idna<4,>=2.5 in
/opt/conda/lib/python3.10/site-packages (from requests->kaggle) (3.6)
Note: you may need to restart the kernel to use updated packages.
```

```
df= pd.read_csv('/kaggle/input/zomato-sales/Zomato data .csv')
print(df.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
```

```
df
```

```
                      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]
```

removing denominator from rate column

```
def handleRate(value):
    value= str(value).split('/')
    value= value[0];
    return float (value)
df['rate']=df['rate'].apply(handleRate)
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
```

info dataframe

```
df.info()
```

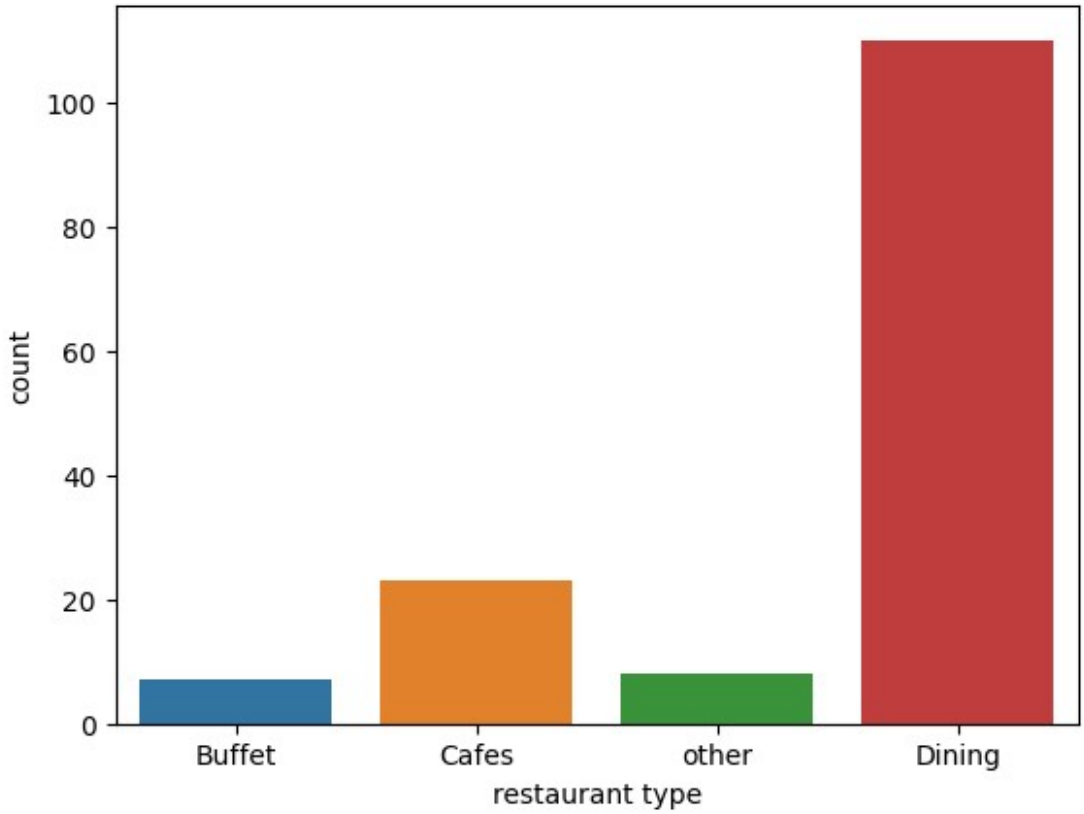
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148 entries, 0 to 147
Data columns (total 7 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   name                         148 non-null    object
 1   online_order                 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
```

no null value here

```
sns.countplot(x=df['listed_in(type)'])
plt.xlabel('restaurant type')
```

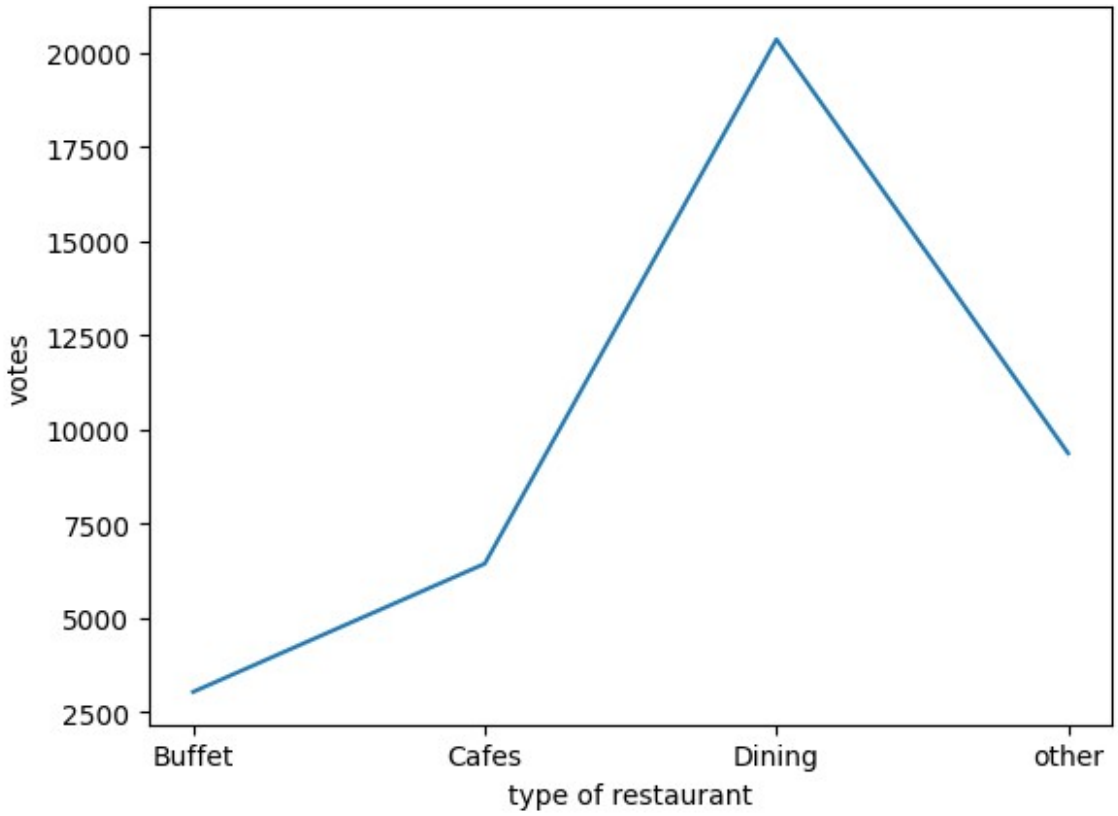
```
Text(0.5, 0, 'restaurant type')
```



dining categary holds max orders

```
grp_data=df.groupby('listed_in(type)')['votes'].sum()
result=pd.DataFrame({'votes': grp_data})
plt.plot(result)
plt.xlabel('type of restaurant')
plt.ylabel('votes')
```

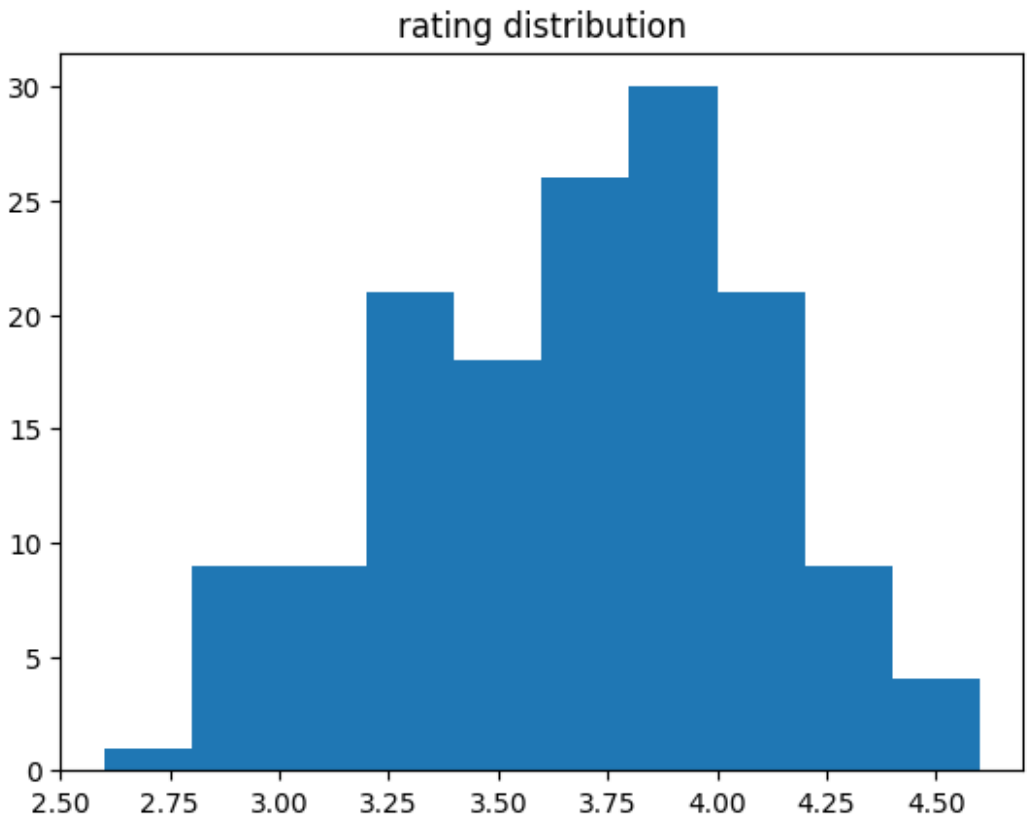
```
Text(0, 0.5, 'votes')
```



majority of restourant recieved votes

for rating distrubution

```
plt.hist(df['rate'])
plt.title('rating distribution')
plt.show()
```

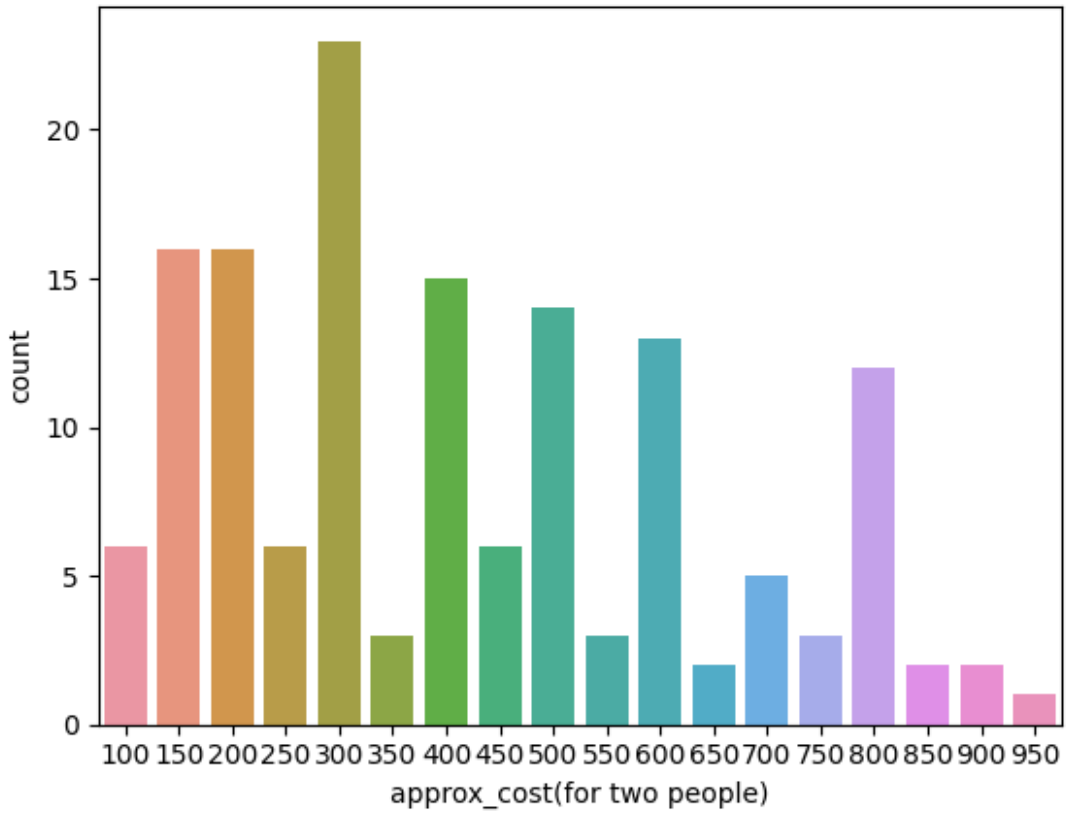


Majority of rest. recieved 3.5 to 5 ratings

for two people/couple data info

```
cpl_data=df['approx_cost(for two people)']
sns.countplot(x=cpl_data)
```

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

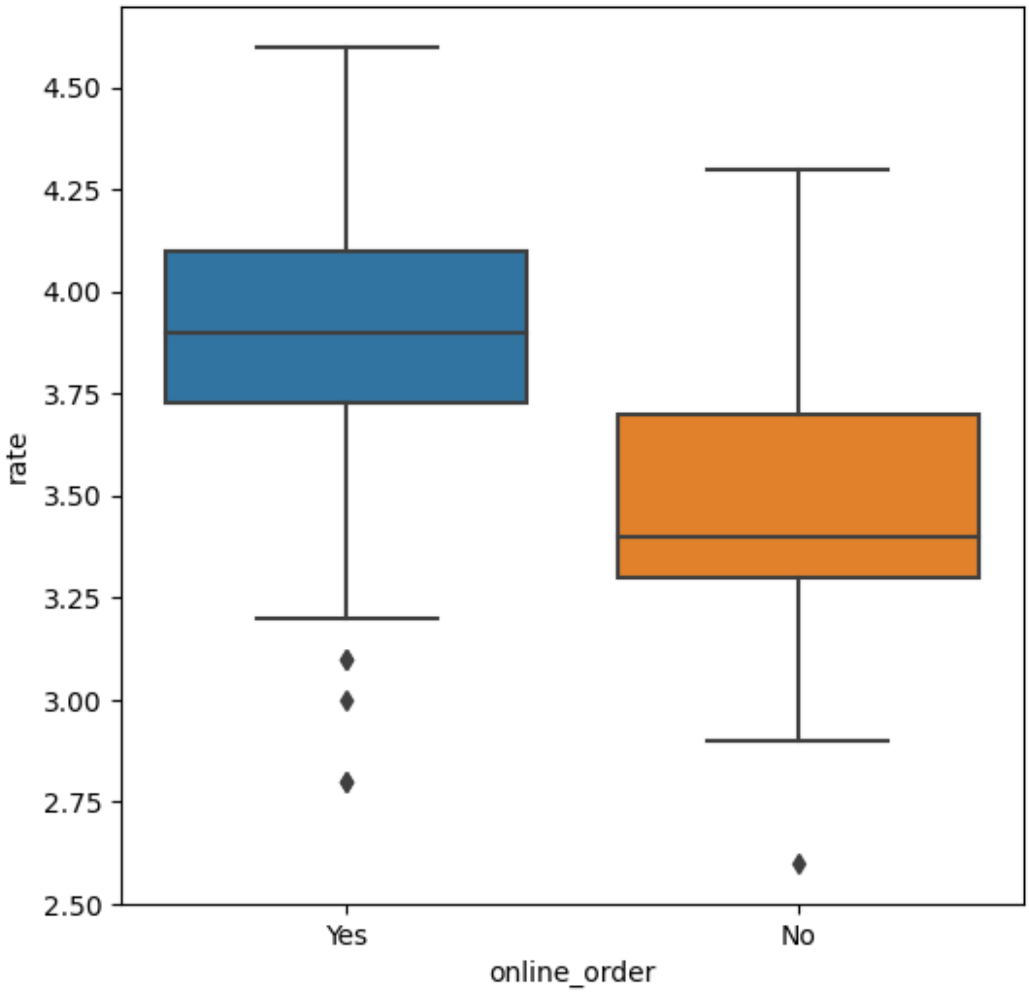


Most couples/ two people prefer cost 300

Online or offline which given higher rating

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

```
<Axes: xlabel='online_order', ylabel='rate'>
```



online orders having higher rating than offline