zomato-data-analysis

October 24, 2024

1. Importing Libraries

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

---

2. Create the Data Frame (df)

```
[39]: dataframe = pd.read_csv(r"H:\DA. Python\Zomato Data Analysis\Zomato data.csv")
```

```
[40]: 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
```

1

```
147                              200            Dining
```

```
[148 rows x 7 columns]
```

[41]: `dataframe`

[41]:
```
                          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]
```

---

3. Cleaning Data

Convert the data type of column 'Rate'

[42]:
```python
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
```

Summary of Dataframe

`[43]:` `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
```

There is no NULL values in the Dataframe

---

4. Exploratory Data Analysis (EDA)

TYPE OF RESTAURANT

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

`[44]:`
```
                    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
```

[45]:
```python
sns.countplot(x=dataframe['listed_in(type)'])
plt.xlabel('type of restaurant')
```

[45]: Text(0.5, 0, 'type of restaurant')
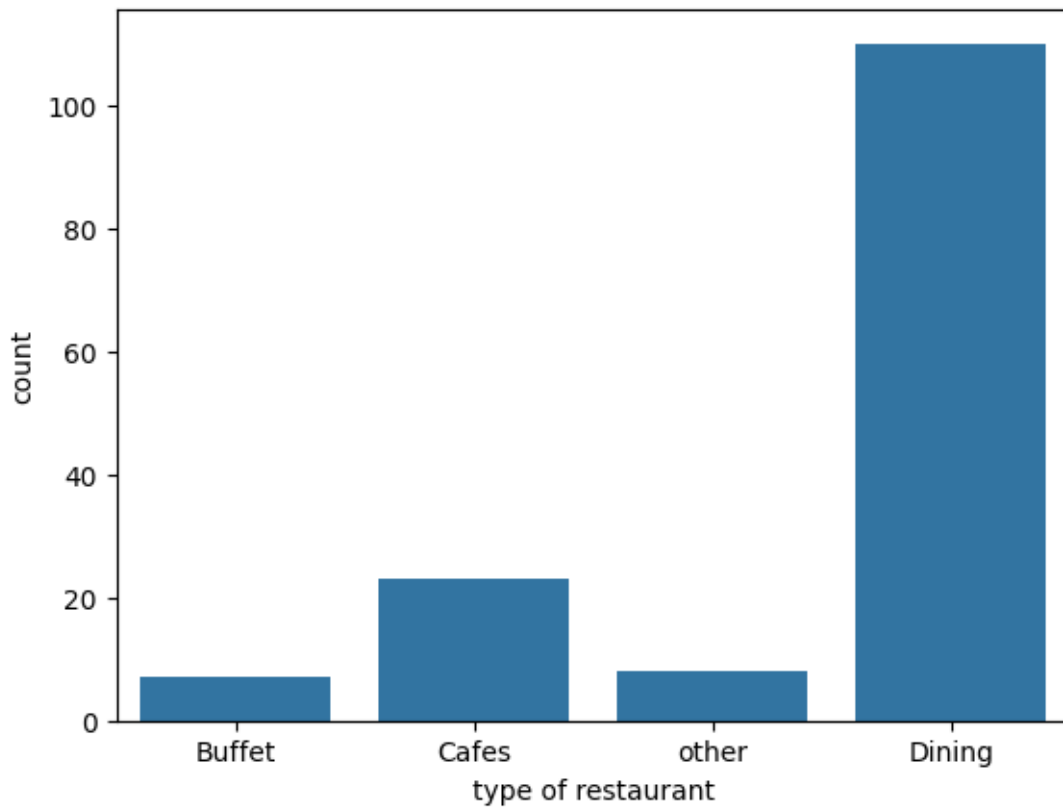
CONCLUSION = Majority of the restaurant falls in dining category.

VOTES RECIEVED BY RESTAURANTS

[46]:
```python
dataframe.head()
```

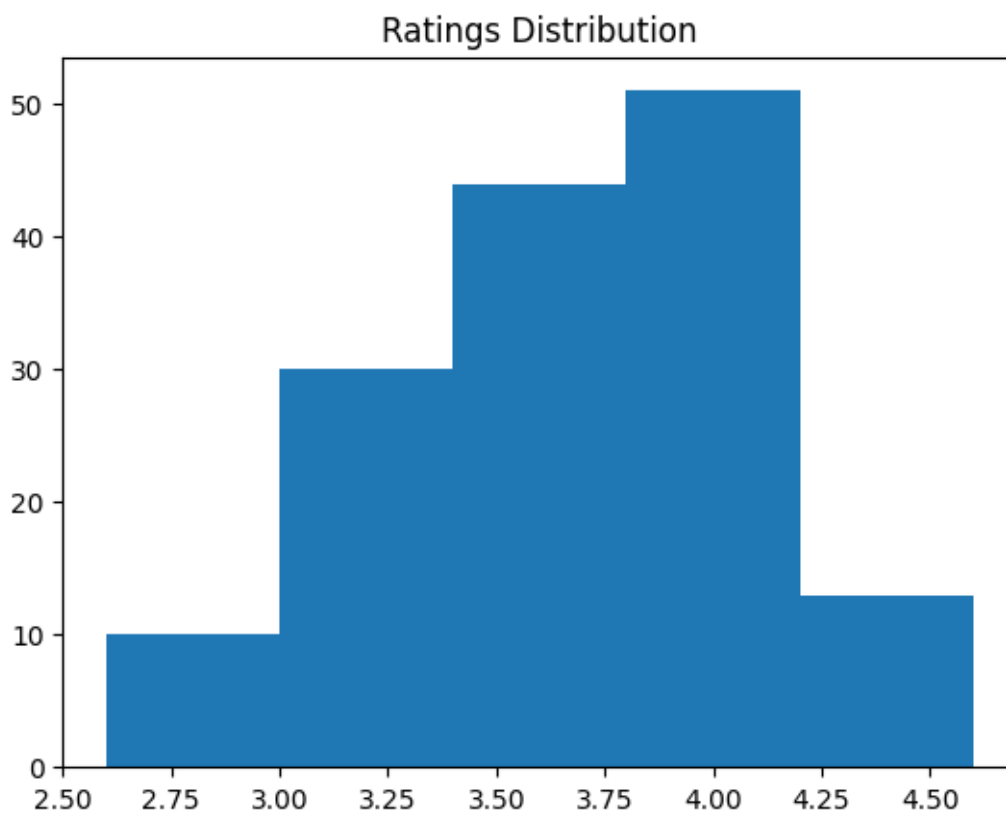[46]:
```
                name online_order book_table  rate  votes  \
0              Jalsa          Yes        Yes   4.1    775
1      Spice Elephant        Yes         No   4.1    787
```

```
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
```

```python
grouped_data = dataframe.groupby('listed_in(type)')['votes'].sum()
result = pd.DataFrame({'votes' : grouped_data})
plt.plot(result, c = "red" , marker = "o")
plt.xlabel('Type of restaurant', c = 'black', size = 15)
plt.ylabel('Votes', c = 'black',  size = 15)
```

[56]: Text(0, 0.5, 'Votes')
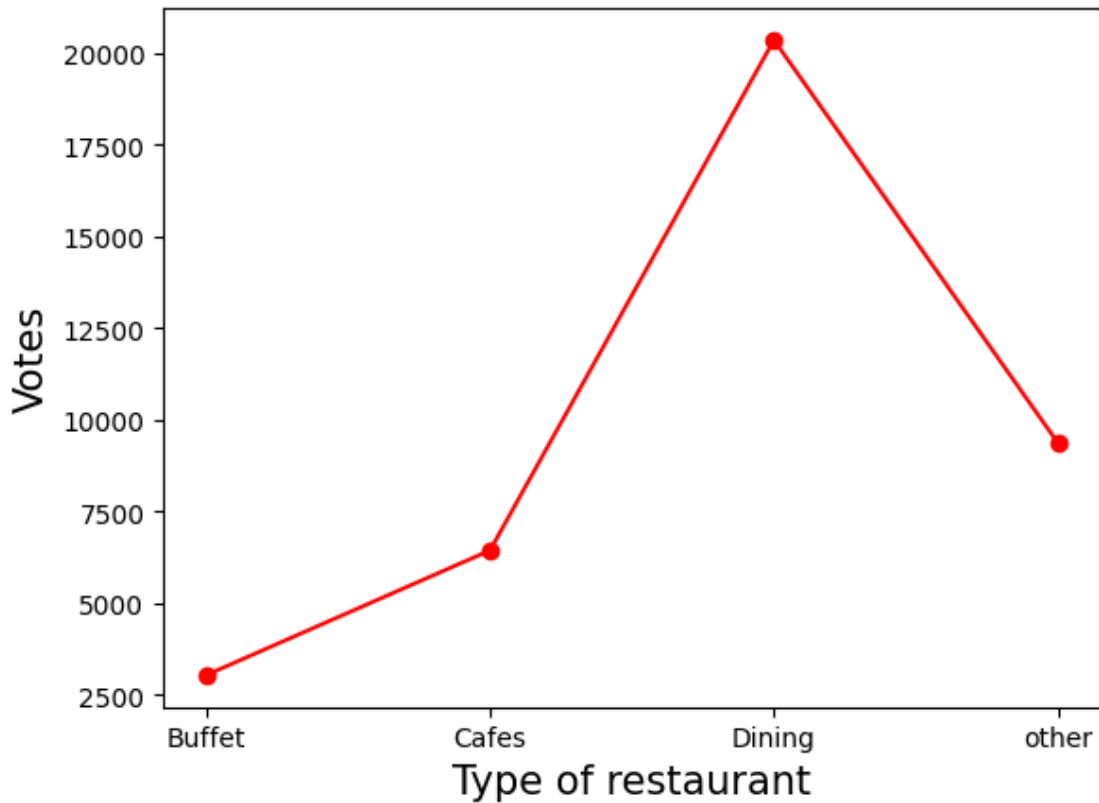
CONCLUSION = Dining restaurant have recieved maximum votes.

_____

RATINGS RECIEVED BY RESTAURANTS

[57]: `dataframe.head()`

[57]:
```
                    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
```

[62]:
```
plt.hist(dataframe['rate'], bins = 5)
plt.title('Ratings Distribution')
plt.show()
```

CONCLUSION = The majority of restaurants recieved ratings from 3.5 to 4.

AVERAGE ORDER SPENDING BY COUPLES

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

```
[63]:                     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
```

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

```
[65]: <Axes: xlabel='approx_cost(for two people)', ylabel='count'>
```
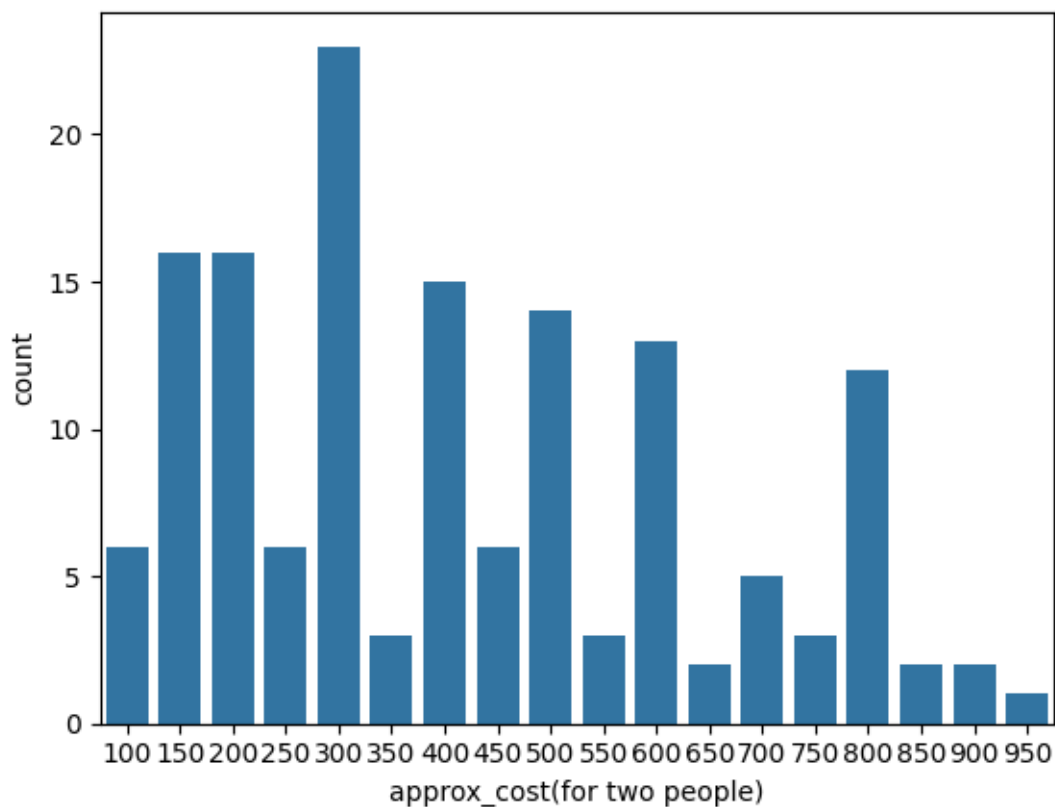
CONCLUSION = The majority of couples prefers restaurant with an approximate cost of 300 rs.

WHICH MODE RECIEVES THE HIGHEST RATINGS (online or offline)

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

```
[66]:                     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
```
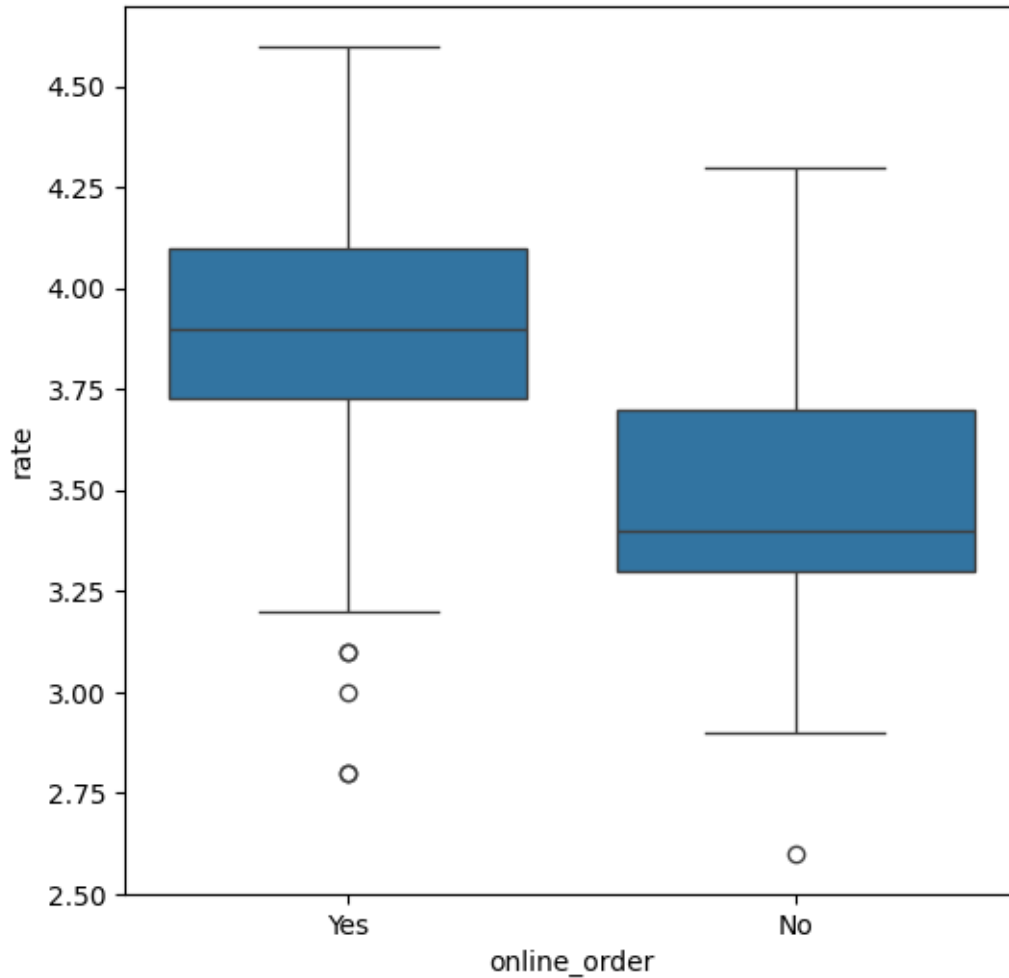
```
[67]: plt.figure(figsize=(6,6))
      sns.boxplot(x = 'online_order', y = 'rate', data = dataframe)
```

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

CONCLUSION = Offline orders recieved lower ratings in comparison to Online order.

WHICH TYPE OF RESTAURANT RECIEVES MORE ORDERS

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

[68]:   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
```

[70]:
```python
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.xlabel("Online Order")
plt.ylabel("Listed_in(Type)")
plt.show()
```


CONCLUSION = Dining restaurants primarily accepts offline orders, whereas cades primarily recieve online orders. This suggests that clients prefers to place orders in person at restaurants, but prefers online ordering at cafes.