# Zomato Data Analysis Project

#### Step-1 Importing Libraries

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

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

#### Step-2 Creating Data Frame

```
dataframe = pd.read_csv("Zomato data .csv")
print(dataframe)
                       name online order book table
                                                               votes
                                                         rate
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 3.7/5
                                                                  88
                                       No
4
              Grand Village
                                       No
                                                   No 3.8/5
                                                                  166
                                       . . .
                                                   . . .
                                                                  . . .
143
          Melting Melodies
                                                   No 3.3/5
                                                                    0
                                       No
144
            New Indraprasta
                                       No
                                                   No 3.3/5
                                                                    0
145
               Anna Kuteera
                                      Yes
                                                   No 4.0/5
                                                                  771
146
                                                   No 3.0/5
                                                                   98
                     Darbar
                                       No
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
                                             Buffet
                               300
4
                                             Buffet
                               600
143
                               100
                                             Dining
144
                               150
                                             Dining
145
                               450
                                             Dining
146
                               800
                                             Dining
147
                               200
                                             Dining
[148 rows x 7 columns]
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
                                                          3.8/5
            San Churro Cafe
                                        Yes
                                                                    918
                                                     No
3
     Addhuri Udupi Bhojana
                                         No
                                                     No
                                                         3.7/5
                                                                     88
4
              Grand Village
                                                     No
                                                          3.8/5
                                         No
                                                                    166
                                        . . .
                                                     . . .
                                                                    . . .
143
           Melting Melodies
                                                          3.3/5
                                                                      0
                                         No
                                                     No
            New Indraprasta
144
                                         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]
```

# Let's convert the data type of 'rate' column to float and remove the denominator.

```
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
                                                       4.1
                                                              775
                                                Yes
1
          Spice Elephant
                                                       4.1
                                    Yes
                                                 No
                                                              787
2
         San Churro Cafe
                                    Yes
                                                 No
                                                       3.8
                                                              918
3
   Addhuri Udupi Bhojana
                                     No
                                                 No
                                                       3.7
                                                               88
           Grand Village
                                     No
                                                              166
                                                 No
                                                       3.8
   approx cost(for two people) listed in(type)
0
                             800
                                           Buffet
1
                             800
                                           Buffet
2
                                           Buffet
                             800
```

	3	300
4 600 Buttet	4	600

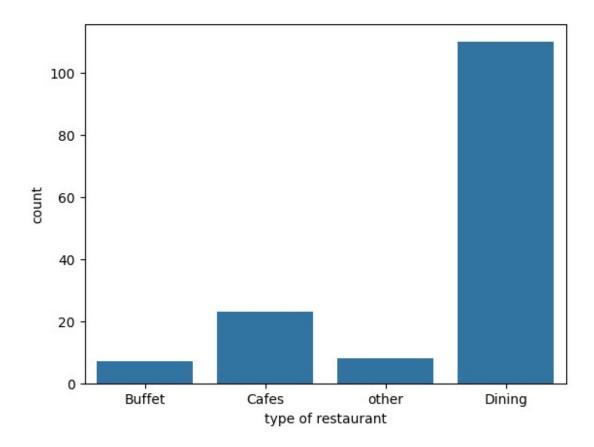
## Summary of the data frame

```
dataframe.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148 entries, 0 to 147
Data columns (total 7 columns):
#
     Column
                                   Non-Null Count
                                                   Dtype
- - -
 0
                                                   object
     name
                                   148 non-null
     online order
 1
                                   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
     listed in(type)
                                   148 non-null
                                                   object
dtypes: float64(1), int64(2), object(4)
memory usage: 8.2+ KB
```

Conclusion there is no NULL values in data frame.

#### Type of restaurant

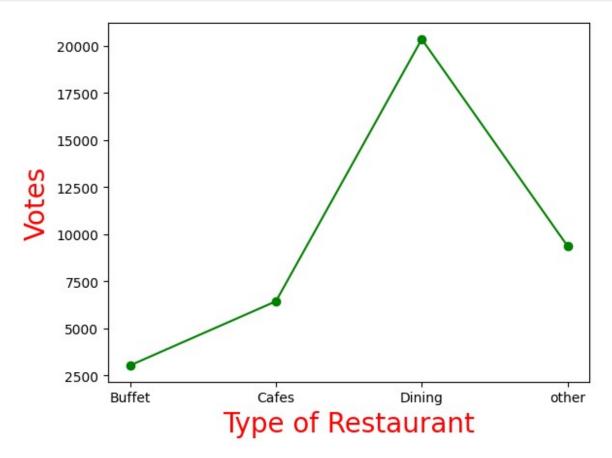
```
dataframe.head()
                     name online order book table
                                                     rate
                                                           votes \
0
                                   Yes
                                                      4.1
                    Jalsa
                                               Yes
                                                             775
1
          Spice Elephant
                                   Yes
                                                      4.1
                                                             787
                                                No
2
         San Churro Cafe
                                   Yes
                                                      3.8
                                                             918
                                                No
3
  Addhuri Udupi Bhojana
                                     No
                                                No
                                                      3.7
                                                              88
           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
sns.countplot(x=dataframe['listed_in(type)'])
plt.xlabel('type of restaurant')
Text(0.5, 0, 'type of restaurant')
```



Conclusion:- majority of the restaurant falls in dining category majority of the customers preffers dinning restaurents.

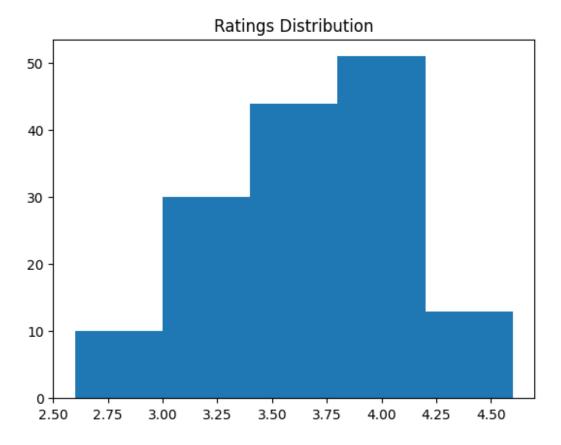
data	frame					
0 1 2 3 4	name Jalsa Spice Elephant San Churro Cafe Addhuri Udupi Bhojana Grand Village	online_order Yes Yes Yes No	book_table Yes No No No No	rate 4.1 4.1 3.8 3.7 3.8	votes 775 787 918 88 166	\
143 144 145 146 147	 Melting Melodies New Indraprasta Anna Kuteera Darbar Vijayalakshmi	No No Yes No Yes	No No No No No	3.3 3.3 4.0 3.0 3.9	 0 0 771 98 47	
0 1 2 3 4	approx_cost(for two pe	eople) listed 800 800 800 800 300 600	_in(type) Buffet Buffet Buffet Buffet Buffet			

```
143
                              100
                                           Dining
144
                              150
                                           Dining
145
                              450
                                           Dining
146
                              800
                                           Dining
147
                              200
                                           Dining
[148 rows x 7 columns]
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 = "red",size = 20)
plt.ylabel("Votes", c="red", size = 20)
Text(0, 0.5, 'Votes')
```



### Majority of restaurants recieved ratings.

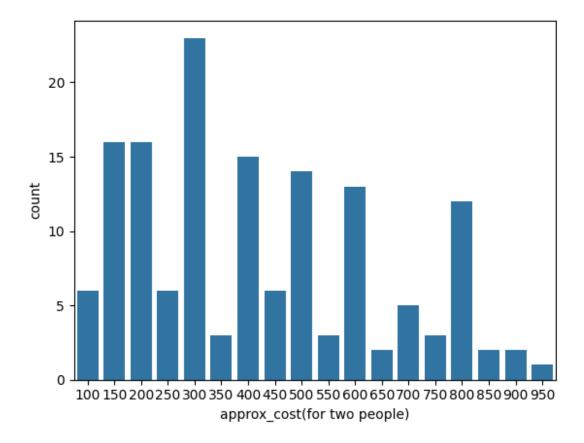
```
plt.hist(dataframe['rate'],bins =5 )
plt.title("Ratings Distribution")
plt.show()
```



Conclusion Majority of the restaurants recieved from rating ranging from 3.5 to 4

Average order spending by couples

```
dataframe.head()
                     name online order book table
                                                      rate
                                                            votes
0
                    Jalsa
                                    Yes
                                                       4.1
                                                              775
                                                Yes
1
          Spice Elephant
                                    Yes
                                                 No
                                                       4.1
                                                              787
2
         San Churro Cafe
                                                       3.8
                                                              918
                                    Yes
                                                 No
3
   Addhuri Udupi Bhojana
                                     No
                                                 No
                                                       3.7
                                                               88
           Grand Village
                                                       3.8
                                                              166
                                     No
                                                 No
   approx cost(for two people) listed in(type)
0
                             800
                                           Buffet
1
                             800
                                           Buffet
2
                             800
                                           Buffet
3
                             300
                                           Buffet
                             600
                                           Buffet
couple_data = dataframe['approx_cost(for two people)']
sns.countplot(x = couple data)
```

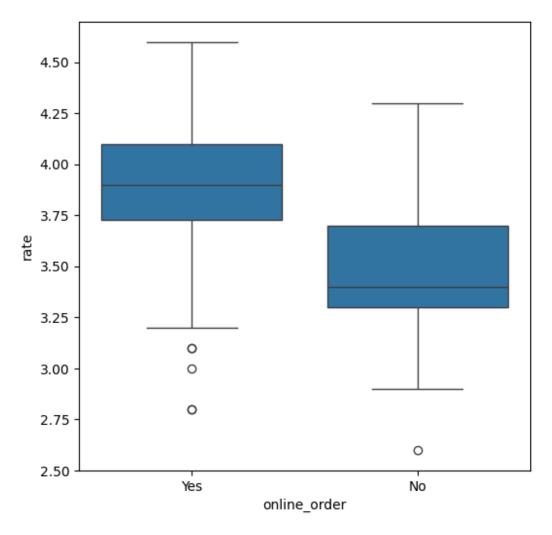


Conclusion: majority of couples prefers restaurant with an approximate cost of 300 rupees.

### Which mode recieves maximum orders?

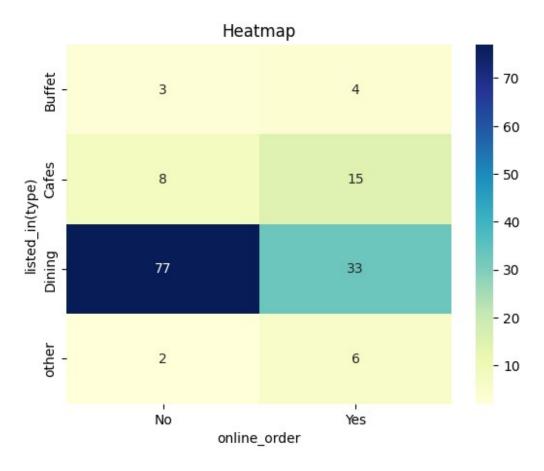
dataframe								
	name	online_order	book_table	rate	votes	\		
0	Jalsa	Yes	Yes	4.1	775			
1	Spice Elephant	Yes	No	4.1	787			
1 2 3	San Churro Cafe	Yes	No	3.8	918			
3	Addhuri Udupi Bhojana	No	No	3.7	88			
4	Grand Village	No	No	3.8	166			
143	Melting Melodies	No	No	3.3	0			
144	New Indraprasta	No	No	3.3	0			
145	Anna Kuteera	Yes	No	4.0	771			
146	Darbar	No	No	3.0	98			
147	Vijayalakshmi	Yes	No	3.9	47			
	approx_cost(for two pe	eople) listed_	_in(type)					

```
0
                                             Buffet
                               800
1
2
                               800
                                             Buffet
                               800
                                             Buffet
3
                                             Buffet
                               300
4
                                             Buffet
                               600
                               . . .
143
                               100
                                             Dining
144
                               150
                                             Dining
145
                               450
                                             Dining
146
                               800
                                             Dining
147
                               200
                                             Dining
[148 rows x 7 columns]
plt.figure(figsize = (6,6))
sns.boxplot(x = 'online_order', y = 'rate', data = dataframe)
<Axes: xlabel='online_order', ylabel='rate'>
```



# conclusion: online orders recieves more rating when compared to offline orders

```
dataframe.head()
                    name online order book table
                                                   rate
                                                         votes \
0
                                  Yes
                                                    4.1
                   Jalsa
                                              Yes
                                                           775
          Spice Elephant
1
                                   Yes
                                                    4.1
                                                           787
                                               No
2
         San Churro Cafe
                                   Yes
                                               No
                                                    3.8
                                                           918
3
  Addhuri Udupi Bhojana
                                                    3.7
                                                            88
                                    No
                                               No
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
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
<function matplotlib.pyplot.show(close=None, block=None)>
```



Conclusion: Dining restaurants primarily accepts offline orders, whereas cafes primarily recieve online orders. This suggests that clients prefer to place orders in person at restaurants, but prefer online orderings at cafes.