Zomato data analysis project

OBJECTIVES

Zomato has an average of 17.5 million monthly transacting customer for it food delivery business.I have a dataset of customers so i need to analyze the data,perform EDA(Exploratory Data Analysis) and visualization.

Importing required libraries

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Create the data frame

In [2]:
```
df = pd.read_csv("Zomato data .csv")
```

In [3]:
```
df.head()
```

Out[3]:

	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

Data cleaning and transformation

Remove '/' from each value of rate column.

In [4]:
```
def rating(value):
    value = str(value).split("/")
    value = value[0]
    return float(value)
```

In [5]:
```
df['rate'] = df['rate'].apply(rating)
df.head()
```

Out[5]:

	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

Check there is any value present in our data

In [6]:
```
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
```

Project Objectives

Q1 What type of resturant do the majority of customer order from?

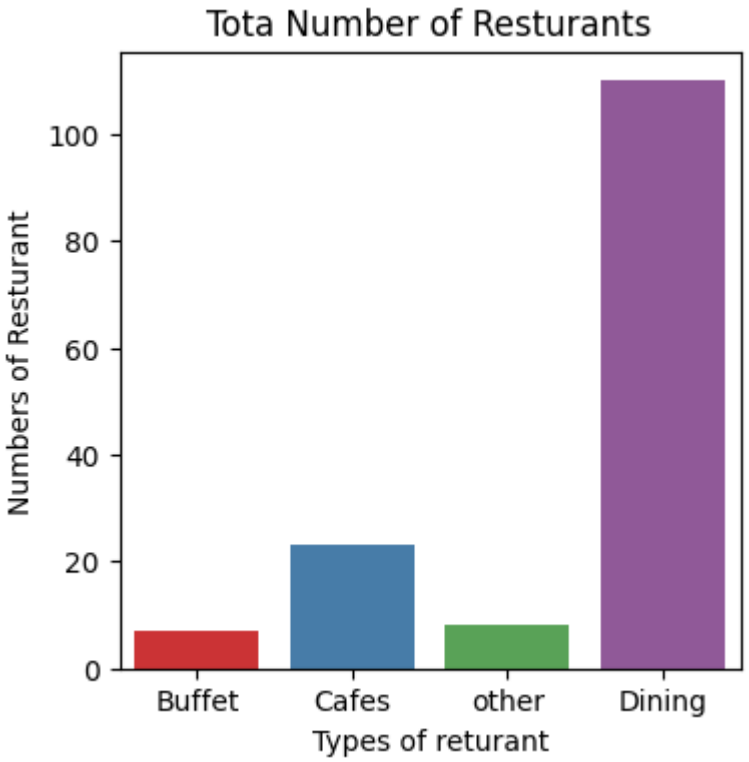
In [7]:
```
df.head()
```

Out[7]:

	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 [58]:
```
plt.figure(figsize = (4,4))
sns.countplot(x = df['listed_in(type)'],legend = False , hue = df['listed_in(type)'],palette = 'Set1')
plt.title('Tota Number of Resturants')
plt.xlabel('Types of returant')
```

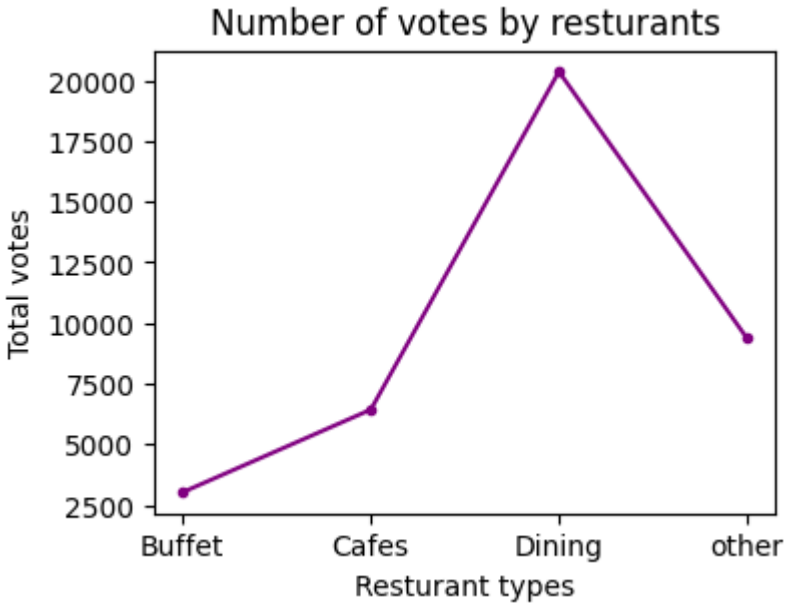
```
plt.ylabel('Numbers of Resturant')
plt.show()
```



Conclusion: Maximum number of resturant falls in dinning category

Q2 How many votes has each type of resturant recieved from customers?

In [59]:
```
plt.figure(figsize = (4,3))
grouped_data = df.groupby('listed_in(type)')['votes'].sum()
result = pd.DataFrame({'votes': grouped_data})
plt.plot(result,color = "purple",marker=".")
plt.title('Number of votes by resturants')
plt.xlabel('Resturant types')
plt.ylabel('Total votes');
```



Conclusion: Dining resturant has maximum votes

Q3 What are the ratings that the majority of resturant have recieved?

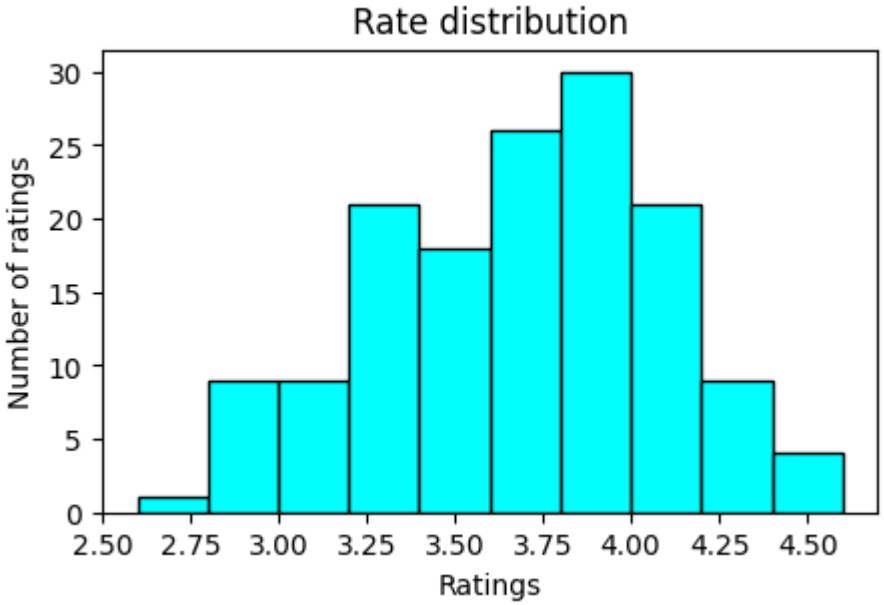
In [84]:
```
df.head()
```

Out[84]:

	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 [60]:
```
plt.figure(figsize = (5,3))
plt.hist(df['rate'],bins = 10, color = 'cyan',edgecolor = 'black')
plt.title('Rate distribution')
plt.xlabel('Ratings')
plt.ylabel('Number of ratings')
```

Out[60]: Text(0, 0.5, 'Number of ratings')



Conclusion: Majority resturant get rating between 3.5 to 4

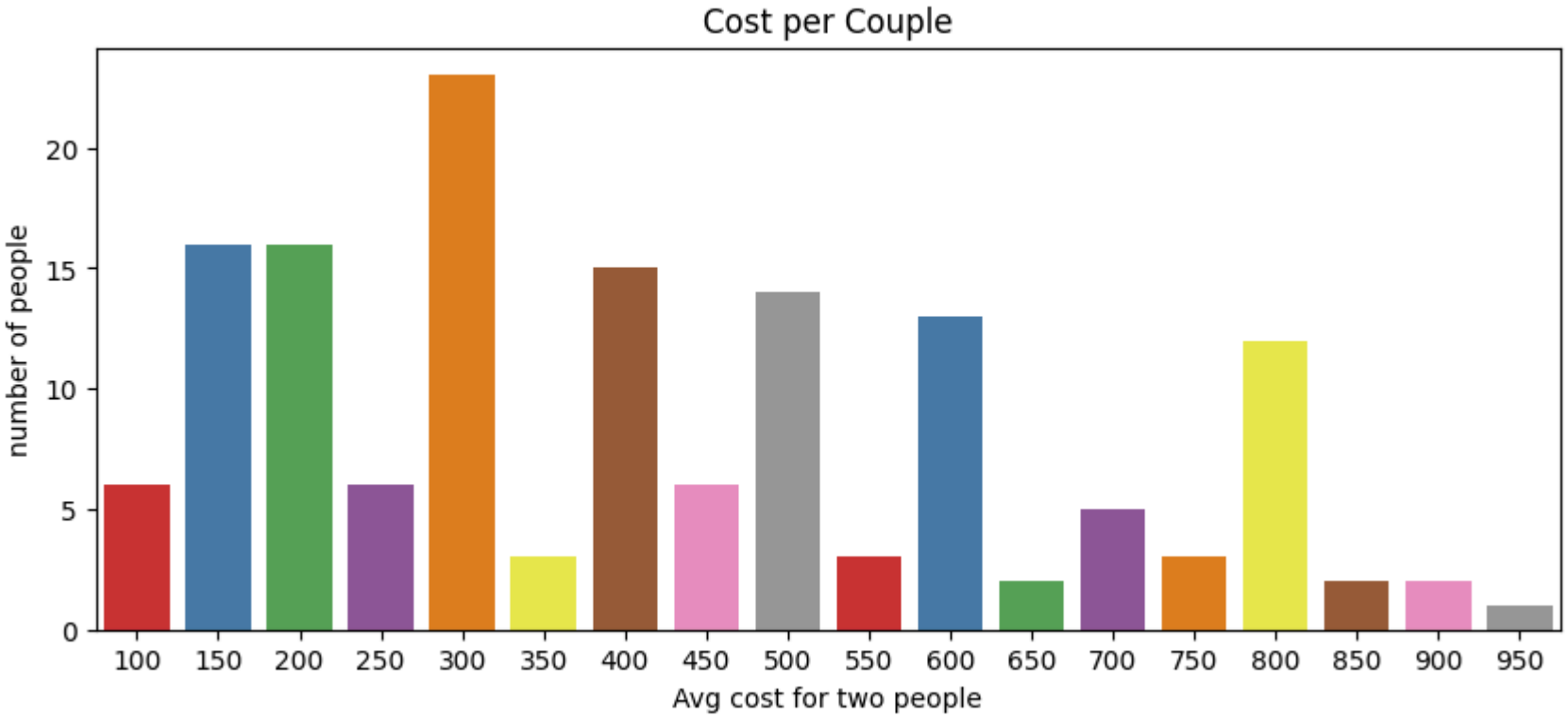
Q4 Zomato has observed the most couples order most of theri food online. What is their average spendingon each order?

In [107…
```
df.head()
```

Out[107…

	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 [61]:
```
plt.figure(figsize = (10,4))
cd = df['approx_cost(for two people)']
sns.countplot(x = cd,legend = False,hue = cd, palette ='Set1')
plt.title('Cost per Couple')
plt.xlabel('Avg cost for two people')
plt.ylabel('number of people');
```



Conclusion: Most ordered by couple is in price range of 300

Q5 Which mode (online or offline) has received the maximum ratings?

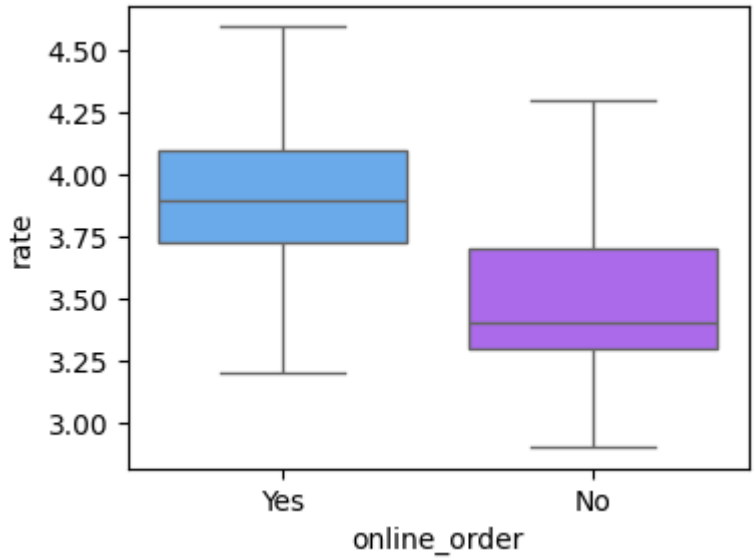
In [121…

```
df.head()
```

Out[121...

	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 [57]:
```
plt.figure(figsize = (4,3))
sns.boxplot(x='online_order',y = 'rate',data = df,showfliers = False,legend = False, hue = 'online_order',palette = 'cool');
```



Conclusion: Offline order get lower rating in comparison of online orders

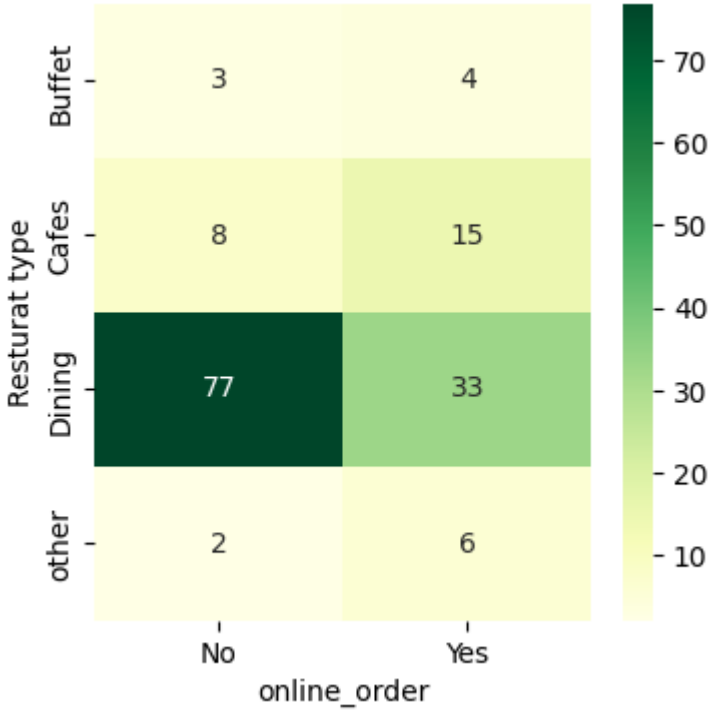
Q6 Which type of resturant received more offline orders ,so that Zomato can pick customers with some good offers.

In [132…
```
df.head()
```

Out[132…

	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 [62]:
```
plt.figure(figsize = (4,4))
pivot_table = df.pivot_table(index = 'listed_in(type)', columns = 'online_order',aggfunc = 'size', fill_value = 0)
sns.heatmap(pivot_table,annot = True,cmap = "YlGn",fmt = 'd')
plt.ylabel('Resturat type');
```



conclusion: Dining resturant primarily accept offline order whereas cafe recives primarily online orders.

KEY INSIGHTS

Maximum number of resturants are in dinning category.

Majority of resturants getting average rating between 3.5 to 4.5

Zomato need to show offers in price range of 300 to 500.

Online orders rating are higher then offline orders.

Dinning resturants accepting offline orders while Cafes gettings online orders.

Thank you

Presented by : RANA BASAK

In [ ]: