zomato data analysis project

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

In [3]:
```
dataframe = pd.read_csv("Zomato data .csv")
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
147                          200          Dining

[148 rows x 7 columns]
```

In [4]:
```
dataframe
```

Out[4]:

	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
...	...	...	...	...	...	...	...
143	Melting Melodies	No	No	3.3/5	0	100	Dining
144	New Indraprasta	No	No	3.3/5	0	150	Dining
145	Anna Kuteera	Yes	No	4.0/5	771	450	Dining
146	Darbar	No	No	3.0/5	98	800	Dining
147	Vijayalakshmi	Yes	No	3.9/5	47	200	Dining

148 rows × 7 columns

data cleaning

In [ ]:
```

```

In [11]:
```
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
```

In [14]:
```
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
```

Type of resturant

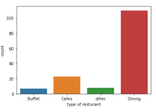
In [18]:
```
dataframe.head()
```

Out[18]:

	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 [21]:
```
sns.countplot(x=dataframe["listed_in(type)"])
plt.xlabel("type of restaurant")
```

Out[21]:
```
Text(0.5, 0, "type of restaurant")
```



conclusion-majority of the restaurant falls in dinning category

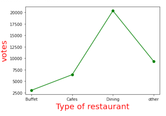
In [23]:
```
dataframe.head()
```

Out[23]:

	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 [25]:
```
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)
```

Out[27]:
```
Text(0, 0.5, 'votes')
```



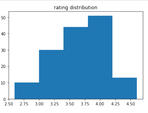
conclusion - dinning restaurant has received maximum votes

In [29]:
```
dataframe.head()
```

Out[29]:

	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 [29]:
```
plt.hist(dataframe['rate'],bins = 5)
plt.title("rating distribution")
plt.show()
```



conclusion - the majority restaurants received rating from 3.5 to 4

In [ ]:
```

```

Average order spending by couples

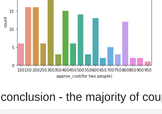
In [31]:
```
dataframe.head()
```

Out[31]:

	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 [32]:
```
couple_data=dataframe['approx_cost(for two people)']
sns.countplot(x=couple_data)
```

Out[32]:
```
<AxesSubplot:xlabel='approx_cost(for two people)', ylabel='count'>
```



conclusion - the majority of couples preferr restaurants with an approximate cost of 300 rupees

In [ ]:
```

```

which mode recevies maximum rating

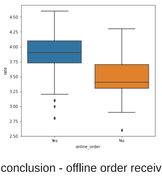
In [34]:
```
dataframe.head()
```

Out[34]:

	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 [35]:
```
plt.figure(figsize = (6,6))
sns.boxplot(x = 'online_order',y= 'rate', data = dataframe)
```

Out[35]:
```
<AxesSubplot:xlabel='online_order', ylabel='rate'>
```



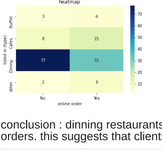
conclusion - offline order received lower rating in comparison to online order

In [36]:
```
dataframe.head()
```

Out[36]:

	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 [39]:
```
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()
```



conclusion : dinning restaurants primarily accept offline orders, whereas cafes primarily receive online orders. this suggests that clients preferes orders in person at restaurants, but prefer online ordering at cafes.

In [ ]:
```

```