

Zomzto Data Analysis Project

Step - 1 (Importing Libraries)

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

Step - 2 (Creat The Data Frame)

```
In [5]: 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 [6]: dataframe

Out[6]:

	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 x 7 columns

Step - 3 (Data Cleaning)

In [7]: #convert of data type of column - rate

```
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 [12]: #checking missing values

```
In [13]: 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
```

In [14]: dataframe.isnull().sum()

Out[14]:

name	0
online_order	0
book_table	0
rate	0
votes	0
approx_cost(for two people)	0
listed_in(type)	0

dtype: int64

Type of Resturant

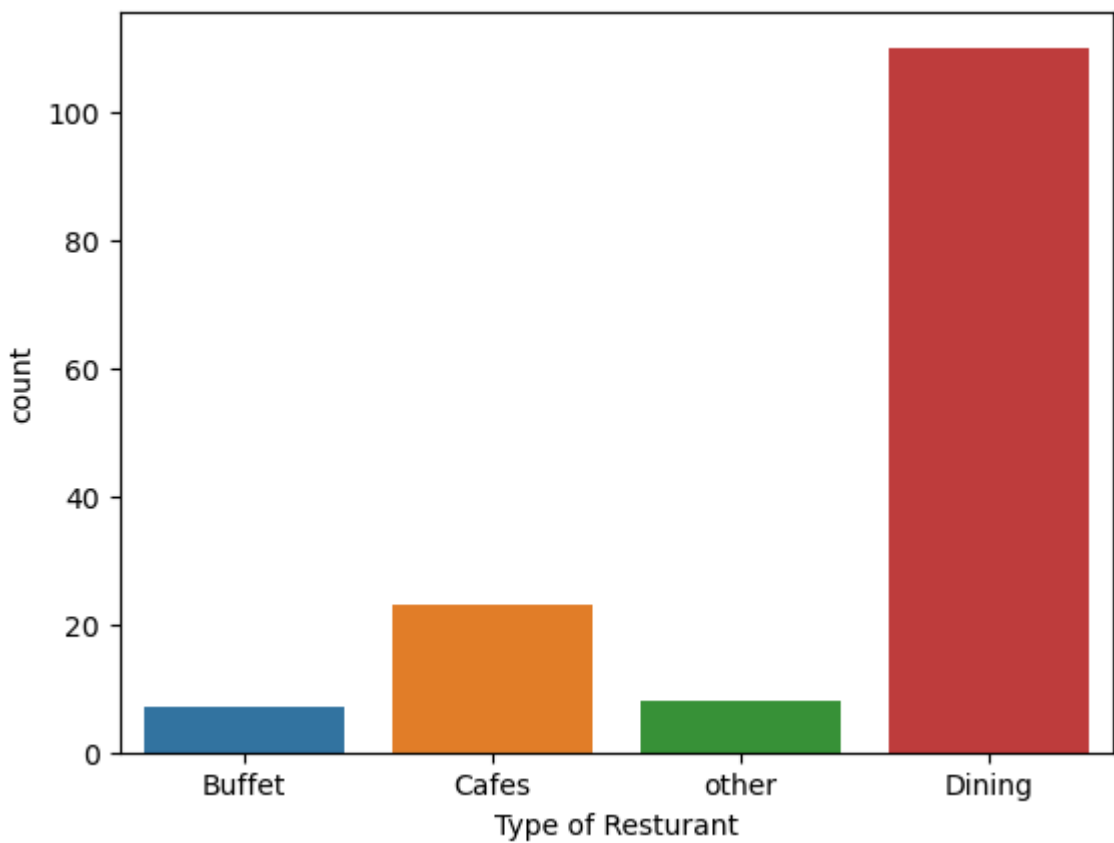
In [15]: dataframe.head()

Out[15]:

	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 [17]: sns.countplot(x=dataframe["listed\_in(type)"])
plt.xlabel("Type of Resturant")

Out[17]: Text(0.5, 0, 'Type of Resturant')



Conclusion - Majarity of the resturant falls in Dining category

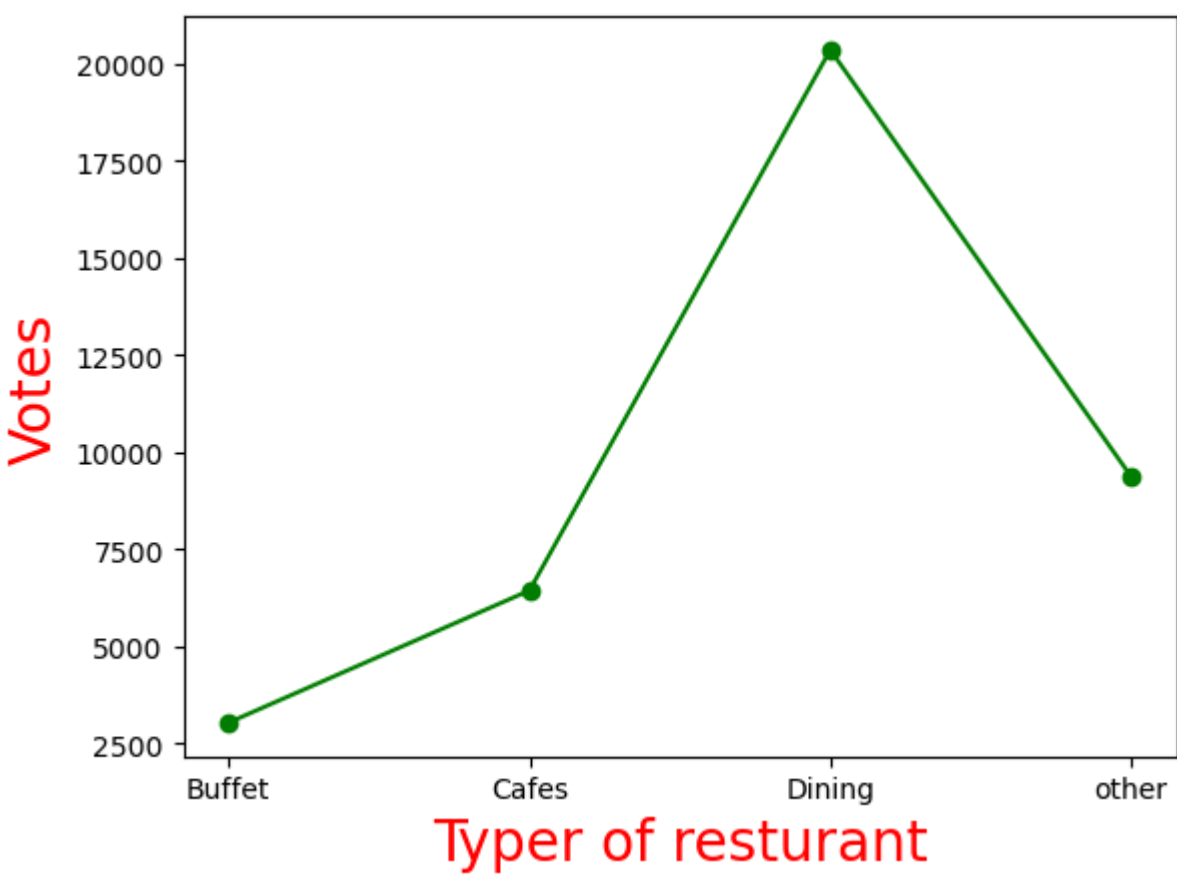
In [19]: dataframe.head()

Out[19]:

	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 [23]: grouped_data = dataframe.groupby("listed_in(type)")[ "votes"].sum()
result = pd.DataFrame({"votes":grouped_data})
plt.plot(result,c="green",marker="o")
plt.xlabel("Typer of resturant",c="red",size=20)
plt.ylabel("Votes",c="red",size=20)
```

Out[23]: Text(0, 0.5, 'Votes')



conclusion - dinning resturants have recieved maximum votes

In [ ]:

