

Zomato data analysis project

Step 1 - Importing Libraries

```
#pandas is used for data manipulation and analysis.  
#numpy is used for numerical operations.  
#matplotlib.pyplot and seaborn are used for data visualization.  
  
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns
```

Step 2 - Create the dataframe

```
#dataframe = pd.read_csv("Zomato data .csv")  
df = pd.read_csv("Zomato data .csv")  
print(df)
```

	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]
```

```
df
```

	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]
```

convert the datatype of column - rate

```
def handleRate(value):  
    value = str(value).split('/')  
    value = value[0];  
    return float (value)  
  
df['rate'] = df['rate'].apply(handleRate)  
print(df.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

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

```

Type of Restaurant

```

df.head()

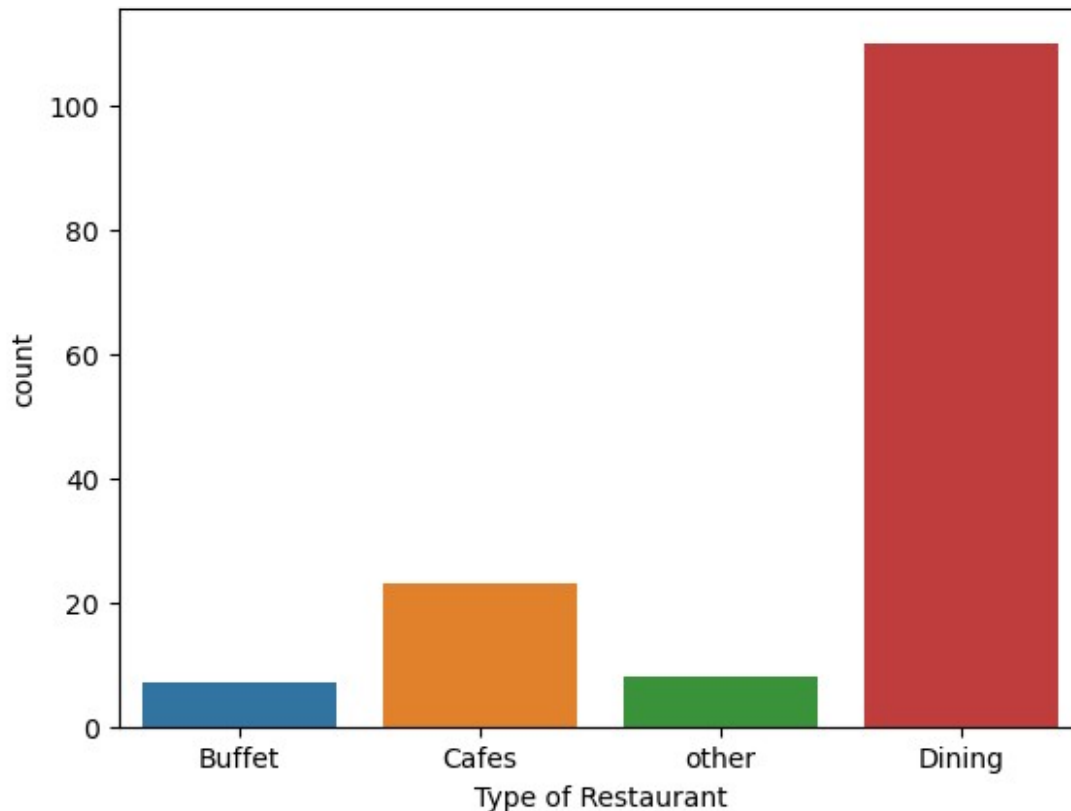
   name online_order book_table rate votes \
0   Jalsa          Yes         Yes  4.1   775
1  Spice Elephant    Yes         No  4.1   787
2  San Churro Cafe   Yes         No  3.8   918
3  Addhuri Udupi    No         No  3.7    88
4  Grand Village     No         No  3.8   166

    approx_cost(for two people) listed_in(type)
0                800          Buffet
1                800          Buffet
2                800          Buffet
3                300          Buffet
4                600          Buffet

sns.countplot(x=df['listed_in(type)'])
plt.xlabel("Type of Restaurant")

Text(0.5, 0, 'Type of Restaurant')

```



conclusion - majority of the Restaurant falls in dining category

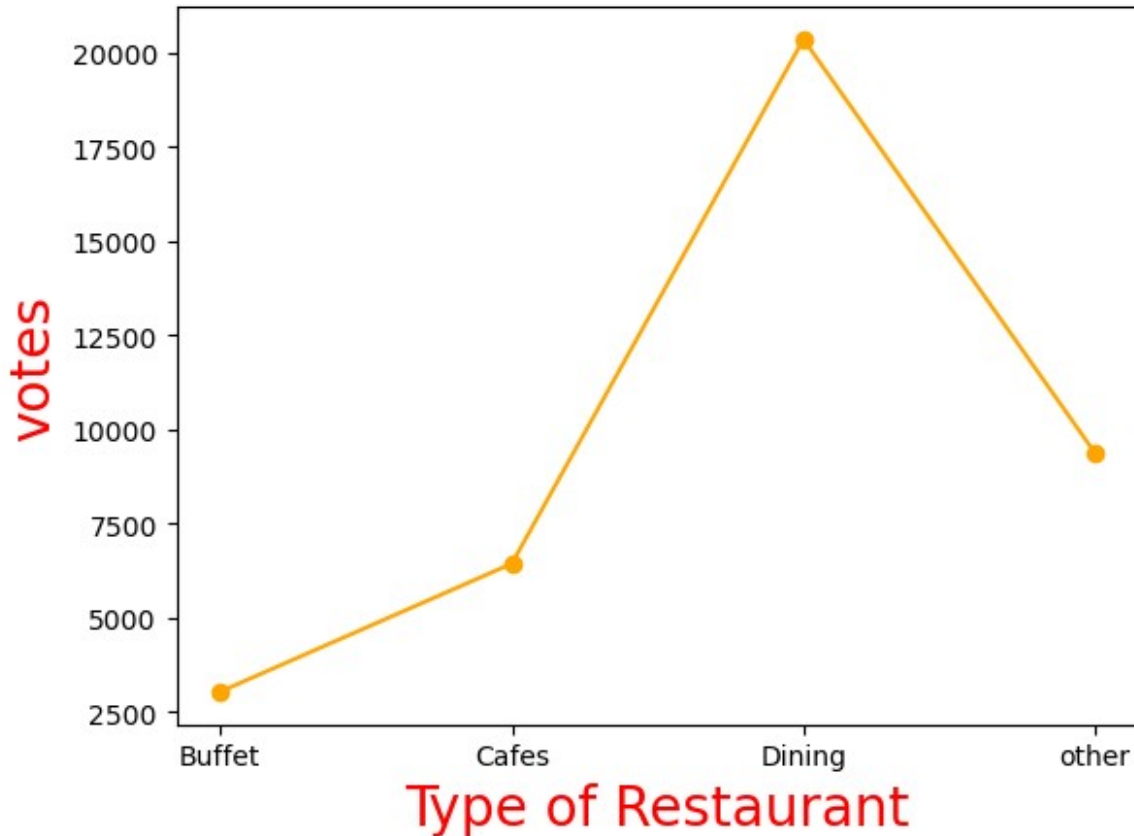
```
df.head()
```

	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 Udipi 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

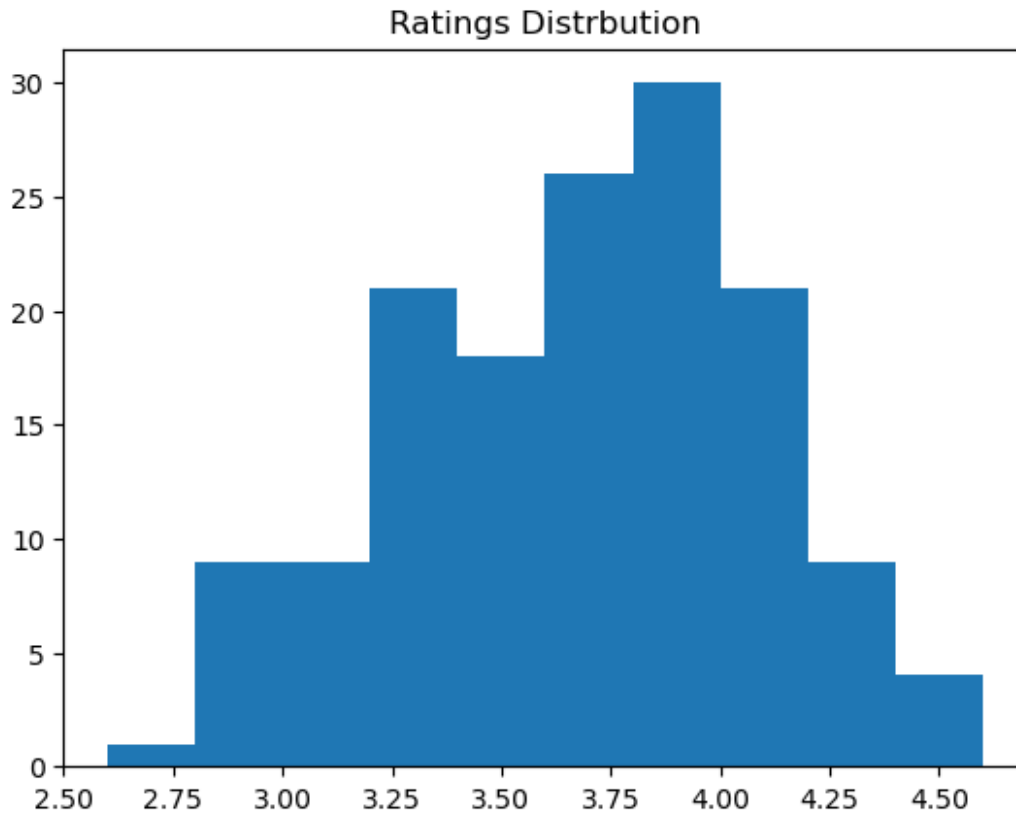
```
grouped_data = df.groupby('listed_in(type)')['votes'].sum()
result=pd.DataFrame({'votes':grouped_data})
plt.plot(result, c="orange", marker="o")
```

```
plt.xlabel("Type of Restaurant", c="red", size=20)
plt.ylabel("votes", c="red", size=20)
Text(0, 0.5, 'votes')
```



conclusion - dining Restaurants has received maximum votes

```
plt.hist(df['rate'], bins=10)
plt.title('Ratings Distrbution')
plt.show()
```



conclusion - the majority Restaurants received ratings from 3.5 to 4

Average order spending by couples

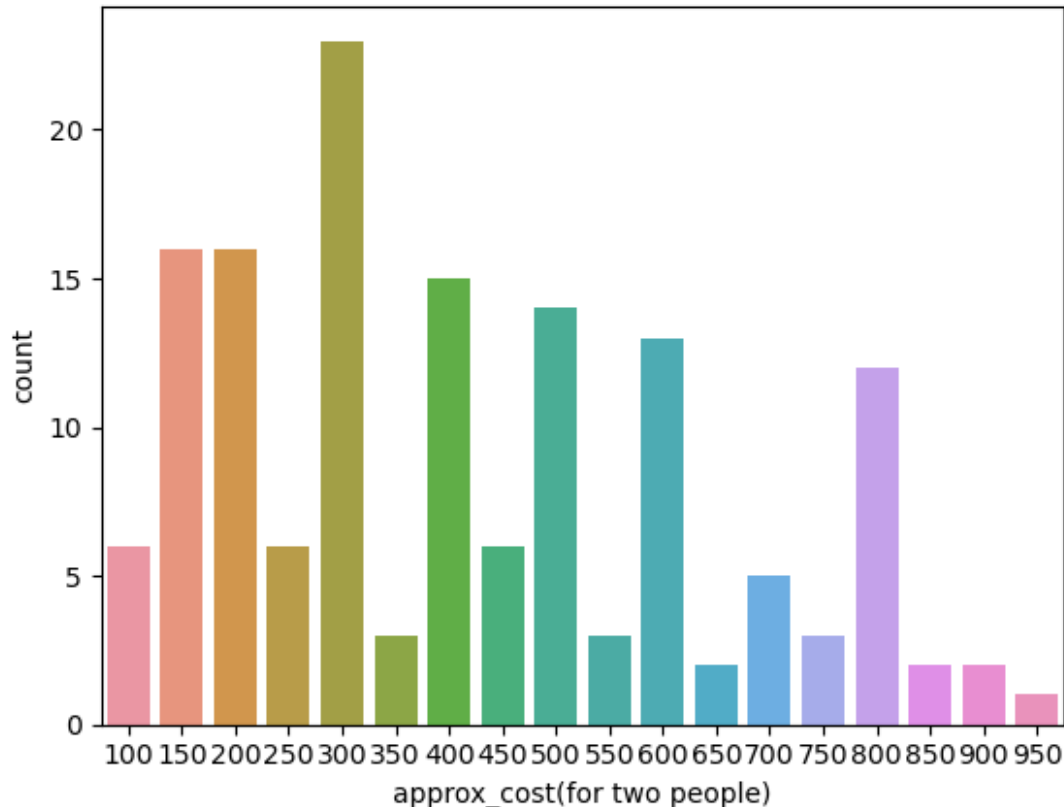
```
df.head()
```

	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 Udipi 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

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

```
<Axes: xlabel='approx_cost(for two people)', ylabel='count'>
```



conclusion - the majority of couples prefers Restaurants with an approximate cost of 300 rupees

```
df.head()
```

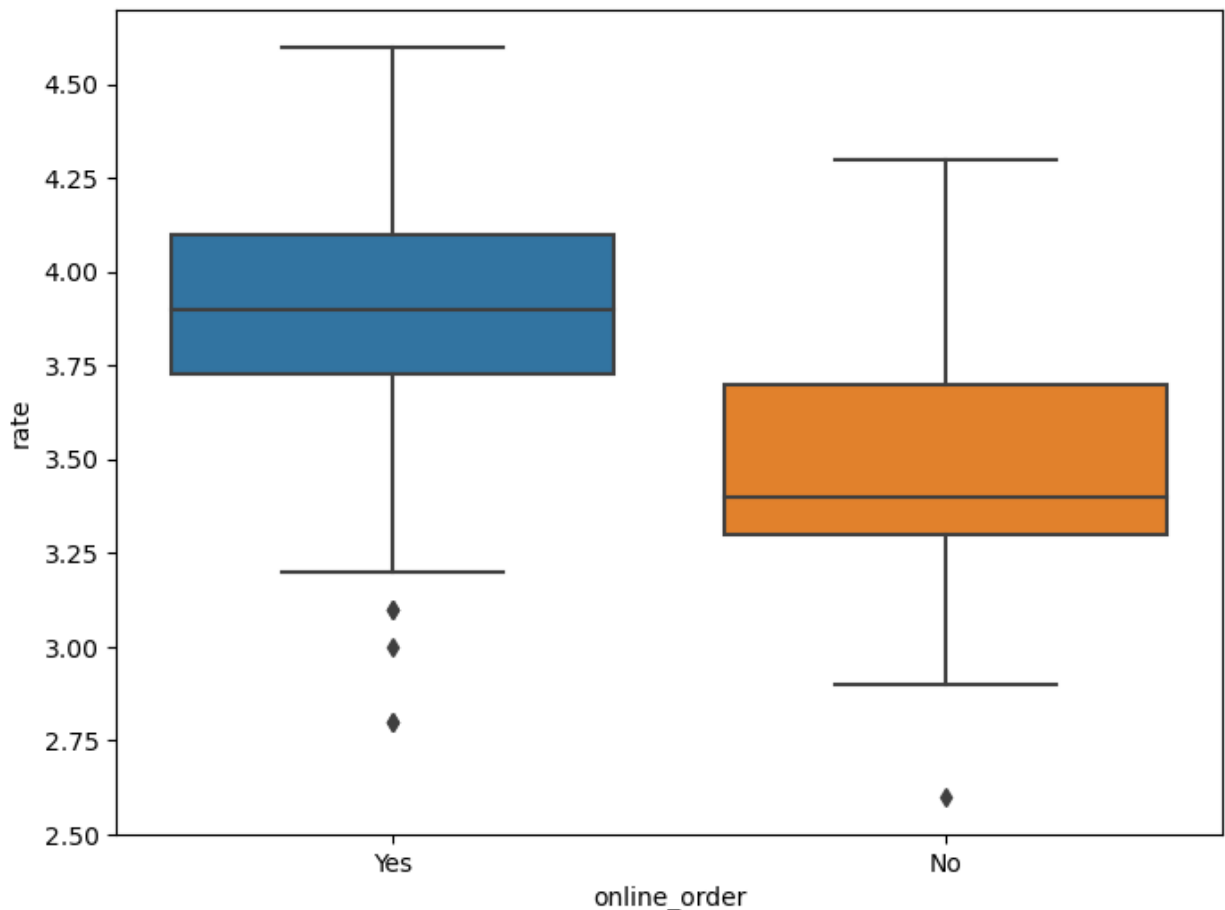
	name	online_order	book_table	rate	votes	\
0	Jalsa	Yes	Yes	4.1	775	
1	Spice Elephant	Yes	No	4.1	787	
2	San Churro Cafe	Yes	No	3.8	918	
3	Addhuri Udipi Bhojana	No	No	3.7	88	
4	Grand Village	No	No	3.8	166	

	approx_cost(for two people)	listed_in(type)
0	800	Buffet
1	800	Buffet

2	800	Buffet
3	300	Buffet
4	600	Buffet

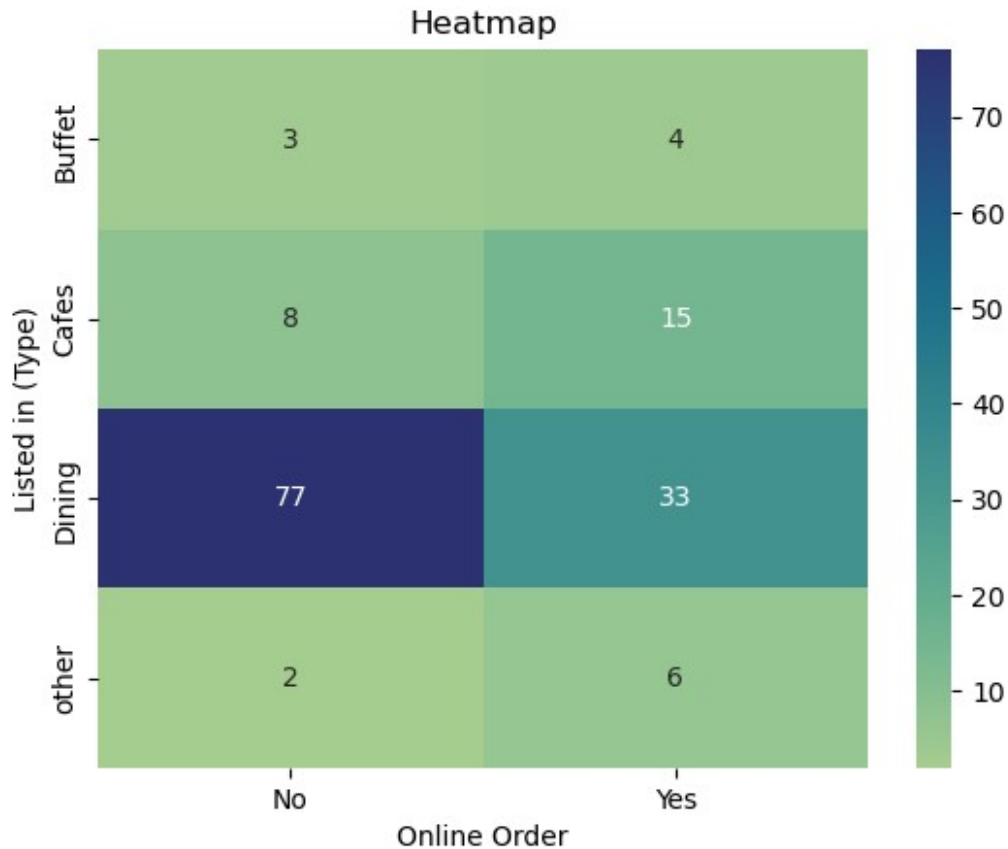
```
plt.figure(figsize=(8,6))
sns.boxplot(x='online_order', y='rate', data = df)
```

```
<Axes: xlabel='online_order', ylabel='rate'>
```



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

```
pivot_table=df.pivot_table(index='listed_in(type)',
columns='online_order', aggfunc='size', fill_value=0)
sns.heatmap(pivot_table,annot=True, fmt='d', cmap="crest")
plt.title("Heatmap")
plt.xlabel("Online Order")
plt.ylabel("Listed in (Type)")
plt.show()
```

CONCLUSION - Dining Restaurants primarily accept offline orders, whereas cafes primarily receive online orders. This suggests that clients prefer to place orders in person at Restaurants, but prefer online ordering at cafes.