Data Analytics Project Using Python

Project Overview

Project Title:
Zomato Data Analysis Using Python

Introduction:
Zomato is one of the largest food delivery platforms globally, with millions of monthly transacting customers and a vast network of restaurant partners. With the rise of data-driven decision-making, analyzing Zomato's data can provide valuable insights into customer behavior, restaurant performance, and overall market trends.

Objective:
The primary goal of this project is to conduct an in-depth Exploratory Data Analysis (EDA) on Zomato's dataset using Python. The analysis aims to uncover key insights, trends, and patterns that can help optimize business strategies, improve customer experience, and drive growth for both Zomato and its restaurant partners.

Dataset Description:
The dataset used in this project contains information on various restaurants listed on Zomato. It includes details such as restaurant names, online orders, book table , ratings, votes, cost for two people, and other relevant features. The data offers a comprehensive view of the restaurant landscape on the Zomato platform.

Key Questions Explored:

- What type of restaurant do the majority of customers order from?
- How many votes has each type of restaurant received from customers?
- What are the ratings that the majority of restaurants have received?
- Zomato has observed that most couples order most of their food online. What is their average spending on each order?
- Which mode(online or offline) has received the maximum rating?
- Which type of restaurants received more offline orders, so that Zomato can provide those customers with some good offers ?

Tools and Techniques:

- **Python Libraries:** Pandas, NumPy, Matplotlib, Seaborn

- **Data Cleaning:** Converting data types.
- **Data Visualization:** Creating insightful visualizations to illustrate key findings, such as bar charts, scatter plots, and heatmaps.
- **Exploratory Analysis:** Investigating correlations, distributions, and trends within the data.

Expected Outcomes: The analysis is expected to provide actionable insights into customer preferences, popular restaurants, cost distribution, and rating trends. These findings can help Zomato enhance its platform's performance, improve customer satisfaction, and support restaurants in making data-driven decisions.

Importing Libraries

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

Loading the dataset

In [20]:
```
data = pd.read_csv(r"C:\Users\ykuma\Downloads\Zomato data .csv")
data
```

Out[20]:

	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 [24]:
```
# Converting the data type of the column 'Rate' to float and remove the denominator

def rating(value):
    value = str(value).split('/')
    value = value[0];
    return float(value)

data['rate'] = data['rate'].apply(rating)
data.head()
```

Out[24]:

	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

Summary of the dataframe

In [25]:
```
data.info()
```

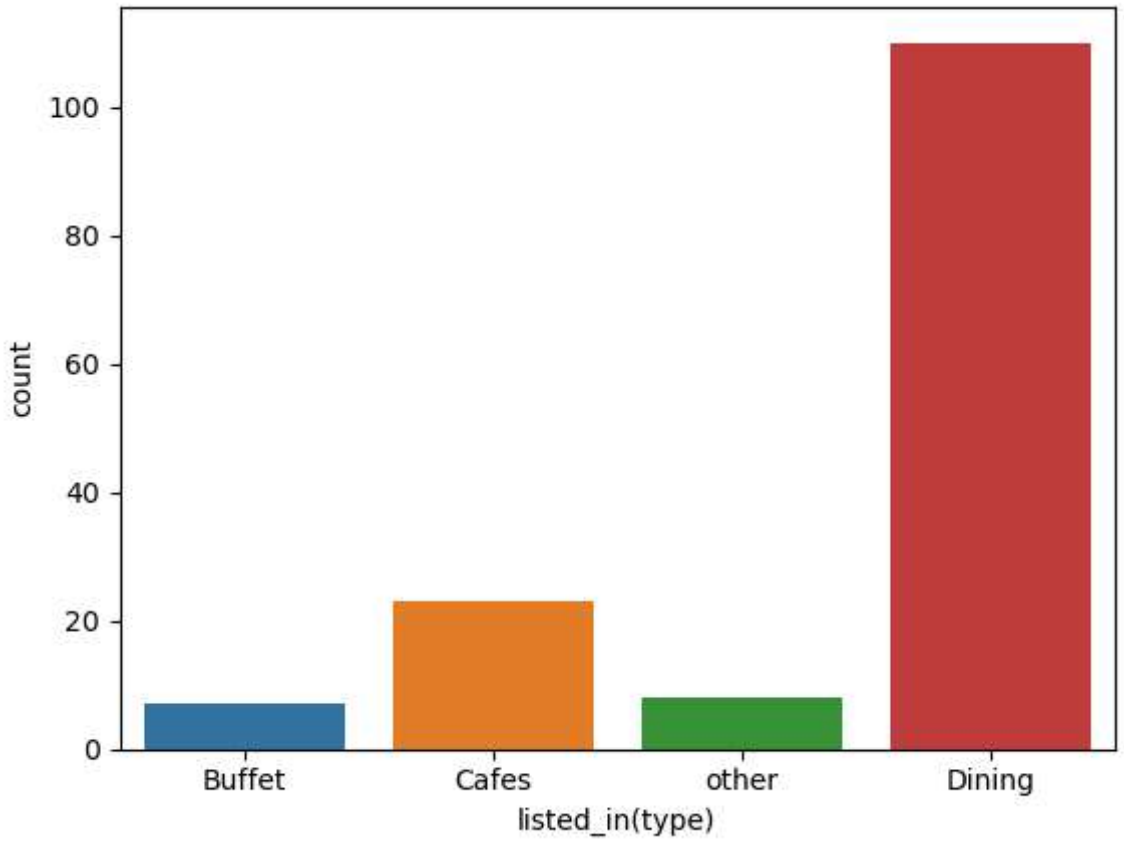
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148 entries, 0 to 147
Data columns (total 7 columns):
 #   Column                       Non-Null Count  Dtype  
---  ------                       --------------  -----  
 0   name                         148 non-null    object 
 1   online_order                 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
```

```
NOTE:- There is no null values in the dataframe
```

Business Problem

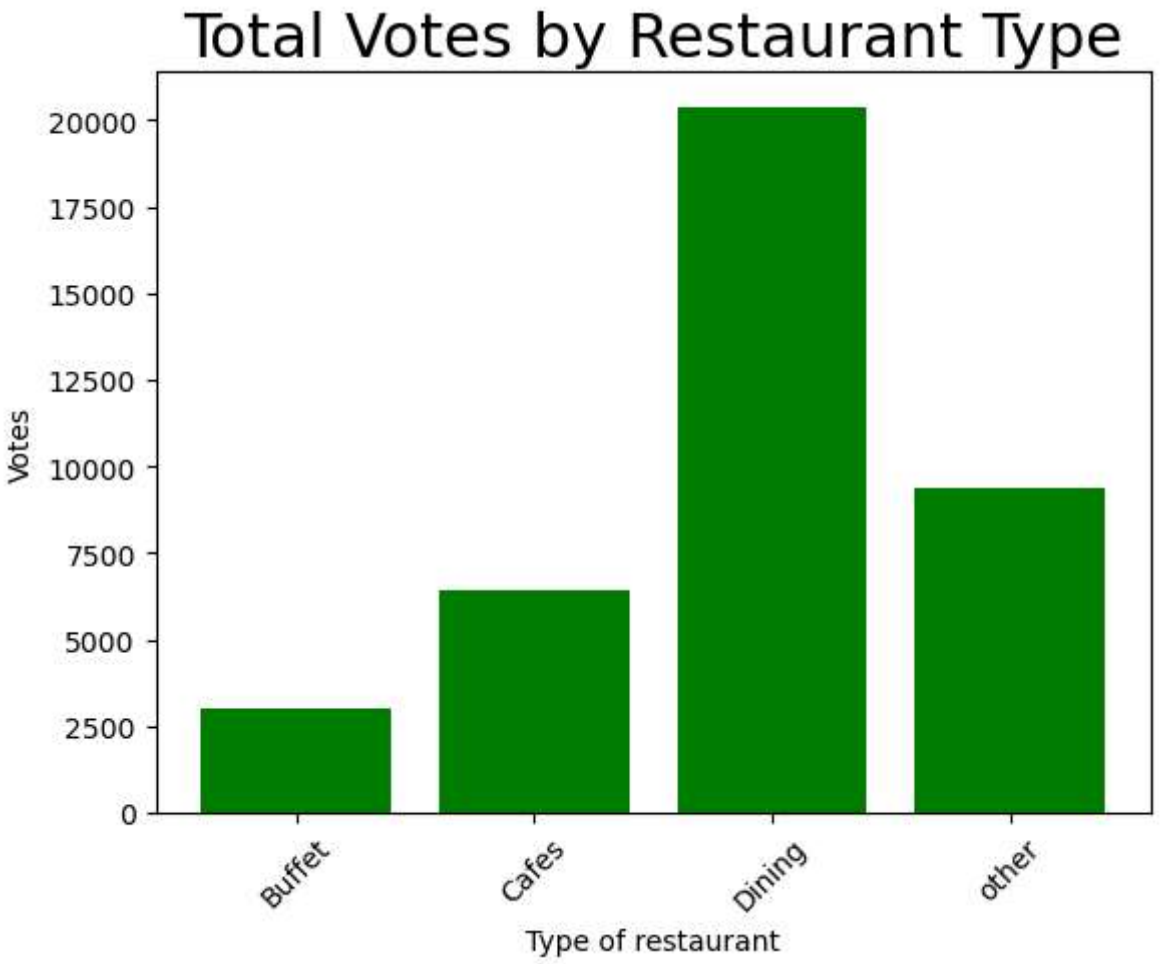
In [27]:
```
# 1. What type of restaurant do the majority of customers order from ?

sns.countplot(data= data, x='listed_in(type)')
plt.xlabel('Type of Restaurant')
```

Out[27]: <Axes: xlabel='listed_in(type)', ylabel='count'>



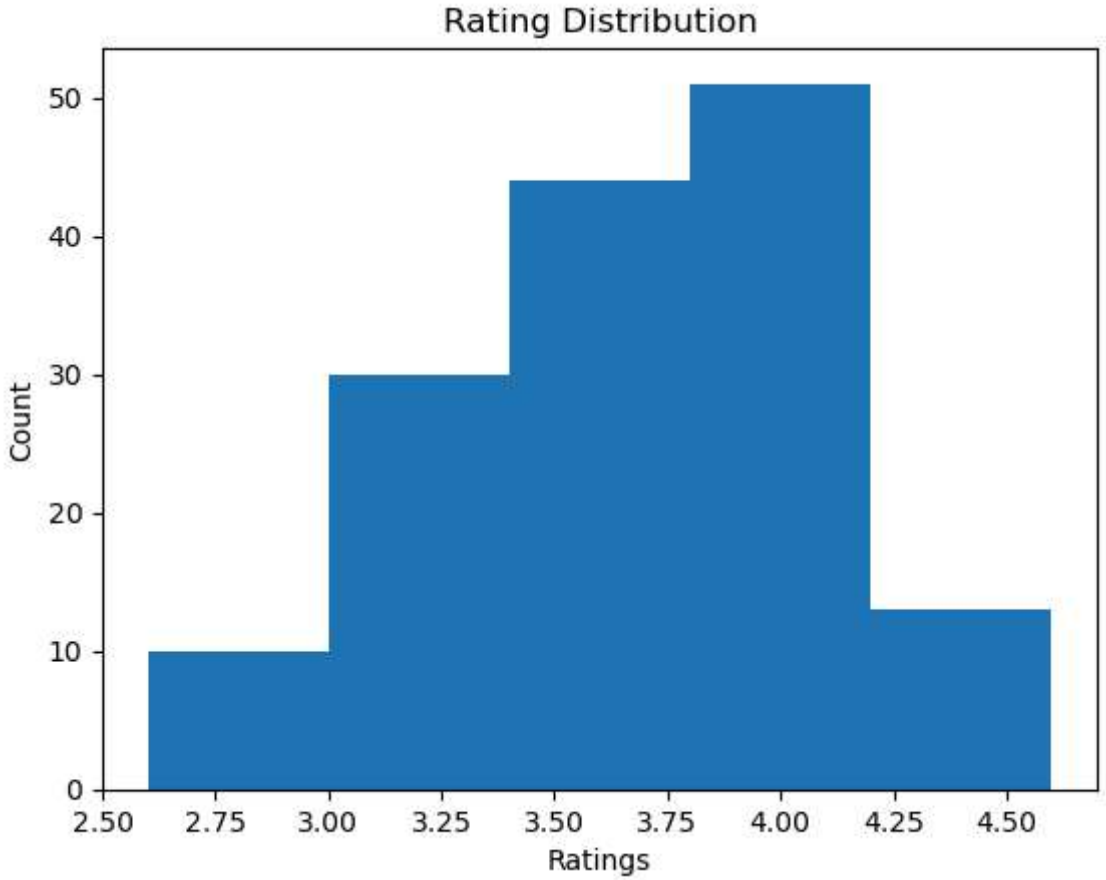
Conclusion:- Dining Restaurants are preferred by a larger number of individuals.

In [86]:
```
# 2. How many votes has each type of restaurant received from customers ?
grouped_data = data.groupby('listed_in(type)')['votes'].sum().reset_index()
#plotting graph
plt.bar(grouped_data['listed_in(type)'], grouped_data['votes'], color='green')
plt.xlabel('Type of restaurant')
plt.ylabel('Votes')
plt.title('Total Votes by Restaurant Type', size=22)
plt.xticks(rotation=45)
plt.show()
```



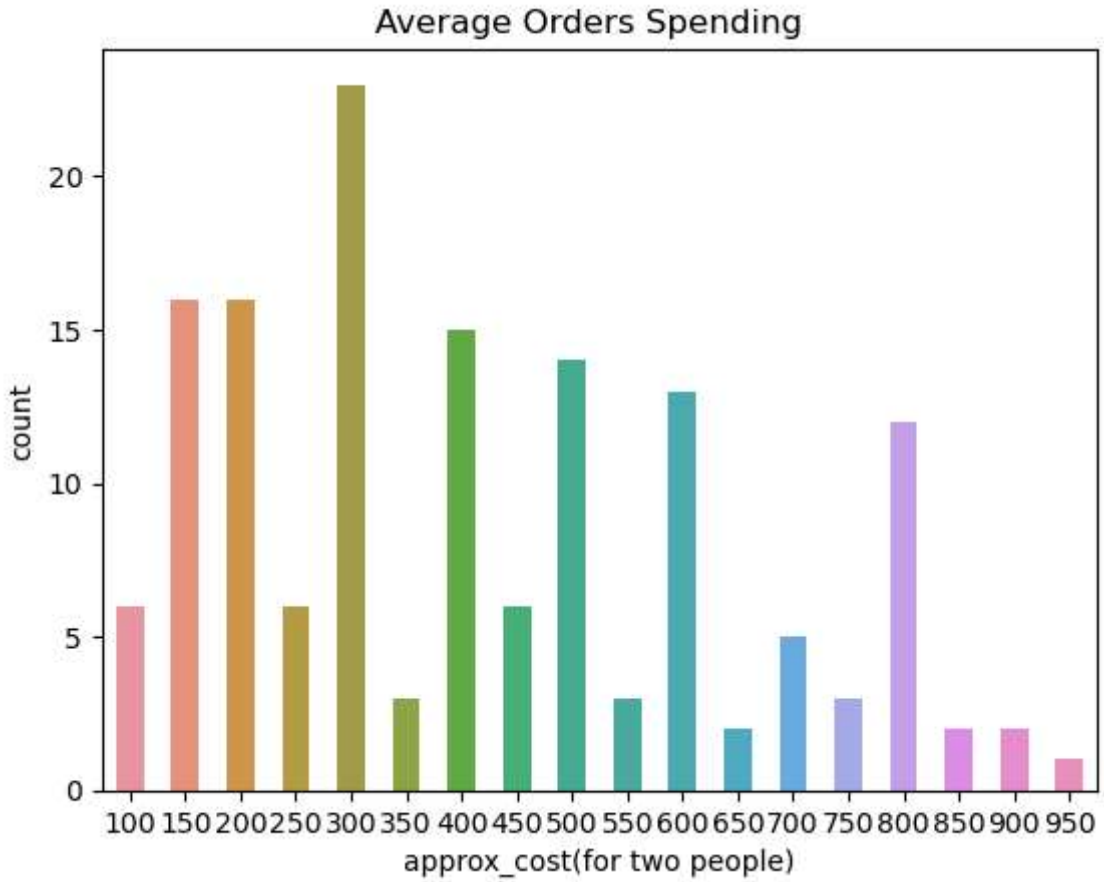
Conclusion:- Dining Restaurants has received maximum votes.

In [49]:
```
# 3. what are the ratings that the majority of restaurants have received ?
plt.hist(data['rate'],bins=5)
plt.title('Rating Distribution')
plt.xlabel('Ratings')
plt.ylabel('Count')
plt.show()
```



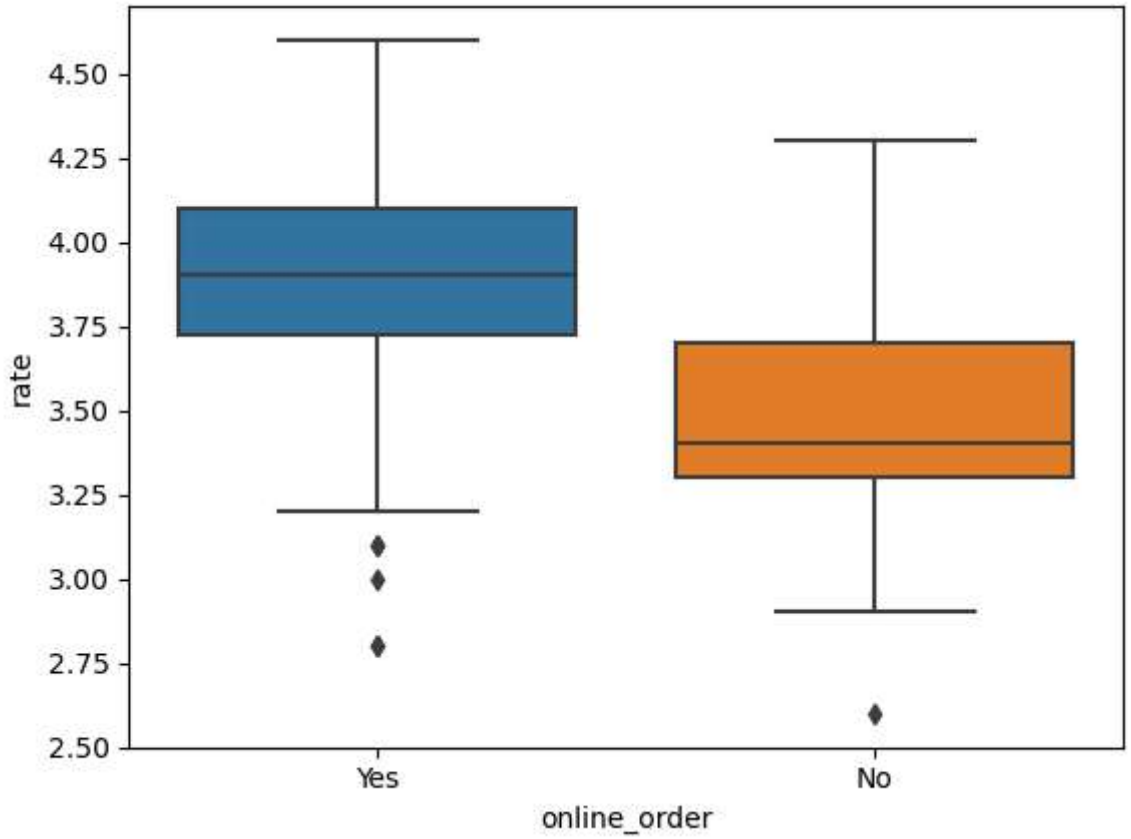
Conclusion:- The majority restaurants received ratings from 3.5 to 4.0

In [69]:
```
# 4. Zomato has observed that most of couples order most of their food online. what is their average spending on each
couple_data= data['approx_cost(for two people)']
plt.title('Average Orders Spending  ')
plt.xlabel('Cost')
sns.countplot(x=couple_data, width=0.5)
plt.show()
```



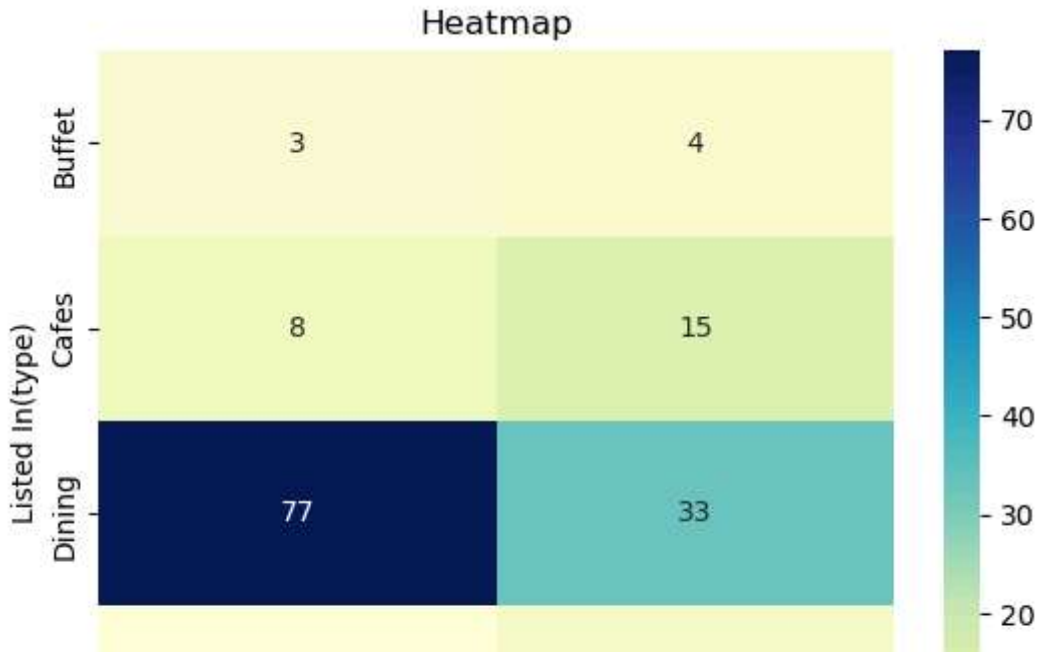
Conclusion:- The majority of couples prefer restaurants with an approximate cost of 300 rupees.

In [70]:
```
# 5. Which mode has received the maximum rating ?
sns.boxplot(data=data, x='online_order', y= 'rate')
plt.show()
```



Conclusion:- Offline orders receive lower rating as compare to onine orders.

In [89]:
```
# 6. Which type of restaurant received more offline orders, so that Zomato can provide those customers with some good
pivot_table= data.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:- Dining Restaurants primarily accept offline orders, whereas cafes primarilty receive online orders.This suggests that clients prefer to place orders in person at restaurants, but prefer online ordering at cafes.

In [ ]: