Capstone Project :- 2 :- (Related To Recommendation System And NLP)

In [ ]:

Capstone Project Name :- Zomato Recommendation System

In [ ]:

Created By :- Jayashri Pacharane, Date :- 15TH Feb.2025

In [ ]:

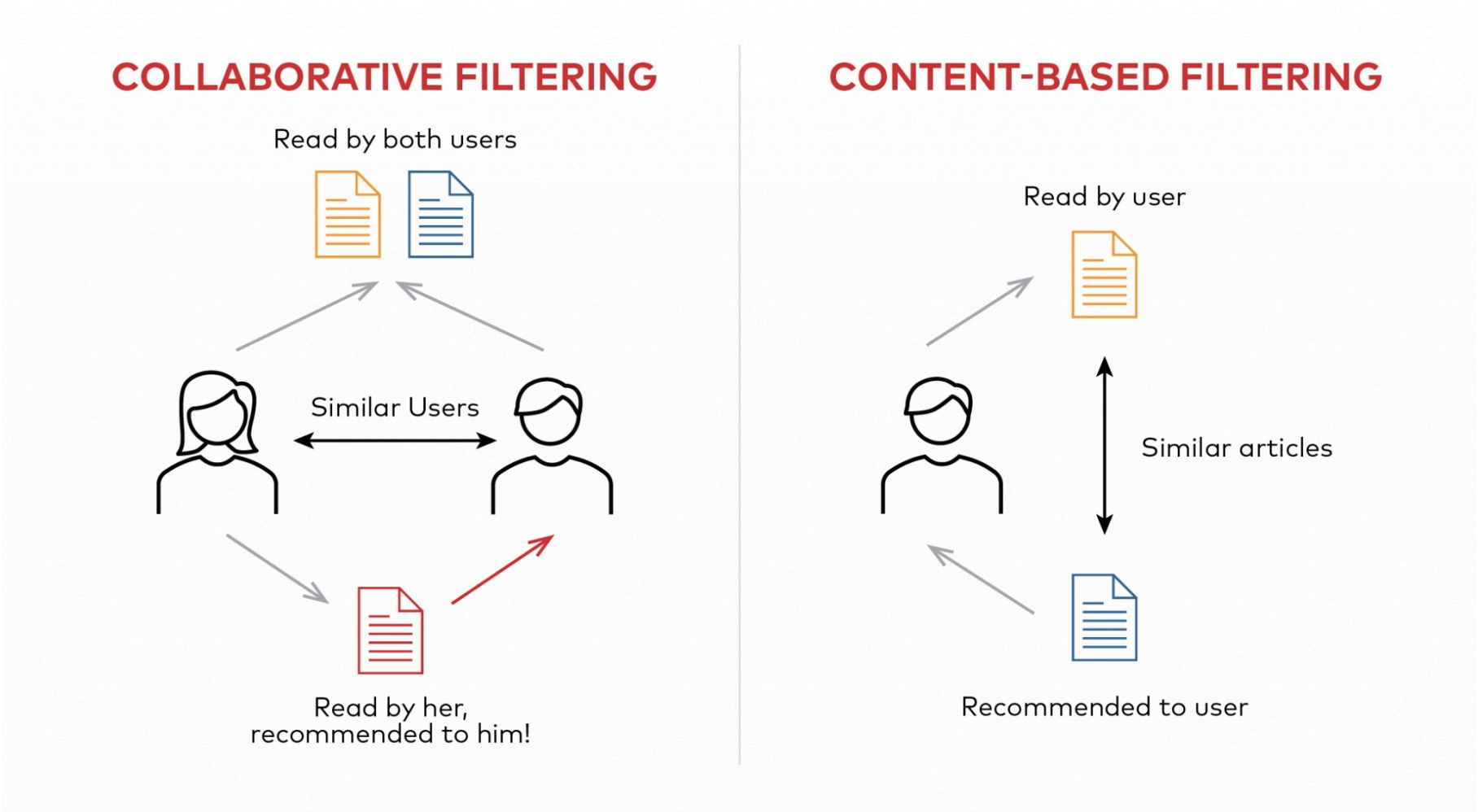
In [1]:
```
from IPython import display
display.Image("C:\\Users\\DELL\\Downloads\\zomato_image.jpg")
```

Out[1]:

In [ ]:

In [2]:
```
from IPython import display
display.Image("C:\\Users\\DELL\\Desktop\\python\\Collaborative-vs-Content-based-filtering.jpg")
```

Out[2]:



In [ ]:

What is Recommendation System?

The rapid growth of data collection has led to a new era of information. Data is being used to create more efficient systems and this is where Recommendation Systems come into play. Recommendation Systems are a type of information filtering systems as they improve the quality of search results and provides items that are more relevant to the search item or are realted to the search history of the user. They are active

information filtering systems which personalize the information coming to a user based on his interests, relevance of the information etc. Recommender systems are used widely for recommending movies, articles, restaurants, places to visit, items to buy etc.

Content Based Filtering- They suggest similar items based on a particular item. This system uses item metadata, such as genre, director, description, actors, etc. for movies, to make these recommendations.

Collaborative Filtering- This system matches persons with similar interests and provides recommendations based on this matching. Collaborative filters do not require item metadata like its content-based counterparts.

In [ ]:

Aim Of The Project:-

The aim is to create a content based recommender system in which when user will write a restaurant name, Recommender system will look at the reviews of other restaurants, and System will recommend us other restaurants with similar reviews and sort them from the highest rated.

In [ ]:

Importing the libraries and Dataset :-

In [3]:
```
#Importing Libraries
import numpy as np
import pandas as pd
import seaborn as sb
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LogisticRegression
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.metrics import r2_score
import warnings
warnings.filterwarnings('always')
warnings.filterwarnings('ignore')
```

```
import re
from nltk.corpus import stopwords
from sklearn.metrics.pairwise import linear_kernel
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
```

In [ ]:

loading the data from csv file to a Pandas DataFrame :-

In [4]:
```
# loading the data from csv file to a Pandas DataFrame

zomato_data = pd.read_csv('zomato.csv')
```

In [5]:
```
zomato_data.head()
```

Out[5]:

	url	address	name	online_order	book_table	rate	votes	phone	loca
0	https://www.zomato.com/bangalore/jalsa-banasha...	942, 21st Main Road, 2nd Stage, Banashankari, ...	Jalsa	Yes	Yes	4.1/5	775	080 42297555\r\n+91 9743772233	Banasha
1	https://www.zomato.com/bangalore/spice-elephan...	2nd Floor, 80 Feet Road, Near Big Bazaar, 6th ...	Spice Elephant	Yes	No	4.1/5	787	080 41714161	Banasha
2	https://www.zomato.com/SanchurroBangalore?cont...	1112, Next to KIMS Medical College, 17th Cross...	San Churro Cafe	Yes	No	3.8/5	918	+91 9663487993	Banasha
3	https://www.zomato.com/bangalore/addhuri-udupi...	1st Floor, Annakuteera, 3rd Stage, Banashankar...	Addhuri Udupi Bhojana	No	No	3.7/5	88	+91 9620009302	Banasha
4	https://www.zomato.com/bangalore/grand-village...	10, 3rd Floor, Lakshmi Associates, Gandhi Baza...	Grand Village	No	No	3.8/5	166	+91 8026612447\r\n+91 9901210005	Basavana

In [ ]:

Check the number of rows and columns :-

In [6]:
```
# number of rows and columns

print("Shape of the dataset : ",zomato_data.shape)
```

```
Shape of the dataset :  (51717, 17)
```

In [ ]:

Getting some informations about the dataset :-

In [7]:
```
# Getting some informations about the dataset

zomato_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51717 entries, 0 to 51716
Data columns (total 17 columns):
 #   Column                       Non-Null Count  Dtype 
---  ------                       --------------  ----- 
 0   url                          51717 non-null  object
 1   address                      51717 non-null  object
 2   name                         51717 non-null  object
 3   online_order                 51717 non-null  object
 4   book_table                   51717 non-null  object
 5   rate                         43942 non-null  object
 6   votes                        51717 non-null  int64 
 7   phone                        50509 non-null  object
 8   location                     51696 non-null  object
 9   rest_type                    51490 non-null  object
 10  dish_liked                   23639 non-null  object
 11  cuisines                     51672 non-null  object
 12  approx_cost(for two people)  51371 non-null  object
 13  reviews_list                 51717 non-null  object
 14  menu_item                    51717 non-null  object
 15  listed_in(type)              51717 non-null  object
 16  listed_in(city)              51717 non-null  object
dtypes: int64(1), object(16)
memory usage: 6.7+ MB
```

In [ ]:

Check the duplicate records :-

In [8]:
```
# Check the duplicate records

zomato_data.duplicated().sum()
```

Out[8]: 0

In [9]:
```
zomato_data.drop_duplicates(inplace=True)
```

In [ ]:

Check the all columns :-

In [10]:
```
# Check the all columns

zomato_data.columns
```

Out[10]:
```
Index(['url', 'address', 'name', 'online_order', 'book_table', 'rate', 'votes',
       'phone', 'location', 'rest_type', 'dish_liked', 'cuisines',
       'approx_cost(for two people)', 'reviews_list', 'menu_item',
       'listed_in(type)', 'listed_in(city)'],
      dtype='object')
```

In [ ]:
```

```

Statistical Measures of the dataset :-

In [11]:
```
# Statistical Measures of the dataset

pd.options.display.float_format = '{:.2f}'.format
zomato_data.describe(include=["int64","float64"]).T
```

Out[11]:

	count	mean	std	min	25%	50%	75%	max
votes	51717.00	283.70	803.84	0.00	7.00	41.00	198.00	16832.00

In [ ]:
```

```

In [12]:
```
zomato_data.describe(include=["object"]).T
```

Out[12]:

	count	unique	top	freq
url	51717	51717	https://www.zomato.com/bangalore/jalsa-banasha...	1
address	51717	11495	Delivery Only	128
name	51717	8792	Cafe Coffee Day	96
online_order	51717	2	Yes	30444
book_table	51717	2	No	45268
rate	43942	64	NEW	2208
phone	50509	14926	080 43334321	216
location	51696	93	BTM	5124
rest_type	51490	93	Quick Bites	19132
dish_liked	23639	5271	Biryani	182
cuisines	51672	2723	North Indian	2913
approx_cost(for two people)	51371	70	300	7576
reviews_list	51717	22513	[]	7595
menu_item	51717	9098	[]	39617
listed_in(type)	51717	7	Delivery	25942
listed_in(city)	51717	30	BTM	3279

In [ ]:

```
```

Data Cleaning and Feature Engineering:-

In [13]:

```
#Deleting Unnnecessary Columns
zomato=zomato_data.drop(['url','dish_liked','phone'],axis=1) #Dropping the column "dish_liked", "phone", "url" and saving the
```

In [ ]:

In [14]:
```
#Removing the Duplicates
zomato.duplicated().sum()
zomato.drop_duplicates(inplace=True)
```

In [ ]:

In [15]:
```
#Remove the NaN values from the dataset
zomato.isnull().sum()
zomato.dropna(how='any',inplace=True)
zomato.info() #.info() function is used to get a concise summary of the dataframe
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 43499 entries, 0 to 51716
Data columns (total 14 columns):
 #   Column                       Non-Null Count  Dtype 
---  ------                       --------------  ----- 
 0   address                      43499 non-null  object
 1   name                         43499 non-null  object
 2   online_order                 43499 non-null  object
 3   book_table                   43499 non-null  object
 4   rate                         43499 non-null  object
 5   votes                        43499 non-null  int64 
 6   location                     43499 non-null  object
 7   rest_type                    43499 non-null  object
 8   cuisines                     43499 non-null  object
 9   approx_cost(for two people)  43499 non-null  object
 10  reviews_list                 43499 non-null  object
 11  menu_item                    43499 non-null  object
 12  listed_in(type)              43499 non-null  object
 13  listed_in(city)              43499 non-null  object
dtypes: int64(1), object(13)
memory usage: 5.0+ MB
```

In [ ]:

In [16]:
```
#Reading Column Names
zomato.columns
```

Out[16]:
```
Index(['address', 'name', 'online_order', 'book_table', 'rate', 'votes',
       'location', 'rest_type', 'cuisines', 'approx_cost(for two people)',
       'reviews_list', 'menu_item', 'listed_in(type)', 'listed_in(city)'],
      dtype='object')
```

In [ ]:

Changing the column names:-

In [17]:
```
#Changing the column names
zomato = zomato.rename(columns={'approx_cost(for two people)':'cost','listed_in(type)':'type',
                                'listed_in(city)':'city'})
zomato.columns
```

Out[17]:
```
Index(['address', 'name', 'online_order', 'book_table', 'rate', 'votes',
       'location', 'rest_type', 'cuisines', 'cost', 'reviews_list',
       'menu_item', 'type', 'city'],
      dtype='object')
```

In [ ]:

In [18]:
```
#Some Transformations
zomato['cost'] = zomato['cost'].astype(str) #Changing the cost to string
zomato['cost'] = zomato['cost'].apply(lambda x: x.replace(',','.')) #Using lambda function to replace ',' from cost
zomato['cost'] = zomato['cost'].astype(float) # Changing the cost to Float
zomato.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 43499 entries, 0 to 51716
Data columns (total 14 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   address       43499 non-null  object 
 1   name          43499 non-null  object 
 2   online_order  43499 non-null  object 
 3   book_table    43499 non-null  object 
 4   rate          43499 non-null  object 
 5   votes         43499 non-null  int64  
 6   location      43499 non-null  object 
 7   rest_type     43499 non-null  object 
 8   cuisines      43499 non-null  object 
 9   cost          43499 non-null  float64
 10  reviews_list  43499 non-null  object 
 11  menu_item     43499 non-null  object 
 12  type          43499 non-null  object 
 13  city          43499 non-null  object 
dtypes: float64(1), int64(1), object(12)
memory usage: 5.0+ MB
```

In [ ]:

Checking the rate of Dataset:-

In [19]:
```
#Reading Rate of dataset
zomato['rate'].unique()
```

Out[19]:
```
array(['4.1/5', '3.8/5', '3.7/5', '3.6/5', '4.6/5', '4.0/5', '4.2/5',
       '3.9/5', '3.1/5', '3.0/5', '3.2/5', '3.3/5', '2.8/5', '4.4/5',
       '4.3/5', 'NEW', '2.9/5', '3.5/5', '2.6/5', '3.8 /5', '3.4/5',
       '4.5/5', '2.5/5', '2.7/5', '4.7/5', '2.4/5', '2.2/5', '2.3/5',
       '3.4 /5', '-', '3.6 /5', '4.8/5', '3.9 /5', '4.2 /5', '4.0 /5',
       '4.1 /5', '3.7 /5', '3.1 /5', '2.9 /5', '3.3 /5', '2.8 /5',
       '3.5 /5', '2.7 /5', '2.5 /5', '3.2 /5', '2.6 /5', '4.5 /5',
       '4.3 /5', '4.4 /5', '4.9/5', '2.1/5', '2.0/5', '1.8/5', '4.6 /5',
       '4.9 /5', '3.0 /5', '4.8 /5', '2.3 /5', '4.7 /5', '2.4 /5',
       '2.1 /5', '2.2 /5', '2.0 /5', '1.8 /5'], dtype=object)
```

In [ ]:

In [20]:
```
#Removing '/5' from Rates

import numpy as np

# Filter out 'NEW' and '-' from the 'rate' column
zomato = zomato.loc[zomato.rate != 'NEW']
zomato = zomato.loc[zomato.rate != '-'].reset_index(drop=True)

# Define a lambda function to remove '/5' and ensure it's applied correctly
remove_slash = lambda x: x.replace('/5', '') if isinstance(x, str) else x

# Apply the function and convert the 'rate' column to float after cleaning
zomato['rate'] = zomato['rate'].apply(remove_slash).str.strip()

# Convert the 'rate' column to float, forcing errors to NaN (if any) and drop those rows if necessary
zomato['rate'] = pd.to_numeric(zomato['rate'], errors='coerce')

# Display the first few rows of the 'rate' column
zomato['rate'].head()
```

Out[20]:
```
0   4.10
1   4.10
2   3.80
3   3.70
4   3.80
Name: rate, dtype: float64
```

In [ ]:

Adjust the column names:-

In [21]:
```
# Adjust the column names
zomato.name = zomato.name.apply(lambda x:x.title())
zomato.online_order.replace(('Yes','No'),(True, False),inplace=True)
zomato.book_table.replace(('Yes','No'),(True, False),inplace=True)
zomato.cost.unique()
```

Out[21]:
```
array([800.  , 300.  , 600.  , 700.  , 550.  , 500.  , 450.  , 650.  ,
       400.  , 900.  , 200.  , 750.  , 150.  , 850.  , 100.  ,   1.2 ,
       350.  , 250.  , 950.  ,   1.  ,   1.5 ,   1.3 , 199.  ,   1.1 ,
         1.6 , 230.  , 130.  ,   1.7 ,   1.35,   2.2 ,   1.4 ,   2.  ,
         1.8 ,   1.9 , 180.  , 330.  ,   2.5 ,   2.1 ,   3.  ,   2.8 ,
         3.4 ,  50.  ,  40.  ,   1.25,   3.5 ,   4.  ,   2.4 ,   2.6 ,
         1.45,  70.  ,   3.2 , 240.  ,   6.  ,   1.05,   2.3 ,   4.1 ,
       120.  ,   5.  ,   3.7 ,   1.65,   2.7 ,   4.5 ,  80.  ])
```

In [ ]:
```
```

In [22]:
```
zomato.head()
```

Out[22]:

	address	name	online_order	book_table	rate	votes	location	rest_type	cuisines	cost	reviews_list	menu_item	typ
0	942, 21st Main Road, 2nd Stage, Banashankari, ...	Jalsa	True	True	4.10	775	Banashankari	Casual Dining	North Indian, Mughlai, Chinese	800.00	[('Rated 4.0', 'RATED\n A beautiful place to ...	[]	Buff
1	2nd Floor, 80 Feet Road, Near Big Bazaar, 6th ...	Spice Elephant	True	False	4.10	787	Banashankari	Casual Dining	Chinese, North Indian, Thai	800.00	[('Rated 4.0', 'RATED\n Had been here for din...	[]	Buff
2	1112, Next to KIMS Medical College, 17th Cross...	San Churro Cafe	True	False	3.80	918	Banashankari	Cafe, Casual Dining	Cafe, Mexican, Italian	800.00	[('Rated 3.0', "RATED\n Ambience is not that ...	[]	Buff
3	1st Floor, Annakuteera, 3rd Stage, Banashankar...	Addhuri Udupi Bhojana	False	False	3.70	88	Banashankari	Quick Bites	South Indian, North Indian	300.00	[('Rated 4.0', "RATED\n Great food and proper...	[]	Buff
4	10, 3rd Floor, Lakshmi Associates, Gandhi Baza...	Grand Village	False	False	3.80	166	Basavanagudi	Casual Dining	North Indian, Rajasthani	600.00	[('Rated 4.0', 'RATED\n Very good restaurant ...	[]	Buff

In [ ]:

Checking City :-

In [23]:
```
zomato['city'].unique()
```

Out[23]:
```
array(['Banashankari', 'Bannerghatta Road', 'Basavanagudi', 'Bellandur',
       'Brigade Road', 'Brookefield', 'BTM', 'Church Street',
       'Electronic City', 'Frazer Town', 'HSR', 'Indiranagar',
       'Jayanagar', 'JP Nagar', 'Kalyan Nagar', 'Kammanahalli',
       'Koramangala 4th Block', 'Koramangala 5th Block',
       'Koramangala 6th Block', 'Koramangala 7th Block', 'Lavelle Road',
       'Malleshwaram', 'Marathahalli', 'MG Road', 'New BEL Road',
       'Old Airport Road', 'Rajajinagar', 'Residency Road',
       'Sarjapur Road', 'Whitefield'], dtype=object)
```

In [ ]:
```

```

In [24]:
```
zomato.head()
```

Out[24]:

	address	name	online_order	book_table	rate	votes	location	rest_type	cuisines	cost	reviews_list	menu_item	typ
0	942, 21st Main Road, 2nd Stage, Banashankari, ...	Jalsa	True	True	4.10	775	Banashankari	Casual Dining	North Indian, Mughlai, Chinese	800.00	[('Rated 4.0', 'RATED\n A beautiful place to ...	[]	Buff
1	2nd Floor, 80 Feet Road, Near Big Bazaar, 6th ...	Spice Elephant	True	False	4.10	787	Banashankari	Casual Dining	Chinese, North Indian, Thai	800.00	[('Rated 4.0', 'RATED\n Had been here for din...	[]	Buff
2	1112, Next to KIMS Medical College, 17th Cross...	San Churro Cafe	True	False	3.80	918	Banashankari	Cafe, Casual Dining	Cafe, Mexican, Italian	800.00	[('Rated 3.0', "RATED\n Ambience is not that ...	[]	Buff
3	1st Floor, Annakuteera, 3rd Stage, Banashankar...	Addhuri Udupi Bhojana	False	False	3.70	88	Banashankari	Quick Bites	South Indian, North Indian	300.00	[('Rated 4.0', "RATED\n Great food and proper...	[]	Buff
4	10, 3rd Floor, Lakshmi Associates, Gandhi Baza...	Grand Village	False	False	3.80	166	Basavanagudi	Casual Dining	North Indian, Rajasthani	600.00	[('Rated 4.0', 'RATED\n Very good restaurant ...	[]	Buff

In [ ]:

In [25]:
```
## Checking Null values
zomato.isnull().sum()
```

Out[25]:
```
address         0
name            0
online_order    0
book_table      0
rate            0
votes           0
location        0
rest_type       0
cuisines        0
cost            0
reviews_list    0
menu_item       0
type            0
city            0
dtype: int64
```

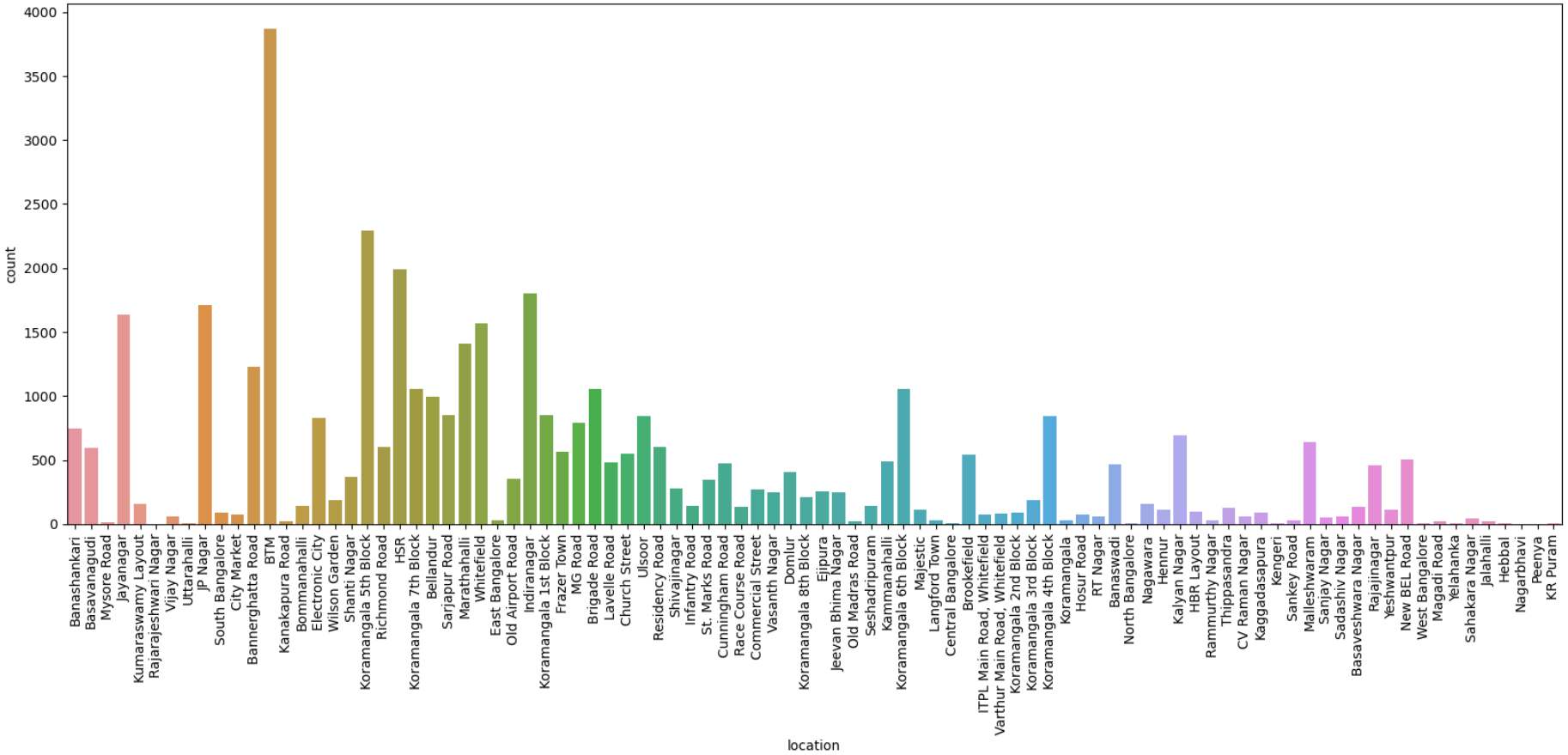
In [ ]:
```

```

Visualization:-

Visualization Of Location:-

In [26]:
```
import matplotlib.pyplot as plt
import seaborn as sns

# Display the value counts of the 'location' column
print(zomato.location.value_counts())

# Create a count plot for the 'location' column
plt.figure(figsize=(20, 7))
plt.xticks(rotation=90)  # Rotate the x-axis labels for better visibility
sns.countplot(x='location', data=zomato)

# Show the plot
plt.show()
```

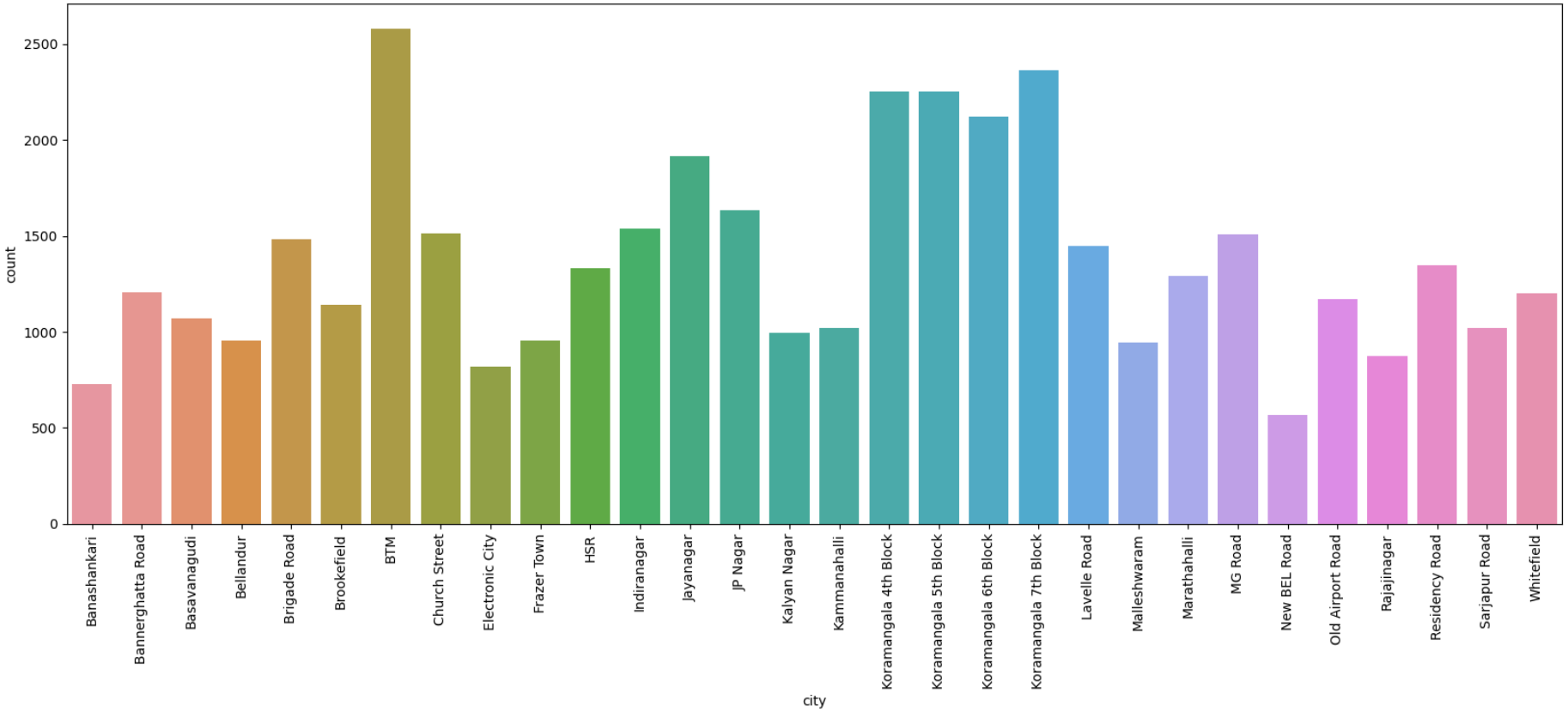
```
location
BTM                      3873
Koramangala 5th Block    2296
HSR                      1993
Indiranagar              1800
JP Nagar                 1710
                         ...
Yelahanka                   4
West Bangalore              3
Rajarajeshwari Nagar        2
Nagarbhavi                  1
Peenya                      1
Name: count, Length: 92, dtype: int64
```



In [ ]:

Visualization Of City:-

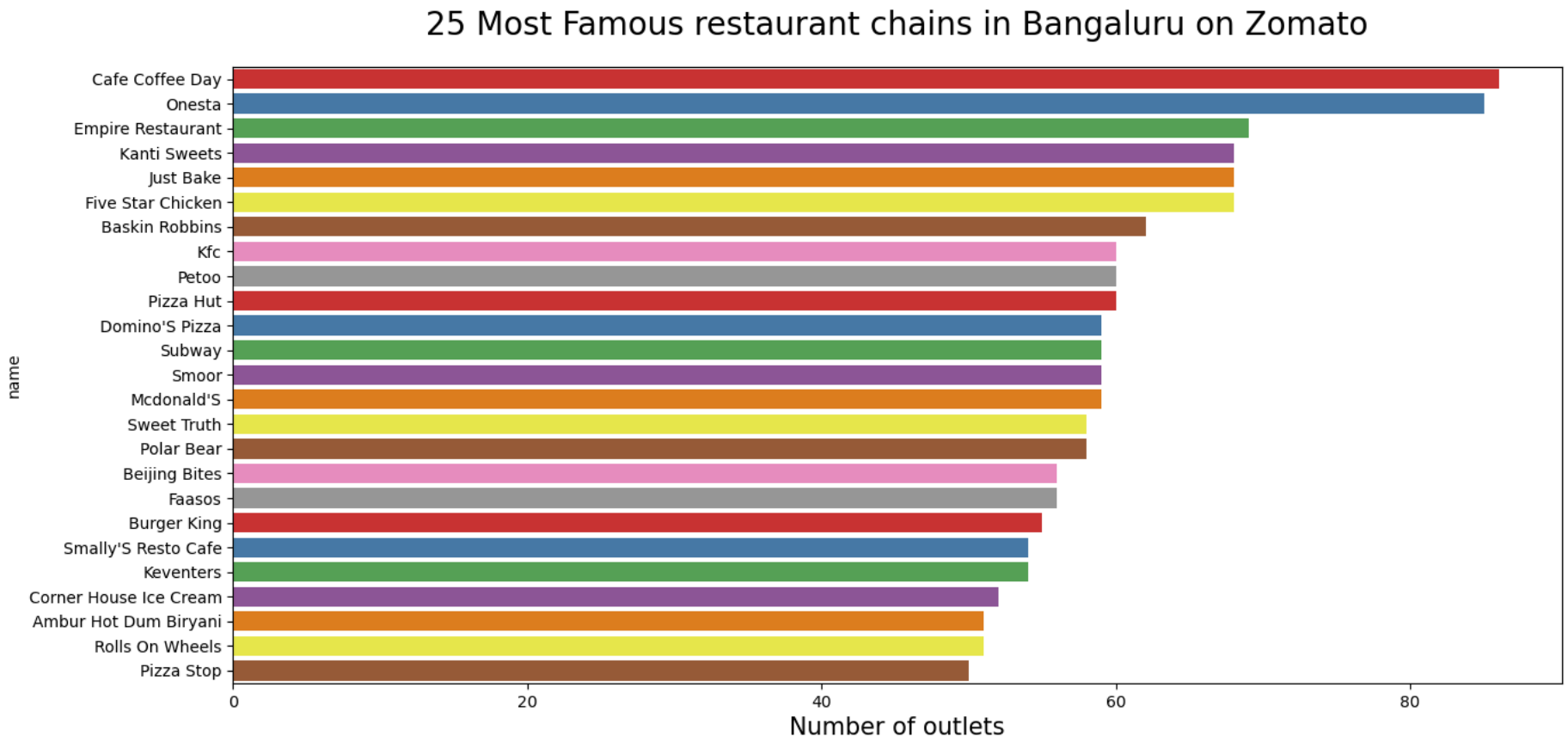
In [27]:
```
import matplotlib.pyplot as plt
import seaborn as sns

# Create a count plot for the 'listed_in(city)' column
plt.figure(figsize=(20, 7))
plt.xticks(rotation=90)  # Rotate the x-axis labels for better visibility
sns.countplot(x='city', data=zomato)

# Show the plot
plt.show()
```



In [ ]:

Visualization Of Famous Restaurant Chains In Bengaluru:-

In [28]:
```
# Famous restaurant chains in Bengaluru
plt.figure(figsize=(15,7))
chains=zomato['name'].value_counts()[:25]
sns.barplot(x=chains,y=chains.index,palette='Set1')
plt.title("25 Most Famous restaurant chains in Bangaluru on Zomato",size=20,pad=20)
plt.xlabel("Number of outlets",size=15);
```

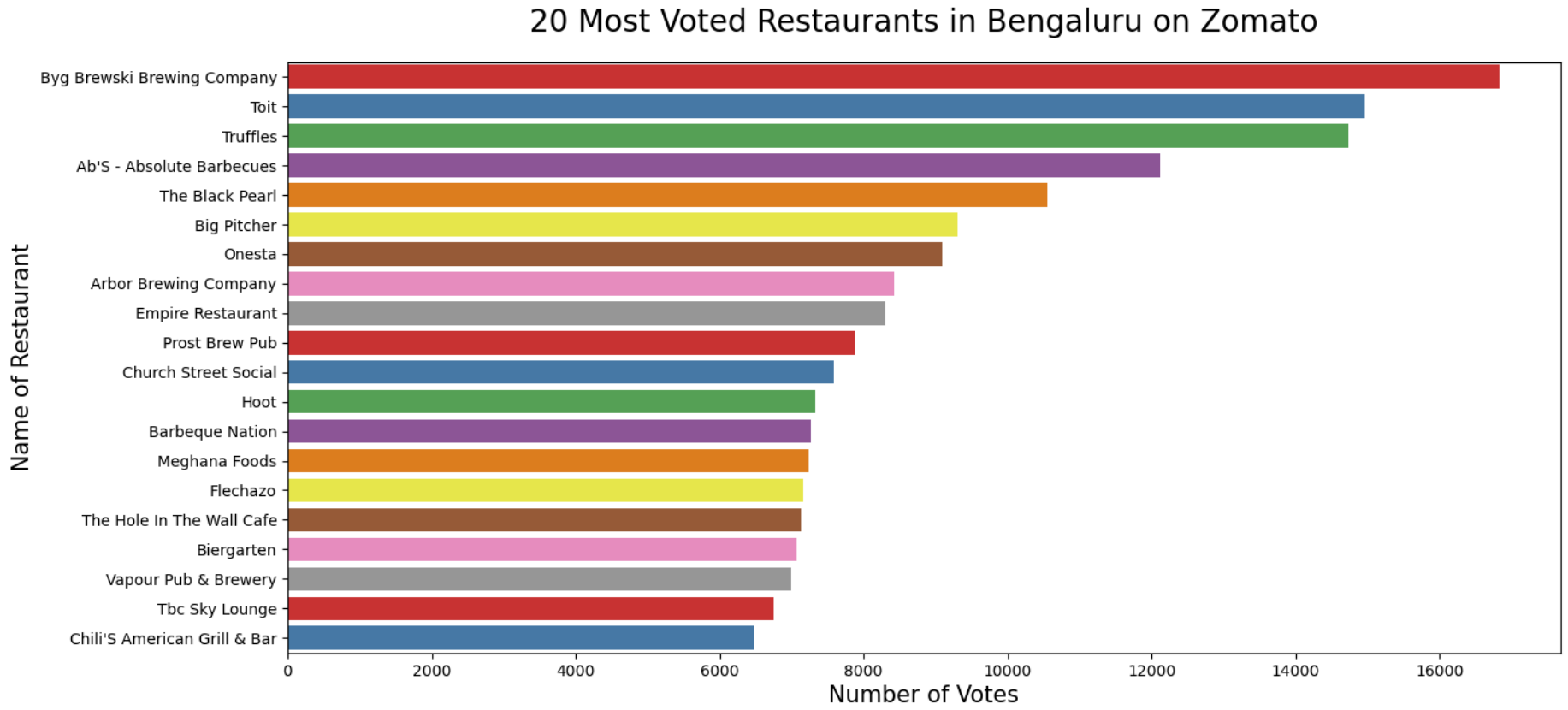


In [ ]:

Visualization Of Highest Voted Restaurant:-

In [29]:
```
import matplotlib.pyplot as plt
import seaborn as sns

# Get the top 20 restaurants by maximum votes
top_20_votes = zomato.groupby('name')['votes'].max().nlargest(20)

# Create the bar plot
plt.figure(figsize=(15,7))
```

```
sns.barplot(x=top_20_votes.values, y=top_20_votes.index, palette='Set1')

# Add titles and labels
plt.title("20 Most Voted Restaurants in Bengaluru on Zomato", size=20, pad=20)
plt.xlabel("Number of Votes", size=15)
plt.ylabel("Name of Restaurant", size=15)

# Show the plot
plt.show()
```

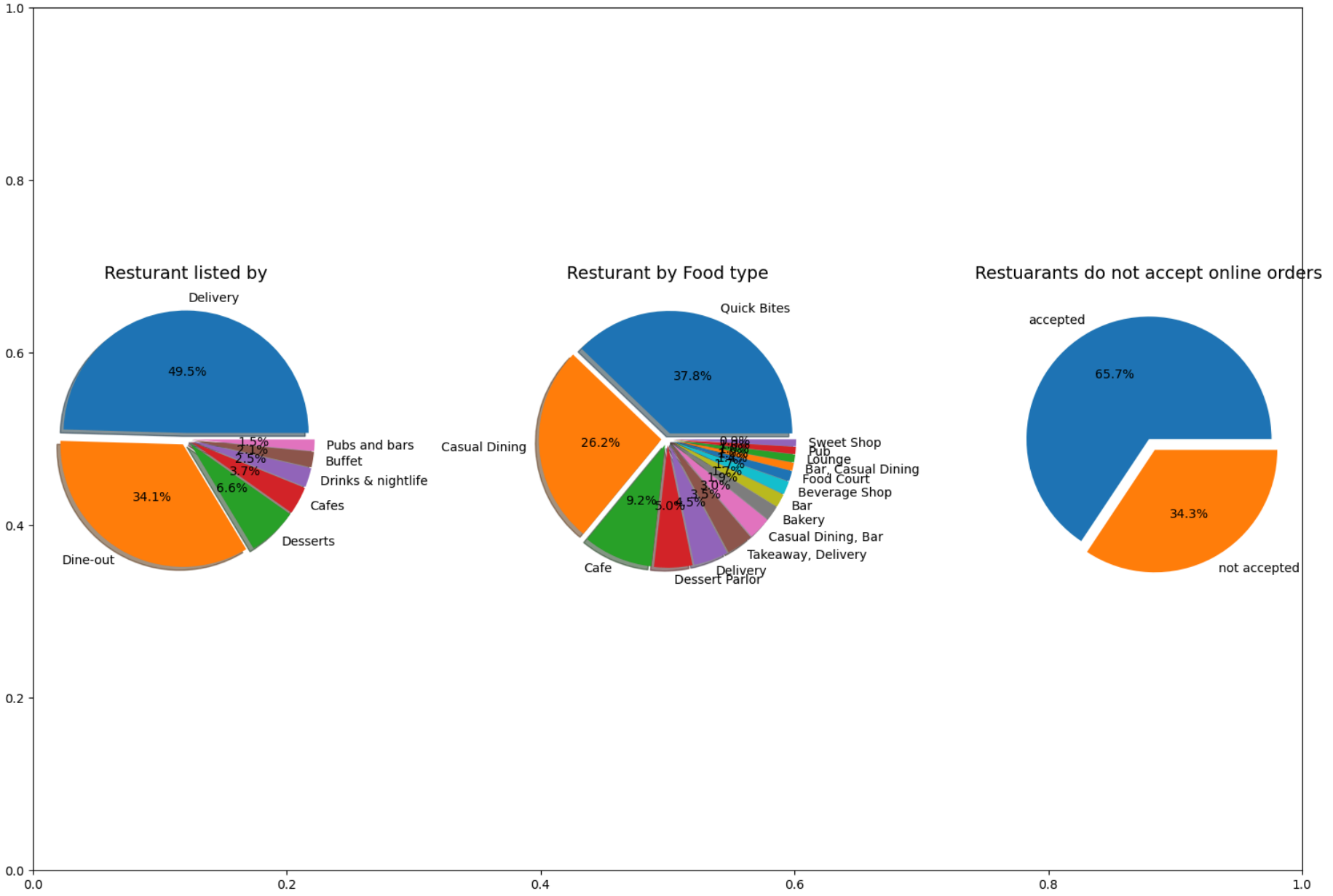


In [ ]:

Pie Chart :-

In [30]:

```
pie, ax = plt.subplots(figsize=[15,10])
plt.subplot(1,3,1)
labels = zomato['type'].value_counts().keys()
plt.pie(x=zomato['type'].value_counts(), autopct="%.1f%%", explode=[0.05]*7, labels=labels, pctdistance=0.5,shadow=True)
plt.title("Resturant listed by", fontsize=14);
pie.savefig("Resturant listed by.png")

plt.subplot(1,3,2)
labels = zomato['rest_type'].value_counts()[:15].keys()
plt.pie(x=zomato['rest_type'].value_counts()[:15], autopct="%.1f%%", explode=[0.05]*15, labels=labels, pctdistance=0.5,shadow=
plt.title("Resturant by Food type", fontsize=14);
pie.savefig("Resturant by Food type.png")

plt.subplot(1,3,3)
x=zomato['online_order'].value_counts()
labels=['accepted','not accepted']
plt.title("Restuarants do not accept online orders", fontsize=14);
plt.pie(x,explode=[0.0,0.1],autopct='%1.1f%%', labels=labels)

plt.tight_layout()
```


Resturant listed by
Delivery
49.5%
Dine-out
34.1%
Desserts
6.6%
Cafes
3.7%
Drinks & nightlife
2.5%
Buffet
2.1%
Pubs and bars
1.5%
Resturant by Food type
Quick Bites
37.8%
Casual Dining
26.2%
Cafe
9.2%
Dessert Parlor
5.0%
Delivery
4.5%
Takeaway, Delivery
3.5%
Casual Dining, Bar
3.0%
Bakery
Bar
Beverage Shop
Food Court
Bar, Casual Dining
Lounge
Pub
Sweet Shop
Restuarants do not accept online orders
accepted
65.7%
not accepted
34.3%
0.0
0.2
0.4
0.6
0.8
1.0

In [ ]:

Content-Based Filtering :-

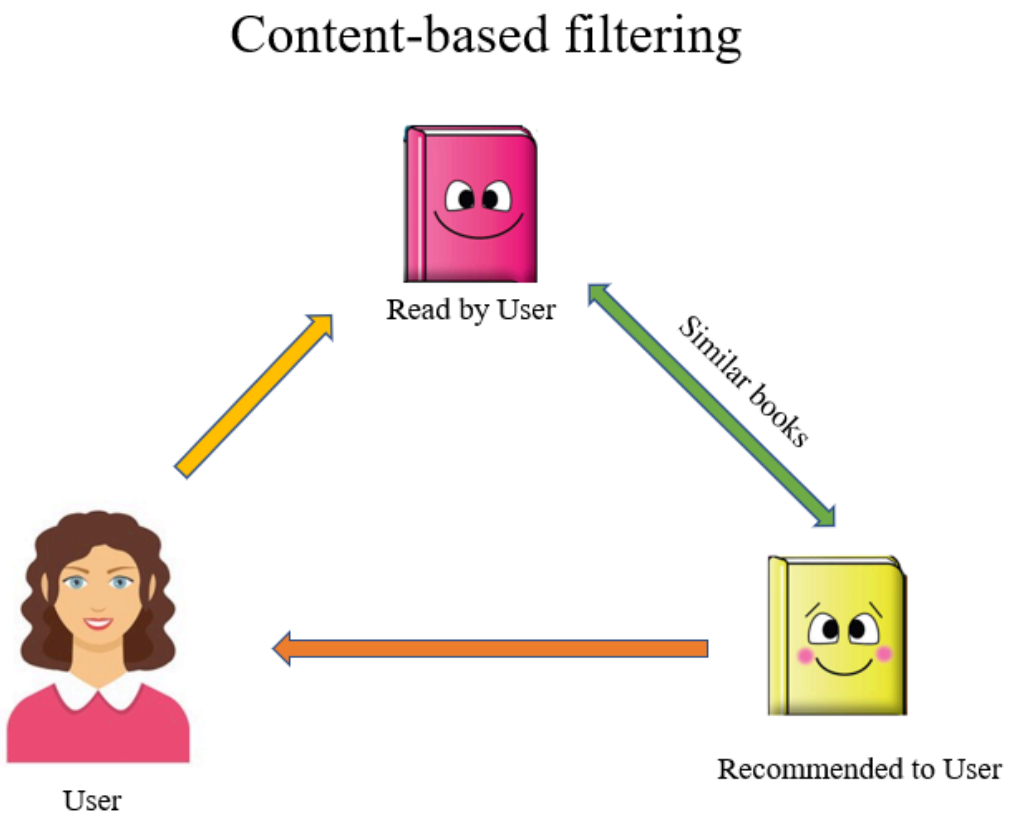
Content Based Filtering:-

They suggest similar items based on a particular item. This system uses item metadata, such as genre, director, description, actors, etc. for movies, to make these recommendations.

In [ ]:

In [31]:
```
from IPython import display
display.Image("C:\\Users\\DELL\\Desktop\\Resume\\Content_Based.png")
```

Out[31]:



In [ ]:

In [32]:
```
## Computing Mean Rating
restaurants = list(zomato['name'].unique())
zomato['Mean Rating'] = 0

for i in range(len(restaurants)):
    zomato['Mean Rating'][zomato['name'] == restaurants[i]] = zomato['rate'][zomato['name'] == restaurants[i]].mean()
```

In [33]:
```
zomato.head()
```

Out[33]:

	address	name	online_order	book_table	rate	votes	location	rest_type	cuisines	cost	reviews_list	menu_item	typ
0	942, 21st Main Road, 2nd Stage, Banashankari, ...	Jalsa	True	True	4.10	775	Banashankari	Casual Dining	North Indian, Mughlai, Chinese	800.00	[('Rated 4.0', 'RATED\n A beautiful place to ...	[]	Buff
1	2nd Floor, 80 Feet Road, Near Big Bazaar, 6th ...	Spice Elephant	True	False	4.10	787	Banashankari	Casual Dining	Chinese, North Indian, Thai	800.00	[('Rated 4.0', 'RATED\n Had been here for din...	[]	Buff
2	1112, Next to KIMS Medical College, 17th Cross...	San Churro Cafe	True	False	3.80	918	Banashankari	Cafe, Casual Dining	Cafe, Mexican, Italian	800.00	[('Rated 3.0', "RATED\n Ambience is not that ...	[]	Buff
3	1st Floor, Annakuteera, 3rd Stage, Banashankar...	Addhuri Udupi Bhojana	False	False	3.70	88	Banashankari	Quick Bites	South Indian, North Indian	300.00	[('Rated 4.0', "RATED\n Great food and proper...	[]	Buff
4	10, 3rd Floor, Lakshmi Associates, Gandhi Baza...	Grand Village	False	False	3.80	166	Basavanagudi	Casual Dining	North Indian, Rajasthani	600.00	[('Rated 4.0', 'RATED\n Very good restaurant ...	[]	Buff

In [ ]:

In [34]:
```
from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler(feature_range = (1,5))

zomato[['Mean Rating']] = scaler.fit_transform(zomato[['Mean Rating']]).round(2)

zomato.sample(3)
```

Out[34]:

	address	name	online_order	book_table	rate	votes	location	rest_type	cuisines	cost	reviews_list	menu_item	
11807	64, MM Road, Frazer Town, Bangalore	Cream Stone	True	False	4.00	80	Frazer Town	Dessert Parlor	Desserts, Beverages	500.00	[('Rated 4.0', 'RATED\n One of the nicest pla...	['Willy Wonka', 'Ferrero Rocher', 'Nutella Bro...	De
34177	G-18, Kedia Arcade, 92, Infantry Road, Bangalore	Peppa Zzing	True	False	4.30	1135	Infantry Road	Cafe	Cafe, Fast Food, Burger	800.00	[('Rated 5.0', 'RATED\n This is a classic bur...	[]	De
22827	30, Bethel, Venkatapura Main Road, 13th Cross,...	New Ambur Dum Biryani	False	False	3.20	7	Koramangala 1st Block	Quick Bites	Biryani, Kebab	200.00	[]	[]	

In [ ]:
```
```

In [35]:
```
zomato.head()
```

Out[35]:

	address	name	online_order	book_table	rate	votes	location	rest_type	cuisines	cost	reviews_list	menu_item	typ
0	942, 21st Main Road, 2nd Stage, Banashankari, ...	Jalsa	True	True	4.10	775	Banashankari	Casual Dining	North Indian, Mughlai, Chinese	800.00	[('Rated 4.0', 'RATED\n A beautiful place to ...	[]	Buff
1	2nd Floor, 80 Feet Road, Near Big Bazaar, 6th ...	Spice Elephant	True	False	4.10	787	Banashankari	Casual Dining	Chinese, North Indian, Thai	800.00	[('Rated 4.0', 'RATED\n Had been here for din...	[]	Buff
2	1112, Next to KIMS Medical College, 17th Cross...	San Churro Cafe	True	False	3.80	918	Banashankari	Cafe, Casual Dining	Cafe, Mexican, Italian	800.00	[('Rated 3.0', "RATED\n Ambience is not that ...	[]	Buff
3	1st Floor, Annakuteera, 3rd Stage, Banashankar...	Addhuri Udupi Bhojana	False	False	3.70	88	Banashankari	Quick Bites	South Indian, North Indian	300.00	[('Rated 4.0', "RATED\n Great food and proper...	[]	Buff
4	10, 3rd Floor, Lakshmi Associates, Gandhi Baza...	Grand Village	False	False	3.80	166	Basavanagudi	Casual Dining	North Indian, Rajasthani	600.00	[('Rated 4.0', 'RATED\n Very good restaurant ...	[]	Buff

In [ ]:

Text Preprocessing :-

Some of the common text preprocessing / cleaning steps are:

Lower casing.

Removal of Punctuations.

Removal of Stopwords.

Removal of URLs.

Spelling correction.

In [ ]:

In [36]:
```
# 5 examples of these columns before text processing:
zomato[['reviews_list', 'cuisines']].sample(5)
```

Out[36]:

	reviews_list	cuisines
37056	[('Rated 3.0', "RATED\n Y this restaurant oft...	Biryani, Fast Food
40302	[('Rated 1.0', 'RATED\n I wanted to have Ruma...	North Indian, Mughlai, Chinese
26747	[('Rated 1.0', "RATED\n Yet another third gra...	Biryani, North Indian, Italian, Chinese, Sandw...
24443	[('Rated 3.0', 'RATED\n I am a big fan of som...	Ice Cream, Bakery, Desserts
30476	[('Rated 3.0', 'RATED\n I ran into this place...	South Indian, North Indian, Chinese, Street Food

In [ ]:

In [37]:
```
## Lower Casing
zomato["reviews_list"] = zomato["reviews_list"].str.lower()
zomato[['reviews_list', 'cuisines']].sample(5)
```

Out[37]:

	reviews_list	cuisines
19650	[('rated 3.0', "rated\n we have been there fo...	Chinese, Malaysian
28598	[('rated 3.0', 'rated\n had ordered a proper ...	North Indian, Chinese
18495	[('rated 2.0', 'rated\n when the "special ins...	Mangalorean, Chinese
22435	[('rated 4.0', 'rated\n great place to have v...	North Indian
10501	[('rated 1.0', 'rated\n food is not tasteful....	North Indian, South Indian

In [ ]:

In [38]:
```
## Removal of Puctuations
import string
PUNCT_TO_REMOVE = string.punctuation
def remove_punctuation(text):
    """custom function to remove the punctuation"""
    return text.translate(str.maketrans('', '', PUNCT_TO_REMOVE))

zomato["reviews_list"] = zomato["reviews_list"].apply(lambda text: remove_punctuation(text))
zomato[['reviews_list', 'cuisines']].sample(5)
```

Out[38]:

	reviews_list	cuisines
37707	rated 30 ratedn overall experience was good a...	North Indian, Mughlai, Hyderabadi
8796	rated 30 ratedn had a nice experience after h...	Biryani, Mughlai, Chinese
15042	rated 30 ratedn a typical kerala style restau...	South Indian, Kerala, Biryani
38349	rated 40 ratedn ah the place offers such good...	Beverages, Fast Food
36120	rated 20 ratedn visiting these guys from the ...	South Indian, North Indian, Continental

In [ ]:

In [39]:
```
import nltk
from nltk.corpus import stopwords
```

```
# Download the stopwords
nltk.download('stopwords')

## Removal of Stopwords
STOPWORDS = set(stopwords.words('english'))
def remove_stopwords(text):
    """custom function to remove the stopwords"""
    return " ".join([word for word in str(text).split() if word not in STOPWORDS])

zomato["reviews_list"] = zomato["reviews_list"].apply(lambda text: remove_stopwords(text))
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\DELL\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

In [ ]:

In [40]:
```
## Removal of URLS
def remove_urls(text):
    url_pattern = re.compile(r'https?://\S+|www\.\S+')
    return url_pattern.sub(r'', text)

zomato["reviews_list"] = zomato["reviews_list"].apply(lambda text: remove_urls(text))
```

In [ ]:

In [41]:
```
zomato[['reviews_list', 'cuisines']].sample(5)
```

Out[41]:

	reviews_list	cuisines
23234	rated 10 ratedn repeatedly ordered wont ever o...	Biryani, North Indian, Charcoal Chicken, Chine...
4887	rated 40 ratedn ambience great friday eve visi...	Chinese, Momos
32327	rated 10 ratedn worst restaurant negative rati...	North Indian, Chinese
36615	rated 10 ratedn pathetic service everordered o...	North Indian, Chinese, Beverages
3798	rated 50 ratedn harvest market easily best pla...	Continental, Pizza

In [ ]:

In [42]:
```
# RESTAURANT NAMES:
restaurant_names = list(zomato['name'].unique())
restaurant_names
```

Out[42]:

```
['Jalsa',
 'Spice Elephant',
 'San Churro Cafe',
 'Addhuri Udupi Bhojana',
 'Grand Village',
 'Timepass Dinner',
 'Rosewood International Hotel - Bar & Restaurant',
 'Onesta',
 'Penthouse Cafe',
 'Smacznego',
 'Cafã\x83Â\x83Ã\x82Â\x83Ã\x83Â\x82Ã\x82Â\x83Ã\x83Â\x83Ã\x82Â\x82Ã\x83Â\x82Ã\x82Â© Down The Alley',
 'Cafe Shuffle',
 'The Coffee Shack',
 'Caf-Eleven',
 'Cafe Vivacity',
 'Catch-Up-Ino',
 "Kirthi'S Biryani",
 'T3H Cafe',
 '360 Atoms Restaurant And Cafe',
 'The Vintage Cafe',
 'Woodee Pizza',
 'Cafe Coffee Day',
 'My Tea House',
 'Hide Out Cafe',
 'Cafe Nova',
 'Coffee Tindi',
 'Sea Green Cafe',
 'Cuppa',
 "Srinathji'S Cafe",
 'Redberrys',
 'Foodiction',
 'Sweet Truth',
 'Ovenstory Pizza',
 'Faasos',
 'Behrouz Biryani',
 'Fast And Fresh',
 'Szechuan Dragon',
 'Empire Restaurant',
 'Maruthi Davangere Benne Dosa',
 'Chaatimes',
```

```
'Havyaka Mess',
"Mcdonald'S",
"Domino'S Pizza",
'Hotboxit',
'Kitchen Garden',
'Recipe',
'Beijing Bites',
'Tasty Bytes',
'Petoo',
'Shree Cool Point',
'Corner House Ice Cream',
'Biryanis And More',
'Roving Feast',
'Freshmenu',
'Banashankari Donne Biriyani',
'Wamama',
'Five Star Chicken',
'Xo Belgian Waffle',
'Peppy Peppers',
'Goa 0 Km',
'Chinese Kitchen',
'1947',
'Cake Of The Day',
'Kabab Magic',
"Namma Brahmin'S Idli",
'Gustoes Beer House',
'Sugar Rush',
'Burger King',
'The Good Bowl',
'The Biryani Cafe',
'Lsd Cafe',
'Rolls On Wheels',
'Sri Guru Kottureshwara Davangere Benne Dosa',
'Devanna Dum Biriyani Centre',
'Kolbeh',
'Upahar Sagar',
'Kadalu Sea Food Restaurant',
'Frozen Bottle',
'Parimala Sweets',
'Vaishali Deluxe',
'The Big O Bakes',
```

```
'Meghana Foods',
'Krishna Sagar',
'Dessert Rose',
'Chickpet Donne Biryani House',
"Thanco'S Natural Ice Creams",
'Nandhini Deluxe',
"Vi Ra'S Bar And Restaurant",
'Kaggis',
'Ayda Persian Kitchen',
'Chatar Patar',
'Polar Bear',
"Kidambi'S Kitchen",
'Mane Thindi',
'Kotian Karavali Restaurant',
'Floured-Baked With Love',
'Cakes & Slices',
'Spice 9',
'Naveen Kabab & Biriyani Mane',
'Katriguppe Donne Biryani',
'Atithi Point Ande Ka Funda',
'Just Bake',
'K27 - The Pub',
'Bengaluru Coffee House',
'New Mangalore Lunch Home',
'Coffee Bytes',
'Parjanya Chat Zone',
"Kwality Wall'S Swirl'S Happiness Station",
'Ruchi Maayaka',
'Anna Kuteera',
'Darbar',
'Vijayalakshmi',
'Sri Udupi Food Hub',
'Udupi Upahar',
'House Of Kebabs',
'Roll N Rock',
'Box8- Desi Meals',
'Kfc',
'Roll Over',
'Imperial Restaurant',
'Lassi Shop',
'The Fortuna',
```

```
'Wahab',
'Al Diwan Biryanis And More',
"New Gowda'S Fried Chicken",
'Canton',
'Diabetics Dezire Sugarless Sweets And Bakes',
'The Blue Wagon - Kitchen',
'Hot Coffee',
'Patio 805',
'Lassi Corner',
'Sagar Deluxe',
'Kanti Sweets',
'Vegetalia',
'Aramane Donne Biryani',
'Ande Ka Funda',
'Cake Ghar',
'Energy Addaa',
'Bhattara Bhojana',
'Tandoori Knight',
'Dev Sagar - Food Street',
'Mitraa Da Pizza',
'Paradise Premium',
'Grazers',
'Shakes Theory',
'@Italy',
'Chilli Flakes',
'Calcutta Cafe',
'Old Mumbai Ice Cream',
'Donne Biriyani House',
'By 2 Coffee',
"Kedia'S Fun Food",
'Davangere Butter Dosa Hotel',
'Just Shawarma',
'Mini Punjabi Dhaba',
'Mulabagilu Dosa',
'Gokul Veg',
'Olive - Era',
'Udupi Ruchi',
'Madhappa Hindu Military Hotel',
'Gama Gama',
'Pizza Hut',
'Mangalore Pearl',
```

```
'Asha Sweets Centre',
'Twiststick House',
'Cool Corner',
'Pizza Mane',
'Dal Tadkaa',
'Chutney Chang',
'Mystique Palate',
'Thamboola',
'Castle Rock',
'Vietnamese Kitchen',
'C Corner',
'Paratha Merchant',
'North Rasoe',
'Toscano',
'Lord Of The Kitchen',
'Stoned Monkey',
'Central Jail Restaurant',
'Bella',
'Vennela Andhra Meals',
'New Prashanth Hotel',
'The Grillo',
'Re Malnad Nati Style Hotel',
'Karma Kaapi',
"Tiwari'S",
'Pizza Stop',
"Biggies Burger 'N' More",
"Kollapuri'S",
'Kadamba Classic',
'Rs Shiv Sagar',
'Kholi Mane',
'Harshi Super Sandwich',
'Cake Yard',
'Sri Udupi Veg',
'Cake Art',
'Potato House',
'Matru Sagar',
'Ugadi',
'Sri Krishna Sweets',
'In Time Cane Juice',
'Subway',
'Daal Roti',
```

```
'The Lassi Park',
'A2B - Adyar Ananda Bhavan',
'Srikrishna Bhavan',
'Green Gardenia',
'J Spice',
'Karavali Family Restaurant',
'Karavali Lunch Home',
'Hatti Kaapi',
"Kolkata Kathi Roll'S",
'Upahara Darshini',
'The Chaat Shop',
'Anda Ka Funda',
'Shri Vinayaka Ice N Juice',
'Ibaco',
'Jalaram Sweets',
'Samskruti - Sanman Gardenia',
'Bendakaluru Bytes',
'Cocoa Bakes',
'Chumma Delicious',
'Dwaraka Grand',
'Gufha - The President Hotel',
"Tanna'S Kitchen",
'Shree Mandarathi Grand',
'Mojo Pizza - 2X Toppings',
'Iceberg Icecreams',
'South Kitchen',
'Chung Wah',
'Shanthi Sagar',
'Millet Mama',
'Bangarpet Chat Express',
'Prems Graama Bhojanam',
'Java City',
'Kamat Bugle Rock',
'Puliyogare Point',
'Bangalore Agarwal Bhavan',
'Rustic Stove',
'Udupi Ruchi Grand',
'De Thaali',
'Just Thindi',
'Vasanth Vihar - Since 1965',
'Sea Spice By 7 Star',
```

```
'Desi Dawat',
'Cafe Aira',
'Mast Punjabi',
'South Grand',
'The Pizzeria',
'Cafe Zone',
'Ruchis Point',
'Firangi Bake',
'The Royal Corner - Pai Viceroy',
'Hara Fine Dine',
'Chocoberry',
'Dakshin Grand',
'Sandwichwallas',
'Chai Mane',
'Slv Upachar',
'Waffle-A-Go Go',
'Cross Road Cafe',
'Anand Donne Biriyani',
'Seven Star',
"Stop 'N' Joy",
'Sri Yaksha Shiv Sagar',
'Fudge',
'Just Bunny',
'Bakers Town',
'Shree Mahalakshmi Sweets',
'Kaapi Kendra',
'Ivy',
'The Airos',
'Chai Kraft',
'South Point Pub',
'The Trundle',
'The Krishna Grand Xpress',
'Sri Krishna Darshini',
'Shringar Sweets & Snacks',
'Udupi Grand',
'Ganesh Grand',
'Shreyas Upahar & Burger Point',
'Little Cafe',
'Mumbai Badam Milk & Lassi Center',
'Sri Krishna Sagar',
'Mylari Biryani Family Restaurant',
```

```
'1980S Games Cafe',
'Kadala Tarangaa',
'Namma Biryani',
"Nandhanu'S Rasoi",
'Hotel Pork Paradise',
'As On Fire',
'J K Fish Land',
'Curry Leaves',
'Magic Meals',
'Desi Cream Junction',
'Drunken Monkey',
"Ruchi'S Corner",
'Tandoori Bytes',
'Bangalore Donne Biriyani',
'Sgs Non Veg - Gundu Pulav',
'Keventers',
'Navi Food Point',
'Shawarma Inc',
'The Cafe Nuts And Bolts',
'Am Wow',
'Lalbagh Grand',
'Funjabi Curries',
'Eat Repeat',
'Salut',
'Nammura Donne Biriyani',
'Lassi Stop',
'Dakshin Kitchen',
'Sri Sai Cafe',
'Sri Sai 99 Variety Dosa',
'Utsav Restaurant',
'Crunch And Munch',
'Food Geek',
'Sri Krishna Aramane',
'Great Indian Rolls',
'Chats Point',
'Food Point',
'Swarga Ruchi',
'Kolkata King',
"Sandwich Mama'S And Frozen Monster",
'New Rajadhani Spicy',
'Big Mishra Pedha',
```

```
'Vegeatz',
'Foodizo',
'Food Springs',
"Chung'S Chinese Corner",
'Eurasia Pasta And Barbeque By Little Italy',
'Goli Vada Pav No. 1',
"Ragoo'S",
'Pure & Natural',
'Maiyas',
'Sip N Dine',
'Cafe Mondo',
'Jalpaan',
"Kataria'S Pakwan",
'Juicy Momos',
'Amande Patisserie',
'Anand Sweets And Savouries',
'Chicken Hunt',
'Swadesh Tadka',
'Cane-O-La',
'New Biryani Mane',
'Karnataka Bhel House',
'Sreeraj Lassi Bar',
'Juice Junction Food Court',
"Bunt'S Biriyani Palace",
'Chai Point',
'Janahaar',
'Utsav',
'Meat And Eat',
'Snacks Bite',
'The Spice Saga',
'Dakshin Cafe',
'Kaulige Millet Corner',
'Ifruit Ice Cream',
'Havmor',
'Grand Food',
'Dining Hut',
'Desi Swaad',
'Yummy Momos',
'Masale Daan',
'Grow Fit',
'Sandwich Hub',
```

'Udupi Sri Krishna Cafe',
'Starlite Bakery & Fastfood',
'Veganbreak24X7',
'Aadhya Hotel',
'Dodda Mane Baaduta',
'Swadd Kitchen',
'Happy Chopsticks',
'Udupi Upachar',
'Davanagere Benne Dose Hut',
'Subz',
'Mint And Mustard',
'Chinese Square',
'Punjabi Raswada',
"Vinny'S",
"Chetty'S Corner",
"Kapoor'S Cafe",
'Donne Biryani & Kabab Corner',
'Sukh Sagar',
'S M V Snacks Corner',
'Andhra Ruchulu',
'Steaming Mugs',
'Rajathadri Food Fort',
'Brew Meister',
'Rasoi',
'Mr. Singh Da Dhaba',
'Kababs N Biryani',
'Ayodhya Upachar',
'Biriyani Mane',
'Cafe Ajfan',
'Brundhavana Food Point',
'Slv Corner Restaurant',
'Hotel New Karavali',
'The Krishna Grand',
'Roti Ghar',
'Kettle & Kegs',
'Baisakhi',
'Poonam Sweets',
"Amma'S Pastries",
'The Lassi And Juice Park',
'Corner Stone',
'Arun Ice Cream',

```
'Sweet N Swirl',
'Sri Venkateshwara Sweet Meat Stall',
'Baskin Robbins',
'Srinivasa Brahmins Bakery',
'Jain Bakes',
'Kc Das - Sweet Spot',
'A2B Veg - Adyar Ananda Bhavan',
'Jcubez',
'Blue Wings Bar & Restaurant',
'New Imperial Restaurant',
'Karavali Fish Center',
"Iyer'S Tiffin Centre",
'Kydz Adda',
'Food Box Cafe',
'Sri Ganesh Juice And Chats',
'Manifest Cafe',
'Taaza Thindi',
'Sri Laxmi Venkateshwara Coffee Bar',
'Messy Bowl',
'Brahmin Cafe',
'Hotel Mangala',
'Simple Thindies',
'Slv Swadishta',
'New Sagar Fast Food',
'Parama Ruchi',
'Thrilok Restaurant',
'Hanumanthanagar Biryani Junction',
'Slv Refreshment',
'Svkp Daily Fresh',
'Srinagar Kabab Corner',
'Sri Venkateshwara Chat Centre',
'Vinay Bhel Corner',
'Nandi Chats And Juice',
'Food Adda',
'Hotel Nisarga',
'Yo Roll Corner',
'New Quality Dum Biryani',
'Sri Lakshmi Kabab Center',
'Tasty Bites',
'Raams Chicken',
'Panchami',
```

```
'Kavali',
'Ranganna Military Hotel',
'Vidyarthi Bhavan',
'Bharjari Oota',
'Bridgeway',
'Soho Bar & Grill',
'Bhavani Restaurant',
'Zephyr',
'Hotel Dwarka',
'Nisarga Garden Restaurant',
'Ma-Arya Family Restaurant',
'Udupi Sri Krishna Bhavan',
'Ice Thunder',
'Mahalaxmi Tiffin Room',
'Basavanagudi Mylari',
'Shrinidhi Military Hotel',
'Pramukh Family Restaurant',
'Vybhava',
'Shree Venkateshwara North Karnataka Hotel',
'Anand Sagar Inn',
'Sangam Military Hotel',
'Belly Squad Food Truck',
'50-50 Eating House',
'Om Shiva Shakthi Chats Centre',
'Rolls Corner',
'New Ambur Hot Dum Biryani',
'Deja Vu Resto Bar',
'Fattoush',
'Abhiruchi Hotel',
'Tulips',
'Barbeque Nation',
'Sattvam',
'24Th Main',
'Zaitoon',
'Mango Greens',
'Oye Amritsar',
'Melt - Eden Park',
'Spice Code',
'The Onyx - The Hhi Select Bengaluru',
'The Pavillion',
'Sankranthi Veg Restaurant',
```

```
'Tisano Cafe',
'Cafe Kabana',
'Butterly',
'Black Mug Cafe',
'#Refuel',
'Wafl',
'Vaho Kafe & Pressery',
'Dreamcatcher',
'Cafe Arabica',
'Starbucks',
'Smoor',
'Kalmane Koffees',
'Shee-Sha Cafe',
'Brews N Bites',
'D2V Cafe',
'Cafe Talk',
'Cafe Choco Craze',
'Slate Cafe',
'Dialogues',
'Mudpipe Cafe',
'Tab - Take A Break',
'Cafe Potpourri',
'De Oxford Cafe',
'Vinaya Coffee Moments',
'Brew Point',
'The Cravery',
"Anju'S Cafe",
'Skytouch Le Cafe',
'Hearts Unlock Cafe',
'Eat.Fit',
'Sai Abhiruchi',
'Capsicum Family Restaurant',
'Box Magic',
"Maa'R Rannaghor",
'Easy Bites',
'Hiyar Majhe Kolkata',
'Dabba Gosht',
'Punjabites',
'Sri Lakshmi Dhaba',
'Swadista Aahar',
'Vegetarea',
```

```
'Al-Bek',
"Aniram'S",
'Punjabi Nawabi',
'Yummrajj',
'Swad Punjab Da',
'Roti Wala',
'Midnight Mania',
'Kitchens@Jp Nagar',
'Krishna Kuteera',
'Apna Punjab',
'Paratha Junction',
'Nellore Bhojanam',
'Momoz',
'Kalingas',
'Kanteen The Eatery',
'Kullad Cafe',
'Litti Twist',
'Cakebuy',
'Delight Food',
'Andhra Kitchen',
'Veg By Nature',
'Chicken Magic',
'Swathi Restaurant',
'Fresh Kitchen',
'Hind Ka Chulah',
'Kuttanad',
'New Mahesh Friends Food Corner',
'Bohra Bohra Cafã\x83Â\x83Ã\x82Â\x83Ã\x83Â\x82Ã\x82Â\x83Ã\x83Â\x83Ã\x82Â\x82Ã\x83Â\x82Ã\x82Â©',
'Shree Krishna Sannidhi',
'Bingejoy!',
'Shiv Sai Hotel',
'Mra',
'Burj Hotel',
'Shaadi Ki Biryani',
'Madeena Hotel',
'Biryani Durbar',
'Mahesh Friends Food Center',
'Juice Shop',
"Dadi'S Dum Biryani",
'Krishna Kuteera South',
'Alankrutha',
```

```
'Paradise',
'Kabab Mehal',
'Sri Punjabi Dhaba',
'Arabian Mexico',
'Cakezone',
'Fujian Express',
'Indian Food',
'Tandoori Paradise',
'Kolkata Kathi Rolls',
'Adithya',
'Cheesiano Pizza',
'Nati Palace',
"Dande'S Hyderabad Biryani",
'Upahara Bhavan',
'Sher-E-Punjab',
'Shuddh Desi Khana',
'Karama Restaurant',
'Jaganath Hotel And Restaurant',
'Aramane Donne Biriyani',
"Mani'S Dum Biryani",
'Amontron',
'A M Biryani Hotel',
'Birinz',
'Hyderabadi Bawarchi',
'Fish Chain',
'Prasiddhi Food Corner',
'Biriyani Bhatti',
'Hyderabad Biryaani House',
"Galito'S",
'C. K. Mega Hot Food',
'Red Chilli Restaurant',
'Rss Donne Biryani',
'Rajdhani Thali Restaurant',
'Phew (Play Hard Eat Wild)',
'New Kabab Zone',
'Bawarchi Paradise',
'Shree Udupi Grand',
'Chicken County Grand',
'Darjeeling Momos & Fast Food Center',
'Veruthe Oru Thattukada',
'Savoury - Sea Shell Restaurant',
```

```
'Warm Oven',
'Kundana',
'Food Ka Masti',
'The Shawarma Shop',
'Momo Junction',
'Antilla Aromas',
'Punjabi Food Corner',
'Mealer.In',
'Pathaan Sir',
'Cold Stone Creamery',
'Hari Super Sandwich',
'Amritsari Kulcha Land',
'Chokha Chowka',
'Gorbandh',
'Grills & Rolls',
'Bathinda Junction',
'Stories',
"New Gongura'S",
'Sagar Grand',
'Ubq By Barbeque Nation',
'Agarwal Food Service',
"Daniyal'S",
'Seasons',
'Chef Delicacies',
'Indiana Burgers',
'Moksha',
'Marwa Restaurant',
'Shanghai Court',
'Akshaya Donne Biriyani',
'Bhojohori Manna',
'Richie Rich',
'Hunger Bee',
'Yum In My Tum',
'Maggi N Maggi House',
'Fresh Dinner',
'B.M.W Bhookh Mitaane Wala',
'Biryani Miya',
'Krispy Kreme',
'Paani Kum Chai',
'Chulha Chauki Da Dhaba',
"Magix'S Parattha Roll",
```

```
'Elegant Dining',
'Aliensip',
'Waffle Head',
'Samruddhi Biryani',
'Basmati Delights',
"Charlee'S Chicken",
'Samosa Singh',
'Cravings',
'Nagas',
'Matka',
'Punjabi Swag',
'Taco Bell',
'Ambur Star Dum Biryani',
"Mother'S Rasoi",
'Dosa Bazaar',
'Babu Moshai',
'The Bong Palate',
'Gowdru Mane Oota',
'Banashankari Nati Style',
"Chandrima'S Kitchen",
'Bikaner Jn',
'Crunch Pizzas',
'Lassi Berg',
'Kakal-Kai Ruchi',
'Manchu Cafe',
'Calorie Express',
'Bangalore Box',
'Hotel Khaaja',
'The Gujarat Express',
'Vishal Foods',
'Lassi Darbar',
'Chavadi',
'Nanna Munna Paratha House',
'Cucumber Town',
'Kolkata Famous Kati Roll',
'Nellore Ruchulu',
'Brewsky',
'Chefeana',
'Bangaliana',
'Gud Dhani',
'The Hunger Room',
```

```
'Parisar Veg Restaurant',
'Deejas Kitchen',
'Pancuzzi',
"Zhang'S - Chinese Restaurant",
'Shagun Sweets & Foods',
'Eat Well',
"Dev'S Gugababa",
'Oogway Express',
'Balaji Bombay Vada Pav Gujrati Dalebi',
'The Cake Ville',
'The Egg Factory',
'Chow San',
'New Udupi Grand',
'New Karawali Lunch Home',
'Sr Choco Station',
'Momo Jojo',
'Simply Indian',
'Delhi Biryani Cafe',
"Rithika'S Kitchen",
'Punjabi Times',
'Cravy Wings',
'Funky Punjab',
'A1 Garam Masala',
'Punjabi Corner',
'Guru Palace',
'Zeeshan',
'Ambara Gardenia',
'Layerbite',
'Bhavani Chats',
'More Pizza',
'Crafted Plate',
"Shetty'S Kitchen",
'Machali Port',
'Naati Manae',
'Mad Over Biryani',
'Garma Garam',
'Sambram Biriyani Paradise',
'Mr. Meetharam',
'The Coastal Crew By Fujian On 24Th',
'Bib - Breakfast In The Box',
'Angel Restaurant',
```

```
'Wow Momo',
'Pizza Da Dhaba',
'Kalpavruksha',
'Juice Land',
'Tandoor And Spice',
'Rock Stone Ice Cream Factory',
'Paratha Plaza',
'Krishna Vijayashree',
'Soup Station',
'Tasty Point',
'Natural Mumbai Kulfi',
'Namma Kudla',
'Triveni',
'Wangs Kitchen',
'Rk Fresh Food',
'Alif Restaurant',
'Dine One One Restaurant',
'Pallavas Veg Cuisine',
'The Food Cottage',
'Brundhavana Pure Veg',
'Spice Up',
'Suryawanshi',
'Tempteys',
'Delibox.In',
'Frybies',
'Crumb Together',
'The Foodyz',
"Bean D'Er Cafe",
'Chaat Chatore',
'Taste Of Kolkata',
'Y Not Restaurant',
'24/7 Food Service',
'Modern Restaurant',
'Burrito Boys',
'Trippy Paradise',
'Biryani Magic',
'Nagarjuna',
'Laddoos',
'Waffle Stories',
'Late Night',
'Waffle Walle',
```

```
'Shake It Off',
'Bombay Kulfi',
'Chaat Central',
'Two Friends Cauldron',
'Bun Town',
'Garden Fresh',
'Dhabeliwala',
"Rayaan'S Bbq",
'Chaat Junction',
'Asharfilal',
'Street Foods By Punjab Grill',
"Tiwari'S Ghee Paratha And Chats",
'Spurthi Foods',
'Melting Pot At Woodrose',
'Pot Biryani',
"Thanco'S Natural Ice Cream",
'Mango Tree- The Beer Garden',
'Dessi Cuppa',
'Halli Mane Mudde Oota',
'Wow Vada Pav',
'The Jade Kitchen',
'Krishna Vaibhava',
'Nandhanus Rasoi',
'Donne Biriyani Mane',
'Fish And Dish',
'Miss Momo',
'Intalia',
'Chicken County Restaurant',
'Velvet Kitchen',
'Hotel Aradhana',
'The Chervil',
'Sri Rajasthani Foods',
'New Royal Treat',
'I Cool',
'Mahek Pizza',
'Kc Das- Sweet World',
'Panther Cafe',
'Roots N Fruits',
'Kaizen Wellness Kitchen',
'Mad Over Donuts',
'Kolkata Famous Kati Rolls',
```

```
'Vinaya Cafe',
'Rbp Greens Garden',
'Dazu Momoz',
'Desi Dhaba',
'World Of Waffles',
"Pika'S Kitchen",
'Lassi Cafe',
'New Mogul Empire',
'Atithi Biryani Corner',
'Oottupura Family Restaurant',
'Yumme Veg',
'Shree Guru Raghavendra Chats Chintamani Special',
'Vyanjan',
'Sri Sankara Cafe',
'Bhaijaan Barbeques',
'Chatter Platter',
'Bikaneri Sweets',
"Watson'S",
'Sri Nandi Grand',
'Chicken Man',
"Chef Baker'S",
'Sree Ganesha Fruits & Juice',
'Nandhana Palace',
'Biryani Nights',
'Thejas Bhavan',
'Graffitea',
'Gappe',
'On The Nose',
'Juicy & Spicy',
'Eat Repeat Express',
'Tfi - The Fresh Ice Cream',
'Mirch Masala',
'Just Chill',
"Bengalooru Tiffany'S",
'Udupi Thaja Thindi',
'Baba Ka Dhaba',
'Hot Burg',
'Northern Bites',
'Hotel Smile',
'Roll Wala',
'Natural Ice Cream',
```

'F3 Food Fun Fiesta',
"Churchill'S",
"Kulkarni'S New Uttara Karnataka Food Speciality Stores",
'The Hungers Zone',
'The Chocolate Room',
'Aramane Restaurant',
'Hunger2Eat',
'Bbq Ride 46',
'Bawarchi Inn',
"Ruh'S Cafe",
'Eatery Have U Been',
'Upahar Banashree',
'Me And My Cake',
'Bhojon Tripti',
'New South Corner',
'Khana Khazana',
'China Tang',
'California Burrito',
'Tandoor Garden',
'Caffe Pascucci',
'Freeze It',
'Meal Square',
'Chaitanya Cafe',
'Tasty Jigarthanda',
'Berrylicious',
'Rcs Kitchen',
'Pappu Da Dhaba',
'Smoke - The Sizzler House',
'Sandwich Shop',
'Joon Restaurant',
'Snack Knack',
'Chung Wah Opus',
'Chef & Dine',
'Once In Nature',
'Manjushree Upahara',
'Hotel Annapoorna',
'Dream A Dozen',
'Bread Crumbs Bakery',
'Nite Out',
'Cupcake Bliss',
'Armaani Caffe',

```
'Muddhe Bytes',
'Bombay Kulfis',
'Willys Top Cafe',
'Mist N Creams',
'Papacream',
"Woody'S",
'Jp Fish Land',
'Pepper Crown Restaurant',
'Pizza Paradise',
'The Curry Hut Plus',
'Malabar Mess',
'Malabar Cafe',
'Tangra Indo - Chinese Cuisine',
'Juice Junction',
'Xpress Chai',
'Zhangs Classic',
'Guru Greens',
'D View Cafe',
'The Cuboidal',
'Delifusion - Hunger Sorted',
'Flavour Of China',
'Gujrati Mess',
'Cake-O-Mania',
'The Park Inn Restaurant',
'Eagle Ridge',
'Vidya Cafe',
'Kabab Treat',
'Limra Hotel',
'Kafe Nook',
'Kumbha Bhojanam',
'Yummerica Fries',
'Casa Piccosa',
'Bake Addiction',
'Tongue Twisters',
'Hot Rolls & House Of Kebabs',
'Aami Kolkata',
'Vaathsalya Millet Cafe',
'Hungry Buddies',
'Thericebowl.In',
'Guru Garden',
'Cakes By M',
```

```
'Brewz Coffee',
'Fritz Haber',
'Athithi',
'Smooth Blender',
'Sri Bhagya Grand',
"Shree Guru Juice 'N' Ice",
'Bhavya Military Hotel',
'Neals Cafe',
'Flavorsome Bakes',
'Bake-Ooh',
'The Yummy Tummy',
'Mast Kalandar',
"Sultan'S Biryani",
'Halli Sogasu',
'Crepe Nation',
'Rice Bar',
'Kanika Biryani Paradise',
'Resto',
'New Tandoori Point',
'Shiv Sagar',
'Little Shangai',
"Nadella'S Kitchen",
"Uncle'S Kitchen",
'Coastal Inn',
"Mom'S Momos",
'Aishwarya Parkland',
'Slice Of Spice',
'Bao And Rolls',
'Andhra Bhojanam',
'Kanchan Dhaba',
'Encyclofoodia',
'Bella?S Kitchen',
'Food Feast Multicusine Restaurant',
'Bangalir Rannaghar',
'Sai Samosa & Chat Corner',
"Lalchee'S Rasoi",
'Kailash Parbat',
"Dream'S Kitchen",
'Tandoor Hut',
'The Chocolatiers',
'Ballava',
```

```
 'Zaika Take Away',
 'Snackiey',
 'Sandwich Mamas',
 'The Shake Factory Originals',
 'Hotel Lakshmi Paradise',
 'Desi Bites',
 'Shiv Sai',
 'The Juicy',
 'Feast And Burp',
 'Zengi Pub & Restaurant',
 'Cafe Indiana',
 'Multi Cakes',
 'Bhukkad',
 'Malhar Maharashtrian Cuisine',
 'Mandi',
 'Pallavi North Indian Veg Restaurant',
 'Restro Cafe',
 ...]
```

In [ ]:

In [43]:
```
def get_top_words(column, top_nu_of_words, nu_of_word):

    vec = CountVectorizer(ngram_range= nu_of_word, stop_words='english')

    bag_of_words = vec.fit_transform(column)

    sum_words = bag_of_words.sum(axis=0)

    words_freq = [(word, sum_words[0, idx]) for word, idx in vec.vocabulary_.items()]

    words_freq =sorted(words_freq, key = lambda x: x[1], reverse=True)

    return words_freq[:top_nu_of_words]
```

In [ ]:

In [44]:
```
zomato.head()
```

Out[44]:

	address	name	online_order	book_table	rate	votes	location	rest_type	cuisines	cost	reviews_list	menu_item
0	942, 21st Main Road, 2nd Stage, Banashankari, ...	Jalsa	True	True	4.10	775	Banashankari	Casual Dining	North Indian, Mughlai, Chinese	800.00	rated 40 ratedn beautiful place dine inthe int...	[]
1	2nd Floor, 80 Feet Road, Near Big Bazaar, 6th ...	Spice Elephant	True	False	4.10	787	Banashankari	Casual Dining	Chinese, North Indian, Thai	800.00	rated 40 ratedn dinner family turned good choo...	[]
2	1112, Next to KIMS Medical College, 17th Cross...	San Churro Cafe	True	False	3.80	918	Banashankari	Cafe, Casual Dining	Cafe, Mexican, Italian	800.00	rated 30 ratedn ambience good enough pocket fr...	[]
3	1st Floor, Annakuteera, 3rd Stage, Banashankar...	Addhuri Udupi Bhojana	False	False	3.70	88	Banashankari	Quick Bites	South Indian, North Indian	300.00	rated 40 ratedn great food proper karnataka st...	[]
4	10, 3rd Floor, Lakshmi Associates, Gandhi Baza...	Grand Village	False	False	3.80	166	Basavanagudi	Casual Dining	North Indian, Rajasthani	600.00	rated 40 ratedn good restaurant neighbourhood ...	[]

In [ ]:

In [45]:
```
zomato.sample(5)
```

Out[45]:

	address	name	online_order	book_table	rate	votes	location	rest_type	cuisines	cost	reviews_list	menu_item
30719	42, Church Street, Bangalore	Blow	True	True	4.30	968	Church Street	Cafe	Cafe, Continental, American, Burger	750.00	rated 30 rated 50 rated 35 rated 50 rated 50 r...	[
5730	31/8, 2nd Phase, Hoodi Road, Opposite HDFC ATM...	Biryani Paradise	True	False	3.80	131	Whitefield	Casual Dining	Andhra, Biryani	700.00	rated 40 ratedn ordered food always wonderful ...	[
32541	161, 8th Cross, Ashwath Nagar, Marathahalli, B...	Eat Eroo - Bombay Vada Pav	False	False	3.40	7	Marathahalli	Takeaway, Delivery	Street Food, Maharashtrian	200.00	rated 40 ratedn place marathahalli find good v...	[
20319	Royal suits hotel, Kalyan Nagar.	The Pakoda Shop	True	False	3.40	12	Kalyan Nagar	Delivery	Street Food	150.00	rated 30 ratedn less quantity price charged qu...	[
20463	412, 3rd Floor, 1st Avenue, 5th Main, 2nd Bloc...	The Pipe Rack	False	False	3.90	35	Kalyan Nagar	Casual Dining	Cafe	700.00	rated 50 ratedn visited place weekend couple f...	[

In [ ]:

In [46]:
```
zomato.shape
```

Out[46]: (41237, 15)

In [ ]:

In [47]:
```
zomato.columns
```

Out[47]:
```
Index(['address', 'name', 'online_order', 'book_table', 'rate', 'votes',
       'location', 'rest_type', 'cuisines', 'cost', 'reviews_list',
       'menu_item', 'type', 'city', 'Mean Rating'],
      dtype='object')
```

In [ ]:

In [48]:
```
zomato=zomato.drop(['address','rest_type', 'type', 'menu_item', 'votes'],axis=1)
```

In [ ]:

In [49]:
```
import pandas

# Randomly sample 60% of your dataframe
zomato_percent = zomato.sample(frac=0.5)
```

In [ ]:

In [50]:
```
zomato_percent.shape
```

Out[50]: (20618, 10)

In [ ]:

Term Frequency-Inverse Document Frequency(TF-IDF):-

Term Frequency-Inverse Document Frequency (TF-IDF) vectors for each document. This will give you a matrix where each column represents a word in the overview vocabulary (all the words that appear in at least one document) and each column represents a restaurant, as before.

TF-IDF is the statistical method of evaluating the significance of a word in a given document.

TF — Term frequency(tf) refers to how many times a given term appears in a document.

IDF — Inverse document frequency(idf) measures the weight of the word in the document, i.e if the word is common or rare in the entire document. The TF-IDF intuition follows that the terms that appear frequently in a document are less important than terms that rarely appear. Fortunately, scikit-learn gives you a built-in TfIdfVectorizer class that produces the TF-IDF matrix quite easily.

In [ ]:

In [51]:
```
# Reverse mapping of indices
indices = pd.Series(zomato_percent.index, index=zomato_percent['name']).drop_duplicates()
```

In [52]:
```
indices
```

Out[52]:
```
name
The Wok Shop                                    10362
Abhiruchi Hotel                                  8749
Nook - Aloft Bengaluru Cessna Business Park      3012
Nandhi Andhra Style Family Restaurant           17392
Fava                                            31234
                                                ...
Palos                                           32448
Nandhini Deluxe                                  3923
The Egg Factory                                 21283
Hotel Oottupura                                 26551
Taantraa - Organic Handbaking                   36184
Length: 20618, dtype: int64
```

In [ ]:

Creating TF-IDF MATRIX:-

In [53]:
```
from sklearn.feature_extraction.text import TfidfVectorizer

# TfidfVectorizer setup
tfidf = TfidfVectorizer(analyzer='word', ngram_range=(1, 2), min_df=2, stop_words='english')
```

```
# Apply the tf-idf transformation to the reviews list column
tfidf_matrix = tfidf.fit_transform(zomato_percent['reviews_list'])
```

In [ ]:

In [55]:
```
tfidf_matrix.shape
```

Out[55]:
```
(20618, 973206)
```

In [ ]:

In [56]:
```
from sklearn.metrics.pairwise import linear_kernel

cosine_similarities = linear_kernel(tfidf_matrix, tfidf_matrix)
```

In [ ]:

In [57]:
```
cosine_similarities[1]
```

Out[57]:
```
array([0.00341901, 1.        , 0.01683436, ..., 0.00735674, 0.00902657,
       0.0163422 ])
```

In [ ]:

In [58]:
```
zomato_percent.index
```

Out[58]:
```
Index([10362,  8749,  3012, 17392, 31234, 26905,  4112, 16278,  8478,  5923,
       ...
        2558,  8078, 40098, 22224, 13335, 32448,  3923, 21283, 26551, 36184],
      dtype='int64', length=20618)
```

In [ ]:

In [59]:
```
def give_recommend(name, cosine_similarities = cosine_similarities):

    # Get the index corresponding to the restaurant name
    idx = indices[name][0]
```

```
    # Get the top 10 similar restaurant indexes(excluding the input restuarant itself)
    score_series = pd.Series(cosine_similarities[idx])
    top10_indexes = score_series.nlargest(11).iloc[1:].index # Exclude the first item(itself)

    # Extract the names of the top 10 recommended restaurants
    recommend_restaurant = zomato_percent['name'].iloc[top10_indexes].tolist()

    # Creating the new dataframe to store the cuisines,ratings and costs of the recommended restaurants
    data=[]
    for each in recommend_restaurant:

        # Get the corresponding row from zomato_percent for each restaurant
        row = zomato_percent.loc[zomato_percent['name'] == each ,['name','cuisines', 'Mean Rating', 'cost']]
        data.append(row)

    #combine all rows into a single DataFrame
    zomato_new = pd.concat(data)

    # Drop duplicates based on the selected columns
    zomato_new = zomato_new.drop_duplicates(subset=['name','cuisines','Mean Rating', 'cost'])

    # Sort by the 'Mean Rating' column in descending order and pick the top 10
    zomato_new = zomato_new.sort_values(by='Mean Rating', ascending=False).head(10)


    print('TOP 10 RESTAURANTS LIKE {name} WITH SIMILAR REVIEWS: ')

    return zomato_new
```

In [ ]:

In [60]:
```
give_recommend('Pai Vihar')
```

```
TOP 10 RESTAURANTS LIKE {name} WITH SIMILAR REVIEWS:
```

Out[60]:

	name	cuisines	Mean Rating	cost
22407	The Black Pearl	North Indian, European, Mediterranean	4.78	1.40
3954	The Black Pearl	North Indian, European, Mediterranean, BBQ	4.78	1.50
3030	Village - The Soul Of India	North Indian, Lucknowi, Gujarati, Maharashtria...	3.85	1.10
8622	Village - The Soul Of India	Rajasthani, Gujarati, North Indian, South Indian	3.85	1.00
13374	Silbatti	North Indian	3.77	750.00
13428	Atithi	North Indian, Chinese, Street Food	3.63	800.00
37307	Atithi	North Indian	3.63	750.00
18380	Nouvelle Garden	North Indian, Continental, Italian	3.45	900.00
39028	Flavours - Octave Hotel & Spa	North Indian, Chinese	2.55	700.00

In [ ]:

In [61]:
```
give_recommend('South Inn')
```

```
TOP 10 RESTAURANTS LIKE {name} WITH SIMILAR REVIEWS:
```

Out[61]:

	name	cuisines	Mean Rating	cost
8663	Dice N Dine	Continental, Cafe, Italian, Fast Food, Steak	4.23	900.00
15966	Mug N Bean	Cafe, Continental, Italian, Burger	3.71	400.00
14209	Drunken Monkey	Beverages, Fast Food	3.65	400.00
31481	Drunken Monkey	Beverages, Fast Food	3.65	500.00
3465	B.M.W - Bhookh Mitaane Wala	North Indian, South Indian, Chinese	3.42	500.00
10733	Godavari Delights	Andhra, Biryani	2.85	700.00
11132	Godavari Delights	Andhra, North Indian, Biryani, Seafood	2.85	450.00
10900	Godavari Delights	South Indian	2.85	400.00

In [ ]:

In [ ]:

Collaborative Filtering:-

Collaborative filtering:-

It is a method of making recommendations based on the preferences of similar users. In the case of creating a Zomato recommendation filtering system using collaborative filtering, we can use user-item interactions (like ratings or likes on particular restaurants) to suggest restaurants to a user based on the ratings or preferences of other similar users.

Similarity Calculation:-

In collaborative filtering, we measure how similar users or items are based on their ratings. There are two main types of collaborative filtering:

User-based Collaborative Filtering:-

User similarity: Compute the similarity between users based on their ratings for the restaurants. You can use cosine similarity to measure how similar users are in terms of their restaurant preferences.

Item-based Collaborative Filtering:-

Restaurant similarity: Compute the similarity between restaurants based on how users rate them. Restaurants that have similar ratings across users are considered similar. Again, you can use cosine similarity or Pearson correlation for this. For example, if User1 likes "Jalsa" and "Spice Elephant", and User2 likes "Spice Elephant" and "San Churro Cafe", we can suggest "San Churro Cafe" to User1 based on their similarity to User2.

In [ ]:

In [ ]:
```
from IPython import display
display.Image("C:\\Users\DELL\\Desktop\\Resume\\Collaborative_filtering.png")
```

In [ ]:

Importing the libraries and Dataset :-

In [141…
```
## Database Phase
import pandas as pd
import numpy as np

#Visualization Phase
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib as mpl
import matplotlib.pylab as pylab
%matplotlib inline
pd.set_option('display.max_columns', 500)
mpl.style.use('ggplot')
sns.set_style('white')
pylab.rcParams['figure.figsize'] = 12,8

#ignore warnings
```

```
import warnings
warnings.filterwarnings('ignore')
```

In [ ]:

loading the data from csv file to a Pandas DataFrame :-

In [142…

```
# loading the data from csv file to a Pandas DataFrame

zomato_data = pd.read_csv('zomato.csv')
```

In [143…

```
zomato_data.head()
```

Out[143...

	url	address	name	online_order	book_table	rate	votes	phone	loca
0	https://www.zomato.com/bangalore/jalsa-banasha...	942, 21st Main Road, 2nd Stage, Banashankari, ...	Jalsa	Yes	Yes	4.1/5	775	080 42297555\r\n+91 9743772233	Banasha
1	https://www.zomato.com/bangalore/spice-elephan...	2nd Floor, 80 Feet Road, Near Big Bazaar, 6th ...	Spice Elephant	Yes	No	4.1/5	787	080 41714161	Banasha
2	https://www.zomato.com/SanchurroBangalore?cont...	1112, Next to KIMS Medical College, 17th Cross...	San Churro Cafe	Yes	No	3.8/5	918	+91 9663487993	Banasha
3	https://www.zomato.com/bangalore/addhuri-udupi...	1st Floor, Annakuteera, 3rd Stage, Banashankar...	Addhuri Udupi Bhojana	No	No	3.7/5	88	+91 9620009302	Banasha
4	https://www.zomato.com/bangalore/grand-village...	10, 3rd Floor, Lakshmi Associates, Gandhi Baza...	Grand Village	No	No	3.8/5	166	+91 8026612447\r\n+91 9901210005	Basavana

In [ ]:

EDA:-

In [144…
```
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style('white')
%matplotlib inline
```

In [ ]:

Getting some informations about the dataset :-

In [145…
```
zomato_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51717 entries, 0 to 51716
Data columns (total 17 columns):
 #   Column                       Non-Null Count  Dtype 
---  ------                       --------------  ----- 
 0   url                          51717 non-null  object
 1   address                      51717 non-null  object
 2   name                         51717 non-null  object
 3   online_order                 51717 non-null  object
 4   book_table                   51717 non-null  object
 5   rate                         43942 non-null  object
 6   votes                        51717 non-null  int64 
 7   phone                        50509 non-null  object
 8   location                     51696 non-null  object
 9   rest_type                    51490 non-null  object
 10  dish_liked                   23639 non-null  object
 11  cuisines                     51672 non-null  object
 12  approx_cost(for two people)  51371 non-null  object
 13  reviews_list                 51717 non-null  object
 14  menu_item                    51717 non-null  object
 15  listed_in(type)              51717 non-null  object
 16  listed_in(city)              51717 non-null  object
dtypes: int64(1), object(16)
memory usage: 6.7+ MB
```

In [ ]:

In [146…
```
#Deleting Unnnecessary Columns
zomato=zomato_data.drop(['url','dish_liked'],axis=1) #Dropping the column "dish_liked", "url" and saving the new dataset as "z
```

In [147…
```
#Removing the Duplicates
zomato.duplicated().sum()
zomato.drop_duplicates(inplace=True)
```

In [ ]:

In [148…
```
#Remove the NaN values from the dataset
zomato.isnull().sum()
```

```
zomato.dropna(how='any',inplace=True)
zomato.info() #.info() function is used to get a concise summary of the dataframe
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 42913 entries, 0 to 51716
Data columns (total 15 columns):
 #   Column                       Non-Null Count  Dtype 
---  ------                       --------------  ----- 
 0   address                      42913 non-null  object
 1   name                         42913 non-null  object
 2   online_order                 42913 non-null  object
 3   book_table                   42913 non-null  object
 4   rate                         42913 non-null  object
 5   votes                        42913 non-null  int64 
 6   phone                        42913 non-null  object
 7   location                     42913 non-null  object
 8   rest_type                    42913 non-null  object
 9   cuisines                     42913 non-null  object
 10  approx_cost(for two people)  42913 non-null  object
 11  reviews_list                 42913 non-null  object
 12  menu_item                    42913 non-null  object
 13  listed_in(type)              42913 non-null  object
 14  listed_in(city)              42913 non-null  object
dtypes: int64(1), object(14)
memory usage: 5.2+ MB
```

In [ ]:

In [149…
```
#Reading Column Names
zomato.columns
```

Out[149…
```
Index(['address', 'name', 'online_order', 'book_table', 'rate', 'votes',
       'phone', 'location', 'rest_type', 'cuisines',
       'approx_cost(for two people)', 'reviews_list', 'menu_item',
       'listed_in(type)', 'listed_in(city)'],
      dtype='object')
```

In [ ]:

In [150…
```
#Changing the column names
zomato = zomato.rename(columns={'approx_cost(for two people)':'cost','listed_in(type)':'type',
                                'listed_in(city)':'city'})
zomato.columns
```

Out[150…
```
Index(['address', 'name', 'online_order', 'book_table', 'rate', 'votes',
       'phone', 'location', 'rest_type', 'cuisines', 'cost', 'reviews_list',
       'menu_item', 'type', 'city'],
      dtype='object')
```

In [ ]:
```
```

In [151…
```
#Some Transformations
zomato['cost'] = zomato['cost'].astype(str) #Changing the cost to string
zomato['cost'] = zomato['cost'].apply(lambda x: x.replace(',','.')) #Using lambda function to replace ',' from cost
zomato['cost'] = zomato['cost'].astype(float) # Changing the cost to Float
zomato.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 42913 entries, 0 to 51716
Data columns (total 15 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   address       42913 non-null  object 
 1   name          42913 non-null  object 
 2   online_order  42913 non-null  object 
 3   book_table    42913 non-null  object 
 4   rate          42913 non-null  object 
 5   votes         42913 non-null  int64  
 6   phone         42913 non-null  object 
 7   location      42913 non-null  object 
 8   rest_type     42913 non-null  object 
 9   cuisines      42913 non-null  object 
 10  cost          42913 non-null  float64
 11  reviews_list  42913 non-null  object 
 12  menu_item     42913 non-null  object 
 13  type          42913 non-null  object 
 14  city          42913 non-null  object 
dtypes: float64(1), int64(1), object(13)
memory usage: 5.2+ MB
```

In [ ]:

In [152…
```
#Reading Rate of dataset
zomato['rate'].unique()
```

Out[152…
```
array(['4.1/5', '3.8/5', '3.7/5', '3.6/5', '4.6/5', '4.0/5', '4.2/5',
       '3.9/5', '3.1/5', '3.0/5', '3.2/5', '3.3/5', '2.8/5', '4.4/5',
       '4.3/5', 'NEW', '2.9/5', '3.5/5', '2.6/5', '3.8 /5', '3.4/5',
       '4.5/5', '2.5/5', '2.7/5', '4.7/5', '2.4/5', '2.2/5', '2.3/5',
       '3.4 /5', '-', '3.6 /5', '4.8/5', '3.9 /5', '4.2 /5', '4.0 /5',
       '4.1 /5', '3.7 /5', '3.1 /5', '2.9 /5', '3.3 /5', '2.8 /5',
       '3.5 /5', '2.7 /5', '2.5 /5', '3.2 /5', '2.6 /5', '4.5 /5',
       '4.3 /5', '4.4 /5', '4.9/5', '2.1/5', '2.0/5', '1.8/5', '4.6 /5',
       '4.9 /5', '3.0 /5', '4.8 /5', '2.3 /5', '4.7 /5', '2.4 /5',
       '2.1 /5', '2.2 /5', '2.0 /5', '1.8 /5'], dtype=object)
```

In [ ]:

In [153…
```
#Removing '/5' from Rates

import numpy as np

# Filter out 'NEW' and '-' from the 'rate' column
zomato = zomato.loc[zomato.rate != 'NEW']
zomato = zomato.loc[zomato.rate != '-'].reset_index(drop=True)

# Define a lambda function to remove '/5' and ensure it's applied correctly
remove_slash = lambda x: x.replace('/5', '') if isinstance(x, str) else x

# Apply the function and convert the 'rate' column to float after cleaning
zomato['rate'] = zomato['rate'].apply(remove_slash).str.strip()

# Convert the 'rate' column to float, forcing errors to NaN (if any) and drop those rows if necessary
zomato['rate'] = pd.to_numeric(zomato['rate'], errors='coerce')

# Display the first few rows of the 'rate' column
zomato['rate'].head()
```

Out[153…
```
0   4.10
1   4.10
2   3.80
3   3.70
4   3.80
Name: rate, dtype: float64
```

In [ ]:

In [154…
```
# Adjust the column names
zomato.name = zomato.name.apply(lambda x:x.title())
zomato.online_order.replace(('Yes','No'),(True, False),inplace=True)
zomato.book_table.replace(('Yes','No'),(True, False),inplace=True)
zomato.cost.unique()
```

Out[154…
```
array([800.  , 300.  , 600.  , 700.  , 550.  , 500.  , 450.  , 650.  ,
       400.  , 900.  , 200.  , 750.  , 150.  , 850.  , 100.  ,   1.2 ,
       350.  , 250.  , 950.  ,   1.  ,   1.5 ,   1.3 , 199.  ,   1.1 ,
         1.6 , 230.  , 130.  ,   1.7 ,   1.35,   2.2 ,   1.4 ,   2.  ,
         1.8 ,   1.9 , 180.  , 330.  ,   2.5 ,   2.1 ,   3.  ,   2.8 ,
         3.4 ,  50.  ,  40.  ,   1.25,   3.5 ,   4.  ,   2.4 ,   2.6 ,
         1.45,  70.  ,   3.2 , 240.  ,   6.  ,   1.05,   2.3 ,   4.1 ,
       120.  ,   5.  ,   3.7 ,   1.65,   2.7 ,   4.5 ,  80.  ])
```

In [ ]:

In [155…
```
zomato.head()
```

Out[155...

	address	name	online_order	book_table	rate	votes	phone	location	rest_type	cuisines	cost	reviews_li
0	942, 21st Main Road, 2nd Stage, Banashankari, ...	Jalsa	True	True	4.10	775	080 42297555\r\n+91 9743772233	Banashankari	Casual Dining	North Indian, Mughlai, Chinese	800.00	[('Rate 4.('RATED\n beautif place to
1	2nd Floor, 80 Feet Road, Near Big Bazaar, 6th ...	Spice Elephant	True	False	4.10	787	080 41714161	Banashankari	Casual Dining	Chinese, North Indian, Thai	800.00	[('Rate 4.('RATED' Had bee here f dir
2	1112, Next to KIMS Medical College, 17th Cross...	San Churro Cafe	True	False	3.80	918	+91 9663487993	Banashankari	Cafe, Casual Dining	Cafe, Mexican, Italian	800.00	[('Rate 3.("RATED' Ambience not that
3	1st Floor, Annakuteera, 3rd Stage, Banashankar...	Addhuri Udupi Bhojana	False	False	3.70	88	+91 9620009302	Banashankari	Quick Bites	South Indian, North Indian	300.00	[('Rate 4.("RATED' Great foo ar prope
4	10, 3rd Floor, Lakshmi Associates, Gandhi Baza...	Grand Village	False	False	3.80	166	+91 8026612447\r\n+91 9901210005	Basavanagudi	Casual Dining	North Indian, Rajasthani	600.00	[('Rate 4.('RATED' Very goo restaura

In [ ]:

In [156…
```
## Checking Null values
zomato.isnull().sum()
```

Out[156…
```
address         0
name            0
online_order    0
book_table      0
rate            0
votes           0
phone           0
location        0
rest_type       0
cuisines        0
cost            0
reviews_list    0
menu_item       0
type            0
city            0
dtype: int64
```

In [ ]:

In [157…
```
## Computing Mean Rating
restaurants = list(zomato['name'].unique())
zomato['Mean Rating'] = 0

for i in range(len(restaurants)):
    zomato['Mean Rating'][zomato['name'] == restaurants[i]] = zomato['rate'][zomato['name'] == restaurants[i]].mean()
```

In [ ]:

Let's create a ratings dataframe with average rating and number of ratings:

In [158…
```
zomato.groupby('name')['rate'].mean().sort_values(ascending=False).head()
```

Out[158…
```
name
SantãÂÃÂÃÂÃ
```

In [ ]:

Now set the number of ratings column:

In [161…
```
ratings['num of ratings'] = pd.DataFrame(zomato.groupby('name')['rate'].count())
ratings.head()
```

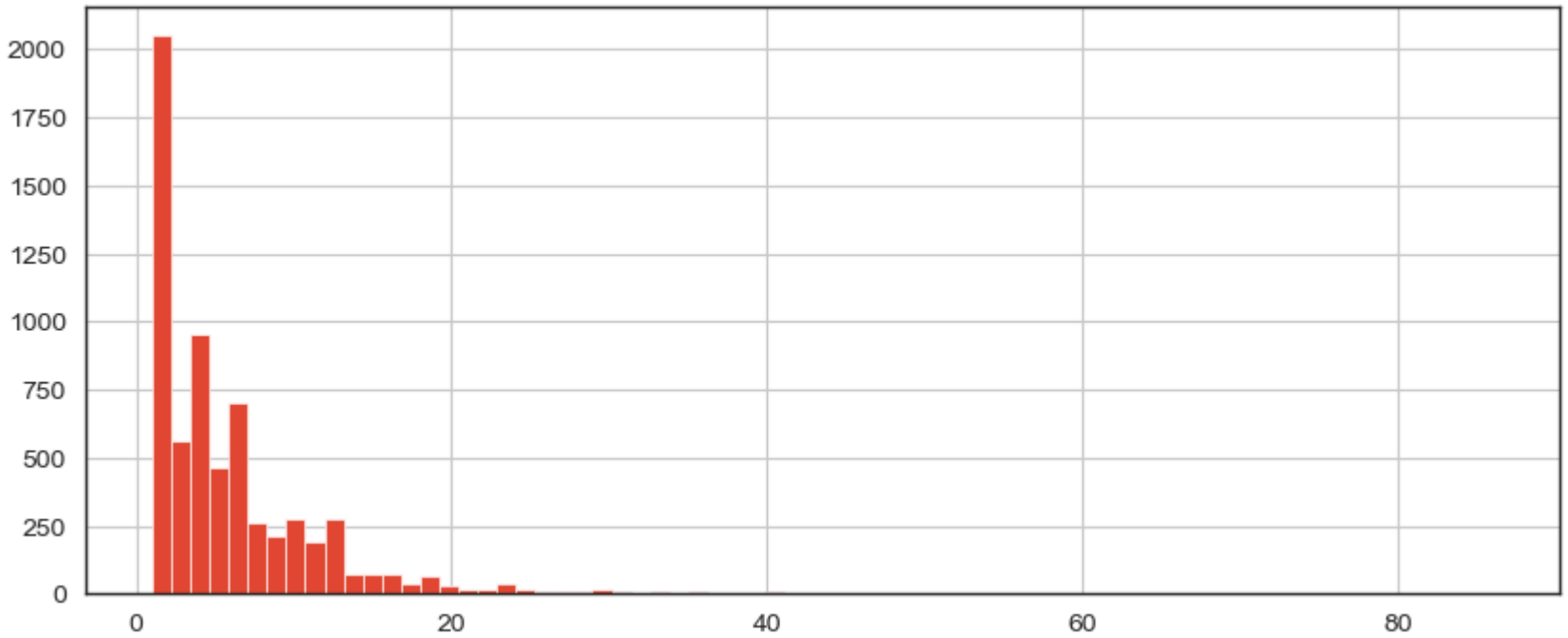
Out[161…

name	rate	num of ratings
#Feeltheroll	3.40	2
#L-81 Cafe	3.90	9
#Refuel	3.70	3
1000 B.C	3.20	6
100Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â°C	3.70	3

In [ ]:

In [162…
```
plt.figure(figsize=(10,4))
ratings['num of ratings'].hist(bins=70)
```

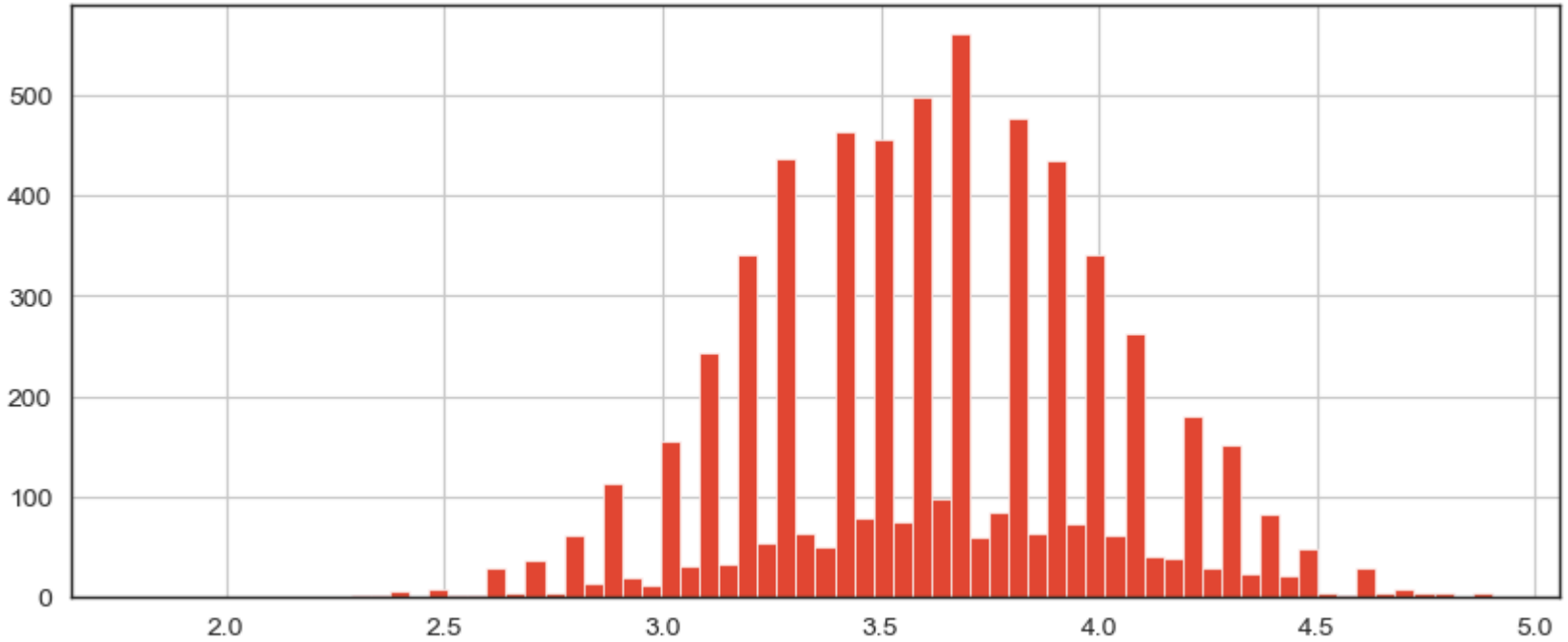
Out[162…
```
<Axes: >
```



In [ ]:

In [163…
```
plt.figure(figsize=(10,4))
ratings['rate'].hist(bins=70)
```

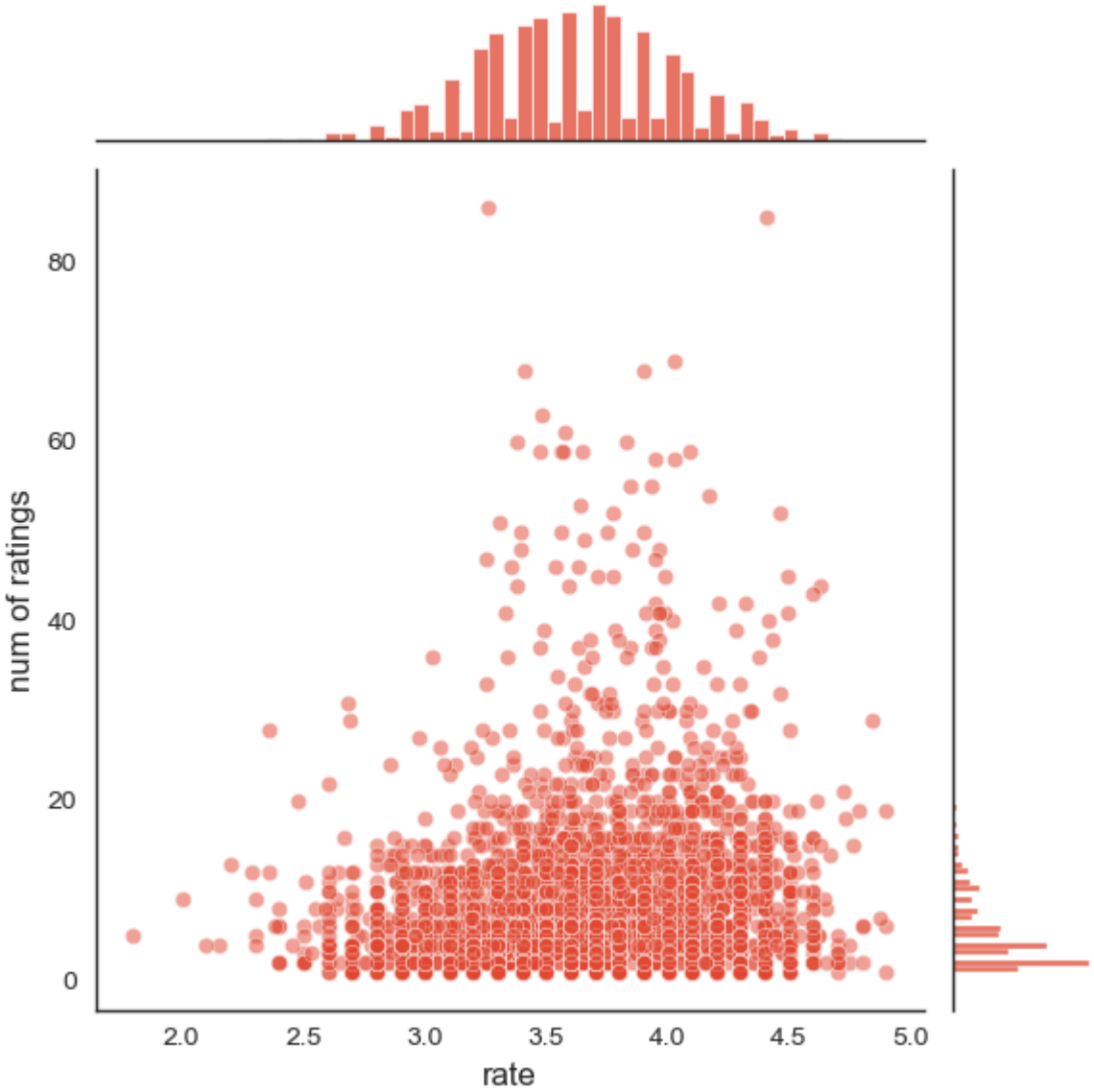
Out[163… <Axes: >



In [ ]:

In [164…
```
sns.jointplot(x='rate',y='num of ratings',data=ratings,alpha=0.5)
```

Out[164…
```
<seaborn.axisgrid.JointGrid at 0x1a7e5027410>
```



In [ ]:

In [165…

```
zomatomat = zomato.pivot_table(index='votes',columns='name',values='rate')
zomatomat.head()
```

Out[165…

name	#Feeltheroll	#L-81 Cafe	#Refuel	1000 B.C	100Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â°C	11 To 11 Express Biriyanis	1131 Bar + Kitchen	12Th Main - Grand Mercure	1441 Pizzeria	1522 - The Pub	Br
votes											
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

5 rows × 6503 columns

In [166…

```
ratings.sort_values('num of ratings',ascending=False).head(10)
```

Out[166…

name	rate	num of ratings
Cafe Coffee Day	3.26	86
Onesta	4.41	85
Empire Restaurant	4.03	69
Kanti Sweets	3.90	68
Just Bake	3.41	68
Five Star Chicken	3.48	63
Baskin Robbins	3.58	61
Pizza Hut	3.38	60
Petoo	3.83	60
Kfc	3.65	59

In [ ]:

In [167…
```
ratings.head()
```

Out[167…

name	rate	num of ratings
#Feeltheroll	3.40	2
#L-81 Cafe	3.90	9
#Refuel	3.70	3
1000 B.C	3.20	6
100Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â°C	3.70	3

In [ ]:

In [168…
```
Cafe_Coffee_Day_user_ratings = zomatomat['Cafe Coffee Day']
Onesta_user_ratings = zomatomat['Onesta']
Cafe_Coffee_Day_user_ratings.head(10)
```

Out[168…
```
votes
0     NaN
1     NaN
2     NaN
4     NaN
5     NaN
6    3.00
7    3.40
8     NaN
9    3.20
10   3.08
Name: Cafe Coffee Day, dtype: float64
```

In [ ]:

In [169…
```
similar_to_Cafe_Coffee_Day = zomatomat.corrwith(Cafe_Coffee_Day_user_ratings)
similar_to_Onesta = zomatomat.corrwith(Onesta_user_ratings)
```

In [ ]:

In [171…
```
corr_Cafe_Coffee_Day = pd.DataFrame(similar_to_Cafe_Coffee_Day,columns=['Correlation'])
corr_Cafe_Coffee_Day.dropna(inplace=True)
corr_Cafe_Coffee_Day.head(10)
```

Out[171…

name	Correlation
Abhiruchi Hotel	1.00
Abhiruchi Restaurant	1.00
Adithya	1.00
Ahimsaamrit Green Cafe	1.00
Al-Taj Restaurant	1.00
Altaj Restaurant	1.00
Amaravati Food Court	-1.00
Ambujam'S Chettinad Mess	-1.00
Ambur Dum Biriyani	-1.00
Ambur Hot Dum Biryani	-0.02

In [ ]:

```
```

In [172…

```
corr_Cafe_Coffee_Day.sort_values('Correlation',ascending=False).head(10)
```

Out[172…

name	Correlation
Abhiruchi Hotel	1.00
Cafe Get Hooked	1.00
Lassi Darbar	1.00
Mad Over Donuts	1.00
Calcutta Chats	1.00
Makers Of Milkshakes	1.00
Matsya Lunch Home	1.00
Cake Yard	1.00
Mayura Grand	1.00
Cake Art	1.00

In [ ]:

```
```

In [180…

```
corr_Cafe_Coffee_Day[corr_Cafe_Coffee_Day['num of ratings']>50].sort_values('Correlation',ascending=False).head()
```

Out[180…

name	Correlation	num of ratings
Cafe Coffee Day	1.00	86
Petoo	0.70	60
Beijing Bites	0.64	53
Baskin Robbins	0.35	61
Keventers	0.33	52

In [ ]:

In [184…
```
corr_Onesta = pd.DataFrame(similar_to_Onesta ,columns=['Correlation'])
corr_Onesta.dropna(inplace=True)
corr_Onesta = corr_Onesta .join(ratings['num of ratings'])
corr_Onesta[corr_Onesta ['num of ratings']>80].sort_values('Correlation',ascending=False).head(10)
```

Out[184…

name	Correlation	num of ratings
Onesta	1.00	85

In [ ]:

Using Phone AS User:-

In [ ]:

In [185…
```
# Convert the 'phone' column to string
zomato['phone'] = zomato['phone'].astype(str)

# Alternatively, you can use the apply method to clean up the phone numbers
zomato['phone'] = zomato['phone'].apply(lambda x: str(x).strip())

# View the result
print(zomato['phone'].head())
```

```
0      080 42297555\r\n+91 9743772233
1                          080 41714161
2                       +91 9663487993
3                       +91 9620009302
4    +91 8026612447\r\n+91 9901210005
Name: phone, dtype: object
```

In [ ]:

In [186…

```
def clean_phone(phone):
    # Removing everything from\n onword ,including the newline character
    return re.sub(r'\n.*', '', phone)

#Apply the function to the 'phones' column
zomato['phone'] = zomato['phone'].apply(clean_phone)

#show the updated dataframe
print(zomato)
```

```
                                                address  \
0      942, 21st Main Road, 2nd Stage, Banashankari, ...
1      2nd Floor, 80 Feet Road, Near Big Bazaar, 6th ...
2      1112, Next to KIMS Medical College, 17th Cross...
3      1st Floor, Annakuteera, 3rd Stage, Banashankar...
4      10, 3rd Floor, Lakshmi Associates, Gandhi Baza...
...                                                  ...
40658  136, SAP Labs India, KIADB Export Promotion In...
40659  139/C1, Next To GR Tech Park, Pattandur Agraha...
40660  Four Points by Sheraton Bengaluru, 43/3, White...
40661  Sheraton Grand Bengaluru Whitefield Hotel & Co...
40662  ITPL Main Road, KIADB Export Promotion Industr...

                                                  name  online_order  \
0                                                Jalsa          True
1                                       Spice Elephant          True
2                                      San Churro Cafe          True
3                                   Addhuri Udupi Bhojana         False
4                                           Grand Village         False
...                                                   ...           ...
40658                          The Farm House Bar N Grill         False
40659                                              Bhagini         False
40660  Best Brews - Four Points By Sheraton Bengaluru...         False
40661  Chime - Sheraton Grand Bengaluru Whitefield Ho...         False
40662                         The Nest - The Den Bengaluru         False

       book_table  rate  votes             phone                   location  \
0            True  4.10    775    080 42297555\r                Banashankari
1           False  4.10    787      080 41714161                Banashankari
2           False  3.80    918    +91 9663487993                Banashankari
3           False  3.70     88    +91 9620009302                Banashankari
4           False  3.80    166  +91 8026612447\r                Basavanagudi
...           ...   ...    ...               ...                         ...
40658       False  3.70     34    +91 9980121279                  Whitefield
40659       False  2.50     81      080 65951222                  Whitefield
40660       False  3.60     27      080 40301477                  Whitefield
40661        True  4.30    236      080 49652769  ITPL Main Road, Whitefield
40662       False  3.40     13    +91 8071117272  ITPL Main Road, Whitefield

                rest_type                                   cuisines  \
0            Casual Dining              North Indian, Mughlai, Chinese
```

```
1            Casual Dining                  Chinese, North Indian, Thai
2      Cafe, Casual Dining                       Cafe, Mexican, Italian
3              Quick Bites                   South Indian, North Indian
4            Casual Dining                     North Indian, Rajasthani
...                    ...                                          ...
40658   Casual Dining, Bar                     North Indian, Continental
40659   Casual Dining, Bar  Andhra, South Indian, Chinese, North Indian
40660                  Bar                                   Continental
40661                  Bar                                   Finger Food
40662   Bar, Casual Dining       Finger Food, North Indian, Continental

        cost                                       reviews_list menu_item  \
0     800.00  [('Rated 4.0', 'RATED\n  A beautiful place to ...        []
1     800.00  [('Rated 4.0', 'RATED\n  Had been here for din...        []
2     800.00  [('Rated 3.0', "RATED\n  Ambience is not that ...        []
3     300.00  [('Rated 4.0', "RATED\n  Great food and proper...        []
4     600.00  [('Rated 4.0', 'RATED\n  Very good restaurant ...        []
...      ...                                                ...       ...
40658 800.00  [('Rated 4.0', 'RATED\n  Ambience- Big and spa...        []
40659 800.00  [('Rated 4.0', 'RATED\n  A fine place to chill...        []
40660   1.50  [('Rated 5.0', "RATED\n  Food and service are ...        []
40661   2.50  [('Rated 4.0', 'RATED\n  Nice and friendly pla...        []
40662   1.50  [('Rated 5.0', 'RATED\n  Great ambience , look...        []

               type          city  Mean Rating
0             Buffet  Banashankari         4.12
1             Buffet  Banashankari         4.10
2             Buffet  Banashankari         3.80
3             Buffet  Banashankari         3.70
4             Buffet  Banashankari         3.80
...              ...           ...          ...
40658  Pubs and bars    Whitefield         3.70
40659  Pubs and bars    Whitefield         2.28
40660  Pubs and bars    Whitefield         3.60
40661  Pubs and bars    Whitefield         4.30
40662  Pubs and bars    Whitefield         3.40

[40663 rows x 16 columns]
```

In [ ]:

In [122…
```
zomatomat = zomato.pivot_table(index='phone',columns='name',values='rate')
zomatomat.head()
```

Out[122…

phone	name	#Feeltheroll	#L-81 Cafe	#Refuel	1000 B.C	100Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â°C	11 To 11 Express Biriyanis	1131 Bar + Kitchen	12Th Main - Grand Mercure	1441 Pizzeria
+91 6281959891\r		NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
+91 6281959891\r\r		NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
+91 6360029505		NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
+91 6360086653		NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
+91 6360116461		NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

5 rows × 6503 columns

In [ ]:
```
```

Most rating Restaurant:-

In [123…
```
ratings.sort_values('num of ratings',ascending=False).head(10)
```

Out[123...

name	rate	num of ratings
Cafe Coffee Day	3.26	86
Onesta	4.41	85
Empire Restaurant	4.03	69
Kanti Sweets	3.90	68
Just Bake	3.41	68
Five Star Chicken	3.48	63
Baskin Robbins	3.58	61
Pizza Hut	3.38	60
Petoo	3.83	60
Kfc	3.65	59

In [ ]:

In [124...
```
ratings.head()
```

Out[124...

name	rate	num of ratings
#Feeltheroll	3.40	2
#L-81 Cafe	3.90	9
#Refuel	3.70	3
1000 B.C	3.20	6
100Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â□Ã□Â°C	3.70	3

In [ ]:

In [125…
```
Cafe_Coffee_Day_user_ratings = zomatomat['Cafe Coffee Day']
Onesta_user_ratings = zomatomat['Onesta']
Cafe_Coffee_Day_user_ratings.head()
```

Out[125…
```
phone
+91 6281959891\r      NaN
+91 6281959891\r\r    NaN
+91 6360029505        NaN
+91 6360086653        NaN
+91 6360116461        NaN
Name: Cafe Coffee Day, dtype: float64
```

In [ ]:

In [126…
```
similar_to_Cafe_Coffee_Day = zomatomat.corrwith(Cafe_Coffee_Day_user_ratings)
similar_to_Onesta = zomatomat.corrwith(Onesta_user_ratings)
```

In [ ]:

In [127…
```
corr_Cafe_Coffee_Day = pd.DataFrame(similar_to_Cafe_Coffee_Day,columns=['Correlation'])
corr_Cafe_Coffee_Day.dropna(inplace=True)
corr_Cafe_Coffee_Day.head()
```

Out[127…

name	Correlation
Cafe Coffee Day	1.00

In [ ]:

In [128…
```
corr_Cafe_Coffee_Day.sort_values('Correlation',ascending=False).head(10)
```

Out[128…

name	Correlation
Cafe Coffee Day	1.00

In [ ]:

In [129…
```
corr_Cafe_Coffee_Day = corr_Cafe_Coffee_Day.join(ratings['num of ratings'])
corr_Cafe_Coffee_Day.head()
```

Out[129…

name	Correlation	num of ratings
Cafe Coffee Day	1.00	86

In [ ]:

In [130…
```
corr_Cafe_Coffee_Day[corr_Cafe_Coffee_Day['num of ratings']>50].sort_values('Correlation',ascending=False).head()
```

Out[130…

name	Correlation	num of ratings
Cafe Coffee Day	1.00	86

In [ ]:

In [131…
```
corr_Onesta = pd.DataFrame(similar_to_Onesta ,columns=['Correlation'])
corr_Onesta.dropna(inplace=True)
corr_Onesta = corr_Onesta .join(ratings['num of ratings'])
corr_Onesta[corr_Onesta ['num of ratings']>80].sort_values('Correlation',ascending=False).head()
```

Out[131…

name	Correlation	num of ratings
Onesta	1.00	85

In [ ]:

In [189…

```
# Convert 'phone' column to string and handle NaN values
df['phone'] = df['phone'].astype(str).apply(lambda x: x.split("\r\n")[0] if isinstance(x, str) and "\r\n" in x else x)

# Convert ratings to float (handle missing values safely)
df['rate'] = pd.to_numeric(df['rate'].str.split('/').str[0], errors='coerce')

# Display the first few rows
print(df[['name', 'phone', 'rate']].head())
```

```
                   name           phone  rate
0                 Jalsa    080 42297555  4.10
1          Spice Elephant    080 41714161  4.10
2        San Churro Cafe  +91 9663487993  3.80
3  Addhuri Udupi Bhojana  +91 9620009302  3.70
4           Grand Village  +91 8026612447  3.80
```

In [ ]:

In [191…

```
# Ensure 'phone' is treated as a string and handle NaN values
df['phone'] = df['phone'].astype(str).apply(lambda x: x.split("\r\n")[0] if isinstance(x, str) and "\r\n" in x else x)

# Convert ratings to float (handling missing values safely)
df['rate'] = pd.to_numeric(df['rate'].str.split('/').str[0], errors='coerce')

# Display first few rows
print(df[['name', 'phone', 'rate']].head())
```

```
                   name           phone  rate
0                 Jalsa    080 42297555  4.10
1         Spice Elephant    080 41714161  4.10
2        San Churro Cafe  +91 9663487993  3.80
3  Addhuri Udupi Bhojana  +91 9620009302  3.70
4          Grand Village  +91 8026612447  3.80
```

In [ ]:

In [194…

```python
#Removing '/5' from Rates

import numpy as np

# Filter out 'NEW' and '-' from the 'rate' column
df= df.loc[df.rate != 'NEW']
df = df.loc[df.rate != '-'].reset_index(drop=True)

# Define a lambda function to remove '/5' and ensure it's applied correctly
remove_slash = lambda x: x.replace('/5', '') if isinstance(x, str) else x

# Apply the function and convert the 'rate' column to float after cleaning
df['rate'] = df['rate'].apply(remove_slash).str.strip()

# Convert the 'rate' column to float, forcing errors to NaN (if any) and drop those rows if necessary
df['rate'] = pd.to_numeric(df['rate'], errors='coerce')

# Display the first few rows of the 'rate' column
df['rate'].head()
```

Out[194…

```
0   4.10
1   4.10
2   3.80
3   3.70
4   3.80
Name: rate, dtype: float64
```

In [ ]:

In [ ]:

```python
import pandas as pd
import numpy as np
```

```python
from sklearn.metrics.pairwise import cosine_similarity

# Load the dataset
df = pd.read_csv('zomato.csv')

# Preprocess the phone number (keep only the first phone number)
df['phone'] = df['phone']

# Convert ratings to float
df['rate'] = df['rate'].str.split('/').str[0].astype(float)

# Create the user-item matrix
pivot_table = df.pivot_table(index='phone', columns='name', values='rate').fillna(0)

# Compute similarity between users
similarity_matrix = cosine_similarity(pivot_table)

# Convert similarity matrix to DataFrame
similarity_df = pd.DataFrame(similarity_matrix, index=pivot_table.index, columns=pivot_table.index)

# Function to recommend restaurants based on user similarity
def get_recommendations(phone, pivot_table, similarity_df):
    if phone not in similarity_df.index:
        return f"No recommendations found for phone: {phone}"

    # Find similar users
    similar_users = similarity_df[phone].sort_values(ascending=False).index[1:]  # Exclude the user itself

    # Aggregate restaurant ratings from similar users
    restaurant_scores = pivot_table.loc[similar_users].mean().sort_values(ascending=False)

    # Recommend top restaurants
    recommended_restaurants = restaurant_scores.index.tolist()
    return recommended_restaurants

# Example usage
phone_number = '080 42297555'  # Replace with any phone number from the dataset
recommendations = get_recommendations(phone_number, pivot_table, similarity_df)

print(f"Recommended restaurants for {phone_number}: {recommendations}")
```

In [ ]:

END OF THE PROJECT

In [ ]: