An Analysis of Online Food Delivery Service: Zomato Restaurant data

TEAM NUMBER -4

TEAM MEMBERS :

MADASU DEEPIKA(19MIA1066)

G.HARINISRI(19MIA1069)

School of Computer Science and Engineering

CERTIFICATE

The project report entitled **"An Analysis of Online Food Delivery Service: Zomato Restaurant data"** is prepared and submitted by **MADASU DEEPIKA (19MIA1066)**
G.HARINISRI(19MIA1069)

It has been found satisfactory in terms of scope, quality, and presentation as partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology – Computer Science and Engineering** in Vellore Institute of Technology, Chennai, India.

Examined by:

Dr. TULASI PRASAD SARIKI

ABSTRACT:

Food technology in a broad area, online food delivery apps are just part of it. This conceptual study will give more insight into emerging innovative technologies in the restaurant industry and strategies followed by online food start-ups Zomato.

From this research paper, we would understand the drivers of online food sites. Different services are given by application that makes consumers happy and satisfied. Comfort and Convenience makes consumer more inclined towards online food ordering.

The research concluded that due to urbanization in the Indian landscape, online food delivery applications are growing with flying colors. The future of online food ordering websites is bright. Facilities, Comfort, User-friendliness are the key features of the success of the online website.

Here we will analysing and studying the zomato dataset with help of exploratory data analysis , visualize them with help graphs and charts, we will be finding three main objectives and provide a recommendation system based on restaurant name and reviews text.

INTRODUCTION:

The food delivery industry is a rapidly growing sector ripe for data-based development. Everything is quantifiable, from delivery times and zip codes to prices and customer satisfaction. These data points can be collated and processed to improve operations and profitability while cutting down on loss.

Zomato, one of the largest players in the industry, uses demographic data to make intelligent predictions about what offerings might interest a particular user, then serves that customer ads for that item.

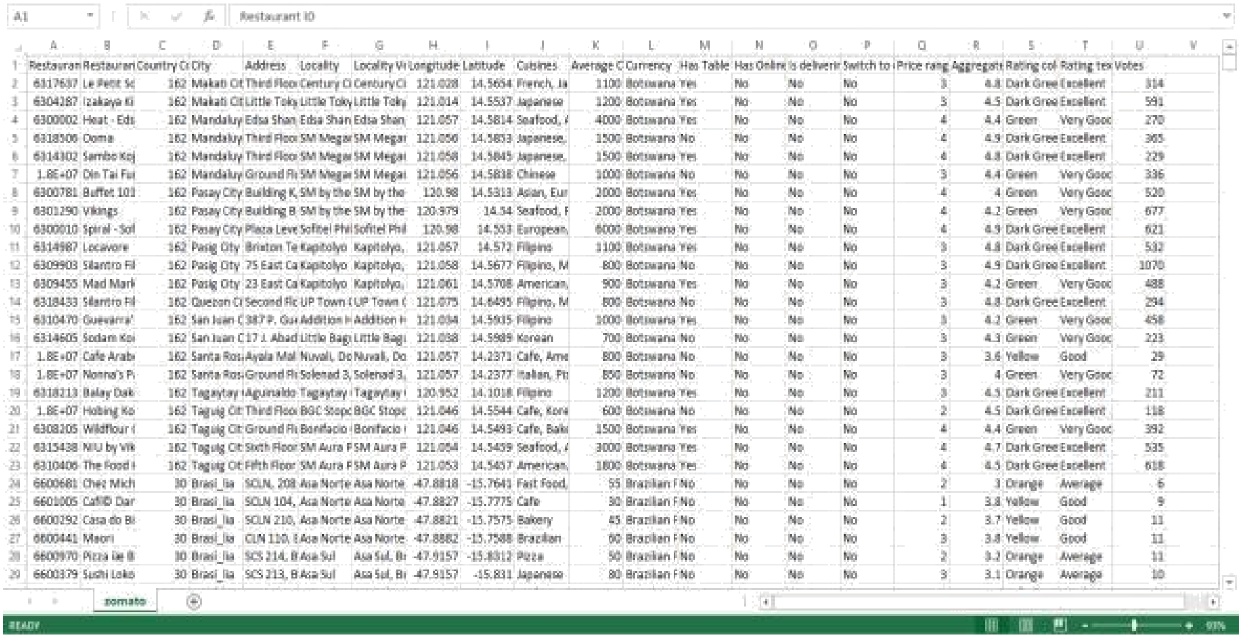
It looks at relationships between variables like time of day, day of the week, and expected prep time of a food item at particular restaurants (along with other factors such as high-profile sports events and weather events) to predict the time a user will have to wait for their food.

Also use predictive analytics to plan how many drivers they will need for particular shifts and offer incentives when drivers are in demand, and they monitor customer satisfaction with particular items to estimate which food items are likely to be in high demand soon.

DATASET DESCRIPTION AND TOOLS:

The dataset used for this project was found on an online public data platform called Kaggle.

It was relatively clean and did not require any major changes.



A1 | Restaurant ID

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Restauran	Restauran	Country Co	City	Address	Locality	Locality Vi	Longitude	Latitude	Cuisines	Average C	Currency	Has Table	Has Online	Is deliveri	Switch to	Price rang	Aggregate	Rating col	Rating tex	Votes
2	6317637	Le Petit So	162	Makati Cit	Third Floor	Century Ci	Century Ci	121.028	14.5654	French, Ja	1100	Botswana	Yes	No	No	No	3	4.8	Dark Gree	Excellent	314
3	6304287	Izakaya Ki	162	Makati Cit	Little Toky	Little Toky	Little Toky	121.014	14.5537	Japanese	1200	Botswana	Yes	No	No	No	3	4.5	Dark Gree	Excellent	591
4	6300002	Heat - Eds	162	Mandaluy	Edsa Shan	Edsa Shan	Edsa Shan	121.057	14.5814	Seafood, A	4000	Botswana	Yes	No	No	No	4	4.4	Green	Very Gooc	270
5	6318506	Ooma	162	Mandaluy	Third Floor	SM Megar	SM Megar	121.056	14.5853	Japanese,	1500	Botswana	No	No	No	No	4	4.9	Dark Gree	Excellent	365
6	6314302	Sambo Koj	162	Mandaluy	Third Floor	SM Megar	SM Megar	121.058	14.5845	Japanese,	1500	Botswana	Yes	No	No	No	4	4.8	Dark Gree	Excellent	229
7	1.8E+07	Din Tai Fu	162	Mandaluy	Ground Fl	SM Megar	SM Megar	121.056	14.5838	Chinese	1000	Botswana	No	No	No	No	3	4.4	Green	Very Gooc	336
8	6300781	Buffet 101	162	Pasay City	Building K,	SM by the	SM by the	120.98	14.5313	Asian, Eur	2000	Botswana	Yes	No	No	No	4	4	Green	Very Gooc	520
9	6301290	Vikings	162	Pasay City	Building B,	SM by the	SM by the	120.979	14.54	Seafood, F	2000	Botswana	Yes	No	No	No	4	4.2	Green	Very Gooc	677
10	6300010	Spiral - Sof	162	Pasay City	Plaza Leve	Sofitel Phi	Sofitel Phi	120.98	14.553	European,	6000	Botswana	Yes	No	No	No	4	4.9	Dark Gree	Excellent	621
11	6314987	Locavore	162	Pasig City	Brixton Te	Kapitolyo	Kapitolyo,	121.057	14.572	Filipino	1100	Botswana	Yes	No	No	No	3	4.8	Dark Gree	Excellent	532
12	6309903	Silantro Fil	162	Pasig City	75 East Ca	Kapitolyo	Kapitolyo,	121.058	14.5677	Filipino, M	800	Botswana	No	No	No	No	3	4.9	Dark Gree	Excellent	1070
13	6309455	Mad Mark	162	Pasig City	23 East Ca	Kapitolyo	Kapitolyo,	121.061	14.5708	American,	900	Botswana	Yes	No	No	No	3	4.2	Green	Very Gooc	488
14	6318433	Silantro Fil	162	Quezon Ci	Second Flo	UP Town C	UP Town C	121.075	14.6495	Filipino, M	800	Botswana	No	No	No	No	3	4.8	Dark Gree	Excellent	294
15	6310470	Guevarra'	162	San Juan C	387 P. Gu	Addition H	Addition H	121.034	14.5935	Filipino	1000	Botswana	Yes	No	No	No	3	4.2	Green	Very Gooc	458
16	6314605	Sodam Ko	162	San Juan C	17 J. Abad	Little Bag	Little Bag	121.038	14.5989	Korean	700	Botswana	No	No	No	No	3	4.3	Green	Very Gooc	223
17	1.8E+07	Cafe Arab	162	Santa Ros	Ayala Mal	Nuvali, Do	Nuvali, Do	121.057	14.2371	Cafe, Ame	800	Botswana	No	No	No	No	3	3.6	Yellow	Good	29
18	1.8E+07	Nonna's P	162	Santa Ros	Ground Fl	Solenad 3,	Solenad 3,	121.057	14.2377	Italian, Pi	850	Botswana	No	No	No	No	3	4	Green	Very Gooc	72
19	6318213	Balay Dak	162	Tagaytay	Aguinaldo	Tagaytay	Tagaytay	120.952	14.1018	Filipino	1200	Botswana	Yes	No	No	No	3	4.5	Dark Gree	Excellent	211
20	1.8E+07	Hobing Ko	162	Taguig Cit	Third Floor	BGC Stopc	BGC Stopc	121.046	14.5544	Cafe, Kore	600	Botswana	No	No	No	No	2	4.5	Dark Gree	Excellent	118
21	6308205	Wildflour (	162	Taguig Cit	Ground Fl	Bonifacio	Bonifacio	121.046	14.5493	Cafe, Bak	1500	Botswana	Yes	No	No	No	4	4.4	Green	Very Gooc	392
22	6315438	NIU by Vik	162	Taguig Cit	Sixth Floor	SM Aura P	SM Aura P	121.054	14.5459	Seafood, A	3000	Botswana	Yes	No	No	No	4	4.7	Dark Gree	Excellent	535
23	6310406	The Food	162	Taguig Cit	Fifth Floor	SM Aura P	SM Aura P	121.053	14.5457	American,	1800	Botswana	Yes	No	No	No	4	4.5	Dark Gree	Excellent	618
24	6600681	Chez Mich	30	Brasí_lia	SCLN, 208	Asa Norte	Asa Norte	-47.8818	-15.7641	Fast Food,	55	Brazilian F	No	No	No	No	2	3	Orange	Average	6
25	6601005	CaffÌÄ Dar	30	Brasí_lia	SCLN 104,	Asa Norte	Asa Norte	-47.8827	-15.7775	Cafe	30	Brazilian F	No	No	No	No	1	3.8	Yellow	Good	9
26	6600292	Casa do Bi	30	Brasí_lia	SCLN 210,	Asa Norte	Asa Norte	-47.8821	-15.7575	Bakery	45	Brazilian F	No	No	No	No	2	3.7	Yellow	Good	11
27	6600441	Maori	30	Brasí_lia	CLN 110, B	Asa Norte	Asa Norte	-47.8882	-15.7588	Brazilian	60	Brazilian F	No	No	No	No	3	3.8	Yellow	Good	11
28	6600970	Pizza ãe B	30	Brasí_lia	SCS 214, B	Asa Sul	Asa Sul, Br	-47.9157	-15.8312	Pizza	50	Brazilian F	No	No	No	No	2	3.2	Orange	Average	11
29	6600379	Sushi Loko	30	Brasí_lia	SCS 213, B	Asa Sul	Asa Sul, Br	-47.9157	-15.831	Japanese	80	Brazilian F	No	No	No	No	3	3.1	Orange	Average	10

zomato

READY

The collected data has been stored in the Comma Separated Value file Zomato.csv.

Each restaurant in the dataset is uniquely identified by its Restaurant Id. Every

Restaurant contains the following variables:

- Restaurant Id: Unique id of every restaurant across various cities of the world
- Restaurant Name: Name of the restaurant
- Country Code: Country in which restaurant is located
- City: City in which restaurant is located
- Address: Address of the restaurant
- Locality: Location in the city
- Locality Verbose: Detailed description of the locality
- Longitude: Longitude coordinate of the restaurant's location
- Latitude: Latitude coordinate of the restaurant's location
- Cuisines: Cuisines offered by the restaurant
- Average Cost for two: Cost for two people in different currencies
- Currency: Currency of the country
- Has Table booking: yes/no
- Has Online delivery: yes/ no
- Is delivering: yes/ no
- Switch to order menu: yes/no
- Price range: range of price of food
- Aggregate Rating: Average rating out of 5
- Rating color: depending upon the average rating color

• Rating text: text on the basis of rating of rating

• Votes: Number of ratings casted by people

The tool that will be used is google colab/jupyter python for data analysis.

OBJECTIVE:

- To evaluate the Highest Rated and Lowest Rated Restaurant of the City in all the countries. Graph plotted only for countries with maximum restaurants (India and U.S.A).
- To evaluate the most popular cuisine of the world sold in a country and which locality in that country has most number of outlets selling that cuisine.
- To evaluate the value for money restaurants in the U.S.A for the best cuisines served in the cities of U.S.A(value for money refers to the restaurants with highest rating and lowest cost).
- To provide recommendation system with the help of natural language processing.

METHODODLOGY:

- Using pandas, grouping 'Country' and 'City', the aggregate rating is calculated and then the top and least rated restaurants are found for every city in that country.
- Using pandas, splitting the multiple cuisines and stacking them up with the location, melting them with locality, grouping the locality and then country code and then forward sorting the average cost for two and reverse sorting the rating in the country with the names of the restaurants
- Using pandas, splitting the multiple cuisines and stacking them up with the location, melting them with locality,

grouping the locality and then country code and then reverse sorting the count of restaurants in the country with their names.

- **For recommendation system :**

Data Cleaning:

- ➢ Deleting redundant columns.
- ➢ Renaming the columns.
- ➢ Dropping duplicates.
- ➢ Cleaning individual columns.
- ➢ Remove the NaN values from the dataset
- ➢ Some Transformations

Text Preprocessing

- Cleaning unnecessary words in the review.
- Removing links and other unncessary items
- Removing Symbols

Recommendation System

We will develop a restaurant recommendation system in Python using Matrix Factorization or Latent Factor Collaborative Filtering by considering the attributes user Id, business Id, rating and review text.

RESULT AND DISCUSSIONS:

STEP 1: IMPORT THE PACKAGES AND LIBRARIES REQUIRED AND THE DATASET IS LOADED.

IMPORT THE PACKAGES AND LIBRARIES REQUIRED

In [1]:

```
import pandas as pd
from pandas import DataFrame
import warnings
warnings.filterwarnings('ignore')
```

IMPORTING THE ZOMATO.CSV DATASET

In [2]:

```
df=pd.read_csv('zomato.csv',sep=',', encoding='latin-1') #read csv
df.head()
```

Out[2]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	...	Currency	Has Table booking	Has Online delivery	Is delivering now	Switch to order menu	Price range	Aggregate rating	Rating color
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City	Century City Mall, Poblacion, Makati City, Mak...	121.027535	14.565443	French, Japanese, Desserts	...	Botswana Pula(P)	Yes	No	No	No	3	4.8	Dark Green
1	6304287	Izakaya Kikufuji	162	Makati City	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...	Little Tokyo, Legaspi Village, Makati City	Little Tokyo, Legaspi Village, Makati City, Ma...	121.014101	14.553708	Japanese	...	Botswana Pula(P)	Yes	No	No	No	3	4.5	Dark Green
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City	Edsa Shangri-La, 1 Garden W...	Edsa Shangri-La, Ortigas, Mandaluyong	Edsa Shangri-La, Ortigas, Mandaluyong	121.056831	14.581404	Seafood, Asian, Filipino,	...	Botswana Pula(P)	Yes	No	No	No	4	4.4	Green

STEP 2: GROUPING THE NECESSARY COLUMNS

In [3]:

```
print((df.groupby("Country Code")["Restaurant Name"].count().sort_values(ascending = False))[:5])
print((df.groupby("City")["Restaurant Name"].count().sort_values(ascending = False))[:5])
```

```
Country Code
1      8652
216     434
215      80
214      60
189      60
Name: Restaurant Name, dtype: int64
City
New Delhi     5473
Gurgaon       1118
Noida         1080
Faridabad      251
Ghaziabad       25
Name: Restaurant Name, dtype: int64
```

In [4]:

```
df.groupby('Cuisines')['Aggregate rating'].mean()
```

Out[4]:

```
Cuisines
Afghani                                    0.725
Afghani, Mughlai, Chinese                  0.000
Afghani, North Indian                      0.000
Afghani, North Indian, Pakistani, Arabian  0.000
African                                    4.700
                                            ...
Western, Asian, Cafe                       4.200
Western, Fusion, Fast Food                 3.200
World Cuisine                              4.900
World Cuisine, Mexican, Italian            4.400
World Cuisine, Patisserie, Cafe            4.200
Name: Aggregate rating, Length: 1825, dtype: float64
```

STEP 3: PRINTING THE SHAPE OF DATASET,DESCRIBING IT AND PROVIDING INFO ABOUT THE DATASET.

```
In [5]: df.shape
Out[5]: (9551, 21)
In [6]: df.describe()
```

Out[6]:

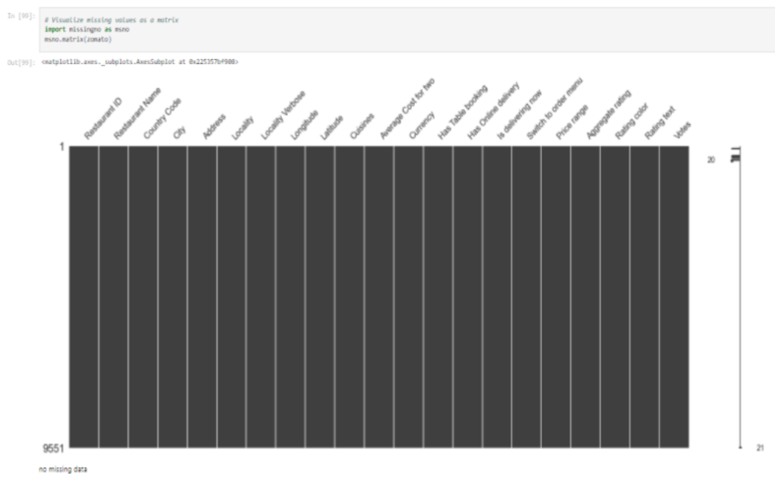
	Restaurant ID	Country Code	Longitude	Latitude	Average Cost for two	Price range	Aggregate rating	Votes
count	9.551000e+03	9551.000000	9551.000000	9551.000000	9551.000000	9551.000000	9551.000000	9551.000000
mean	9.051128e+06	18.365616	64.126574	25.854381	1199.210763	1.804837	2.666370	156.909748
std	8.791521e+06	56.750546	41.467058	11.007935	16121.183073	0.905609	1.516378	430.169145
min	5.300000e+01	1.000000	-157.948486	-41.330428	0.000000	1.000000	0.000000	0.000000
25%	3.019625e+05	1.000000	77.081343	28.478713	250.000000	1.000000	2.500000	5.000000
50%	6.004089e+06	1.000000	77.191964	28.570469	400.000000	2.000000	3.200000	31.000000
75%	1.835229e+07	1.000000	77.282006	28.642758	700.000000	2.000000	3.700000	131.000000
max	1.850065e+07	216.000000	174.832089	55.976980	800000.000000	4.000000	4.900000	10934.000000

```
In [7]: df.info()
&lt;class 'pandas.core.frame.DataFrame'&gt;
RangeIndex: 9551 entries, 0 to 9550
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Restaurant ID         9551 non-null   int64
 1   Restaurant Name       9551 non-null   object
 2   Country Code          9551 non-null   int64
 3   City                  9551 non-null   object
 4   Address               9551 non-null   object
 5   Locality              9551 non-null   object
 6   Locality Verbose      9551 non-null   object
 7   Longitude             9551 non-null   float64
 8   Latitude              9551 non-null   float64
 9   Cuisines              9542 non-null   object
```

STEP 4: TO CHECK IF THE DATA HAS ANY NULL VALUES AND DOING ANALYSIS TO CHECK THE MISSING VALUES.

```
In [8]: df.isnull().sum()

Out[8]: Restaurant ID           0
        Restaurant Name         0
        Country Code            0
        City                    0
        Address                 0
        Locality                0
        Locality Verbose        0
        Longitude               0
        Latitude                0
        Cuisines                9
        Average Cost for two    0
        Currency                0
        Has Table booking       0
        Has Online delivery     0
        Is delivering now       0
        Switch to order menu    0
        Price range             0
        Aggregate rating        0
        Rating color            0
        Rating text             0
        Votes                   0
        dtype: int64
```

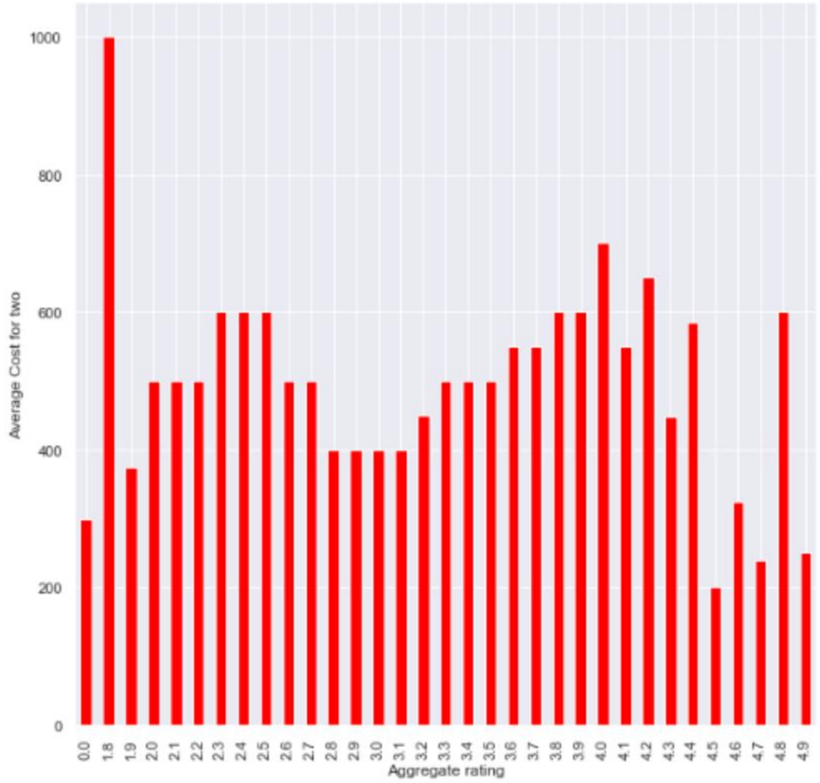


From this we can find that there are no missing data and no special preprocessing is required.

STEP 5: TO VISUALISE THE DATASET

In [9]:

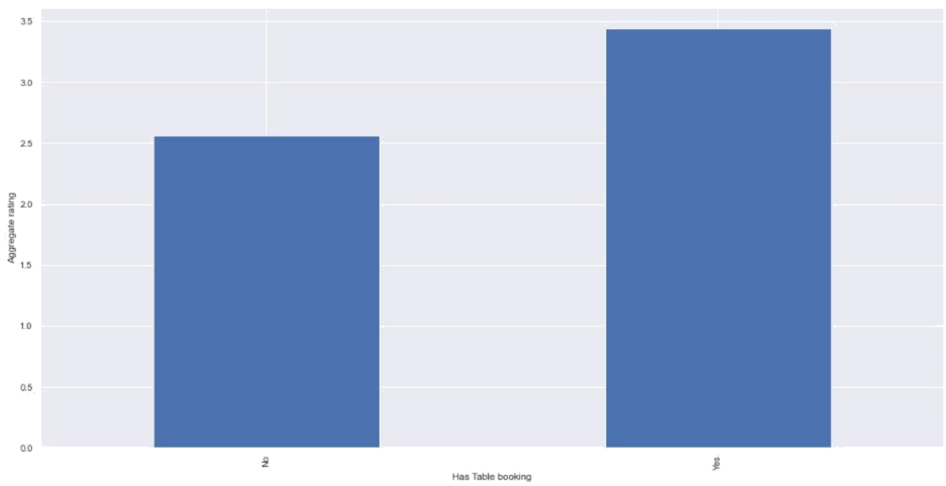
```
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(rc={'figure.figsize':(10, 8)});
df.groupby('Aggregate rating')['Average Cost for two'].median().plot(kind='bar', figsize=(10,10), color='red')
plt.xlabel('Aggregate rating')
plt.ylabel('Average Cost for two')
plt.figure();
```



From the above graph we can find the in aggregrating rating 1.5 rating is highest with average cost of two.

In [10]:

```
df.groupby('Has Table booking')['Aggregate rating'].mean().plot(kind='bar', figsize=(20,10))
plt.ylabel('Aggregate rating')
plt.show();
```



From the above graph we can find the who has table booking with the help of aggregrate rating and it happens to be that with the aggregrating rate of 3.5 many people optes for table booking.

We will understanding the rating in a simpler way ,

```
ratings = df.groupby(['Aggregate rating','Rating color', 'Rating text']).size().reset_index().rename(columns={0:'Rating Count'}) ratings
```

The information helps us to understand the relation between 'Aggregate rating', 'color' and 'text'. We conclude the following color assigned to the ratings: Rating 0 - White - Not rated

Rating 1.8 to 2.4 - Red - Poor Rating 2.5 to

3.4 - Orange – Average Rating 3.5 to 3.9 -

Yellow - Good Rating 4.0 to 4.4 - Green -

Very Good Rating 4.5 to 4.9 - Dark Green -

Excellent

Out[13]:

	Aggregate rating	Rating color	Rating text	Rating Count
0	0.0	White	Not rated	2148
1	1.8	Red	Poor	1
2	1.9	Red	Poor	2
3	2.0	Red	Poor	7
4	2.1	Red	Poor	15
5	2.2	Red	Poor	27
6	2.3	Red	Poor	47
7	2.4	Red	Poor	87
8	2.5	Orange	Average	110
9	2.6	Orange	Average	191
10	2.7	Orange	Average	250
11	2.8	Orange	Average	315
12	2.9	Orange	Average	381
13	3.0	Orange	Average	468
14	3.1	Orange	Average	519
15	3.2	Orange	Average	522
16	3.3	Orange	Average	483
17	3.4	Orange	Average	498
18	3.5	Yellow	Good	480
19	3.6	Yellow	Good	458
20	3.7	Yellow	Good	427
21	3.8	Yellow	Good	400
22	3.9	Yellow	Good	335
23	4.0	Green	Very Good	266
24	4.1	Green	Very Good	274
25	4.2	Green	Very Good	221
26	4.3	Green	Very Good	174

We will be understanding the currency in simpler way ,

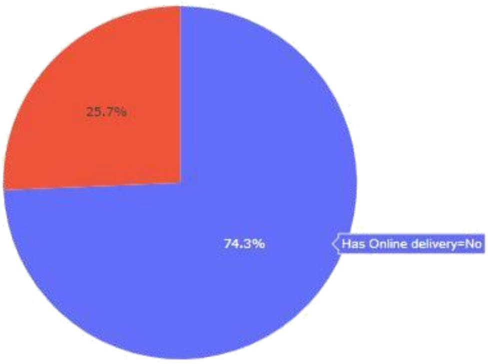
```
country_currency = df[['Country Code','Currency']].groupby(['Country
Code','Currency']).size().reset_index(name='count').drop('count', axis=1,
inplace=False)
country_currency
```

Out[14]:

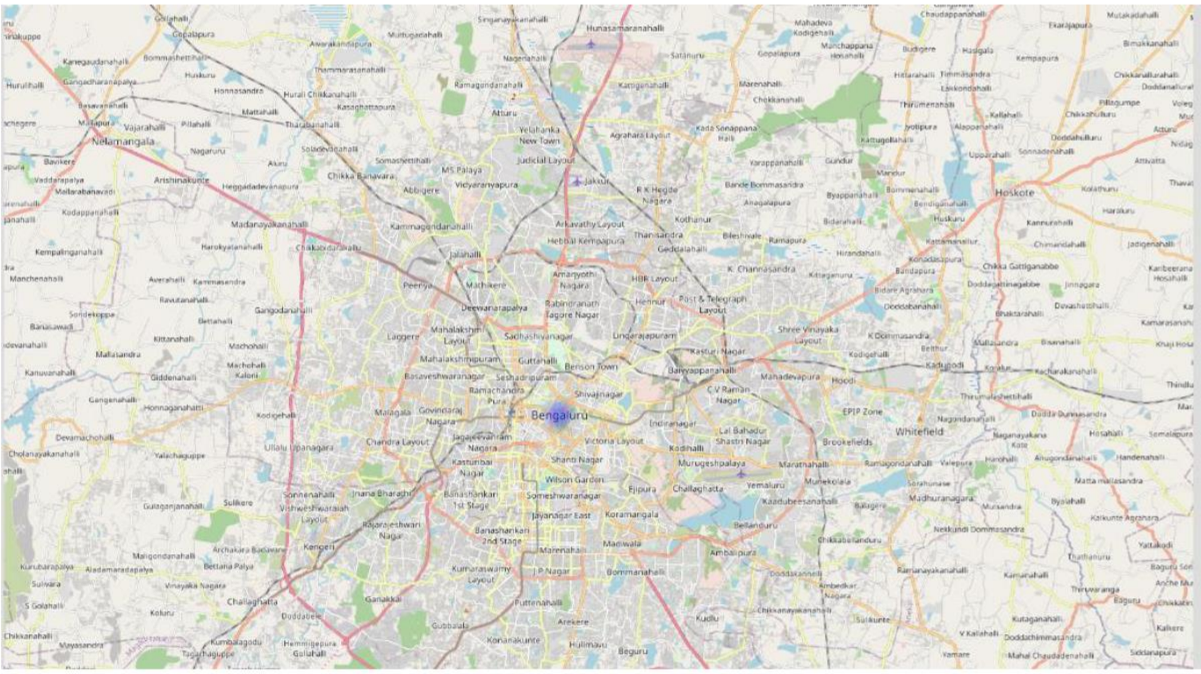
	Country Code	Currency
0	1	Indian Rupees(Rs.)
1	14	Dollar($)
2	30	Brazilian Real(R$)
3	37	Dollar($)
4	94	Indonesian Rupiah(IDR)
5	148	NewZealand($)
6	162	Botswana Pula(P)
7	166	Qatari Rial(QR)
8	184	Dollar($)
9	189	Rand(R)
10	191	Sri Lankan Rupee(LKR)
11	208	Turkish Lira(TL)
12	214	Emirati Diram(AED)
13	215	Pounds(£)
14	216	Dollar($)

Whether online delivery available or not?

Blue indicates no that occupies 74.3% and red indicates yes that occupies 25.7% . Therefore we are able to find that many people didn't opt for online delivery.



With the help of heat map and geo cluster we are able to find the zomato restaurants in any part of world with the help of markers.



Using wordcloud we will be finding it for

(i) Rating text:

(ii) Restaurant name:

(iii) Locality verbose :

(iv) City:

STEP 6:

Objective1 :To evaluate the Highest Rated and Lowest Rated Restaurant of the City in all the countries.

After many preprocessing we will be finding the highly rated and lowest rated restaurant of city in all countries.

	Country Code	City	Highest Rated Restaurant	Rating Max	Lowest Rated Restaurant	Rating Min
106	216	Albany	Jimmie's Hot Dogs	3.9	BJ's Country Buffet	3.3
107	216	Athens	Sr. Sol 1	4.6	The Grill	3.7
108	216	Augusta	Rae's Coastal Cafe	4.9	Sconyers Bar B Que	3.5
109	216	Boise	Flatbread Neapolitan Pizzeria	4.6	Chandlers Steakhouse	3.9
110	216	Cedar Rapids/Iowa City	Shorts Burger and Shine	4.9	Bluebird Diner	3.6
111	216	Clatskanie	Berry Patch Restaurant	4.3	Berry Patch Restaurant	4.3
112	216	Cochrane	Sakura Sushi & Grill	3.1	Sakura Sushi & Grill	3.1
113	216	Columbus	Cafe Le Rue @ The Landings	4.6	Uptown Vietnam cuisine	3.3
114	216	Dalton	Oakwood Cafe	4.9	Southern Bliss Bakery	3.7
115	216	Davenport	Tantra Asian Bistro	4.9	Frick's Tap	0.0
116	216	Des Moines	The Cafe	4.9	Malo	3.2
117	216	Dubuque	L. May Eatery	3.8	Catfish Charlie's	3.3
118	216	Fernley	Jehova es Mi Pastor Tacos y Burritos	3.7	Jehova es Mi Pastor Tacos y Burritos	3.7
119	216	Gainesville	Atlanta Highway Seafood Market	4.9	Troll Tavern	2.2
120	216	Lakeview	Burger Queen Drive In	3.6	Burger Queen Drive In	3.6
121	216	Lincoln	Blue Orchid Thai Restaurant	4.5	Blue Orchid Thai Restaurant	4.5
122	216	Macon	Ingleside Village Pizza	4.9	Benson's Steak and Sushi	3.7
123	216	Mc Millan	Triangle Restaurant	2.4	Triangle Restaurant	2.4
124	216	Miller	HI Lite Bar & Lounge	3.4	HI Lite Bar & Lounge	3.4
125	216	Monroe	Vince's Restaurant & Pizzeria	3.6	Vince's Restaurant & Pizzeria	3.6
126	216	Ojo Caliente	The Artesian Restaurant	3.6	The Artesian Restaurant	3.6
127	216	Orlando	Yellow Dog Eats	4.9	The Coop	3.6

126	216	Ojo Caliente	The Artesian Restaurant	3.6	The Artesian Restaurant	3.6
127	216	Orlando	Yellow Dog Eats	4.9	The Coop	3.6
128	216	Pensacola	McGuire's Irish Pub & Brewery	4.9	Hemingway's Island Grill	3.5
129	216	Pocatello	El Herradero	4.1	Nosh Mahal	0.0
130	216	Potrero	Barrett Junction Cafe	3.3	Barrett Junction Cafe	3.3
131	216	Princeton	Blue Point Grill	4.0	Blue Point Grill	4.0
132	216	Rest of Hawaii	Mama's Fish House	4.9	Lulu's Waikiki	3.9
133	216	Savannah	Green Truck Pub	4.7	The Lady & Sons	3.3
134	216	Sioux City	Diamond Thai Cuisine	4.0	Kahill's Steak-Fish Chophouse	3.5
135	216	Tampa Bay	Mazzaro's Italian Market	4.9	Edison: Food+Drink Lab	3.9
136	216	Valdosta	Smok'n Pig B-B-Q	4.1	El Toreo Mexican Restaurant	3.1
137	216	Vernonia	Blue House Cafe	4.3	Blue House Cafe	4.3
138	216	Waterloo	Tokyo Japanese Steak House	3.9	Masala Grill & Coffee House	3.2
139	216	Weirton	Theo Yianni's Authentic Greek Restaurant	3.9	Theo Yianni's Authentic Greek Restaurant	3.9
140	216	Winchester Bay	Fishpatrick's Crabby Cafe	3.2	Fishpatrick's Crabby Cafe	3.2

STEP 7:

Objective 2:To evaluate the most popular cuisine of the world sold in a country and which locality in that country has most number of outlets selling that cuisine.

Popular cuisines around the world:

Out[81]:

	Country Code	Locality	Cuisines	Number of restaurants in the country
0	1	Connaught Place	North Indian	48
14	216	Dubuque	American	9
5	148	Te Aro	Cafe	5
2	30	Ipanema	Brazilian	3
6	162	Kapitolyo	Filipino	2
8	184	Marina Centre, Downtown Core	Seafood	2
9	189	Green Point	Grill	2
11	208	Kí_í_í_k Esat	Kebab	2
12	214	Najda	Indian	2
1	14	Victor Harbor	Australian	1
3	37	Yorkton	Asian	1
4	94	Tanjung Duren	Seafood	1
7	166	The Gate, Dafna	Arabian	1

Popular cuisines in india:

Out[82]:

	Cuisines	Locality Verbose	Number of restaurants in the locality
0	North Indian	Connaught Place, New Delhi	48
1	Cafe	Satyaniketan, New Delhi	31
2	Chinese	Chanakyapuri, New Delhi	13
3	Mughlai	Jama Masjid, New Delhi	12
4	Bakery	Vasant Vihar, New Delhi	11
5	Chinese	Majnu ka Tila, New Delhi	10

STEP 8:

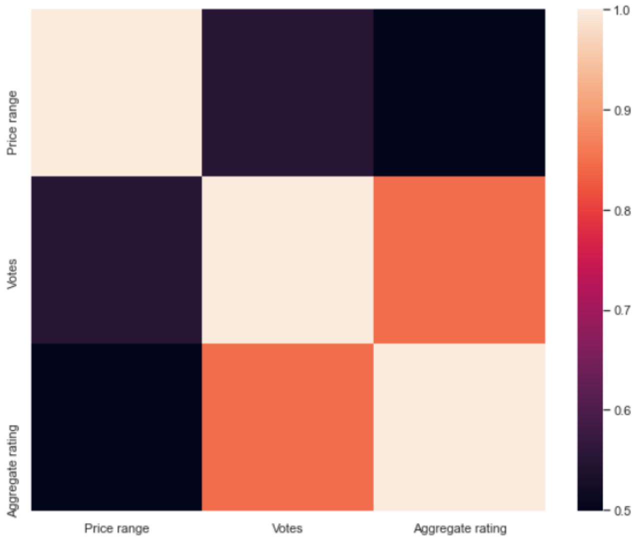
Objective 3:To evaluate the value for money restaurants in the U.S.A (value for money refers to the restaurants with highest rating and lowest cost)

Out[94]:

	City	Cuisine	Locality Verbose	Country Code	Count	Latitude	Longitude	Aggregate rating	Average Cost for two	Restaurant Name
29	Pensacola	Greek	Perdido Key, Pensacola	216	1	30.319982	-87.421896	4.7	10	Original Georgios Authentic Greek Food
47	Cedar Rapids/Iowa City	American	Iowa City, Cedar Rapids/Iowa City	216	3	41.663849	-91.531414	4.5	10	The Hamburg Inn No. 2 Inc.
54	Cedar Rapids/Iowa City	Pizza	Marion, Cedar Rapids/Iowa City	216	1	42.033100	-91.599500	4.7	25	Zoeys Pizzeria
56	Rest of Hawaii	Seafood	Kahuku, Rest of Hawaii	216	1	21.677078	-157.948486	4.5	25	Giovanni's Shrimp Truck
57	Lincoln	Thai	Haymarket, Lincoln	216	1	40.814300	-96.707200	4.5	25	Blue Orchid Thai Restaurant
62	Orlando	Vegetarian	Winter Park, Orlando	216	1	28.601088	-81.322631	4.7	25	Tibby's New Orleans Kitchen
64	Orlando	Tea	The Milk District, Orlando	216	1	28.543571	-81.351467	4.9	25	Pom Pom's Teahouse and Sandwicheria

STEP 9: correlation map with help of spearmen between the price range , votes, aggregrate rating and found that the relation between themselves gives the value of 1.

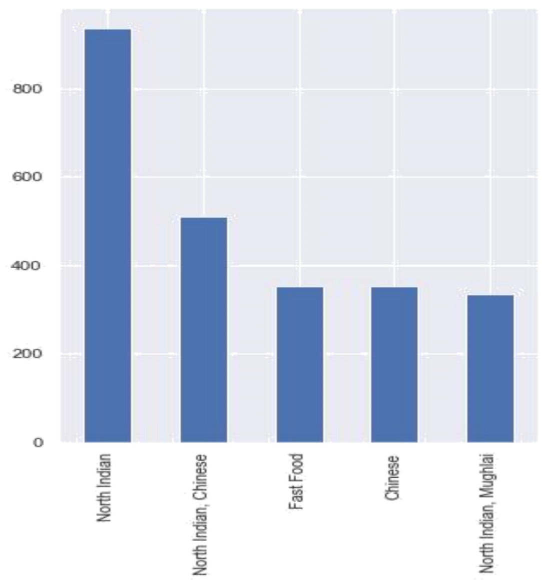
```
In [95]: numeric = ['Price range', 'Votes', 'Aggregate rating']
         sns.heatmap(zomato[numeric].corr(method='spearman'));
```



To find which type of cuisines has highest rating . Here is it found to be north indian with more than 800 people like it.

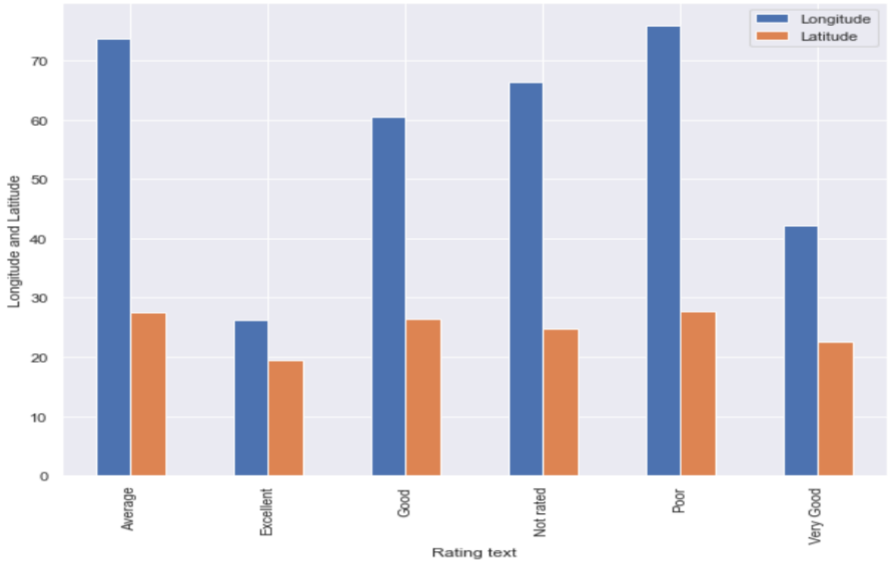
```
In [97]: zomato['Cuisines'].value_counts().head(5).plot(kind='bar', figsize = (5,7))

Out[97]: <matplotlib.axes._subplots.AxesSubplot at 0x22536b55608>
```



With the help of below graph we can find the bar graph between rating text and longitude and latitude and found that poor has more value longitude wise.

```
In [98]:
zomato.groupby('Rating text')['Longitude', 'Latitude'].mean().plot(kind='bar')
plt.ylabel('Longitude and Latitude')
plt.show();
```



Building a Recommendation system:

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

Using the reviews_list and cuisines column

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

	reviews_list	cuisines
25274	[('Rated 3.0', "RATED\n An average bar, nothi...	Continental, American, Chinese, BBQ, Salad
21244	[('Rated 3.0', 'RATED\n Just ordered shahi pa...	North Indian, Biryani, Chinese
18372	[('Rated 2.0', "RATED\n Ambiance - They have ...	South Indian, North Indian, Chinese
23189	[('Rated 5.0', 'RATED\n Ordered plain maggi\U...	Cafe, Continental
15559	[('Rated 3.0', 'RATED\n Had been to this plac...	Cafe, Burger, Continental, Italian, Desserts

Appling text preprocessing/ cleaning steps that are:

- Lower casing
- Removal of Punctuations

- Removal of Stopwords
- Removal of URLs
- Spelling correction

In [23]:
```python
## Lower Casing
zomato["reviews_list"] = zomato["reviews_list"].str.lower()
zomato[['reviews_list', 'cuisines']].sample(5)
```

Out[23]:

	reviews_list	cuisines
7791	[('rated 5.0', "rated\n very very tasty dabel...	Pizza, Burger
33096	[('rated 1.0', "rated\n this place is located...	Continental, Biryani, Pizza, North Indian, Chi...
30288	[('rated 3.0', "rated\n though it's located i...	Cafe, American, Italian, Burger
7583	[('rated 1.0', "rated\n i've visited this hot...	Chinese
24351	[('rated 5.0', 'rated\n you ask for it and it...	Desserts

In [24]:
```python
## Removal of Punctuations
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[24]:

	reviews_list	cuisines
35205	rated 50 ratedn good food top quality packin...	North Indian, Chinese
21758	rated 50 ratedn awesome food taste and cost e...	South Indian, Chinese
14823	rated 30 ratedn a famous vegetarian place hen...	North Indian, Chinese, South Indian
11758	rated 40 ratedn this place is better than the...	South Indian, North Indian, Chinese, Street Fo...
17859	rated 30 ratedn wasnt as great as we thought ...	North Indian, Chinese

In [25]:
```python
## Removal of Stopwords
from nltk.corpus import 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))
```

In [26]:
```python
## 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 [27]:
```python
zomato[['reviews_list', 'cuisines']].sample(5)
```

Out[27]:

	reviews_list	cuisines
12344	rated 40 ratedn wifiyesnfriendly staff710ntast...	North Indian
35257	rated 20 ratedn pizzas taste meh located right...	Pizza, Fast Food
10881	rated 40 ratedn ordered egg biryani one paneer...	North Indian, South Indian, Chinese, Andhra, B...
23596	rated 50 ratedn nice food lot see food verity ...	Mangalorean, Chinese, North Indian
24680	rated 50 ratedn paratha fabulousnbakasura reco...	North Indian

Model building:

Using Term Frequency-Inverse Document Frequency

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 [37]:
```
df_percent.set_index('name', inplace=True)
```

In [38]:
```
indices = pd.Series(df_percent.index)
```

In [39]:
```
# Creating tf-idf matrix
tfidf = TfidfVectorizer(analyzer='word', ngram_range=(1, 2), min_df=0, stop_words='english')
tfidf_matrix = tfidf.fit_transform(df_percent['reviews_list'])
```

In [48]:
```
cosine_similarities = linear_kernel(tfidf_matrix, tfidf_matrix)
```

In [41]:
```
def recommend(name, cosine_similarities = cosine_similarities):

    # Create a list to put top 10 restaurants
    recommend_restaurant = []

    # Find the index of the hotel entered
    idx = indices[indices == name].index[0]

    # Find the restaurants with a similar cosine-sim value and order them from biggest number
    score_series = pd.Series(cosine_similarities[idx]).sort_values(ascending=False)

    # Extract top 30 restaurant indexes with a similar cosine-sim value
    top30_indexes = list(score_series.iloc[0:31].index)

    # Names of the top 30 restaurants
    for each in top30_indexes:
        recommend_restaurant.append(list(df_percent.index)[each])

    # Creating the new data set to show similar restaurants
    df_new = pd.DataFrame(columns=['cuisines', 'Mean Rating', 'cost'])

    # Create the top 30 similar restaurants with some of their columns
    for each in recommend_restaurant:
        df_new = df_new.append(pd.DataFrame(df_percent[['cuisines','Mean Rating', 'cost']][df_percent.index == each].sample()))

    # Drop the same named restaurants and sort only the top 10 by the highest rating
    df_new = df_new.drop_duplicates(subset=['cuisines','Mean Rating', 'cost'], keep=False)
    df_new = df_new.sort_values(by='Mean Rating', ascending=False).head(10)

    print('TOP %s RESTAURANTS LIKE %s WITH SIMILAR REVIEWS: ' % (str(len(df_new)), name))

    return df_new
```

To print the recommended restaurants based on similar ratings:

```
In [44]: # HERE IS A RANDOM RESTAURANT. LET'S SEE THE DETAILS ABOUT THIS RESTAURANT:
df_percent[df_percent.index == 'Sri Krishna Kitchen'].head()
```

Out[44]:

name	online_order	book_table	rate	location	cuisines	cost	reviews_list	city	Mean Rating
Sri Krishna Kitchen	True	False	3.6	BTM	Fast Food	400.0	rated 20 ratedn place known unlimited thali ge...	Koramangala 5th Block	3.34
Sri Krishna Kitchen	True	False	3.6	BTM	Fast Food	400.0	rated 20 ratedn place known unlimited thali ge...	Koramangala 6th Block	3.34
Sri Krishna Kitchen	True	False	3.7	Bellandur	North Indian, Chinese	400.0	rated 30 ratedn ordered mutton boneless biryan...	Sarjapur Road	3.34
Sri Krishna Kitchen	True	False	3.6	BTM	Fast Food	400.0	rated 30 ratedn food taste much good rated 10 ...	BTM	3.34
Sri Krishna Kitchen	True	False	3.6	BTM	Fast Food	400.0	rated 20 ratedn place known unlimited thali ge...	Koramangala 6th Block	3.34

```
In [45]: recommend('Sri Krishna Kitchen')
```

TOP 10 RESTAURANTS LIKE Sri Krishna Kitchen WITH SIMILAR REVIEWS:

Out[45]:

	cuisines	Mean Rating	cost
Shree Thali	North Indian	3.94	150.0
3 Spice	North Indian, Chinese	3.71	450.0
Cinnamon	North Indian, Chinese, Biryani	3.62	550.0
Raichur Biryani House	Biryani, North Indian, Chinese	3.58	400.0
Desi Doze	North Indian, Fast Food	3.58	400.0
Kakaji	North Indian	3.45	350.0
Sri Krishna Kitchen	North Indian, Chinese	3.34	400.0
Sri Sai Mango Tree Restaurant	North Indian, Biryani, Chinese	3.32	600.0
Red Chilliez	North Indian, Chinese, Seafood, Mangalorean	3.26	650.0
Red Chilliez	North Indian, South Indian, Chinese, Seafood	3.26	550.0

To check the accuracy:

```
pip install pickle5
```

```
Collecting pickle5Note: you may need to restart the kernel to use updated packages.

  Downloading pickle5-0.0.12-cp36-cp36m-win_amd64.whl (124 kB)
Installing collected packages: pickle5
Successfully installed pickle5-0.0.12
```

```
import pickle
```

```
file = open("res.pkl", 'wb')
pickle.dump(df_new2, file)
```

```
In [45]: df_new2.to_dict()
```

```
Out[45]: {'name': {0: 'Cinnamon',
  1: 'Kakaji',
  2: 'Agarwal Food Service',
  3: "Swad 'E' Punjab",
  4: 'Desi Dhaba',
  5: 'Punjabi Tasty Khana',
  6: 'Taza Khaana',
  7: 'Sri Lakshmi Dhaba',
  8: 'Yummy Punjabi',
  9: 'Indian Food'},
 'cuisines': {0: 'North Indian, Chinese, Biryani',
  1: 'North Indian',
  2: 'North Indian, Chinese, Biryani',
  3: 'North Indian, Chinese, Mughlai',
  4: 'Chinese, North Indian',
  5: 'North Indian, Chinese, Biryani',
  6: 'Chinese, North Indian',
  7: 'North Indian',
  8: 'North Indian, Chinese',
  9: 'North Indian, Biryani, Chinese, Momos'},
 'Mean Rating': {0: 3.62,
  1: 3.45,
  2: 3.39,
  3: 3.32,
  4: 3.19,
  5: 2.68,
  6: 2.63,
  7: 2.5,
  8: 2.5,
  9: 2.42},
 'cost': {0: 550.0,
  1: 350.0,
  2: 400.0,
  3: 500.0,
  4: 300.0,
  5: 450.0,
  6: 450.0,
  7: 250.0,
  8: 400.0,
  9: 450.0}}
```

Conclusion:

This paper has analysed of various characteristics of current restaurants in different localities of a city in a particular country and also analyses them to predict the restaurant ratings related to particular food. This makes it an important thing to take into consideration before making a dining in or online ordering decision. Finally using cosine similarities we were able to find the user-user (Restaurant to restaurant) with similar ratings.

References:

Product and Brand Strategy of Zomato* (ijert.org)

1803_Kundan_FoodRecommedationSystemBasedOnContentFilteringAlgorithm.pdf

(PDF) Diet-Right: A Smart Food Recommendation System (researchgate.net)

SawantPai-YelpFoodRecommendationSystem.pdf (stanford.edu)

zomato recommendation system (satoevyemekleri.com)