VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

EXPLORATORY DATA ANALYSIS ON ZOMATO

by

Vaasu Bhatnagar	**19BCE1420**
Rahul Garg	**19BCE1431**

A project report submitted to

Proff. Joshan Athanesious J

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

in partial fulfilment of the requirements for the course of

CSE4058– Business Intelligence

in

B. Tech. COMPUTER SCIENCE AND ENGINEERING (CPS)

Abstract

Data analysis utilising visual methods is called exploratory data analysis (EDA). With the use of statistical summaries and graphical representations, it is used to identify trends, patterns, or to verify assumptions.

EDA helps with a better understanding of the variables in the data collection and their relationships, and is usually used to investigate what data might disclose beyond the formal modelling or hypothesis testing assignment. It can also assist in determining the suitability of the statistical methods you are considering using for data analysis. EDA approaches, which were initially created by American mathematician John Tukey in the 1970s, are still a frequently employed strategy in the data discovery process.

Introduction

Businesses are looking for a competitive advantage in bringing products and services to crowded markets. Data-driven predictive models can assist businesses in solving long-standing problems in novel ways.

Equipment manufacturers, for example, may find it difficult to innovate solely in hardware. To add value to existing solutions, product developers can incorporate predictive capabilities. Predictive analytics for equipment maintenance, also known as predictive maintenance, can predict equipment failures, forecast energy requirements, and lower operating costs. Sensors that measure vibrations in automotive parts, for example, can alert the driver to the need for maintenance before the vehicle fails on the road.

Predictive analytics is frequently discussed in relation to big data. Engineering data, for example, is collected from sensors, instruments, and networked systems located throughout the world. A company's business system data may include transaction data, sales results, customer complaints, and marketing information. Businesses are increasingly making data-driven decisions based on this valuable resource.

Predictive analytics is the process of using data analytics to make data-driven predictions. This method creates a predictive model for forecasting future events by combining data, analysis, statistics, and machine learning techniques.

Local hotels, restaurants, chefs, and canteens provide takeout and meal parcels to customers' doorsteps through online food ordering apps. This kind of thought is easily spreading due to the rise in the working young generation in major cities and the frantic work life culture. Working people's kitchens now have a new dimension thanks to this system. Nowadays, customers are more drawn to internet ordering services than they are to restaurant home delivery. Since there is no human involvement in the process of ordering food online, it is more private. Numerous restaurants and chefs' kitchens are included in apps along with their individual menus. Therefore, for further orders, customers do not need to carry menus and brochures.

These apps are more accessible because they may be downloaded immediately to a smartphone. You can create a payment information account by providing your address and profile. Customers must, however, download the software to their smartphones and sign up for an account before using it. Their address and payment information are included when creating an app profile. Different payment methods, including credit cards, debit cards, cashless accounts, and free home delivery, are available through apps. Various applications provide various services, offers, features, or restaurants. An app that was downloaded offered some coupons for discounts, information on prior orders, palette ideas, and recent patron reviews of both restaurants and food.

According to BCG, the market value of the Indian food industry is predicted to reach $420 billion by 2020. The Indian food industry is estimated to be worth $350 billion in 2019. To provide their customers with ease, satisfaction, and customer retention, this industry is coming up with creative innovations. Free home delivery offered by specific restaurants and online meal ordering applications compete fiercely with one another.

Literature Survey

In this section, we discuss related works. We reviewed the related work in two different domains: technical and financial, respectively.

1) **Serhat Murat Alagoz & Haluk Hekimoglu (2012)**, opined that while e-commerce is rapidly expanding around the world, the food industry is also expanding. They proposed using the Technology Acceptance Model (TAM) to investigate the acceptance of online food ordering apps. According to their data analysis, the attitude toward online food ordering is due to the ease

and usefulness of the process, and it also varies according to their innovativeness with regard to information technology, their trust in e commerce websites, and a few external influences.

2) **Ashoutosh bhargve (2013)** Foodpanda, an online food ordering app, was launched in the Indian market in May 2012. Foodpanda's first significant move was the acquisition of TastyKhana, which was founded in Pune in 2007. It is now available in over 200 cities and has delivery partnerships with over 12,000 restaurants, thanks to the acquisition of TastyKhana and JUST EAT. JUST EAT, which debuted in Denmark in 2001 and is now publicly traded on the London Stock Exchange, is also mentioned. In 2006, they launched their Indian venture as Hungry Bangalore. It was reintroduced in 2011 after JUST EAT purchased a majority stake in the company. The company now works with over 2,000 restaurants.
3) **H.S. Sethu & Bhavya Saini (2016)**, Their plan was to examine students' perceptions, behaviour, and satisfaction with online food ordering and delivery apps. According to their findings, online food ordering apps save them time because they are easily accessible. It has also been discovered that the main reasons for using the apps are the availability of their favourite food at all times as well as constant access to the internet and free data.
4) According to **Sheryl E. Kimes (2011)**,Their plan was to examine students' perceptions, behaviour, and satisfaction with online food ordering and delivery apps. According to their findings, online food ordering apps save them time because they are easily accessible. It has also been discovered that the main reasons for using the apps are the availability of their favourite food at all times as well as constant access to the internet and free data.
5) According to **Leong Wai Hong (2016)**, Many industries' business models have changed as a result of technological advancement. Efficient systems can help a restaurant's productivity and profitability. The use of an online food delivery system is thought to help restaurants' businesses grow over time and to facilitate major business transactions online.
6) According to **Varsha Chavan, et al, (2015)**, The use of a smart phone mobile interface for consumers to view orders and follow has aided restaurants in quickly delivering orders from customers. The growing use of smart phones and computers is providing a platform for the service industry. Their analysis concluded that this process is convenient, effective, and simple to use, and that it is expected to improve day by day in the future.s.

Data Set

```
[ ] #read dataset
    df=pd.read_csv("C:\\Users\\Dell\\Desktop\\vit\\zomato\\zomato.csv")
    df.head()
```

	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)
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	Banashankari	Casual Dining	Pasta, Lunch Buffet, Masala Papad, Paneer Laja...	North Indian, Mughlai, Chinese	800	[('Rated 4.0', 'RATED\n A beautiful place to ...	[]	Buffet
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	Banashankari	Casual Dining	Momos, Lunch Buffet, Chocolate Nirvana, Thai G...	Chinese, North Indian, Thai	800	[('Rated 4.0', 'RATED\n Had been here for din...	[]	Buffet
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	Banashankari	Cafe, Casual Dining	Churros, Cannelloni, Minestrone Soup, Hot Choc...	Cafe, Mexican, Italian	800	[('Rated 3.0', 'RATED\n Ambience is not that ...	[]	Buffet
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	Banashankari	Quick Bites	Masala Dosa	South Indian, North Indian	300	[('Rated 4.0', 'RATED\n Great food and proper...	[]	Buffet
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	Basavanagudi	Casual Dining	Panipuri, Gol Gappe	North Indian, Rajasthani	600	[('Rated 4.0', 'RATED\n Very good restaurant ...	[]	Buffet

Columns description

1. url contains the url of the restaurant in the zomato website
2. address contains the address of the restaurant in Bengaluru
3. name contains the name of the restaurant
4. online_order whether online ordering is available in the restaurant or not
5. book_table table book option available or not
6. rate contains the overall rating of the restaurant out of 5
7. votes contains total number of rating for the restaurant as of the above mentioned date
8. phone contains the phone number of the restaurant
9. location contains the neighborhood in which the restaurant is located
10. rest_type restaurant type
11. dish_liked dishes people liked in the restaurant
12. cuisines food styles, separated by comma
13. approx_cost(for two people) contains the approximate cost for meal for two people
14. reviews_list list of tuples containing reviews for the restaurant, each tuple
15. menu_item contains list of menus available in the restaurant
16. listed_in(type) type of meal
17. listed_in(city) contains the neighborhood in which the restaurant is listed

Methodology

Cleaning of Dataset:

Data cleaning is the process of editing, correcting, and structuring data within a data set so that it's generally uniform and prepared for analysis. This includes removing corrupt or irrelevant data and formatting it into a language which is easy to understand. To clean the dataset, we checked for null values and deleted it from the data frame.

getting all NAN features

```
feature_na=[feature for feature in df.columns if df[feature].isnull().sum()>0]
feature_na
```

```
['rate',
 'phone',
 'location',
 'rest_type',
 'dish_liked',
 'cuisines',
 'approx_cost(for two people)']
```

```
#% of missing values
for feature in feature_na:
    print('{} has {} % missing values'.format(feature,np.round(df[feature].isnull().sum()/len(df)*100,4)))
```

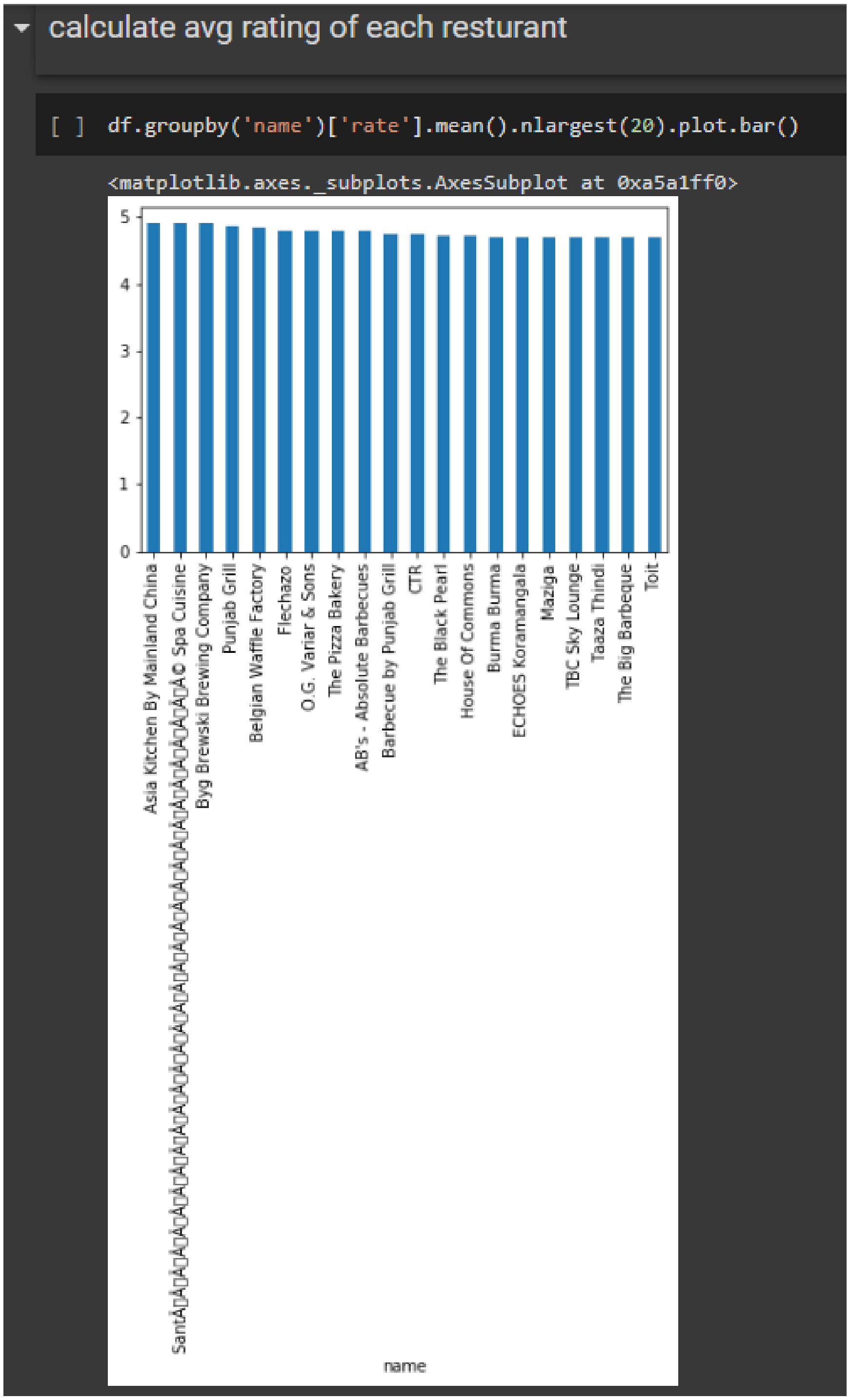
```
rate has 15.0337 % missing values
phone has 2.3358 % missing values
location has 0.0406 % missing values
rest_type has 0.4389 % missing values
dish_liked has 54.2916 % missing values
cuisines has 0.087 % missing values
approx_cost(for two people) has 0.669 % missing values
```

```
df['rate'].unique()
```

```
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', nan, '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)
```

```
df.dropna(axis='index',subset=['rate'],inplace=True)
```

Results



calculate avg rating of each resturant
[ ] df.groupby('name')['rate'].mean().nlargest(20).plot.bar()
<matplotlib.axes._subplots.AxesSubplot at 0xa5a1ff0>
5
4
3
2
1
0
Asia Kitchen By Mainland China
Spa Cuisine
Byg Brewski Brewing Company
Punjab Grill
Belgian Waffle Factory
Flechazo
O.G. Variar & Sons
The Pizza Bakery
AB's - Absolute Barbecues
Barbecue by Punjab Grill
CTR
The Black Pearl
House Of Commons
Burma Burma
ECHOES Koramangala
Maziga
TBC Sky Lounge
Taaza Thindi
The Big Barbeque
Toit
name

alternative is create a list in which u have all the restaurants & in another list,we have all ratings & then using zip we can create a dataframe

```
'''restaurant=[]
avg_rating=[]
for key,name_df in df.groupby('name'):
    restaurant.append(key)
    avg_rating.append(np.mean(name_df['rate']))'''
```

```
"restaurant=[]\navg_rating=[]\nfor key,name_df in df.groupby('name'):\n    restaurant.append(key)\n    avg_rating.append(np.mean(name_df['rate']))"
```

```
'''df_rate=pd.DataFrame(zip(restaurant,avg_rating))
df_rate.columns=['restaurant','rating']
df_rate.head(20)'''
```

```
"df_rate=pd.DataFrame(zip(restaurant,avg_rating))\ndf_rate.columns=['restaurant','rating']\ndf_rate.head(20)"
```

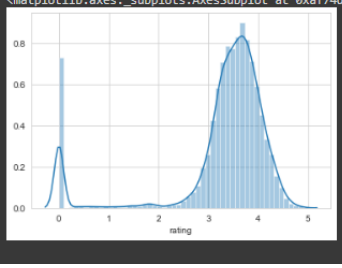
```
'''df_rate.shape'''
```

```
'df_rate.shape'
```

Rating distribution

```
sns.set_style(style='whitegrid')
sns.distplot(df_rate['rating'])
```

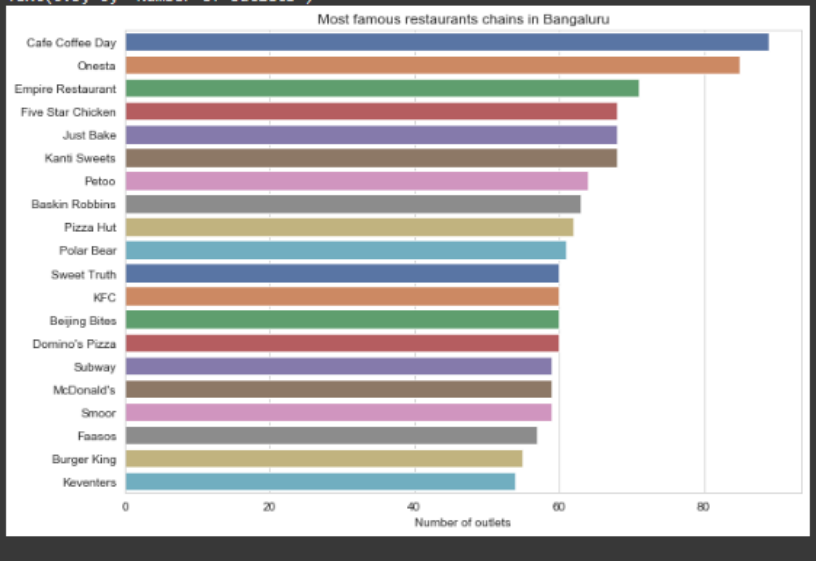
```
<matplotlib.axes._subplots.AxesSubplot at 0xaf74d30>
```



Which are the top restaurant chains in Bangaluru?

```
plt.figure(figsize=(10,7))
chains=df['name'].value_counts()[0:20]
sns.barplot(x=chains,y=chains.index,palette='deep')
plt.title("Most famous restaurants chains in Bangaluru")
plt.xlabel("Number of outlets")
```

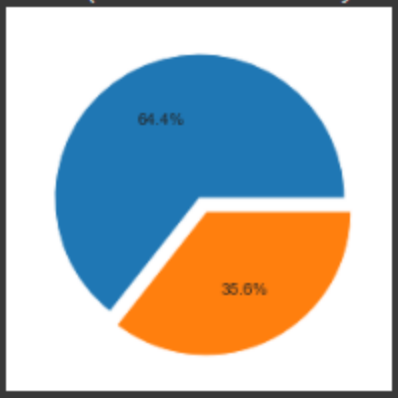
```
Text(0.5, 0, 'Number of outlets')
```



How many of the restuarants do not accept online orders?

```
x=df['online_order'].value_counts()
labels=['accepted','not accepted']
plt.pie(x,explode=[0.0,0.1],autopct='%1.1f%%')
```

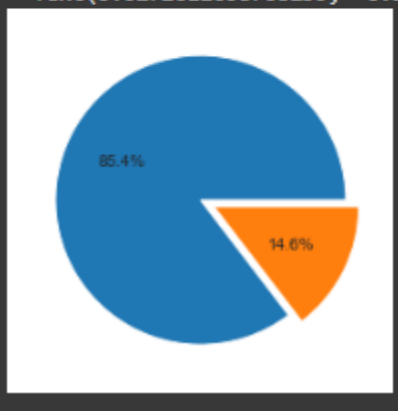
```
([<matplotlib.patches.Wedge at 0xbf27b70>,
  <matplotlib.patches.Wedge at 0xbf27df0>],
 [Text(-0.481488774517003, 0.9890240442042423, ''),
  Text(0.5252605823084886, -1.0789352717716423, '')],
 [Text(-0.26263024064563795, 0.5394676604750411, '64.4%'),
  Text(0.30640200634661835, -0.629378908533458, '35.6%')])
```



What is the ratio b/w restaurants that provide and do not provide table booking ?

```
x=df['book_table'].value_counts()
labels=['not book','book']
plt.pie(x,explode=[0.0,0.1],autopct='%1.1f%%')
```

```
([<matplotlib.patches.Wedge at 0x2ba31230>,
  <matplotlib.patches.Wedge at 0x2ba31550>],
 [Text(-0.9856961240827233, 0.4882654513359477, ''),
  Text(1.0753048626356982, -0.5326532196392153, '')],
 [Text(-0.537652431317849, 0.26632660981960776, '85.4%'),
  Text(0.6272611698708239, -0.31071437812287556, '14.6%')])
```



How many types of restaurants we have?

```
df['rest_type'].isna().sum()
```

```
151
```

```
df['rest_type'].dropna(inplace=True)
```

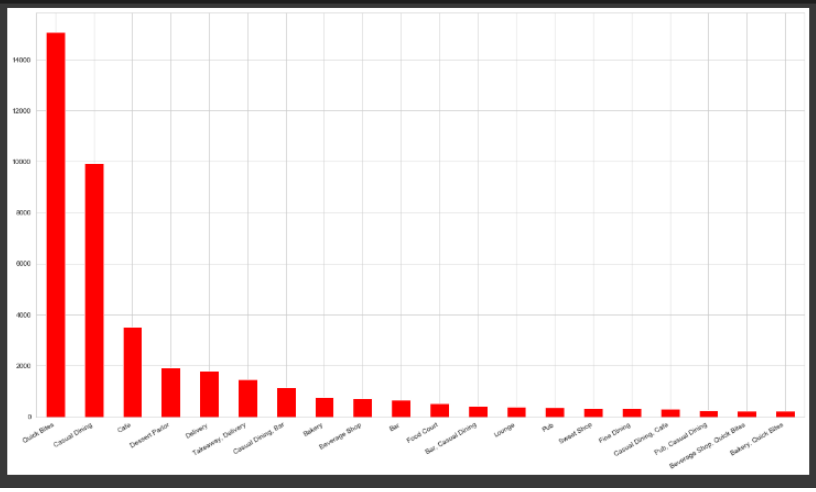
```
df['rest_type'].isna().sum()
```

```
0
```

```
len(df['rest_type'].unique())
```

```
87
```

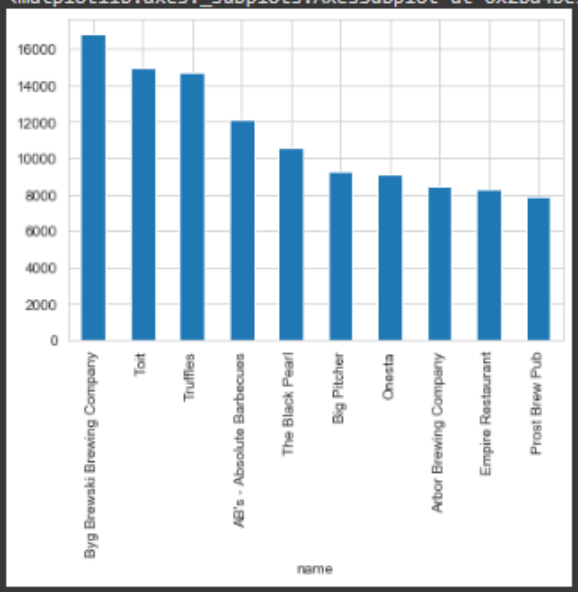
```
plt.figure(figsize=(20,12))
df['rest_type'].value_counts().nlargest(20).plot.bar(color='red')
plt.gcf().autofmt_xdate()
```



highest voted restaurant

```
df.groupby('name')['votes'].max().nlargest(10).plot.bar()
```

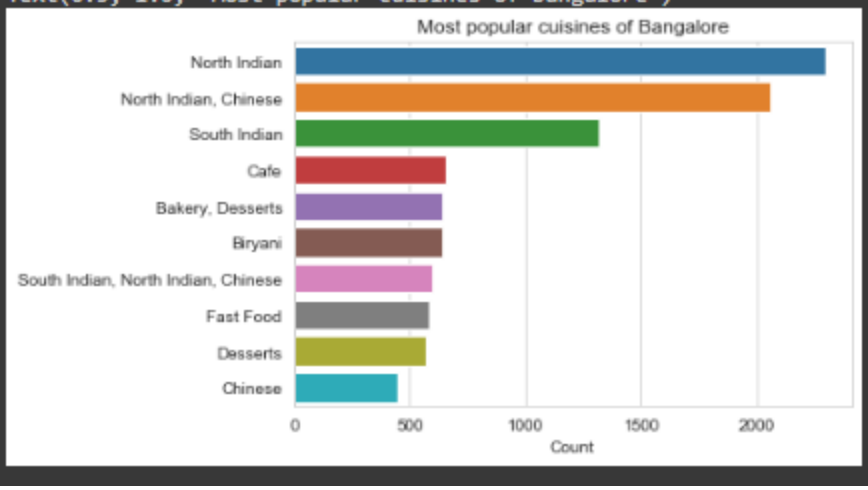
```
<matplotlib.axes._subplots.AxesSubplot at 0x2ba4be30>
```



Total number of variety of restaurants ie north indian,south Indian

```
[ ]
    cuisines=df['cuisines'].value_counts()[:10]
    sns.barplot(cuisines,cuisines.index)
    plt.xlabel('Count')
    plt.title("Most popular cuisines of Bangalore")
```

```
Text(0.5, 1.0, 'Most popular cuisines of Bangalore')
```



remove ,

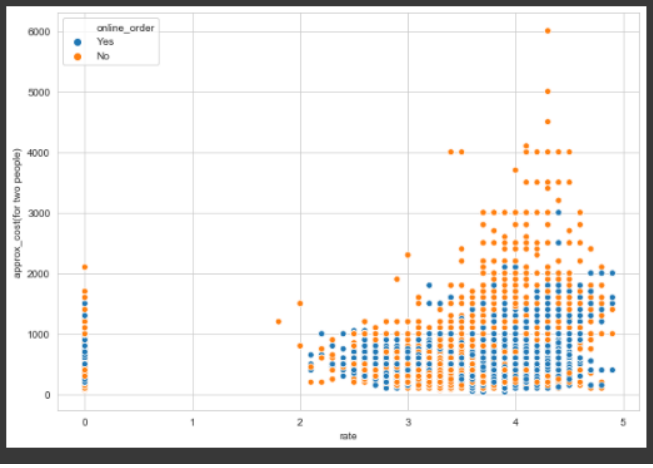
```
[ ] df['approx_cost(for two people)'] = df['approx_cost(for two people)'].apply(lambda x: x.replace(',',''))
```

```
[ ] df['approx_cost(for two people)']=df['approx_cost(for two people)'].astype(int)
```

```
[ ]
```

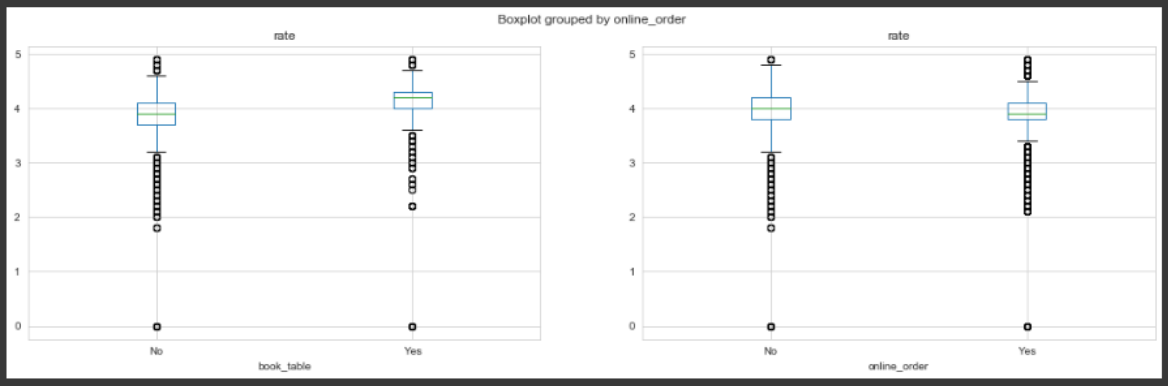
cost vs rating

```
[ ] plt.figure(figsize=(10,7))
    sns.scatterplot(x="rate",y='approx_cost(for two people)',hue='online_order',data=df)
    plt.show()
```



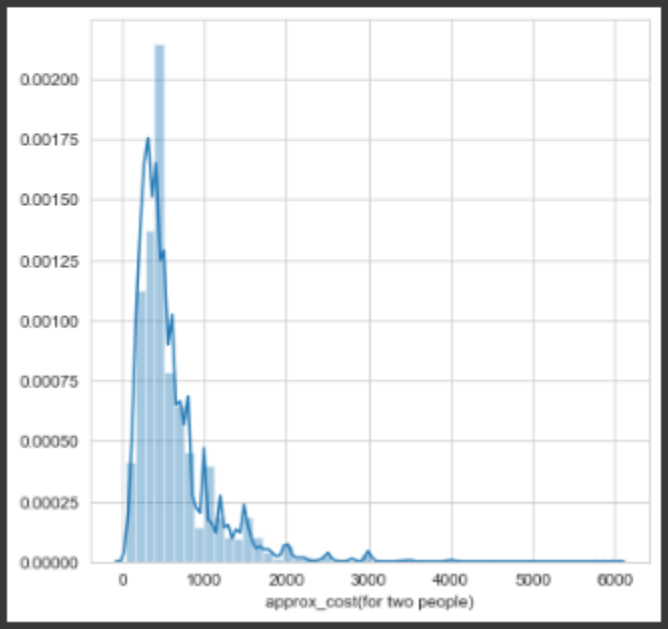
Box plot against rating

```
[ ]  colls = ['book_table','online_order']

     import matplotlib.pyplot as plt
     fig, plott=plt.subplots(nrows=1, ncols=len(colls), figsize=(18,5))

     for x,i in zip(colls,range(len(colls))):
         df.boxplot(column='rate',by=x,figsize=(10,10),vert=True,ax=plott[i])
```



distribution of cost for 2 people

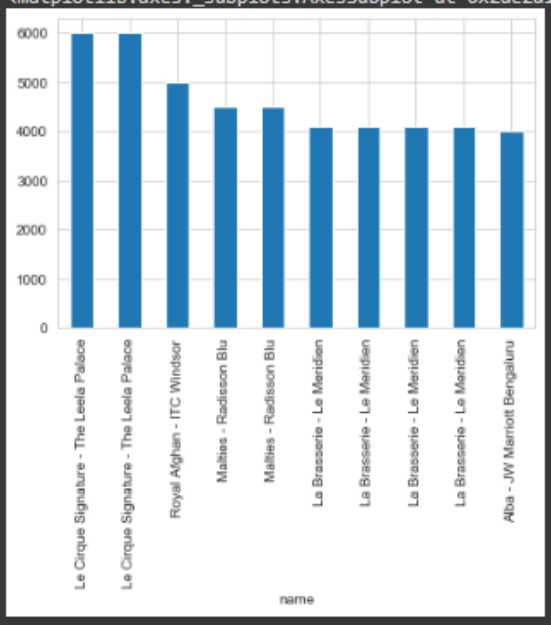
```
plt.figure(figsize=(6,6))
sns.distplot(df['approx_cost(for two people)'])
plt.show()
```



Top 10 Most Expensive restaurant with approx cost for 2 people

```
[ ]  data['approx_cost(for two people)'].nlargest(10).plot.bar()
```

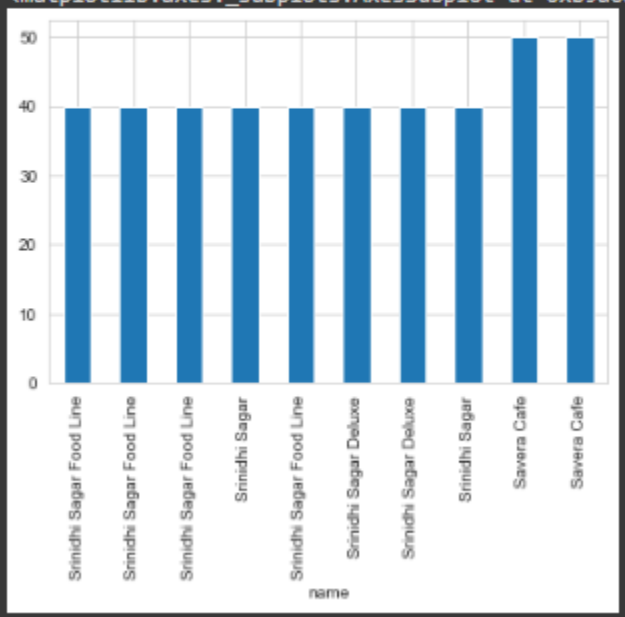
```
<matplotlib.axes._subplots.AxesSubplot at 0x2de2d150>
```



Top 10 Cheapest restaurant with approx cost for 2 people

```
data['approx_cost(for two people)'].nsmallest(10).plot.bar()
```

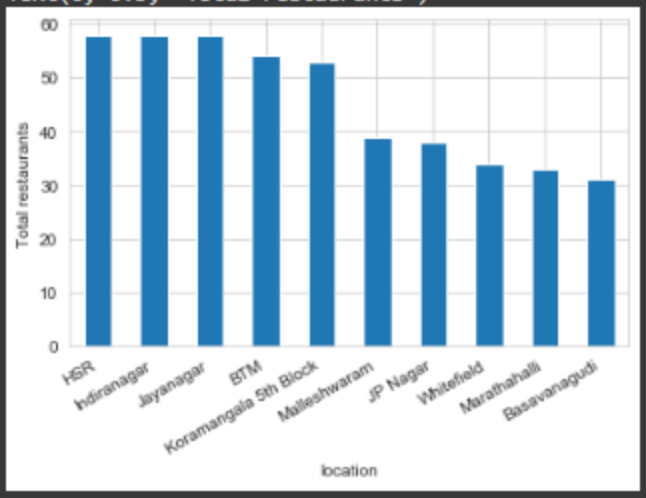
```
<matplotlib.axes._subplots.AxesSubplot at 0xb9d6a30>
```



```
list
```

```
[ ]
location_df['restaurant'].nlargest(10).plot.bar()
plt.gcf().autofmt_xdate()
plt.ylabel('Total restaurants')
```

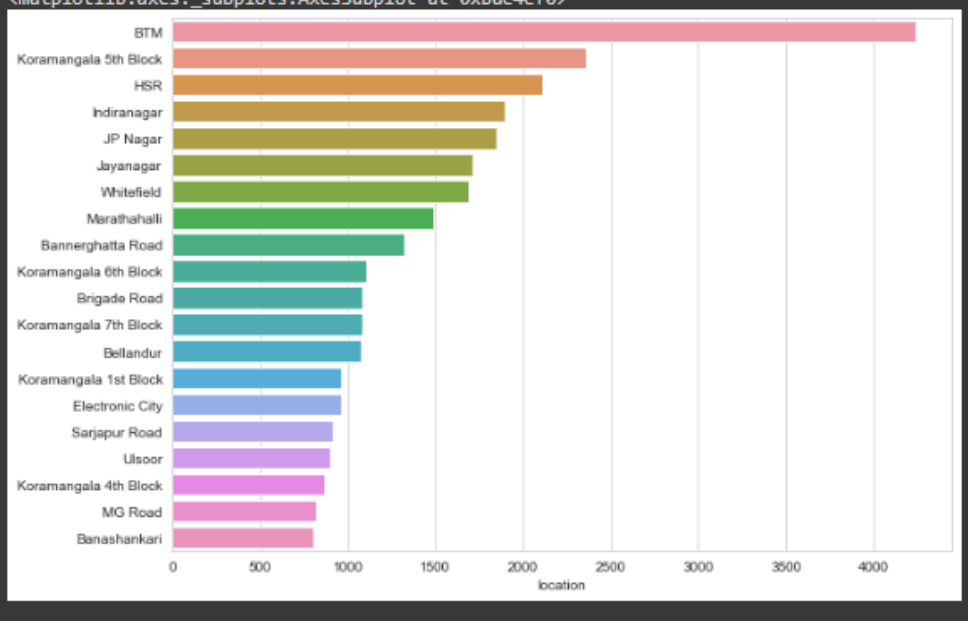
```
Text(0, 0.5, 'Total restaurants')
```



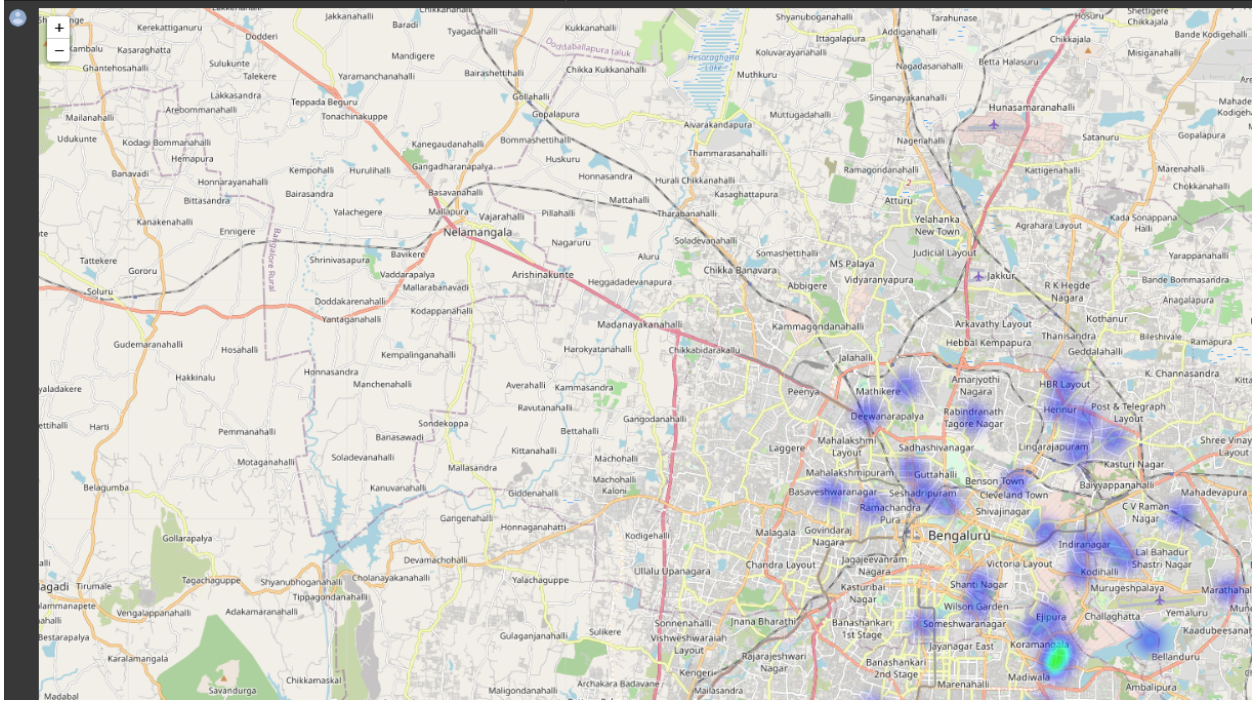
Which are the foodie areas?

```
plt.figure(figsize=(10,7))
Restaurant_locations=df['location'].value_counts()[:20]
sns.barplot(Restaurant_locations,Restaurant_locations.index)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0xbae4ef0>
```



```
basemap=generateBaseMap()
HeatMap(north_india[['lan','lon','count']].values.tolist(),zoom=20,radius=15).add_to(basemap)
basemap
```




```
[ ] stopwords = set(STOPWORDS)
    produce_wordcloud(rest)
```

Conclusion and future work

The feature that most appeals to customers is Doorstep Delivery at any time and from any location. Consumers are most motivated when they receive any Rewards & Cashbacks, followed by loyalty points or benefits.Bad past experience, reviews, and word of mouth are the factors that prevent customers from trying online food delivery apps.This concept with innovation can thrive by providing consistent and effective services.Companies can target Tier 2 cities for business expansion in the future because these cities have a large number of working youth.

References

- Ashutosh Bhargave, Niranjan Jadhav, Apurva Joshi, Prachi Oke, Prof. Mr. S. R Lahane (2013) "Digital ordering system for Restaurant using Android" International Journal of Scientific and Research Publications, Volume 3, Issue 4, April 2013
- H.S. Sethu & Bhavya Saini (2016), "Customer Perception and Satisfaction on Ordering Food via Internet, a Case on Foodzoned.Com, in Manipal", Proceedings of the Seventh AsiaPacific Conference on Global Business, Economics, Finance and Social Sciences (AP16Malaysia Conference) ISBN: 978-1-943579-81-5. Kuala Lumpur, Malaysia. 15-17, July 2016. Paper ID: KL631
- Sheryl E. Kimes Ph.D. (2011), "Customer Perceptions of Electronic Food Ordering", Cornell Hospitality Report, 11(10), pp. 6-15.
- Leong Wai Hong (2016), "Food Ordering System Using Mobile Phone", A report submitted to BIS (Hons) Information Systems Engineering. Faculty of Information and Communication Technology (Perak Campus), UTAR.
- Varsha Chavan, Priya Jadhav, Snehal Korade and Priyanka Teli (2015), "Implementing Customizable Online Food Ordering System Using Web Based Application", International Journal of Innovative Science, Engineering & Technology, Vol 2 Issue 4, April 2015.