STARTLY

MINOR PROJECT

ANSHIKA CHAUDHARY
17001011007

MENTOR
DR. REWA

OUTLINE

1. INTRODUCTION
2. METHODOLOGY
3. CONCLUSION
4. FUTURE SCOPE

DO YOU EVER THINK WHY A RESTAURANT FAILS ?

Obviously there are so many reasons .
But one of the main reason is lack of proper research of the surroundings and the factors affecting the smooth and profitable functioning of the restaurant.

It has become difficult for the new restaurants to compete with already established restaurants. The key issues that continue to pose a challenge to them include high real estate costs, rising food costs, shortage of quality manpower, fragmented supply chain and over-licensing.

INTRODUCTION

This project aims at analyzing demography, its food culture of the location. Most importantly it will help new restaurants in deciding their theme, menus, cuisine, cost etc for a particular location. It also aims at finding similarity between neighborhoods of Bengaluru on the basis of food. With the analysis the project also will help people in choosing the restaurant based on several other factors. The project will mainly try to answer the questions based on restaurants and foodies interest. And what factors should be kept in mind if someone wants to open new restaurant.

The Process

Data is collected from Kaggle.

Data is cleaned using appropriate techniques.

Analysis is done using various plots and t-sne

Here we train our model for sentiment analysis using LSTM

Understanding The Data

Data Analysis is the process of bringing order and structure to collected data. It turns data into information teams can use. Analysis is done using systematic methods to look for trends, groupings, or other relationships between different types of data.

1. Loading the dataset using pandas.
2. Checking for NULL Values.
3. Visualizing NULL values using HEAT Map.
4. Dropping the unnecessary columns.
5. Listing the columns.

Columns Description

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

EDA and Visualization

Data visualization is the process of putting data into a chart, graph, or other visual format that helps inform analysis and interpretation. Data visuals present the analyzed data in ways that are accessible to and engage different stakeholders. Data visuals are also used to communicate MEAL results to meet key stakeholder needs. Multiple visuals will likely be needed to understand the larger change process and inform data use. Common data visual formats include:

Distance Plots
Scatter Plots
Bar charts
Box Plots
Pie charts
Heat Maps
BaseMaps

EDA is different from initial data analysis(IDA) which focuses more narrowly on checking assumptions requires for model fitting and hypothesis testing, and handling missing values and making transformations of variables as needed. EDA encompasses IDA.

The objectives of EDA are to:

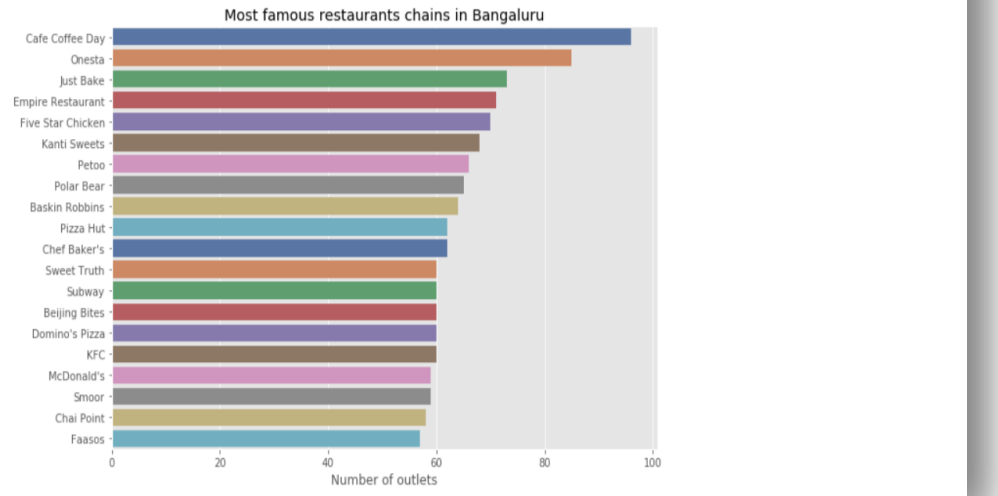
1. Suggest hypothesis about the causes of observed phenomena.
2. Assess assumptions on which statistical inference will be based.
3. Support the selection of appropriate statistical tools and techniques.
4. Provide a basis for further data collection through surveys or experiments.

Some Questions and their Visual Answers

1. What are the Top chains in Bangalore ?

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

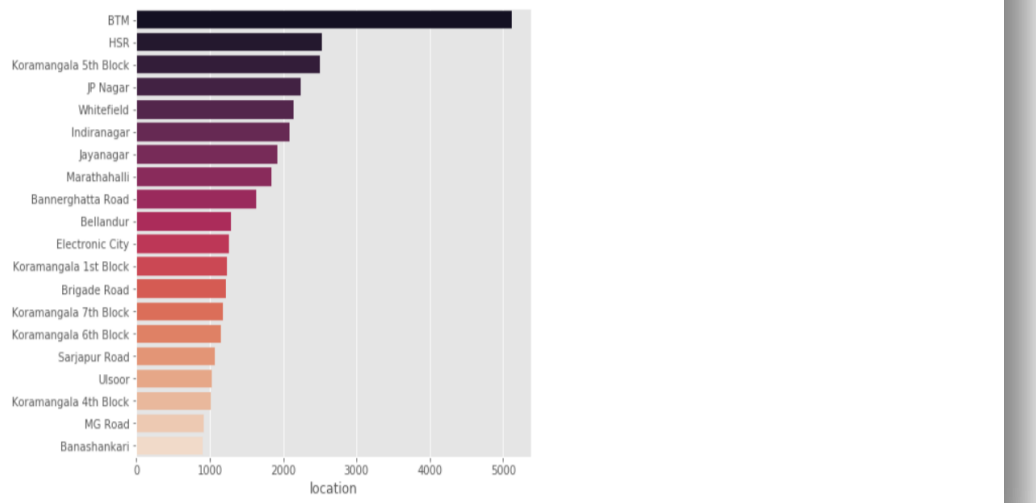
```
Text(0.5, 0, 'Number of outlets')
```



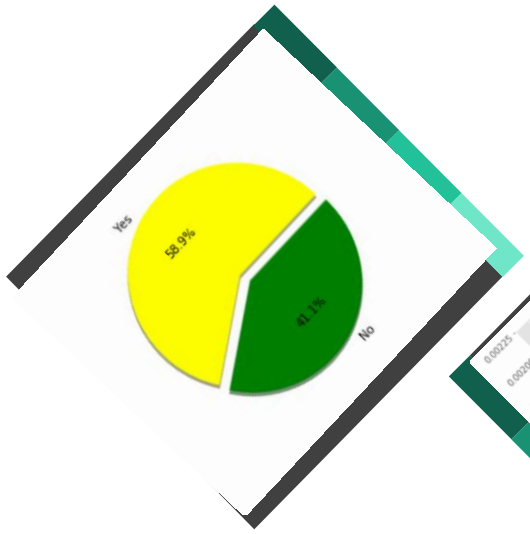
2. What are the most Foodie Areas ?

```
plt.figure(figsize=(7,7))
Rest_locations=df['location'].value_counts()[:20]
sns.barplot(Rest_locations,Rest_locations.index,palette="rocket")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x12d80a5d668>
```



SOME INFORMATIVE PLOTS OF THE PROJECT



Pie Chart

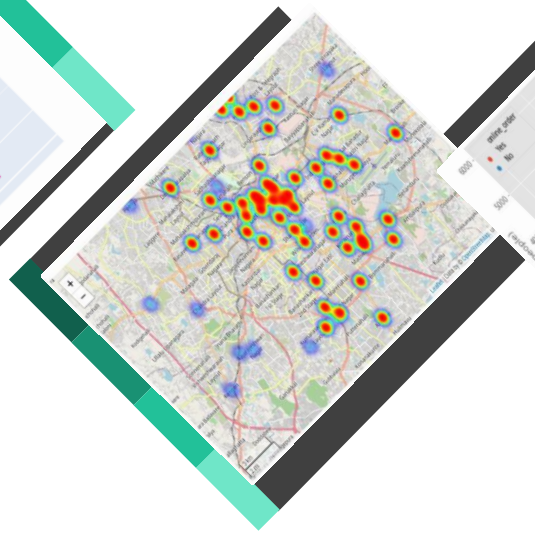
Most of the restaurants are accepting online orders but there is also a huge number of not accepting.

Dist Plot

This means almost 90percent of restaurants serve food for budget less than 1000 INR.

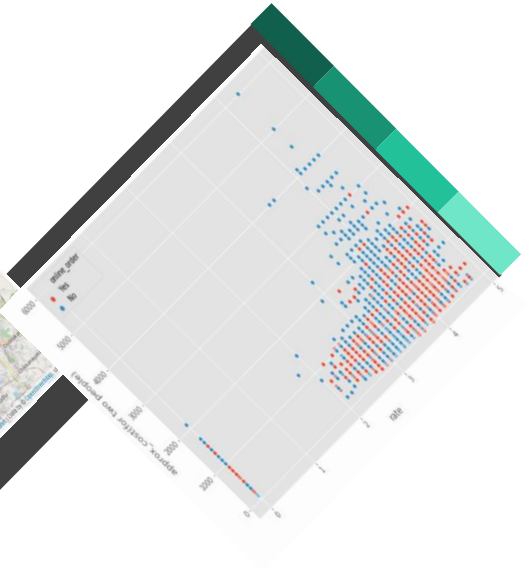
Box Plot

50 percent of restaurants charge between 300 and 650 for single meal for two people.



Heat Map

The clutter of restaurants lowers as we move away from central.

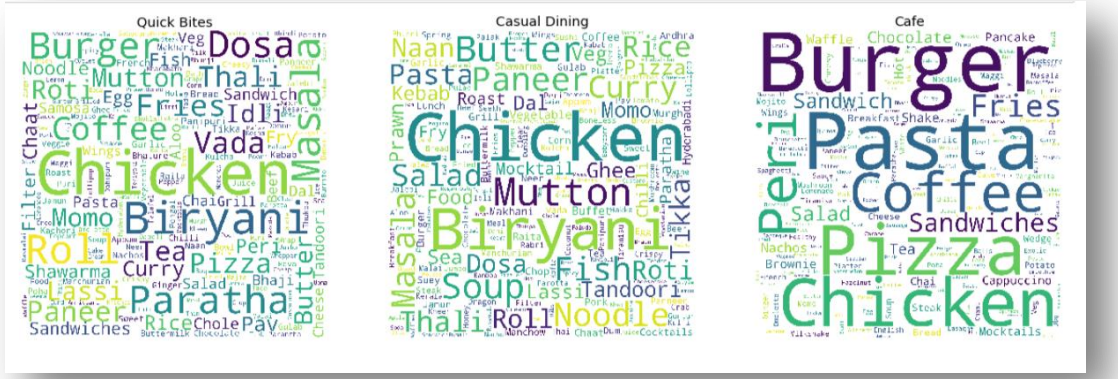


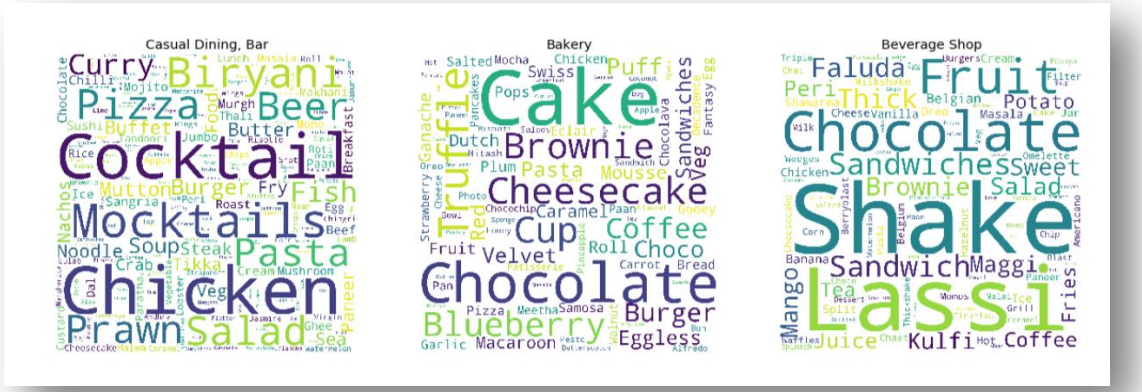
Scatter Plot

This is a cost vs rating plot depicting the relationship between cost for two people and rating given by the customers.

Word Cloud

Word Cloud is an interesting way of visualization which highlights the most common word and the size is relative to the count of their occurrence in the data.





T-SNE of Reviews

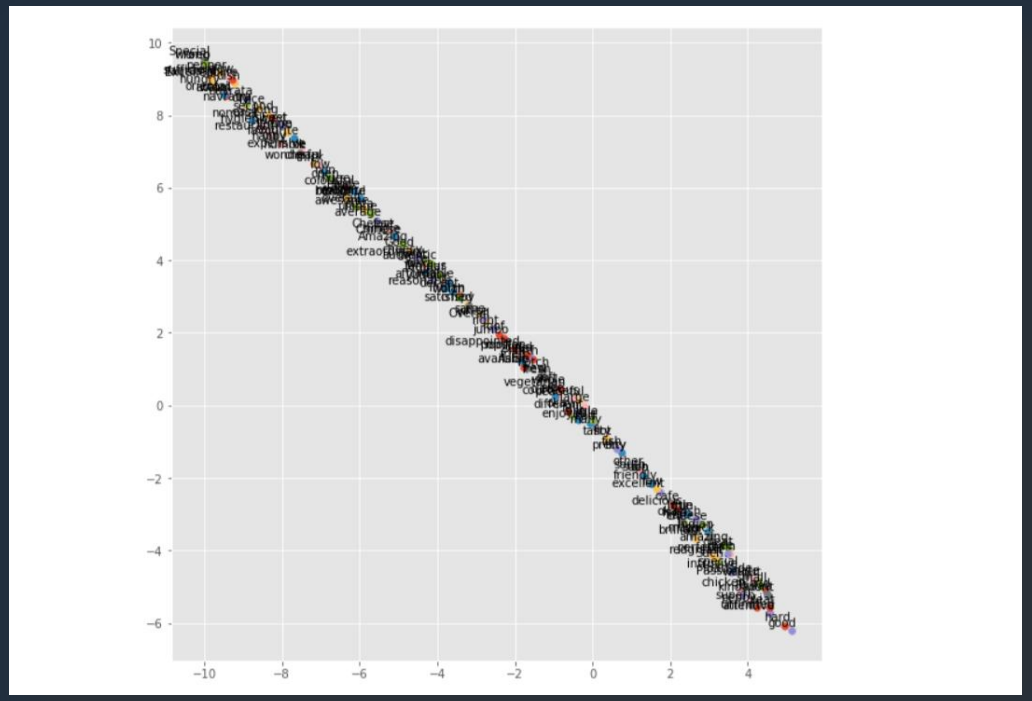
(t-SNE) t-Distributed Stochastic Neighbor Embedding is a non-linear dimensionality reduction algorithm used for exploring high-dimensional data. It maps multi-dimensional data to two or more dimensions suitable for human observation. With help of the t-SNE algorithms, you may have to plot fewer exploratory data analysis plots next time you work with high dimensional data.

In this section we will visualize words used in reviews in a 2 dimensional space.

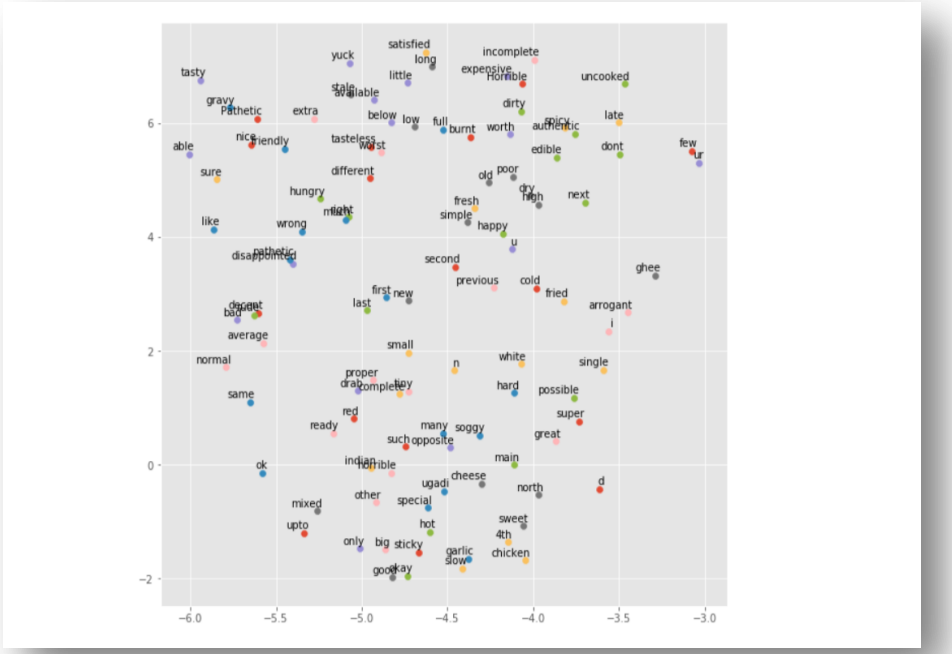
i. For that we will first lemmatize and tokenize each reviews and build a corpus out of it.
ii. Now we will use word2vec to represent each word as a vector.
iii. Ploting the t-sne plot.

T-SNE of adjectives used in positive reviews

T-SNE of adjectives used in positive reviews



T-SNE of adjectives used in negative reviews



Sentimental Analysis

Sentiment Analysis is the process of computationally determining whether a piece of writing is positive, negative or neutral. It's also known as opinion mining, deriving the opinion or attitude of a speaker. Sentiment analysis refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.

Sentiment analysis is important because companies want their brand being perceived positively, or at least more positively than the brands of competitors. **Sentiment analysis**, if accurate, can be a very valuable tool for this specific use case.

Data preparation

For doing sentimental analysis on reviews provided by the users. We have to prepare our data in appropriate format. We will map reviews to positive and negative on the basis of the ratings provided by each user. So, we will map reviews to negative if the rating given is less than 2.5 and positive if rating is greater than 2.5.

Next, we will tokenize the data and vectorize the reviews to feed it to our model.

Building our model

Sequential is the easiest way to build a **model** in Keras. It allows you to build a **model** layer by layer. Each layer has weights that correspond to the layer the follows it. We use the 'add()' function to add layers to our **model**. ... 'Activation' is the activation function for the layer.

Long short-term memory(LSTM) is an artificial recurrent neural network architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points, but also entire sequences of data

RESULTS

1

If a person closely analyze the data of the past of that particular area . Then he would be able to withstand the competition of existence in the market with lots of alternate choices available already.

2

This boxplot represents the average cost people used to spend on the meal for two . So opening a restaurant one should kept in mind this price. One who wants to start the restaurant should take into consideration this amount . If the person charges lot then the customers may swipe to another alternatives present

3

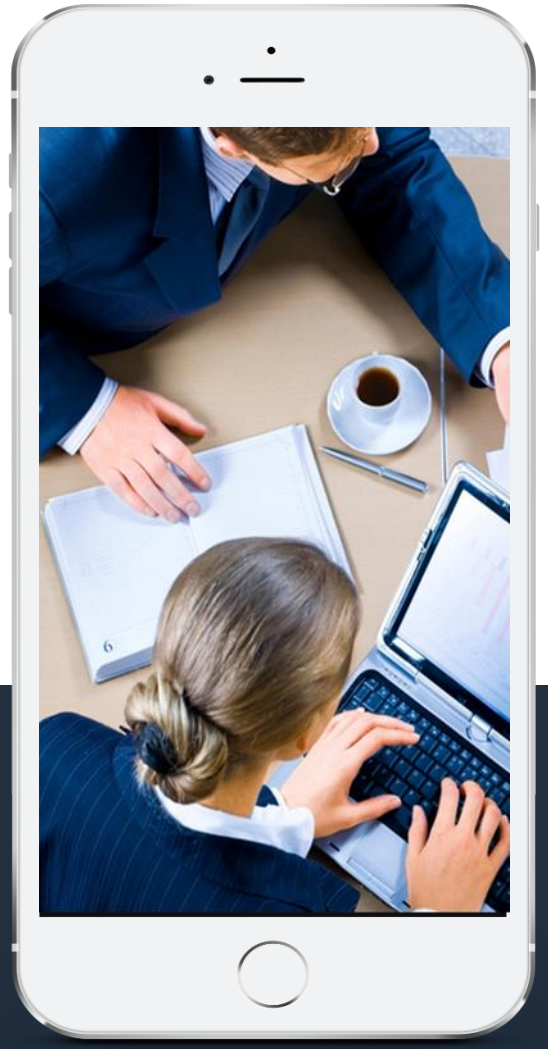
There is a heatmap build using live longitude and latitude data which can help interpreting the best suited location for the restaurant . The more crowded locations depicts the popularity of that place for the people to enjoy outside meals.

CONCLUSION

After all the analysis we did some of the point which were worth noting that there are more number foreign food chains than Indian restaurants in Bengaluru. Cafe Coffee day has its outlet in most of the neighborhood of Bengaluru. The Indian food service market has come a long way from the early Nineties when it was dominated by unorganised players and few brands. The revolution began in 1996 with McDonalds, Pizza Hut, Dominos Pizza, Subway and Yo!China, among others, setting up shop in the country. Since then, the food services market has been continuously growing. Quick service restaurants are a mainstay of the Indian food service market, and are growing fast. Fine dining is gaining prominence too. Both multi-cuisine and single-cuisine establishments have shown tremendous growth. Gone are the days when you would go to Bangalore and only indulge in dosas or at best Kodava food from the region of Coorg. Bangalore is an inspiring mish-mash of old and new, and has plenty of options on offer, regardless of whether you are visiting only for a few days or live there.

One thinking of opening a restaurant should plan according to the data to withstand the competition in the market. Just starting the restaurant without proper research may lead to failure of it. To gain popularity and make a unique place in the market restaurant features should be planned according to the data fetched from the past patterns.

And as the visuals can be easily understandable by the people not involved in technology field.We as engineers can do the same for the business people and help them understanding the patterns and make better decisions for the business. This project is more of a analytical one.



FUTURE SCOPE

This can be developed as a full fledged business model predicting future customers also. We as engineers can do the same for the business people and help them understanding the patterns and make better decisions for their upcoming restaurant projects. A website can be thought of doing the same for the clients taking the subscription or personal assistance can also be provided.

DATA DRIVEN APPROACH

PATTERN ANALYSING

MORE PROMISING THAN ASSUMPTIONS


THANK YOU