ABSTRACT

"ZOMATO- Never have a bad meal"

ZOMATO is India's largest and most successful startup company where food industry meets technology and connects thousands of restaurant with one thread. ZOMATO is the brainchild of Mr. Deepinder Goyal and Mr.Pankaj Chadda who were colleagues in '**Bain And Company'**. The idea to start ZOMATO came from the demand for menu cards to order food among their colleagues. ZOMATO provide local restaurants and hotel search services, they collect data on food menus, contacts and relevant information to users. Key resource for ZOMATO is large database of restaurants across cities. The value proposition that ZOMATO provide to its customers in one-stop shop for dines and offers restaurants and a way to differentiate among them and it also removes the bridge gap between the customers and the restaurants by efficient technology which reduces delivery time. It also provides rating based pricing model for foods by which a customer can choose the best dish or a restaurant according to their needs.

ZOMATO provides information of all restaurants about their food menu, detailed analysis of restaurant, its branches and locations, service hours and reviews & comments from other users who have already experience of having food on respective restaurant. The business model of ZOMATO is affiliate marketing model which work with website and mobile apps. The website enables user to search for food in different ways. Such as by name of food, general search in a specific location, search for available restaurant open in particular time of a day and even they can find out if the restaurant has a bar, buffet, live music, happy hours and other services. This is a great platform for people who love to eat and very choosy in terms of choice of food.

In today's world large amount of data is being generated on daily basis, this data can be any data for example the data of a company or data of government and there is a need for this data to mined and analyzed. Every company for its success wants to analyze the data of their company to gain more profit.

In our project we have tried to analyze the data of ZOMATO which has various fields of data which include restaurant name, address, contact details, user reviews, cost and etc. and we have tried to visualize various fields of data through various graphs and tried to find the trend available in the data.

CONTENTS

PROJECT OVERVIEW

- The data set consists of 51717 rows and 17 columns which contain the data about various restaurants that are working with zomato.
- There are many fields about a restaurants like the name of the restaurants, website of the restaurants, address of the restaurants and etc.
- The data set also contains the ratings that a customer provides about the restaurants.
- It also has column "votes' that stores the votes received to a restaurant provided by the customers
- According to the data present in the data set we can do the analysis and by the analysis we can find which types of restaurants a customer prefers.
- The main purpose of the project is to analyze the data and by the analysis find which restaurants are prefer by the customers and by the analysis help zomato to increase its business.

DATA MINING

Data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. Data mining tools allow enterprises to predict the future trends.

Data mining parameters

In data mining, association rules are created by analyzing data for frequent if/then patterns, then using the support and confidence criteria to locate the most important relationship within the data. Support is how frequently the items appear in the database, while confidence is the number of times if/then statements are accurate.

Other data mining parameters include Sequence or Path Analysis, Classification, Clustering and Forecasting, Sequence or Path Analysis parameters look for patterns where one event leads to another later event. A Sequence is an order lists of set of items, and it is a common type of data structure found in many databases. A Classification parameter looks for new patterns and might result in a change in the way the data is organized. Classification algorithms predict variables based on other factors within the database.

Four stages of data mining

1 DATA SOURCES	2 DATA EXPLORATION/ GATHERING	3 MODELING	4 DEPLOYING MODELS
These range from databases to news wires, and are considered a problem definition	This stage involves the sampling and transformation of data	Users create a model, test it, and then evaluate	Take an action based on the results from the models

Clustering parameters find and visually document groups of facts that were previously unknown. Clustering groups a set of objects and aggregates them based on how similar they are to each other.

There are different ways a user can implement the cluster, which differentiate between each clustering model. Fostering parameters within data mining can discover patterns in data that can lead to reasonable predictions about future, also known as predictive analysis.

Data mining tools and techniques

Data mining techniques are used in many research areas, including mathematics, cybernetics, genetics and marketing. While data mining techniques are a mean to drive efficiencies and predict customer behavior, if used correctly, a business can set itself apart from its competition through the use of predictive analysis.

Web mining, a type of data mining used in customer relationship management, integrates information gathered by traditional data mining methods and techniques over the web. Web mining aims to understand customer behavior and to evaluate how effective a particular website is.

Other data mining techniques include network approaches based on multitask learning for classifying patterns, ensuring parallel and scalable execution of data mining algorithms, the mining of large databases, the handling of relational and complex data types, and machine learning. Machine learning is a type of data mining tool that designs specific algorithms from which to learn and predict.

Benefits of data mining

In general, the benefits of data mining come from the ability to uncover hidden patterns and relationships in data that can be used to make predictions that impact businesses.

Specific data mining benefits vary depending on the goal and the industry. Sales and marketing departments can mine customer data to improve lead conversion rates or create one-to-one marketing campaigns. Data mining information on historical sales patterns and customer behaviors can be used to build prediction models for future sales, new products and services.

Companies in the financial industry use data mining tools to build risk models and detect fraud. The manufacturing industry uses data mining tools to improve operations.

Jupyter Notebook Tool

Jupyter Notebook (formerly IPython Notebooks) is a web-based interactive computational environment for creating Jupyter notebook documents. The "notebook" term can colloquially make reference to many different entities, mainly the Jupyter web application, Jupyter Python web server, or Jupyter document format depending on context. A Jupyter Notebook document is a JSON document, following a versioned schema, and containing an ordered list of input/output cells which can contain code, text (using Markdown), mathematics, plots and rich media, usually ending with the ".ipynb" extension.

Jupyter Notebook can connect to many kernels to allow programming in many languages. By default Jupyter Notebook ships with the IPython kernel. As of the 2.3 release (October 2014), there are currently 49 Jupyter-compatible kernels for many programming languages, including Python, R, Julia and Haskell.

What is data Visualization

Data visualization is a general term that describes any efforts to help people understand the significance of data by placing it in visual context. Patterns, trends and correlations that might go undetected in text-based data can be exposed and recognized easier with data visualization software.

Today's data visualization tools go beyond the standards charts and graphs used in Microsoft Excel spreadsheets, displaying data in more sophisticated ways such as infographics, dials and gauges, geographic maps, sparklines, heat maps, and detailed bar, pie and fever charts. The images may include interactive capabilities, enabling users to manipulate them to drill into the data for querying and analysis. Indicators designed to alert users when data has been updated or predefined conditions occur can also be included.

Importance of Data Visualization

Data visualization has become the de facto standard for modern business intelligence (BI). The success of the two leading vendors in the BI space, Tableau and Qlik – both of which heavily emphasize visualization has moved vendors towards a more visual approach in their software. Virtually all B I software has strong data visualization functionality.

Data visualization tools have been important in democratizing data and analytics and making data driven insights available to workers throughout an organization. They are typically easier to operate than traditional statistical analysis software or earlier versions of BI software. This has led to a rise in lines of business implementing data visualization tools on their own, without support from IT.

Data visualization software also plays an important role in big data and advance analytics projects. As business accumulated massive troves of data during the early years of the big data trend, they needed a way to quickly and easily get an overview of their data. Visualization tools were a natural fit.

Visualization is central to advance analytics for similar reasons. When a data scientists is writing advanced predictive analytics or machine learning algorithms, it becomes important to visualize the outputs to monitor results and ensure that models are performing as intended. This is because visualizations of complex algorithms are generally easier to interpret than numerical outputs.

Examples of Data Visualization

Data visualization tools can be used in a variety of ways. The most common use today is as a BI reporting tool. Users can setup visualization tools to generate automatic dashboards that track company performance across key performance indicators and visually interpret the results.

As a data visualization vendors extend the functionality of these tools, they are increasingly being used as front ends for more sophisticated bif data environments. In this setting, data visualization software helps data engineers and scientists keep track of data sources and do basic exploratory analysis of data sets prior to or after more detailed advanced analyses.

Setting up the Environment and Data Import

1). Loading Packages:-

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import matplotlib as mpl

pd.options.mode.chained_assignment = None

2). Data Import:-

data=pd.read_csv('zomato.csv')

data.head()

Out[3]:	url	address	name	online_order	book_table	rate	votes	phone	location	rest_type	dish_liked	cuisines
	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
	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
	omato.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
	v.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
	ww.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

Fig. 1(Data set Overview)

3). Data Cleaning:-

data['rate'].fillna(0,inplace=True)

data['rest_type'].fillna(0,inplace=True)

data['approx_cost(for two people)'].fillna(0,inplace=True)

data['cuisines'].fillna(0,inplace=True)

data['dish_liked'].fillna(0,inplace=True)

4).Data Preparation:-

We apply feature engineering and change the data from the data set into usable form like converting 'True' to 1 and 'False' to 2 and so on. This is an important step for data analysis.

5).Data Exploration and Visualization:-

corrMatrix=data.corr()

f,ax=plt.subplots(figsize=(9,8))

ax=sns.heatmap(corrMatrix,annot=True,cmap='icefire')

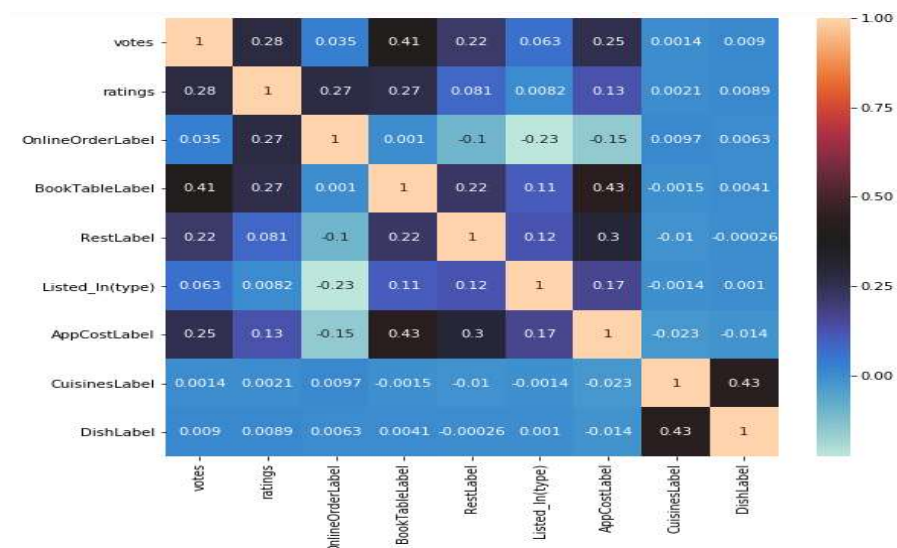Fig. 2.1 (Correlation Matrix)

Count Plot:-

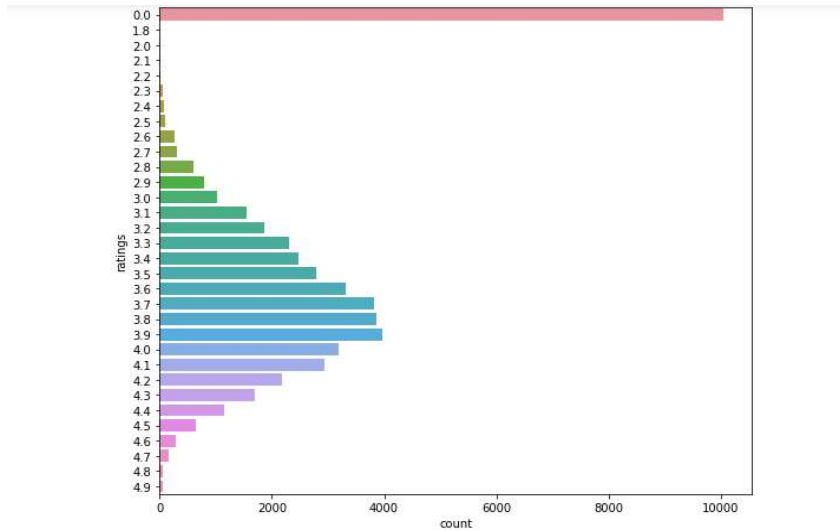plt.figure(figsize=(9,8))

sns.countplot(y='ratings',data=data)

Fig. 2.2 (Countplot 'ratings')

Bar Plot:-

plt.figure(figsize=(9,8))

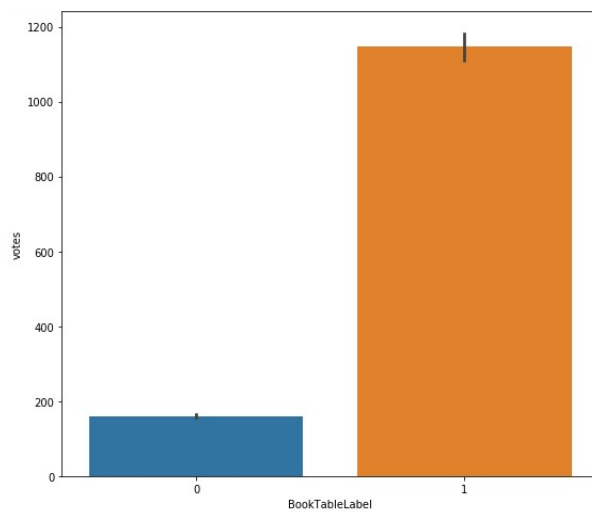sns.barplot(x='BookTableLabel',y='votes',data=data)

Fig. 2.3 (Barplot 'BookTableLabel' vs 'votes')

Need for Choosing the Project

The purpose of the business needs analysis is to clearly understand the business and its needs. A BNA might be completed by a company after identifying they would like to implement a new strategy. In this case, the company would conduct a BNA to understand the current state, identify the key drivers for change, and then make recommendations or plans for how to implement the new strategy.

The major need for choosing this project is to business analysis and on the basis of analysis analyze customer demands and try to provide them with best food as well as from company point of view help them to grow and provide better services to their customer and try to gain as much profit they can.

CONCLUSION

The project is complete and the analysis of the data present in the dataset of zomato is done and visualization of data is done by various graphs as shown above.

Hereby we can conclude the following things after analyzing the data:-

1) On reading the dataset and analyzing the dataset we found a that there are total 51717 rows and 17 columns in our dataset.

2) We found the all the null values present in our dataset and tried to handle them by filling the null values by '0'. We found a field 'menu_item' and found it does not contains any useful data therefore we dropped the column from our data.

3) After that we found a lot of categorical data present in our dataset, therefore we had apply feature eengineering to convert the data into useful data.

4) Then our next step was to explore and visualize the data.

5) We found the co-relation matrix and found co-relations among the following:-

    a) ratings and votes

    b) ratings and OrderOnlineLabel

    c) votes and BookTableLabel

    d) ratings and AppCostLabel

    e) OnlineOrder and votes and etc.

6) Then we plotted various graphs and from the graphs we can conclude the following observations:-

    a) The restaurants with 4.7 customer ratings got 6000 votes (approx)

    b) The restaurants with 4.9 customer ratings got 4000 votes (approx)

    c) The restaurants which had the facility of booking the tables had more number of votes

    d) The restaurants which had the facility of online food booking had approx 300 votes.

e) The restaurants which had more cost for 2 persons had more ratings therefore we can conclude the resturants which are expensive provide good service

7) We can finally conclude from our visualizations and analysis of data that:-

a) User perfer more votings and ratings restaurants

b) Restaurants with more ratings and votes are expensive

c) The restaurants which have table booking facility are more prefered

d) The restaurants which have online ordering facility are more popular

e) The restaurants with more ratings and votes have more good food and are more popular

REFERENCES

IEEE Reference Format – look into any IEEE paper

By Date- 09/09/2019

https://en.wikipedia.org/wiki/Project_Jupyter

https://www.kaggle.com/shrutimehta/zomato-restaurants-data

https://www.analyticsvidhya.com/blog/2019/01/learning-path-data-scientist-machine-learning-2019/

https://searchbusinessanalytics.techtarget.com/definition/business-analytics-BA