**Banglore Eatery Zomato Dataset**

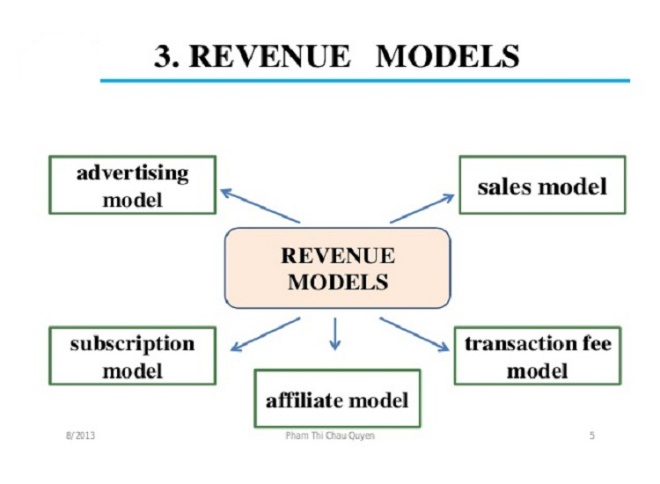
Shiva sai karne (1099134)

skarne@my.bridgeport.edu

Master of Science [Computer Science]

**Introduction**

Zomato is a stage to look through Indian eateries and their data. It was begun in the year 2008 by Deepinder Goyal and Pankaj Chaddah. Beforehand it was named as foodiebay later it was renamed as Zomato in 2010 November. Zomato was propelled in various states in India. In spite of having overwhelming rivalries from different sellers, Zomato was getting one of the main food conveyance administrations in India.

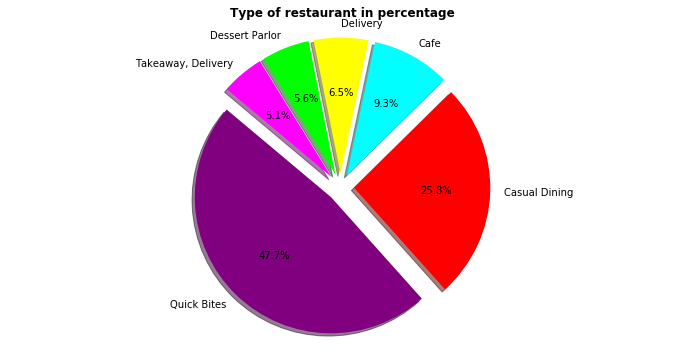
Revenue of the Zomato depends on various factors like quality of the food,

Hospitality, ratings from the customers, However the revenue of the restaurant can be improved by different sources such as restaurant advertising, event advertising, event ticket sale, consulting services, online food ordering and subscription.

**Main Body**

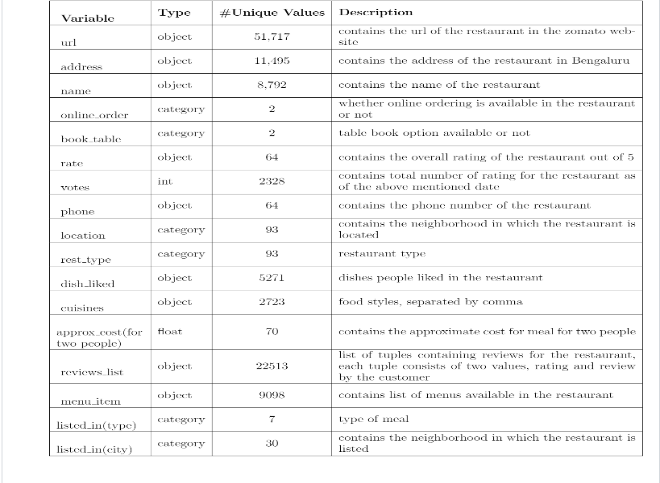
Zomato is where different cafés and their data could be found. This makes the client to pick the café appropriately. Every information that assumes a significant job for the café's business could be found on zomato. Proprietors of the eateries can promote their business and improve showcasing and for the client or clients it's anything but difficult to pick a café as indicated by their need. As there is part of data that could be dissected or could be taken a shot at in the dataset of zomata, this dataset has been decided for the undertaking.

There are several aspects or scenarios that can be worked on and draw a conclusion out of the dataset of zomato from customers point of view and from owners of the restaurant point of view. Zomato is almost all over the world. More the clients, more the difficult articulations come up to break down. The dataset that has been picked is from an area i.e Bengaluru, India. The following is the accompanying example information indicating the Analysis of Zomato Bengaluru eateries. It would be examined in this undertaking utilizing Hadoop.



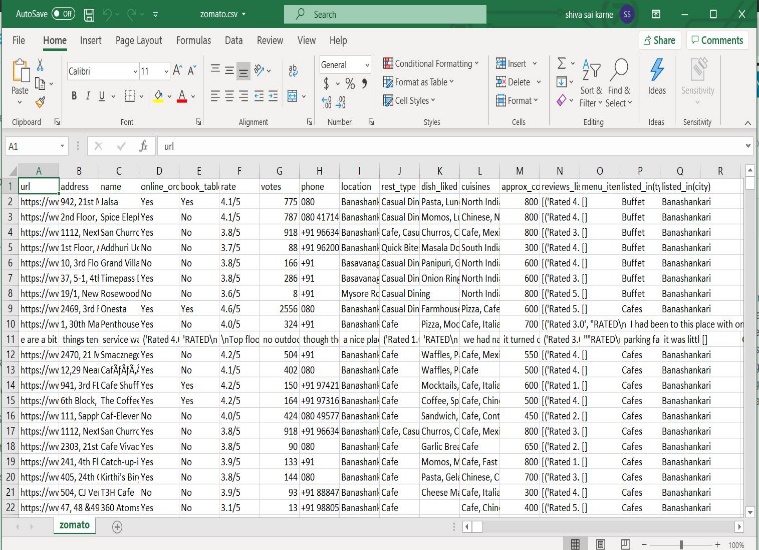
The above pie graph shows the kind of eatery in rates i.e the conveyance choices of cafés is about 6.5%, bistro accessibility is 9.3%, ordinary eat in is 25.8 on generally rate, remove accessibility is 5.1%, for the café and the light meals is 47.7%, dessert parlor with level of 5.1%. A large portion of the cafés incline toward feast in, remove and conveyance choices, as the majority of the benefit rely upon these three choices.

Analysis of Zomato Bengaluru restaurants is shown in the below table



The above data set contains following details of restaurant like url of the restaurants, location details, name of the restaurant, online order availability , booking table, cost of the each food item, votes based on food, phone, location, rest type, dish liked, cuisines, approx. cost of food, reviews from customers, list of various foods, type listed in, city in which restaurant located.

Below is the sample screenshot of the dataset



**Some Hadoop components used:**

Hive: It is an open source information to peruse, compose and oversee enormous dataset records which are spared legitimately in HDFS or other information stockpiling frameworks.

Flash: It is general group processing structure which gives scheme to programming whole bunch with circuitous likeness and adaptation to internal failure.

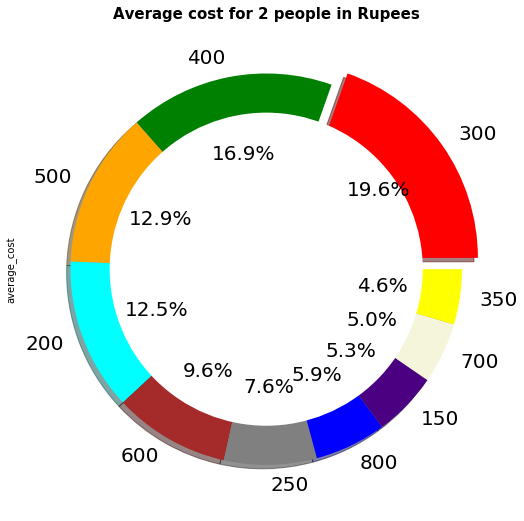
BigQuery: It is a webservice which approves interactional examination of enormous dataset acting in simultaneousness with google stockpiling.

Mapreduce: It is a programming model and a related execution to process and create huge informational collections with equal, scatter calculation on a group.

**Result and Evaluation:**

The following statements can be analyzed using the Zomato dataset of Bengaluru are as follows

* Restaurant with highest rating i.e restaurants whose rating is over 5
* Average cost for two people dining in one location irrespective of the number of restaurants.
* Food order availability through online sources
* Restaurants which has food delivery option and dine in option
* Restaurants with rating below 2.



The above pie chart shows the average cost for 2 people in most of the restaurants. Here majority of the restaurants i.e, 19.6% restaurants charge an average cost of Rs 300 per two people where as least number of restaurants i.e, 4.6% restaurants charge Rs.350 for two people. Ratings will be decided not only based on quality, hospitality of restaurants but also the average cost of the food.

By using the Hadoop components like Map reduce, Pig, Hive, the project can be implemented. So, here by using map reduce the rating of the restaurant can be found out and also I just want to know the average cost of dining for two people in a respective location, though there can be man y restaurants in one location. using hive the restaurants with delivery options can be found. With the help of spark the dine in availability restaurants can be found out and with the google big query it can list all the restaurant name, type and also the restaurants with rating less than 2.

**Evaluation**:

* All the difficult proclamations will be dissected by utilizing various segments of the Hadoop in this undertaking. To make the Hadoop biological system to be utilized to the most extreme to take care of the issue articulations various parts will be utilized relying upon the necessity to tackle the issue.
* The code will be written to analyze the data using the clourera’s VM ware in which Hadoop ecosystem has been installed.
* HDFS will be used to store the input data, output data.
* Input data will be mounted to the HDFS, then code of the different components will be implemented on the input and then respective output from different components will be stored on HDFS again.

**References**:

* Kaggle (<https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants>)
* Google