## DATA MANAGEMENT PROJECT REPORT



Zomato Restaurant

Submitted By:

Regumudi Saikumar

11801885

KM073

Under the guidance of: Miss Ashu

School of Computer Science and Engineering

Lovely Professional University, Phagwara

## CERTIFICATE

This is to certify that Regumudi Saikumar bearing Registration number 11801885 has completed Data Management (INT 217) project titled, “Zomato Restaurant” under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

Signature and Name of the Supervisor School of Computer Science and Engineering Lovely Professional University

Date:

## DECLARATION

I, Regumudi Saikumar, student of Computer Science and Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: Signature

Registration No.: 11801885

## ACKNOWLWDGEMENT

I take this opportunity to present our votes of thanks to all those guideposts who really acted as lightening pillars to enlighten my way throughout this Project that has led to successful and satisfactory completion of this Project. I am grateful to Lovely Professional University for providing us with an opportunity to undertake this Project and providing us with all the facilities. I am highly thankful to All for their active support, valuable time and advice, whole-hearted guidance, sincere cooperation and painstaking involvement during the project and in completing the assignment of preparing the said project within the time stipulated. Lastly, I am thankful to all those, particularly the various friends, who have been instrumental in creating proper, healthy and conductive environment and including new and fresh innovative ideas for me during the project, without their help, it would have been extremely difficult for me to complete the project in a time bound framework.

## INDEX

### Introduction

1. Scope of the Analysis
2. Existing System

 Drawbacks

1. Source of dataset
2. ETL process
3. Analysis on dataset
   1. Restaurants with high aggregate ratings
   2. City having more restaurants
   3. Restaurants having online delivery facility
   4. Restaurants having table booking facility
   5. Restaurants with number of cuisines
   6. City got most votes
4. Dashboard
5. Future Scope
6. References
7. Bibliography
8. INTRODUCTION

Data analysis is a process of inspecting, cleansing, transforming and modelling data with the goal of discovering useful information conclusions and supporting decision making. Data analysis has multiple facts and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science and social science domains.

Zomato analysis is one of the most useful analysis for food who want to taste the best cuisines of every part of the world which lies in their budget. This analysis can also be useful to find the most preferable restaurants in various parts of the country.

The collected data has been stored in the comma separated value in Zomato.csv. Every restaurant contains the following variables.

|  |  |  |
| --- | --- | --- |
|  | Restaurant Name | : Name of the restaurant |
|  | City | : City in which restaurant is located |
|  | Address | : Address of the restaurant |
|  | Locality | : Location in the city |
|  | Locality Verbose | : Detailed description of the locality |
|  | Cuisines | : Food items provided by the restaurant |
|  | Has Table Booking | : If the restaurant has table booking or not |
|  | Has Online Delivery | : If the restaurant has online delivery or not |
|  | Aggregate Rating | : Average rating out of 5 |
|  | Rating Text | : Ratings of the customer as text |
|  | Votes | : Total number of votes given by the customers |

### SCOPE OF THE ANALYSIS

* 1. Analysing the best restaurant based on aggregate rating.
  2. Analysing the restaurants from the customer votes.
  3. Analysing the city with a greater number of restaurants.
  4. Check whether restaurants have online delivery facility or not.
  5. Analyse the number of currencies paid in the restaurants.
  6. Check whether restaurants have table booking facility or not.
  7. Analyse the most preferable cuisines.
  8. Analyse the cities which get a greater number of votes.

### EXISTING SYSTEM

The existing system is currently used to make the dashboard on “Microsoft Excel for Office 365 MSO 64-bit” of version “16.0.11328.20438” and for data cleaning “Tableau Prep Builder” of version “19.31.19.0923.1415” is used.

 DRAWBACKS

* 1. Manual Effort: It is more time consuming and takes lots of efforts of the user to manipulate and analyse tons of data in an excel workbook with lots of sheets. With huge amount of data excel slows down the process and sometimes it does not respond.
  2. Human Error: As the amount of data grows within an Excel spreadsheet, there is increased room for human error and formula errors because of the manual process of copying and pasting data into the workbook.
  3. Security: User gets the full access to the dashboard and data within it. This can become a major issue when sensitive information is being shared both internally and externally because there is no way to secure this data.

### SOURCE OF DATASET

The source of the Zomato dataset is Kaggle. Link is provided below:

https://[www.kaggle.com/shrutimehta/zomato-restaurants-data#zomato.csv](http://www.kaggle.com/shrutimehta/zomato-restaurants-data#zomato.csv)

### ETL PROCESS

ETL is defined as a process that extracts the data from different source systems, then transforms the data and finally loads the data into the Data Warehouse system. ETL is known as Extract, Transform and Load.

* + 1. Extraction:

For data extraction we need to know which variables are not required to fulfil the analysis. In this case Zomato.csv has such columns like restaurant ID, country code, longitude, latitude does not need for the analysis. So, we discard or extract it.

* + 1. Transformation:

In this process we need to check for the data mistakes like spelling mistakes such as “Aalishaan” and “Aalishan” or “Aashirwad” and “Ashirbad”, removing null values, grouping same category data such as “Delhi” and “New Delhi” or “Bengal” and “West Bengal”. We need to transform the data in such a way that it is suitable for the required analysis. All these processing is done in “Tableau Prep Builder”.

* + 1. Load:

After finishing the transformation of the data, we need to get the output of the cleaned dataset and new data is saved in a new excel file named Zomato\_Dataset.csv. We need to save the data in xlsx format so that we can use all the features of the excel. New file name is saved with named Zomato\_Dashboard.xlsx. After that we need to open the dataset in Excel for further analysis.

### ANALYSIS ON DATASET

* + 1. Restaurants with High Aggregate Ratings
       - Introduction

In order to get the best restaurants with customer review is by their given ratings. In this analysis we get the restaurants with high aggregate ratings.

* + - * General Description

We have the columns of Restaurant Name and Aggregate Ratings for this analysis. We need to make a table for top 10 restaurants. So, we create a table with column names Rank, Restaurant Name and Aggregate Ratings with 10 rows.

* + - * Specific Requirements

We need three functions to get this job done. In column Aggregate Rating we apply LARGE function and for the Restaurant Name we apply INDEX and MATCH functions.

* + - * Analysis Result

Here we analysed top 10 restaurants having highest aggregate ratings.

* + - * Visualization

From the bar chart, we can conclude that “Café Coffee Day” has the highest aggregate

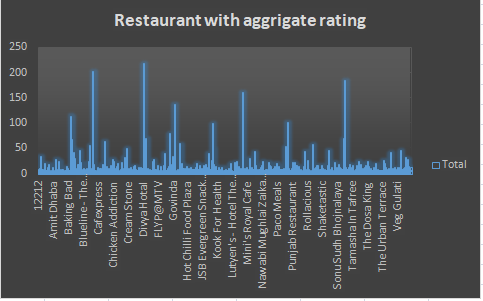
ratings over value

350. After “Café Coffee day”, “Burger King” and Barbeque Nation are the next

restaurant having higher aggregate rating.

We can also select Restaurant Names manually using Slicers. In this slicer we select randomly 10 Restaurant Names and it reflected on the bar chart as well.

From here we can conclude that “Apni Rasoi” has the highest aggregate ratings from the selected restaurants.



* + 1. City Having More Restaurants
       - Introduction

To get to know about which city has the highest number of restaurants we need to analysis this one. In this dataset we are provided with city names.

* + - * General Description

We have the columns of City and Restaurant name for this analysis. Let’s create 5 rows for the top 5 cities with restaurants.

* + - * Specific Requirements

We need LARGE function to find the largest value from the total number of restaurants. INDEX function is also needed to match the cities corresponding to the value of the Restaurant columns.

* + - * Analysis Result

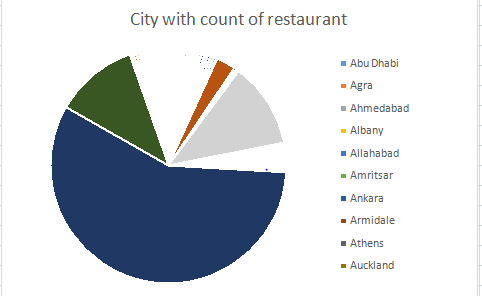
Here are the top 5 cities with its number of restaurants.

* + - * Visualization

From the pie chart we can conclude that New Delhi has the greatest number of

restaurants in the following dataset. We can also select city names with the help of slicers and get the cities

with maximum number of restaurants from the selected cities.



* + 1. Restaurant Having Online Delivery Facility
       - Introduction

To get to know that how many restaurants have Online Delivery Facility and how many have not. We need to count the total number of restaurants that has online delivery facility and same for which has not.

* + - * General Description

We have two columns of Has Online Delivery and Count of Restaurant Names. We need to create two rows to calculate.

* + - * Specific Requirements

We need SUM function to calculate the total number of restaurants which have and have not online delivery.

* + - * Analysis Result

Here we get the two rows with the corresponding value.

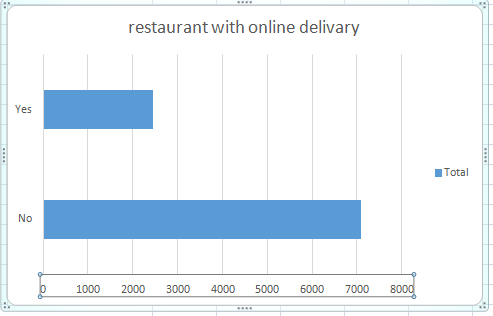


Visualization

After analysis we get that 7085 restaurants do booking and the rest 2451 restaurants do have

facility for the customers.

not have table online delivery



* + 1. Restaurant Having Table Booking Facility
       - Introduction

To get to know that how many restaurants have Table booking Facility and how many have not. We need to count the total number of restaurants that has table booking facility and same for which has not.

* + - * General Description

We have two columns of Has Table Booking and Count of Restaurant Names. We need to create two rows to calculate.

* + - * Specific Requirements

We need SUM function to calculate the total number of restaurants which have and have not table booking.

* + - * Analysis Result

Here we get the two rows with the corresponding value.

* + - * V

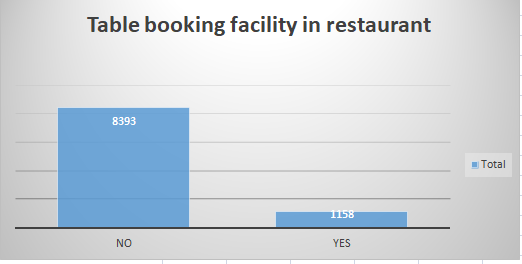
isualization

After analysis we get that 8373 restaurants do not ha ooking and the rest 1158 restaurants do have table

ve table

b booking

facility for the customers.



* + 1. Restaurant with number of Cuisine
       - Introduction

Many restaurants provide different types of food items. We need to look which food item is mostly available in market. In order to get the result, we need to perform this analysis.

* + - * General Description

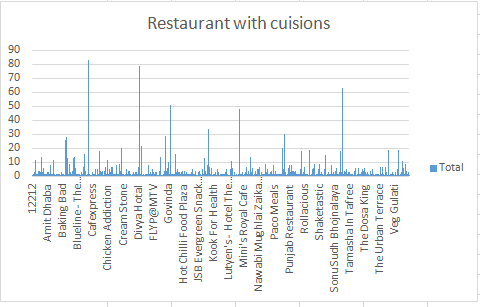
We have the column name of Restaurant Name and Count of Cuisines. Let’s create a table of 10 rows for the top 10 restaurants having highest cuisines.

* + - * Specific Requirements

We need LARGE function to find the largest value from the table and INDEX function to get the restaurant name corresponding to the value of number of cuisines.

* + - * Analysis Result

Here is the top 10 Restaurant with high value of cuisines. “Café express” has the highest number of cuisines.



* + - * Visualization

After analysing we can see in the area graph that “Café express” hast the greatest number of Cuisines along with “Divya hotel”. We can manually select restaurant names to get the cuisines from the slicers.

* + 1. City Got Most Votes
       - Introduction

Sometimes its better to know which city is preferable to the customers. In this case we need to study and analyse the cities based on the votes of the customers.

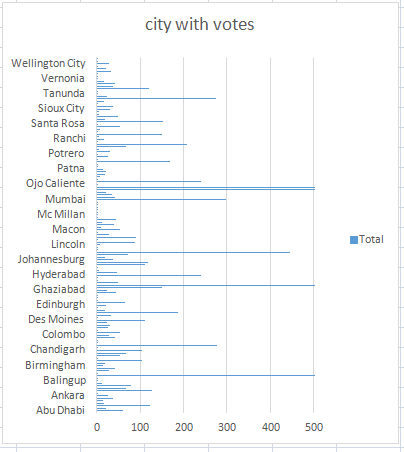
* + - * General Description

We have two columns named City and Sum of Votes. Let’s create 15 rows to see the top 15 cities with greater number of votes given by the people.

* + - * Specific Requirements

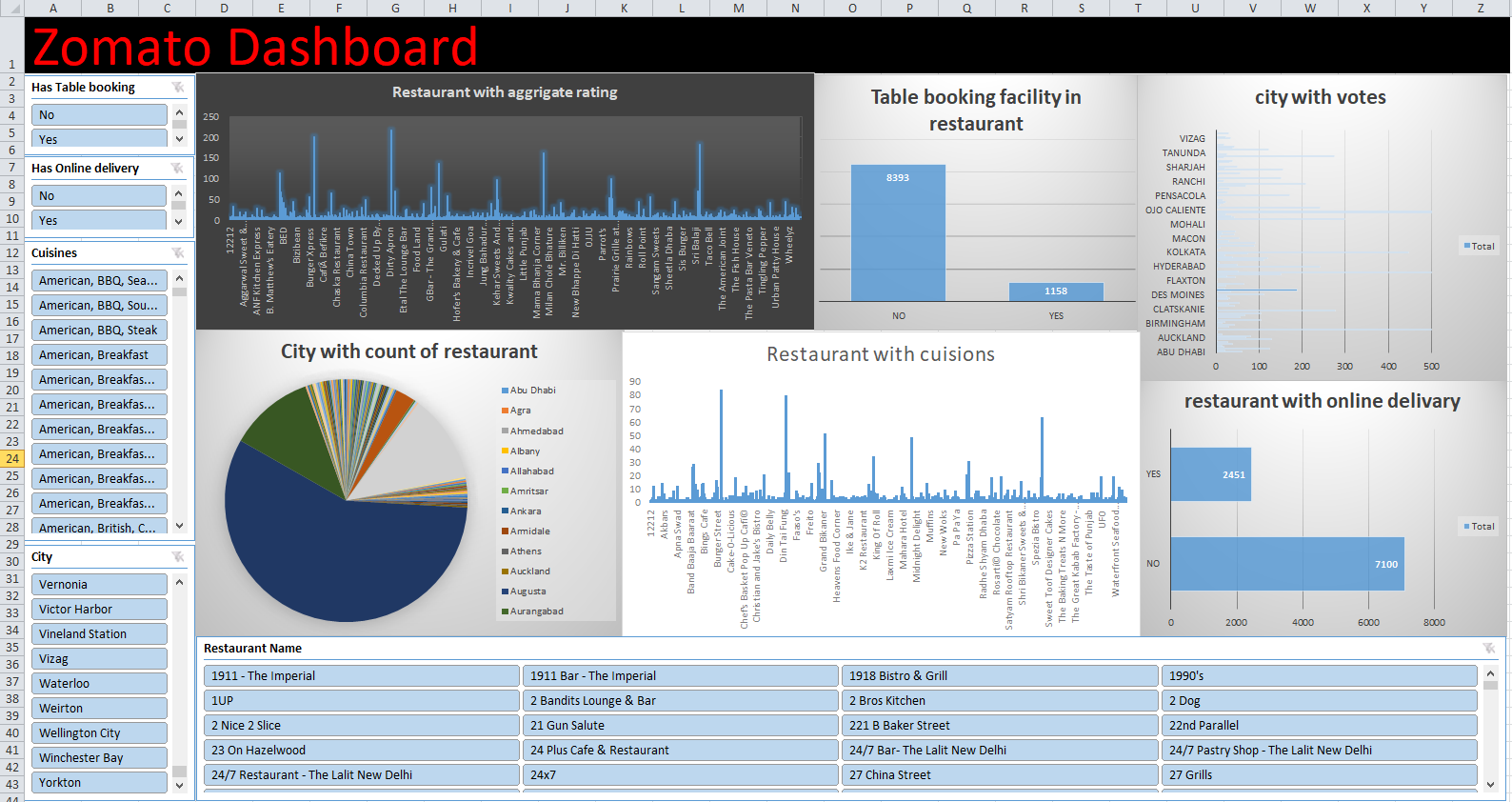
We need to apply LARGE function to pick up the greatest value from the votes column and apply the INDEX function to match the cities with the corresponding value of sum of votes. We can apply SMALL function to

get the city names with lowest votes.



### DASHBOARD

An Excel Dashboard is one pager that helps managers and business leaders in tracking key KPIs or metrics and take a decision based on it. It contains charts or tables or slicers that are backed by data.

Here is Zomato Restaurant Dashboard below.

# ZOMATO RESTAURANT DASHBOARD

### FUTURE SCOPE

Making dashboard is a way to predict and analyse the future scopes of the data. It makes the manager or the business leaders easy to find problems and get the efficient solution. It helps us to visualize the data in tabular formats or different kinds of charts. In this dashboard one can predict that the city “New Delhi” will be the preferable for food restaurants. It can be also predictable that in future there will be more online delivery facility or not.

### REFERENCES

I have taken help from my Data Science teacher Miss. Ashu Mam. I have also taken help from my classmates to complete this project. I am very thankful to them. Apart from these I have researched on the internet for more knowledge about Excel and Dashboard.

### BIBLIOGRAPHY

* https://en.wikipedia.org/wiki/Dashboard\_(business)
* https://[www.guru99.com/etl-extract-load-process.html](http://www.guru99.com/etl-extract-load-process.html)
* https://[www.youtube.com/watch?v=K74\_FNnlIF8&t=1564s](http://www.youtube.com/watch?v=K74_FNnlIF8&amp;t=1564s)
* https://[www.youtube.com/watch?v=sM2mdBu0Owc&t=106s](http://www.youtube.com/watch?v=sM2mdBu0Owc&amp;t=106s)