**PROJECT SYNOPSIS**

**of**

**Zomato Restaurants Data Analysis**

**Abstract:**

We really get fascinated by good quality food being served in the restaurants and would like to help community find the best cuisines around their area.

For looking of our greatest restaurants we frequently goes for various websites and apps to induce an overall idea of restaurants service. The foremost important criteria for all this is often rating and reviews of the those that have already got experience in these restaurants. People see for rating and compare these restaurants with one another and choose for his or her best. The data in this dataset is restricted to Delhi-NCR.

This Zomato dataset provides us with enough information in order that one can decide which restaurants is suitable at which place and what kind of food they must serve so as get maximum profit. The data consist of **9552** rows and **22** columns during this dataset. We'd wish to find the most affordable restaurant in Delhi-NCR.

Since it's a true time data we might start first with **data cleaning** like **cleaning** **spaces**, **garbage texts** etc and then performing **Exploratory Data Analysis**.

**Data Storage:**

The collected data has been stored in the Comma Separated Value file Zomato.csv. Each restaurant in the dataset is uniquely identified by its Restaurant Id. Every Restaurant contains the following variables:

• Restaurant Id: Unique id of every restaurant across various cities of the world

• Restaurant Name: Name of the restaurant

• Country Code: Country in which restaurant is located

• City: City in which restaurant is located

• Address: Address of the restaurant

• Locality: Location in the city

• Locality Verbose: Detailed description of the locality

• Longitude: Longitude coordinate of the restaurant's location

• Latitude: Latitude coordinate of the restaurant's location

• Cuisines: Cuisines offered by the restaurant

• Average Cost for two: Cost for two people in different currencies.

• Currency: Currency of the country

• Has Table booking: yes/no

• Has Online delivery: yes/ no

• Is delivering: yes/ no

• Switch to order menu: yes/no

• Price range: range of price of food

• Aggregate Rating: Average rating out of 5

• Rating color: depending upon the average rating color

• Rating text: text on the basis of rating of rating

• Votes: Number of ratings casted by people

**OBJECTIVES:**

**The study has the following objectives or business use cases:**

* Merge the country sheet and Zomato file to get the country code
* Get the Top rated restaurants in each city in India (On the basis of rating and votes).
* Finding out relationship between the rating and votes.
* No of Restaurants in each country.
* Top 5 restaurants with online delivery
* cheap but best restaurants available in a city
* Top cuisines in each region.
* Aggregate rating of all the restaurants in each city in a country.
* Does rating influences the cost of restaurant? (Draw boxplot for this).
* Top Percentage cover of restaurants in a city.(Draw pie graph for this).
* Top cuisines in Indian restaurants.( Draw pie graph for this)

**Methods Used:**

**1. Data Collection**

The process of gathering and analyzing accurate data from various sources to find answers to research problems, trends and probabilities, etc., to evaluate possible outcomes is Known as **Data Collection**. Knowledge is power, information is knowledge, and data is information in digitized form, at least as defined in IT. Hence, data is power. But before you can leverage that data into a successful strategy for your organization or business, you need to gather it. That’s the first step

**2.** **Data Pre-Processing** :

The average cost for two column has some 0 values which is improper. Hence the two records are removed.

**3. Exploratory Data Analysis**

**Exploratory Data Analysis (EDA)** is an approach that is used to analyze the data and discover trends, patterns, or check assumptions in data with the help of statistical summaries and graphical representations.

**Types of EDA**

Depending on the number of columns we are analyzing we can divide EDA into two types.

* [**Univariate Analysis**](https://www.geeksforgeeks.org/univariate-bivariate-and-multivariate-data-and-its-analysis/) – In univariate analysis, we analyze or deal with only one variable at a time. The analysis of univariate data is thus the simplest form of analysis since the information deals with only one quantity that changes. It does not deal with causes or relationships and the main purpose of the analysis is to describe the data and find patterns that exist within it.
* **Bivariate analysis –** This type of data involves two different variables. The analysis of this type of data deals with causes and relationships and the analysis is done to find out the relationship between the two variables
* **Multivariate analysis**  – When the data involves three or more variables, it is categorized under multivariate.