DATA MINING AND ANALYTICS

MINI PROJECT REPORT

CHENNAI RESTAURANT RECOMMENDATION SYSTEM

Nithish Kumar (RA1911003010217)

Aryan Agarwal (RA1911003010224)

Rohan Aditya Ram (RA1911003010240)

CONTENTS

Contents

[Introduction 3](#_Toc87541008)

[Motivation 4](#_Toc87541009)

[Novelty 4](#_Toc87541010)

[Modules 5](#_Toc87541011)

[**Architecture Diagram** 5](#_Toc87541012)

[**Dataset** 5](#_Toc87541013)

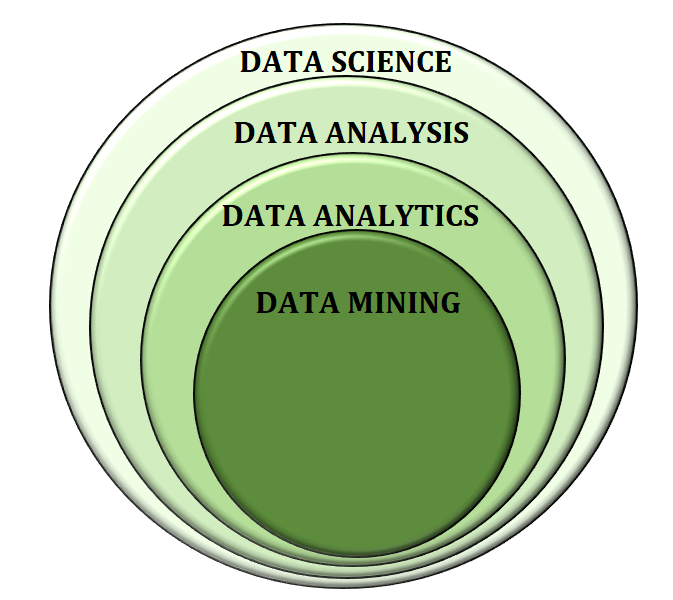
[Implementation 7](#_Toc87541014)

[Output 9](#_Toc87541015)

[Conclusion 11](#_Toc87541016)

# Introduction

Data can be simply defined as facts or figures, or information that's stored in or used by a computer. It is actually much more than that. Today’s Data Analysts and Tech companies feed on the abundant data that is available. Data is the fodder for their algorithms. It is what makes their algorithms better and more efficient. The study of data and making sense out of it can be thought of as Data Analytics. Data Analytics is a field of computing that has gained a lot of popularity and interest in the last 10 years. Data mining is the process of extracting and discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.

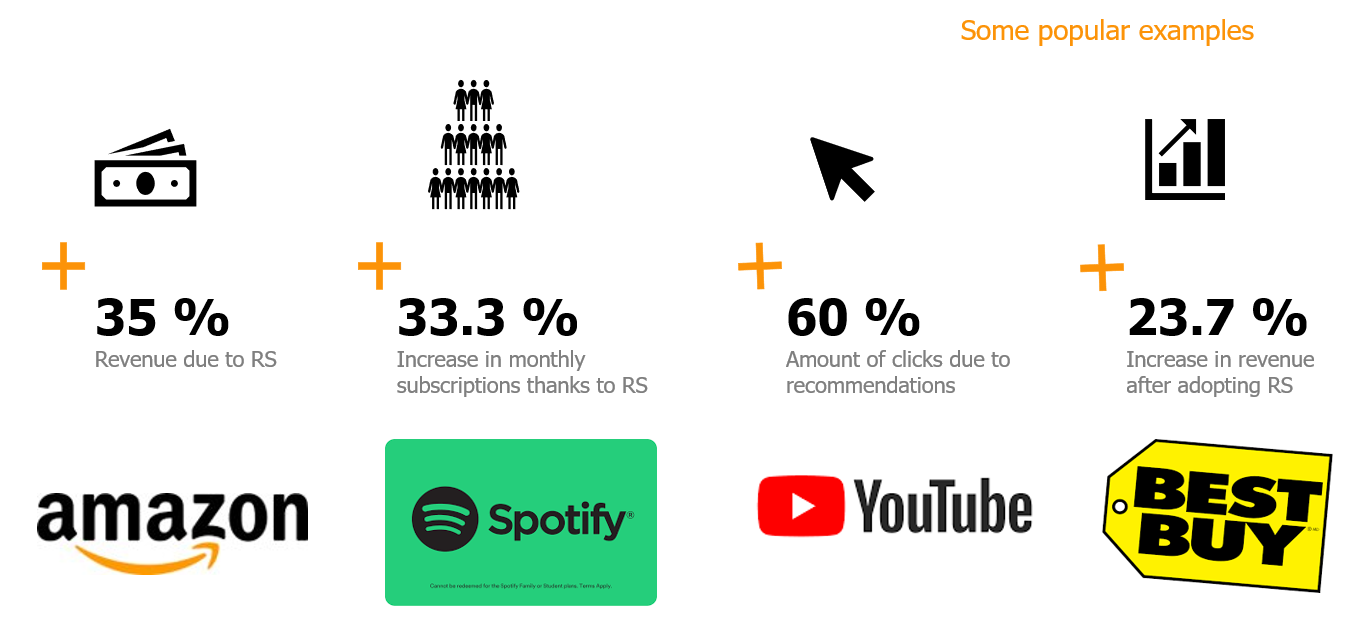


The above image gives a perfect insight of what data mining is and what data analytics is. Data Mining can be thought of a part of Data Analytics.

Recommendation Systems is an application of data mining and analytics. They are designed to recommend things to the user based on many factors. The systems predict the most probable thing(s) you would like based on your past and present preferences. The recommendation system deals with a large volume of information present by filtering the most important information based on the data provided by a user and other factors that take care of the user’s preference and interest. It finds out the match between user and item and imputes the similarities between users and items.

# Motivation

Recommendation systems are being used by a lot of companies to increase their customer attraction. The Recommendation system with a combination of Machine Learning algorithms gives rise to self-learning applications. It is very frustrating to sit and search for something you would like. If an application could recommend what you would like based on your past liking and preferences, your job would become easier. Here is where the Restaurant Recommendation system comes in handy. Based on your liking and past choices, you are given a list of restaurants to pick from that are selected carefully for you based on your taste.



This image gives a clear idea as to how recommendation systems have increased a company’s revenue. Users tend to prefer application which make their job easier and the Restaurant Recommendation System is definitely one such application.

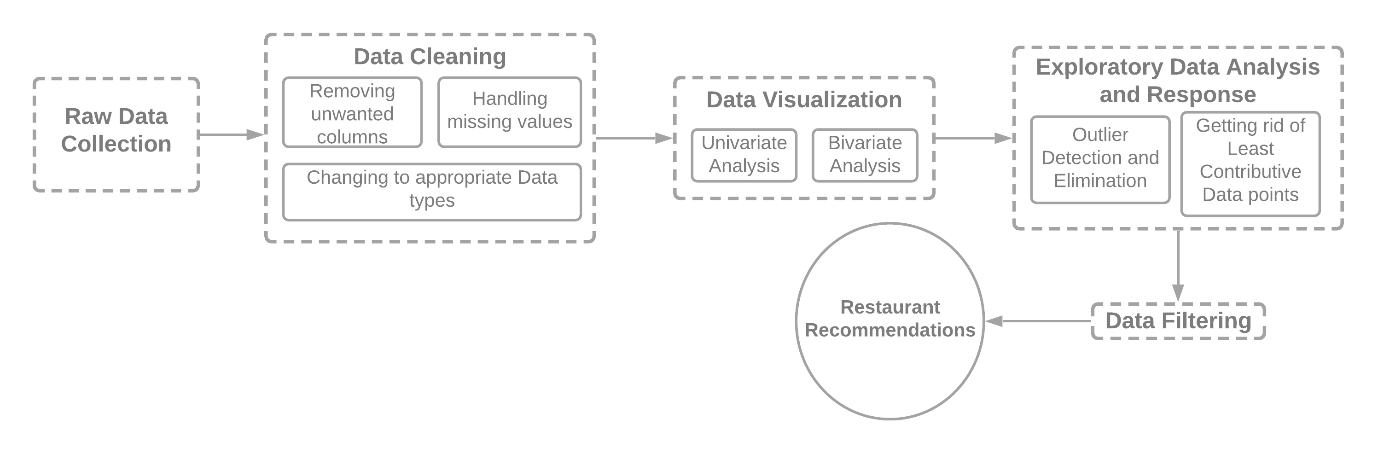
# Novelty

The new features our Restaurant Recommendation system has include:

* Best recommendations of restaurants in the city of Chennai.
* User specific recommendations.
* Over 4500 restaurants for user to explore.

# Modules

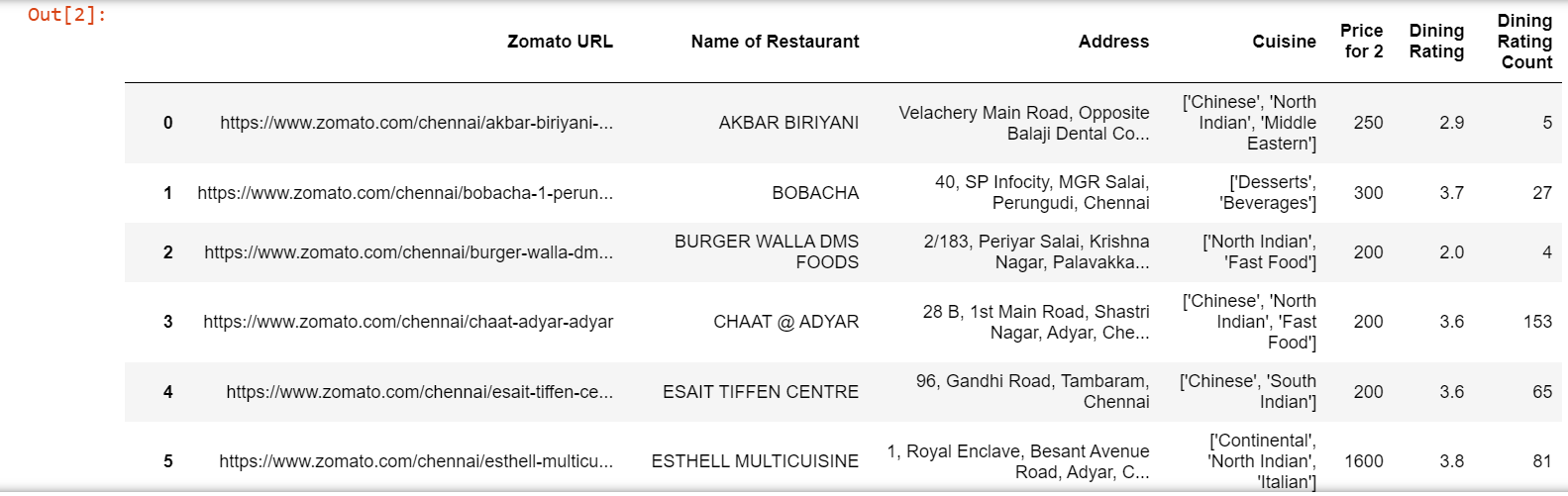
## **Architecture Diagram**



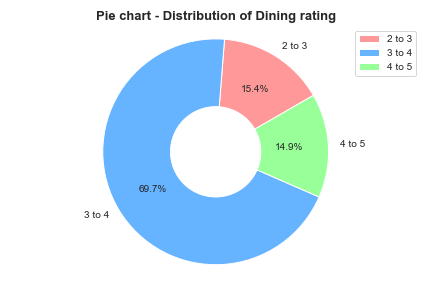
* Raw data is collected.
* Data Cleaning – Using the main restaurant branch. Categorizing the restaurant cuisine properly. Correction of data types and handling missing values.
* Data Visualization – Gaining insights of the dataset by applying univariate and bivariate analysis.
* EDA – Outlier Detection
* Data Filtering – Restaurant data filtered according to user choice.
* Restaurant Recommendations – The top recommendations based on user choice are displayed in the page (created using streamlit)

## **Dataset**

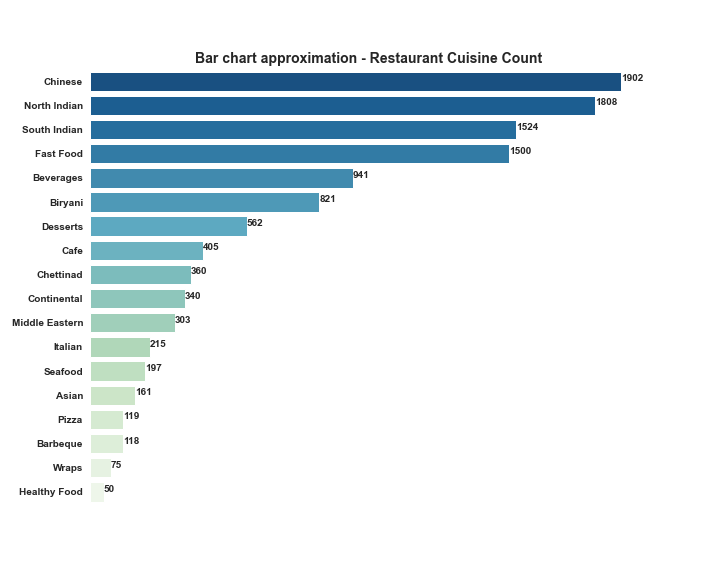
Dataset link - <https://www.kaggle.com/phiitm/chennai-zomato-restaurants-data>



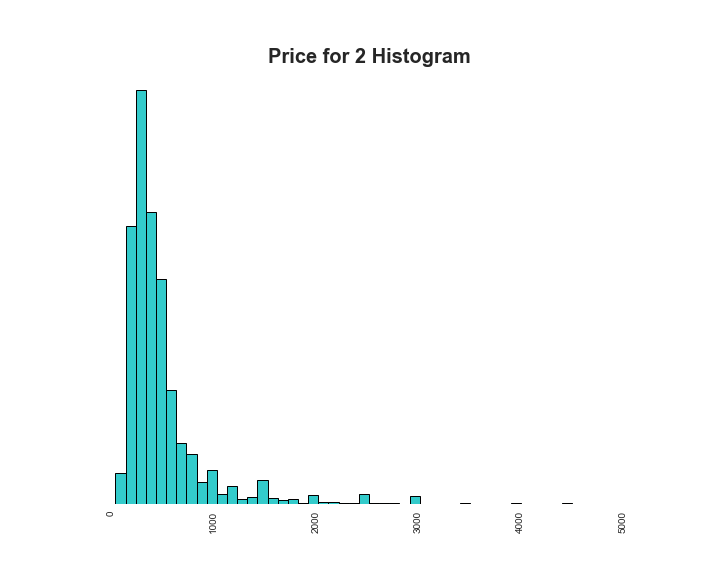
The dataset contains 12032 rows and 12 columns. The 12 columns are: Zomato URL, Name of Restaurant, Address, Location, Cuisine, Top Dishes, Price for 2, Dining Rating, Dining Rating Count, Delivery Rating, Delivery Rating Count and Features.



The above pie chart shows the percentage of data in the different dining rating range.



The above bar chart shows the number of restaurants which serve a particular cuisine. We can infer that 1902 restaurants in Chennai, serve Chinese food.

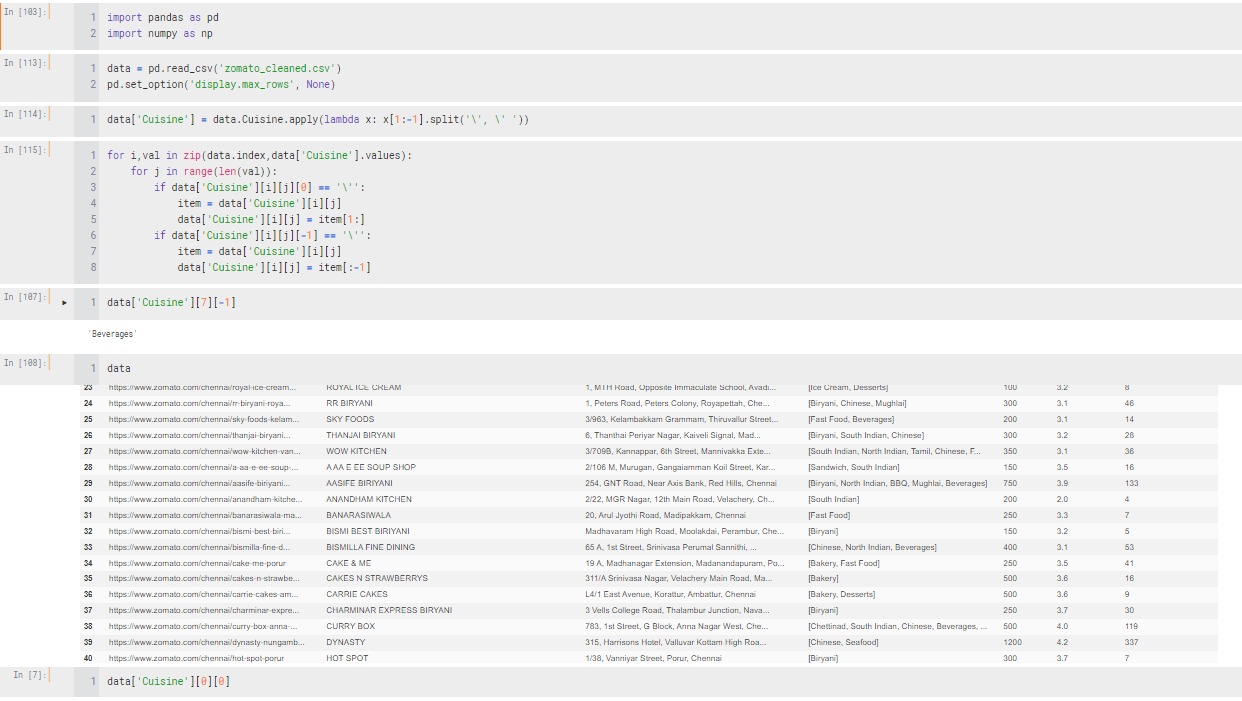


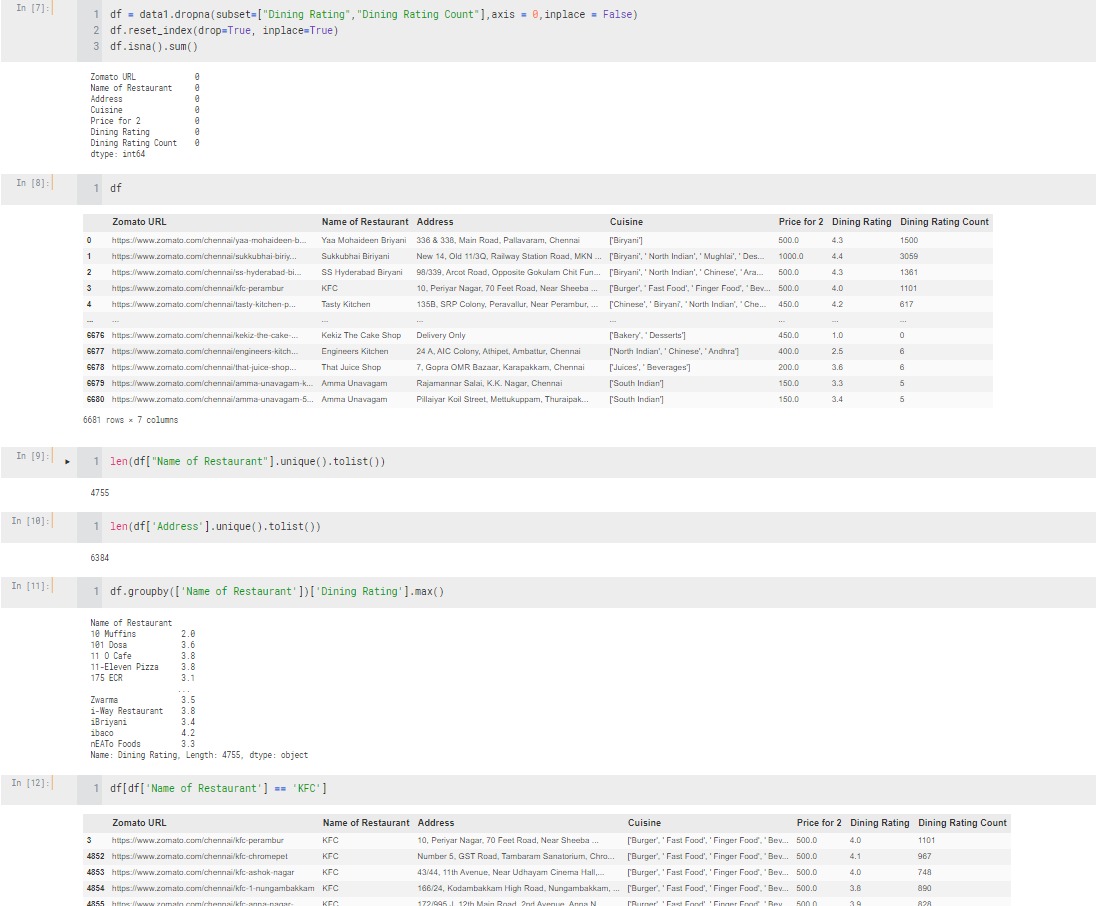
This histogram shows the distribution of cost for two people at a restaurant.

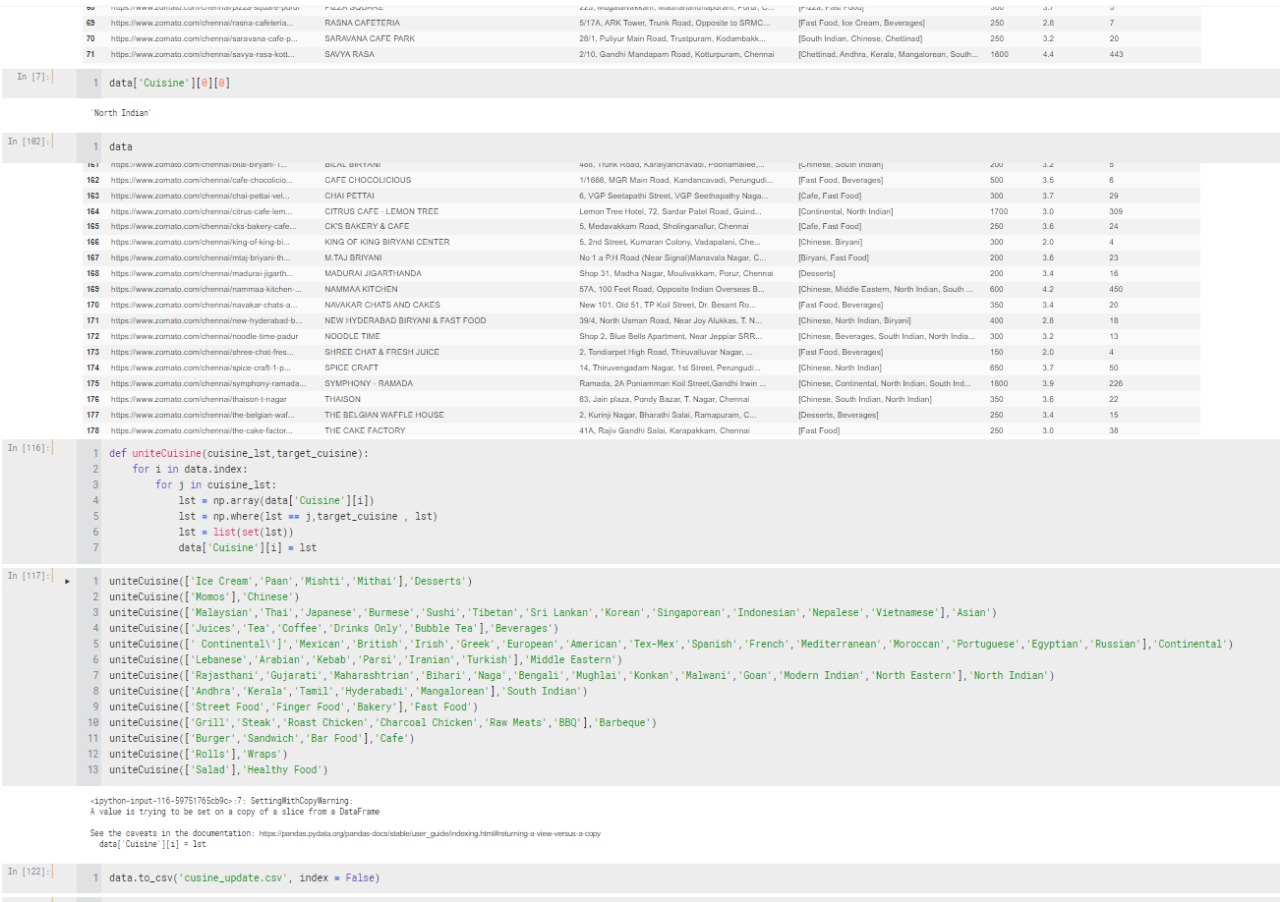
# Implementation

Tools used:

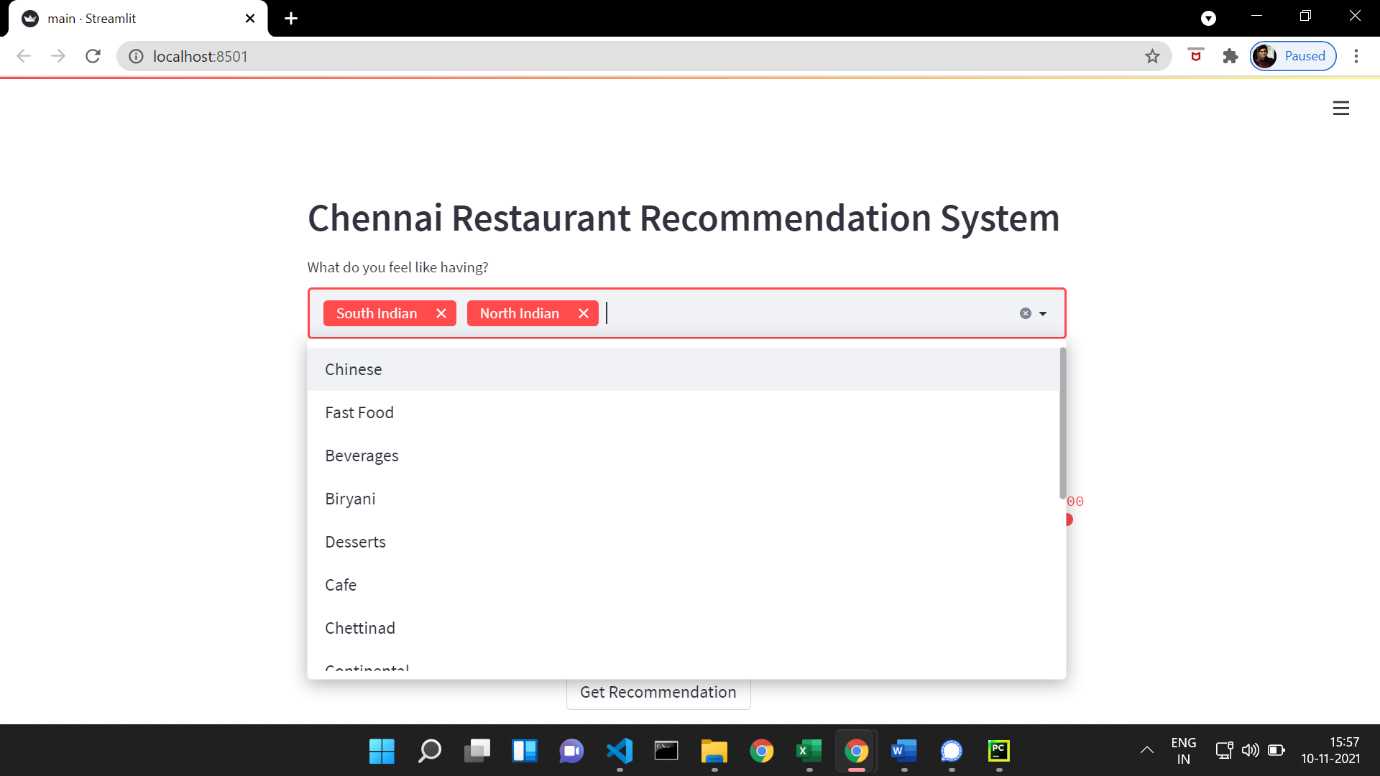
* Pycharm
* Jupyter-Notebook
* Matplotlib
* Pandas
* seaborn
* numpy
* pySpark
* streamlit.



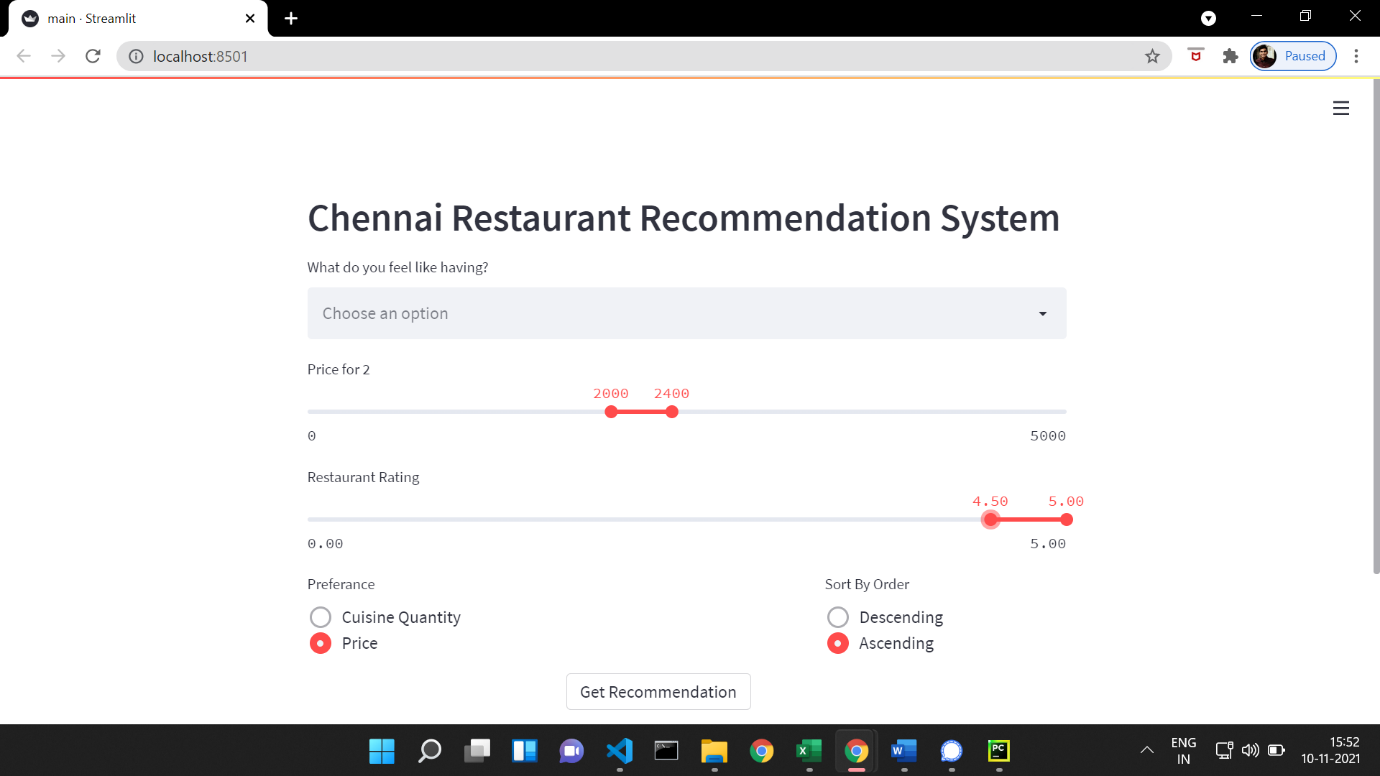




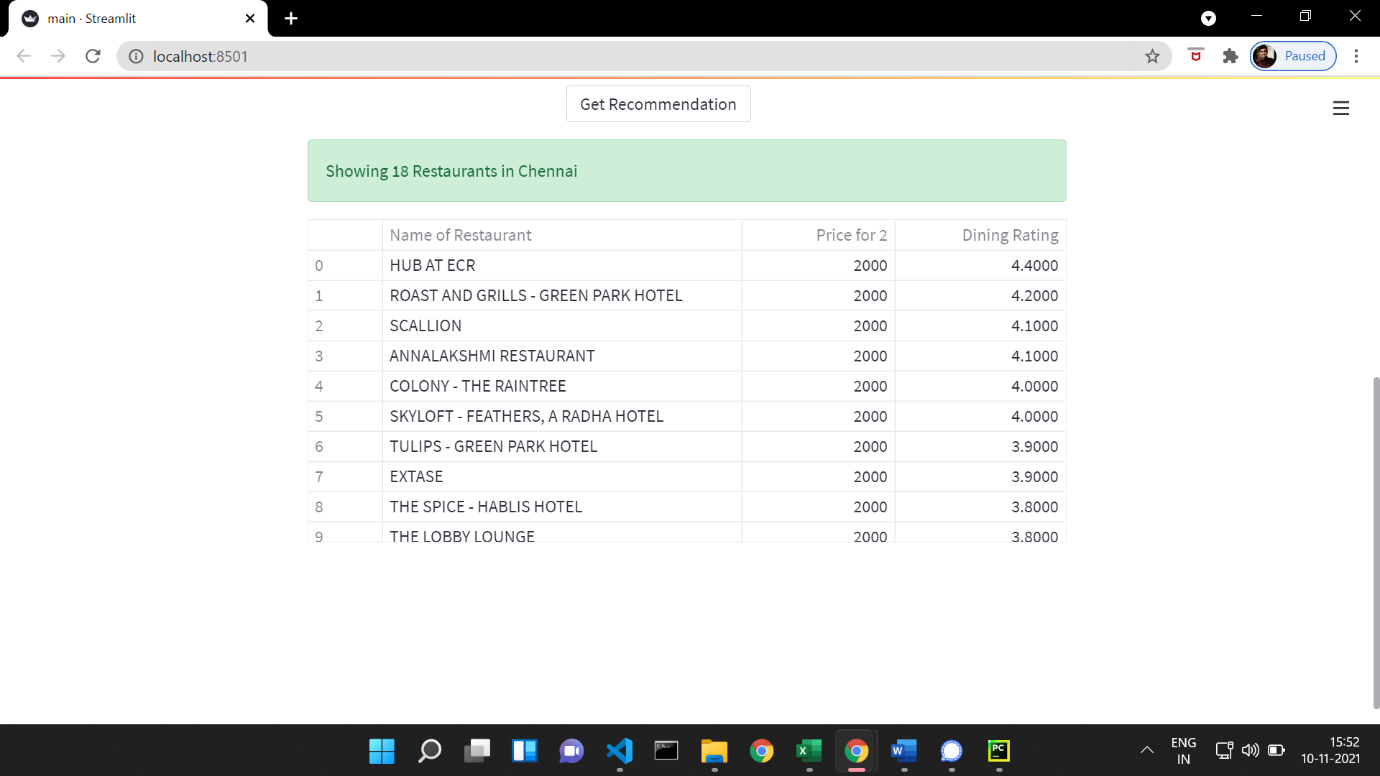
# Output



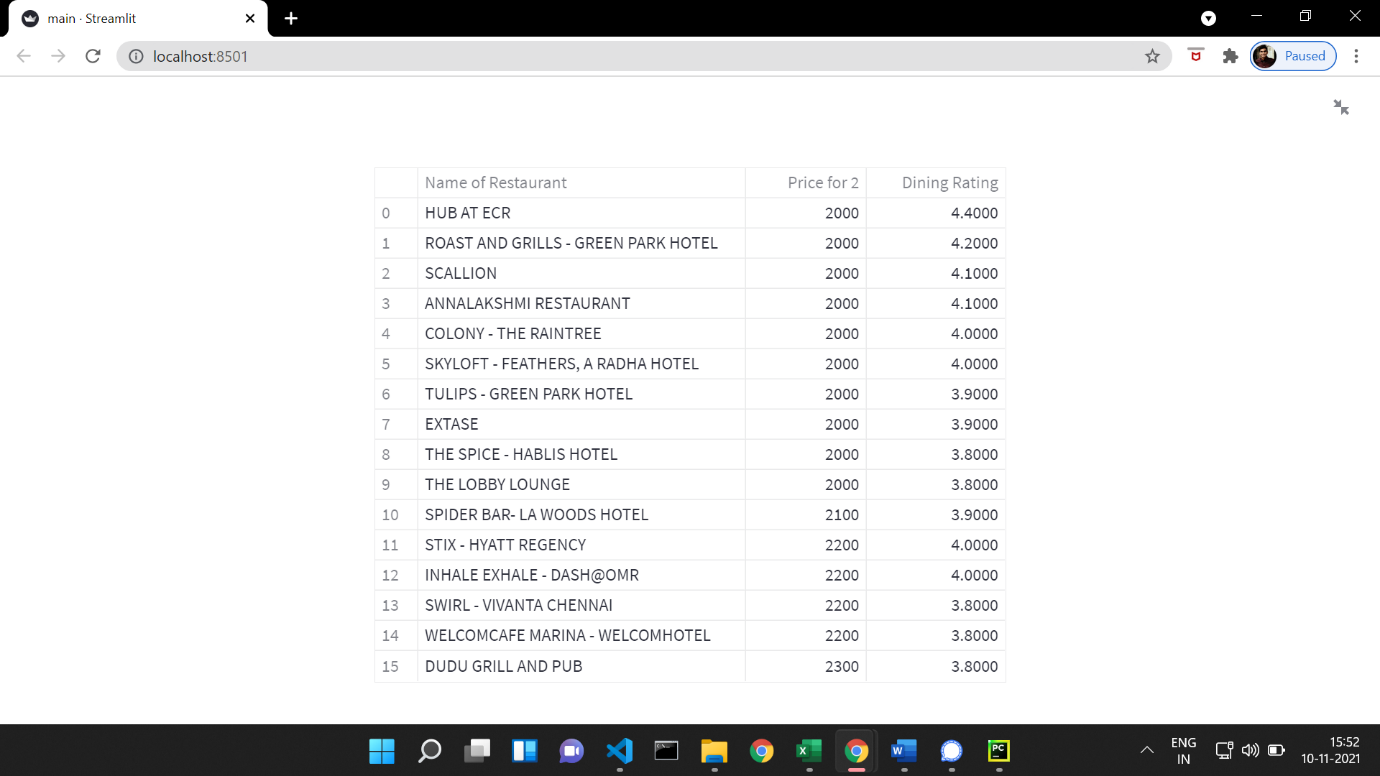
User chooses cuisine from the list.



User adds other constraints and choices.



Restaurant Recommendations displayed to user.



Restaurant Recommendations in the form of a data frame.

# Conclusion

* Recommendation Systems help the growth of a business.
* Good Recommendations can increase customers.
* Food is subjective. So, a recommendation system that understand your liking and recommends can be trusted.

Limitation of the project:

* Only restaurants in Chennai
* Location parameter was not considered for the recommendation.