**Capstone Project Submission**

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| * Anjali Tidke ( [anjalitidke123@gmail.com](mailto:anjalitidke123@gmail.com) ) * Importing Libraries * Loading Dataset * Data Inspection * Data Preprocessing * Data Cleaning and Preparation * Clustering * Sentiment analysis * Models to choose for hyperparameter tuning * Conclusions * Shubham Dukare ([shubhamalmabetter@gmail.com](mailto:shubhamalmabetter@gmail.com)) * Importing Libraries * Loading datasets * Exploratory Data Analysis * Feature Engineering * One hot encoding * Feature scaling * Hyperparameter tuning * Conclusions |
| **Please paste the GitHub Repo link.** |
| Github Link (Shubham Dukare):-  <https://github.com/shubham4955/Zomato-Restaurant-Clustering-and-Sentiment-Analysis>  Github Link (Anjali Tidke):-  <https://github.com/AnjaliTidke/Zomato-Restaurant-clustering-analysis>  Anjali Tidke Drive Link: -  <https://drive.google.com/drive/folders/1Tlb8048fqKEZGtMwq3nxfq9rkCjYQOV2?usp=sharing>  Shubham Drive Link: -  <https://drive.google.com/drive/folders/13mMPQgPc8ThbU33wCXP4jI9yoSicBmKx?usp=sharing> |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**   |  | | --- | | Problem Statement: The problem statement is to analyze the sentiment of reviews given by the customers of the Zomato and make some meaningful insights from the data that would solve some business cases  Introduction: Zomato is an Indian restaurant aggregator and food delivery start-up founded by Deepinder Goyal and Pankaj Chaddah in 2008. Zomato provides information, menus and user-reviews of restaurants, and also has food delivery options from partner restaurants in select cities.  India is quite famous for its diverse multi cuisine available in a large number of restaurants and hotel resorts, which is reminiscent of unity in diversity. Restaurant business in India is always evolving. More Indians are warming up to the idea of eating restaurant food whether by dining outside or getting food delivered. The growing number of restaurants in every state of India has been a motivation to inspect the data to get some insights, interesting facts and figures about the Indian food industry in each city. So, this project focuses on analysing the Zomato restaurant data for each city in India.  The Project focuses on Customers and Company, you have to analyze the sentiments of the reviews given by the customer in the data and made some useful conclusion in the form of Visualizations. Also, cluster the zomato restaurants into different segments. The data is vizualized as it becomes easy to analyse data at instant. The Analysis also solve some of the business cases that can directly help the customers finding the Best restaurant in their locality and for the company to grow up and work on the fields they are currently lagging in.  This could help in clustering the restaurants into segments. Also the data has valuable information around cuisine and costing which can be used in cost vs. benefit analysis  Data could be used for sentiment analysis. Also the metadata of reviewers can be used for identifying the critics in the industry.  Project work flow:   1. Importing Libraries 2. Loading the Dataset 3. Data Inspection 4. Data Preprocessing 5. Data Cleaning and Preparation 6. Exploratory Data Analysis 7. Feature Engineering 8. One hot encoding 9. Feature scaling 10. Clustering 11. Sentiment analysis 12. Conclusion   EDA work Result:  From the EDA, we found out the Top 10 most expensive restaurants, top 10 cheapest restaurants, best restaurants with rating and price, Top cuisines served y restaurants.  Total models used are –   * 1. Decision Tree Classifier   2. Random Forest Classifier   3. K Nearest Neighbours Classifier   4. SVC   5. Logistic Regression   6. Multinomial Naïve Bayes   7. XGB Classifier   8. LGBM Classifier   Starting with loading the data so far we have done EDA, null values treatment, encoding of categorical columns, feature selection, and then model building.  We observed that Logistic regression model is working fine as compared to other models. Its accuracy and recall is maximum. Finally, our conclusion is that the Logistic regression model is the best suitable model for this sentimental analysis of the project. | |