**“Unlocking Insights in Zomato Data: An Analysis using Regression, Classification, and Clustering”**

***Introduction***

Food delivery platforms like Zomato provide a wealth of information about restaurants, menus, and customer reviews. To make the most of this data, it’s important to have the right tools and techniques in place for analyzing it. In this blog post, we’ll be showcasing how we used regression, classification, and clustering techniques to analyze Zomato data and unlock insights that can help improve the platform.



***Regression for Rating Predictions***

One of the key challenges in analyzing Zomato data is predicting the ratings for restaurants. To tackle this challenge, we applied regression algorithms, including linear regression, lasso and ridge regression, decision tree regression, and random forest regression. The goal was to find the best model for predicting the ratings based on factors such as cuisine, location, price, no of votes, customer reviews etc.

After comparing the performance of the models using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared, we found that random forest regression was the best model, providing the highest accuracy in rating predictions. This model was able to take into account the complex interactions between the different factors that influence the ratings, making it more accurate than other regression algorithms.

***Classification for Two-Rating Classes***

Another challenge in analyzing Zomato data is classifying restaurants into two classes based on their ratings. To do this, we used classification algorithms, including logistic regression, KNN, support vector machines, decision tree and random forest classifier. Restaurants with a rating greater than or equal to 4 were classified as class 1, while restaurants with a rating less than 4 were classified as class 0.

We compared the performance of the models using accuracy and F1-score, and found that decision tree and random forest classifier provided the best results. These models were able to accurately classify the restaurants into two classes, making it easier to understand the distribution of ratings on the platform and identify areas for improvement.

***Clustering for Zomato Insights***

Finally, we used clustering techniques to cluster restaurants based on their cuisines, ratings, and cost. We used clustering techniques such as K-means and PCA and found that PCA provided the best results. We used the elbow method to determine the optimal number of clusters and created an interactive dashboard to present the insights we uncovered.

The clustering techniques allowed us to group similar restaurants together based on their cuisines, ratings, and cost, making it easier to understand the distribution of restaurants on the platform and identify areas for improvement. The interactive dashboard we created provides a visual representation of the insights we uncovered, making it easier for users to understand and explore the data.

***Conclusion***

In conclusion, we were able to use regression, classification, and clustering techniques to analyze Zomato data and uncover valuable insights. The random forest regression model provided the highest accuracy in rating predictions, while decision tree and random forest classifier provided the best results for two-rating classes. The PCA clustering model was the best for clustering restaurants based on their cuisines, ratings, and cost. By using these techniques, we hope to have shown how data analysis can help improve platforms like Zomato and provide valuable insights to customers and restaurants. Whether you’re a beginner or an experienced data analyst, the techniques we used in this project are a great starting point for analyzing food delivery data and unlocking valuable insights.