Cognifyz

17/02/2025

INTERNSHIP REPORT

LEVEL 3

PREPARED FOR :

Cognifyz Technologies

COGNIFYZ TECHNOLOGIES DATA SCIENCE INTERNSHIP LEVEL 3 REPORT

Level 3 Objectives

Level 3 of the Cognifyz Data Science Internship focuses on the following tasks:

1. Predictive Modelling
2. Customer Preference Analysis, and
3. Data Visualization.

Task 1: Predictive Modelling

- Build a regression model to predict the aggregate rating of a restaurant based on available features.
- Split the dataset into training and testing sets and evaluate the model's performance using appropriate metrics.
- Experiment with different algorithms (e.g., linear regression, decision trees, random forest) and compare their performance.

Task 2: Customer Preference Analysis

- Analyse the relationship between the type of cuisine and the restaurant's rating.
- Identify the most popular cuisines among customers based on the number of votes.
- Determine if there are any specific cuisines that tend to receive higher ratings.

Task 3: Data Visualization

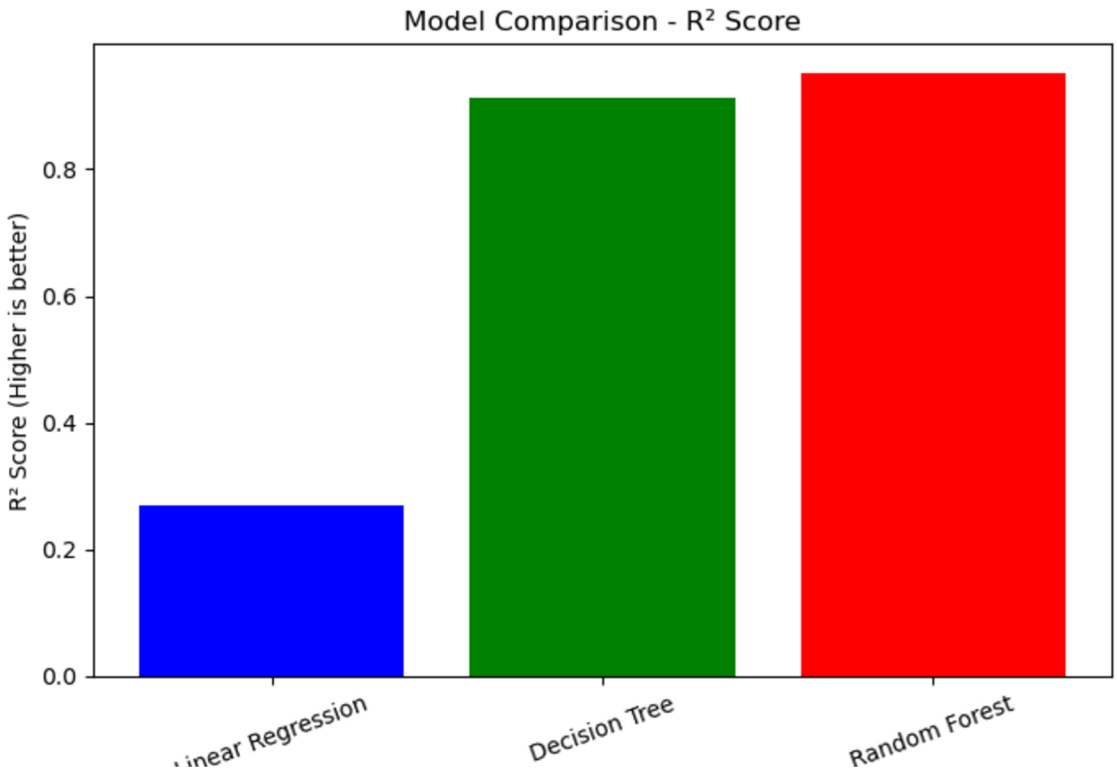
- Create visualizations to represent the distribution of ratings using different charts (histogram, bar plot, etc.).
- Compare the average ratings of different cuisines or cities using appropriate visualizations.
- Visualize the relationship between various features and the target variable to gain insights.

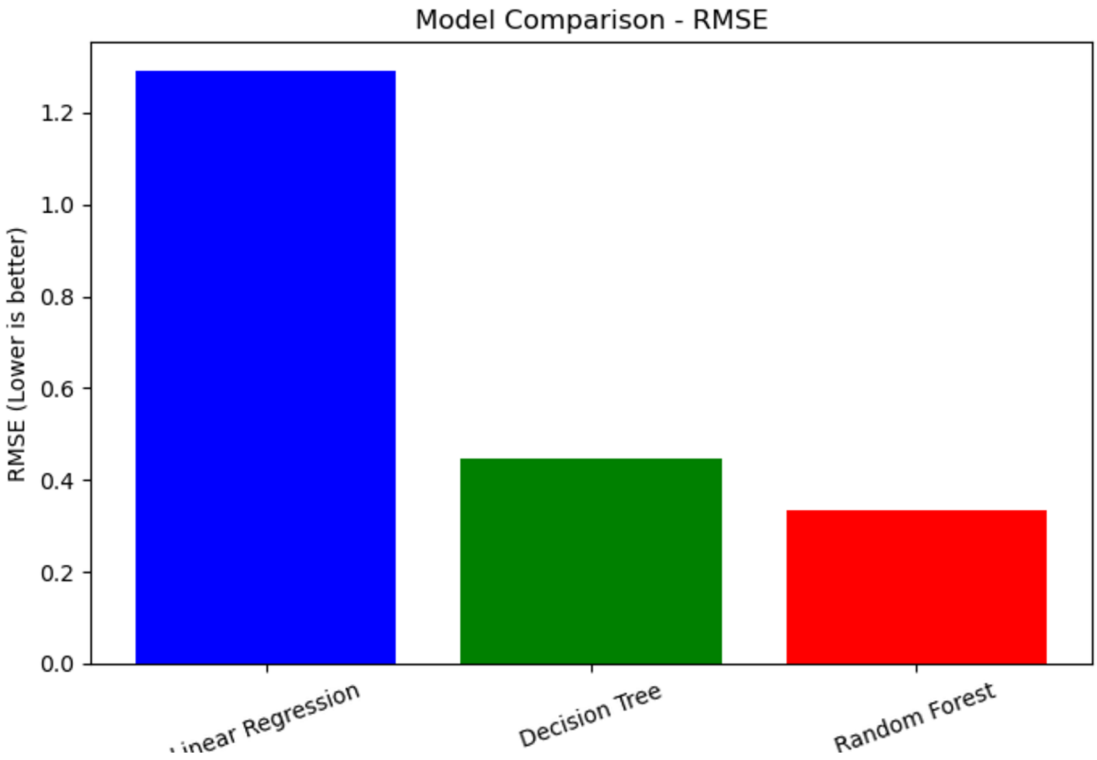
RESULTS

Task 1: Predictive Modelling

Four different regression models were built namely Linear Regression, Decision Tree, Random Forest and Support Vector Machine to predict the aggregate rating of a restaurant based on the available features. The features are: Average Cost for two, Votes, Price range, Has Table Booking and Has Online Delivery. The performance of the models was evaluated and their rmse and r-squared are recorded in the table below for comparison.

Linear Regression	1.29479	0.263445
Decision Tree	0.451976	0.910249
Random Forest	0.366079	0.941122





Task 2: Customer Preference Analysis

Based on the number of votes, North Indian, Mughlai and Chinese cuisines are the most popular cuisines. Also, most cuisines such as American, BBQ, Sandwich, Burger, Grill Caribbean, Seafood, Coffee and Tea among others have the same average rating of 4.9.

Comparison of Linear Regression, Decision Tree, and Random Forest models completed.
Results saved to model_results.txt

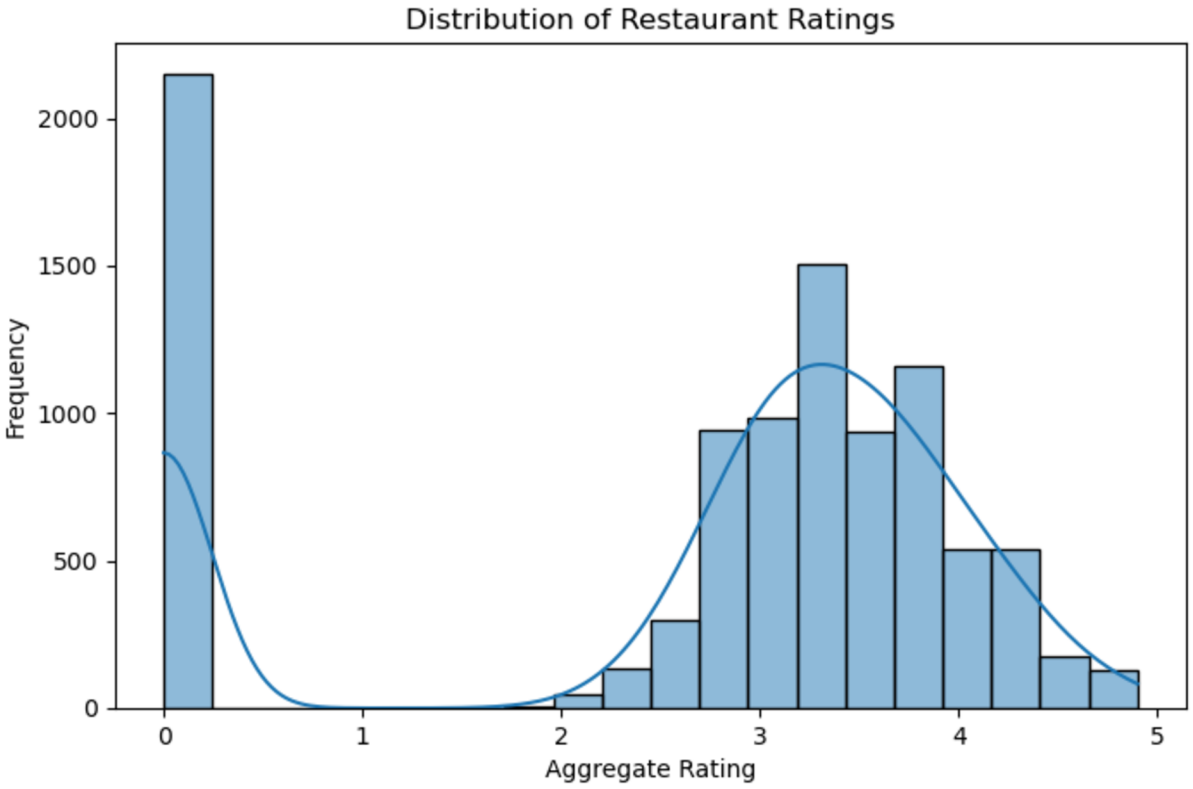
Customer Preference Analysis
Most Popular Cuisines based on Votes:
Cuisines
North Indian 595981
Chinese 364351
Italian 329265
Continental 288255
Fast Food 184058
American 183117
Cafe 177568
Mughlai 151946
Desserts 105889
Asian 104303
Name: Votes, dtype: int64

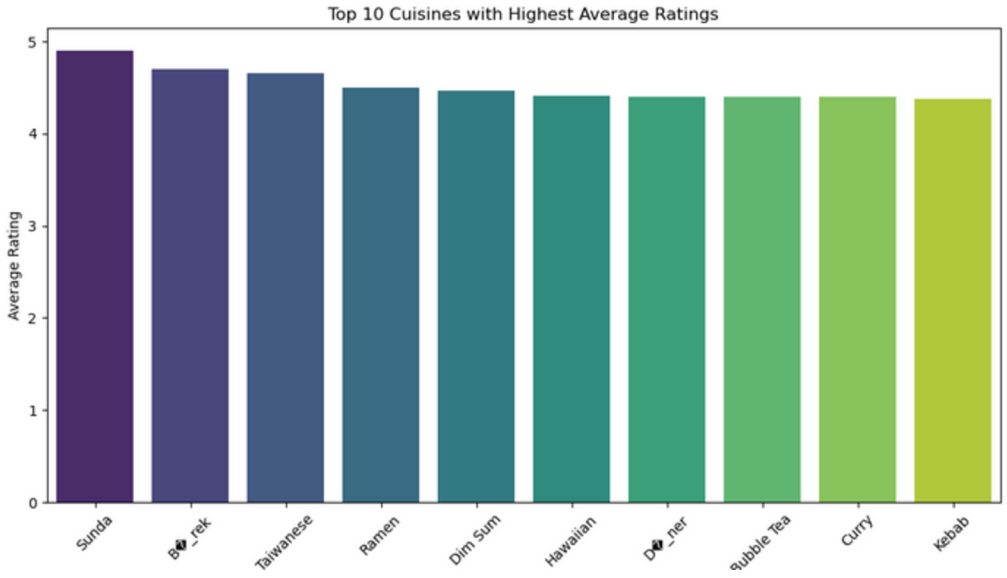
Cuisines with Highest Average Ratings:
Cuisines
Sunda 4.900000
Burger 4.700000
Taiwanese 4.650000
Ramen 4.500000
Dim Sum 4.466667

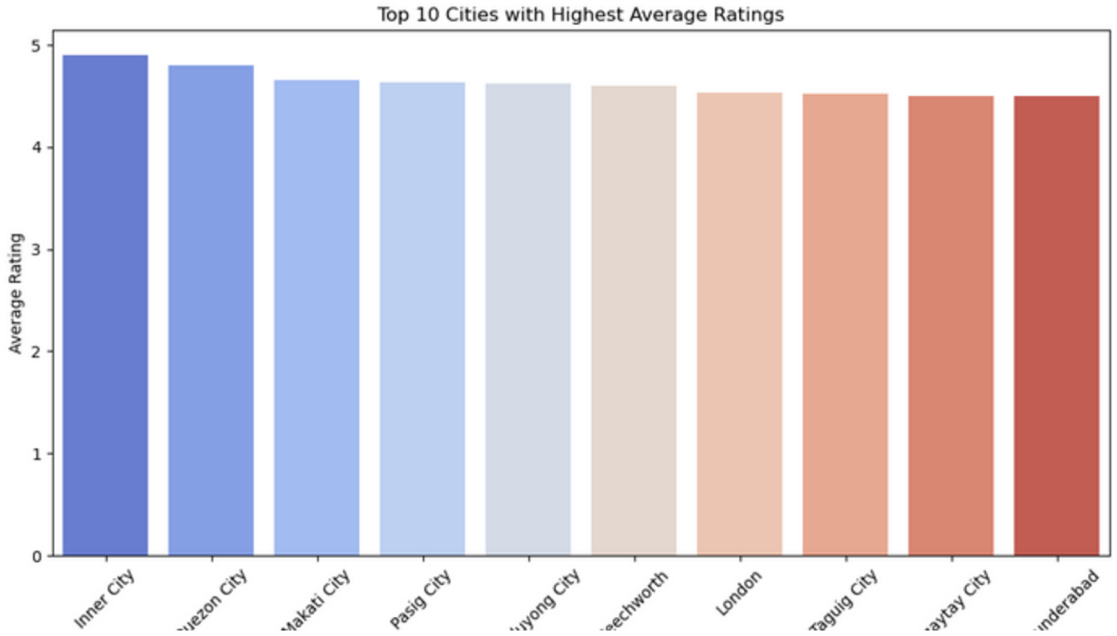
Hawaiian 4.412500
Dinner 4.400000
Bubble Tea 4.400000
Curry 4.400000
Kebab 4.380000
Name: Aggregate rating, dtype: float64

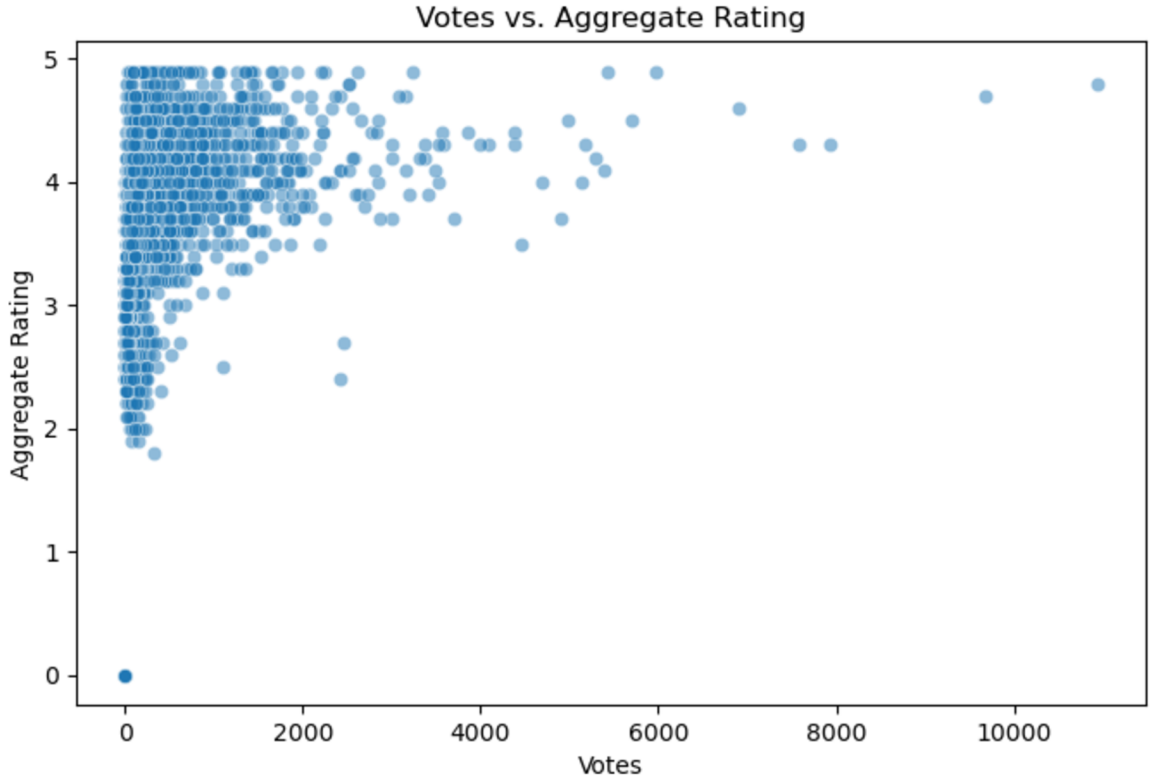
Task 3: Data Visualization

The majority of the rating given is between 3-4.









Conclusion

- This segment of the project has highlighted the critical role of predictive modelling, customer preference analysis, and data visualization in uncovering actionable insights and driving strategic decision-making.
- The customer preference analysis has served as a cornerstone in understanding the needs and preferences of the target audience.
- Also, the utilization of data visualization techniques has helped to communicate complex insights in a clear, concise, and impactful manner.