

**INT234 -PREDICITIVE ANALYTICS**

Insights from Using a Car Dataset

Submitted by

Bhumireddy Sudarshan Reddy

Registration No: 12210632

Programme and Section: Computer Science and Engineering-K22DP44

Course Code: INT234

**Under the Guidance of**

**Aashima:28968**

**Discipline of CSE**

**Lovely Professional University, Phagwara**

**School of Computer Science and Engineering**

# DECLARATION

I hereby declare that I have successfully completed my project titled "Insights from

Using a Car Dataset Report Using R: Insights from the World Happiness Report 2024." This project was conducted as part of my academic requirements for the degree of B.Tech. in Computer Science & Engineering at Lovely Professional University, Phagwara.

The project commenced on “09-10-2024” and concluded on “09-11-2024”.

Throughout this period, I dedicated myself fully to the training and implementation of various machine learning algorithms using R, including Linear Regression, Logistic Regression`, Visualization: Distribution of Car Prices Plotting histogram of car prices, Random Forest and K-Means K-Nearest. My work involved extensive data analysis, model training, and performance evaluation to derive insights from the dataset.

I affirm that the learning outcomes achieved during this project meet the necessary requirements for my degree program. I appreciate the support and guidance provided by my faculty and peers throughout this endeavour.

B. Sudarshan Reddy 12210632

## Acknowledgement

I would like to express my sincere gratitude to my faculty and peers for their unwavering support and guidance throughout the duration of this project. Their insights and encouragement were invaluable in helping me navigate the complexities of predictive analysis. I am especially thankful to my mentor, Aashima, whose expertise and constructive feedback significantly enhanced the quality of my work.

I would also like to extend my appreciation to the developers of the R programming language, whose powerful tools and libraries facilitated my data analysis and modelling efforts. The resources available through the R community were instrumental in overcoming challenges encountered during the project. Additionally, I am grateful to my fellow students for their collaborative spirit and for sharing knowledge that enriched my learning experience.

With Regards B. Sudarshan Reddy

**Introduction:**

The dataset **Car\_Data.csv** contains comprehensive information about a diverse range of vehicles, including their specifications and pricing. It features 293 entries, each characterized by several attributes. ID: A unique identifier for each vehicle. Brand: The manufacturer of the vehicle (e.g., Mazda, Ford, BMW). Model: The specific model of the vehicle (e.g., Generic Model 3, Focus). Year: The year of manufacture, indicating the vehicle's age. Colour: The colour of the vehicle. Mileage: The total distance the vehicle has traveled, typically measured in kilometers or miles. Price: The selling price of the vehicle. Condition: The state of the vehicle, categorized as either "New" or "Used".

This dataset serves as a valuable resource for analyzing trends in the automotive market, enabling various applications such as price prediction modeling and market research. By examining factors like mileage, year, and condition, one can derive insights into how these variables influence vehicle pricing. Additionally, the diversity in brands and models allows for comparative analysis across different manufacturers and types of vehicles. Overall, this dataset is instrumental for researchers, marketers, and automotive enthusiasts interested in understanding car sales dynamics

**Objectives / Scope of the Analysis :**

The primary objectives of this analysis are:

1. Data Preprocessing: Clean and normalize the dataset for effective modeling.
2. Model Implementation: Apply multiple machine learning algorithms to classify happiness levels.
3. Performance Evaluation: Compare the accuracy of different models to identify the most effective approach for predicting happiness levels.
4. Visualization: Present findings through visual representations for better understanding.

**Sources of Data :**

The data used for this analysis comes from the Insights from Using a Car

Dataset. The dataset includes attributes such as ID, Brand, Model, Year, Colour, Mileage, Price, and Condition. This structured data can be used for analysis and modeling in various domains related to automotive sales and market trends.

**Analysis on Dataset :**

Data Preparation

1. Loading Libraries: The following libraries were utilized:

* + Class for K
  + e1071 for Naive Bayes
  + gmodels for cross-tabulation
  + Ggplot2 for data visualization
  + randomForest for Random Forest modeling

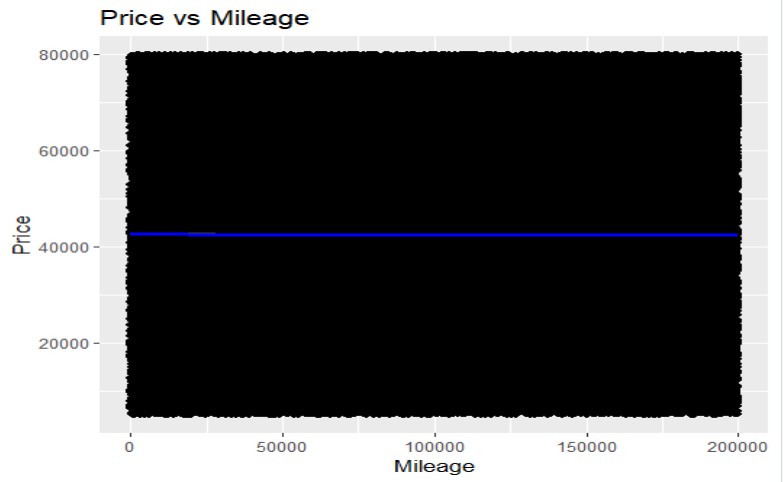
**Model Implementation :**

1. Linear Regression Load necessary libraries.
2. Logistic Regression Convert Condition to a factor.
3. Visualization: Distribution of Car Prices Plotting histogram of car prices.
4. Random Forest Load necessary library for random forest. 5. K-Means Clustering Scale the data for clustering.

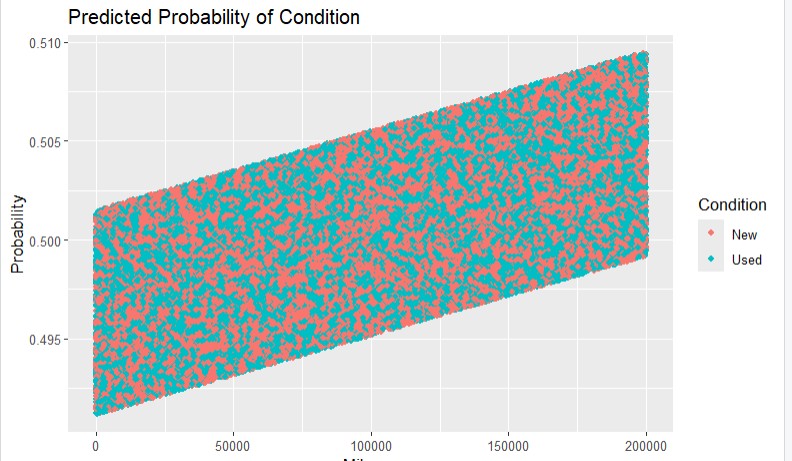
**Performance Comparison :**

The accuracies of all models were compared using a bar plot and line plot, illustrating their performance visually:

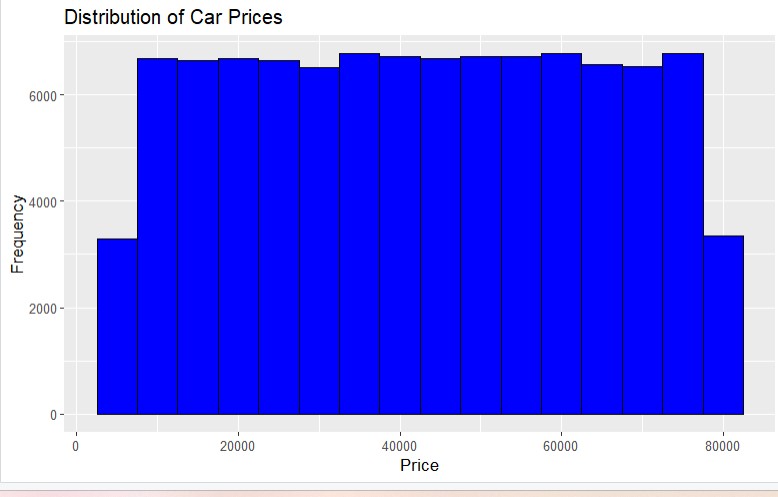
1)Linear Regression



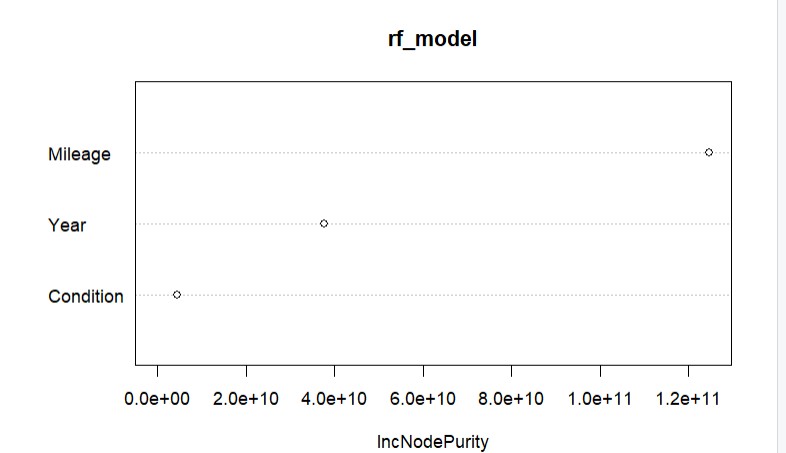
2)Logistic Regression



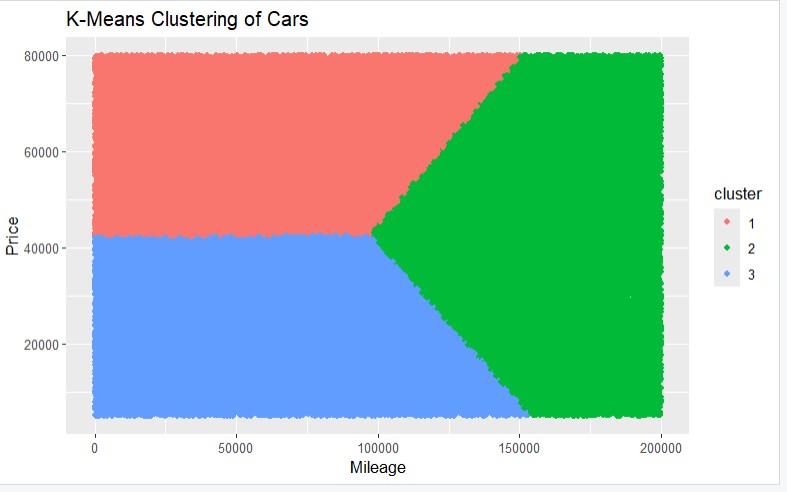
1. Visualization: Distribution of Car Prices Plotting histogram of car prices



1. Random Forest

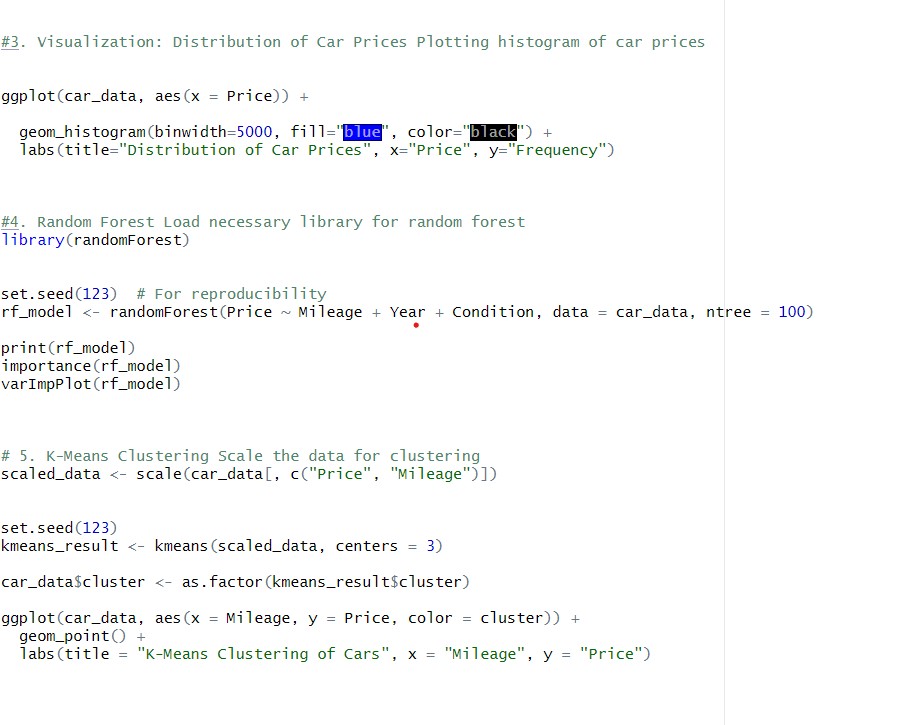


1. K-Means

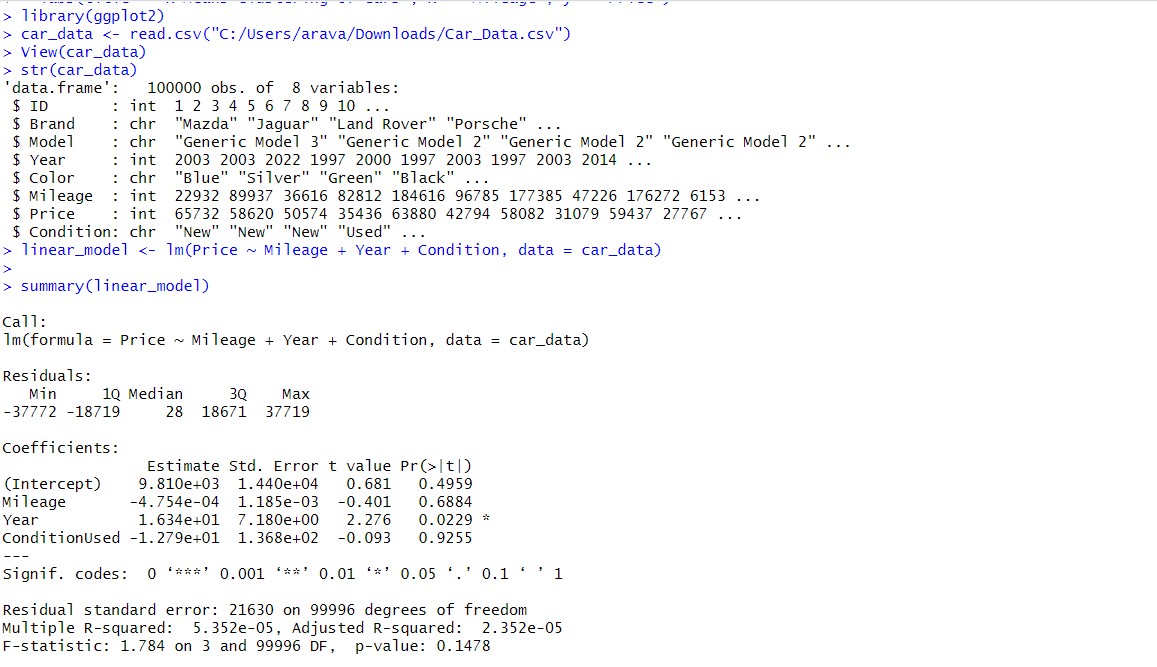


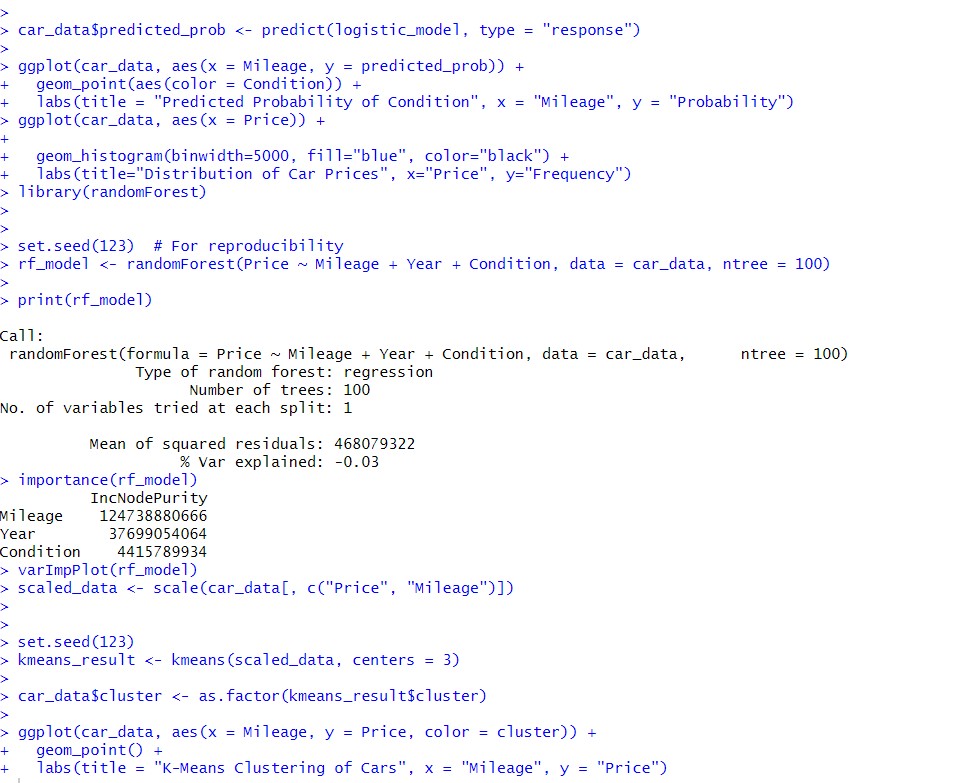
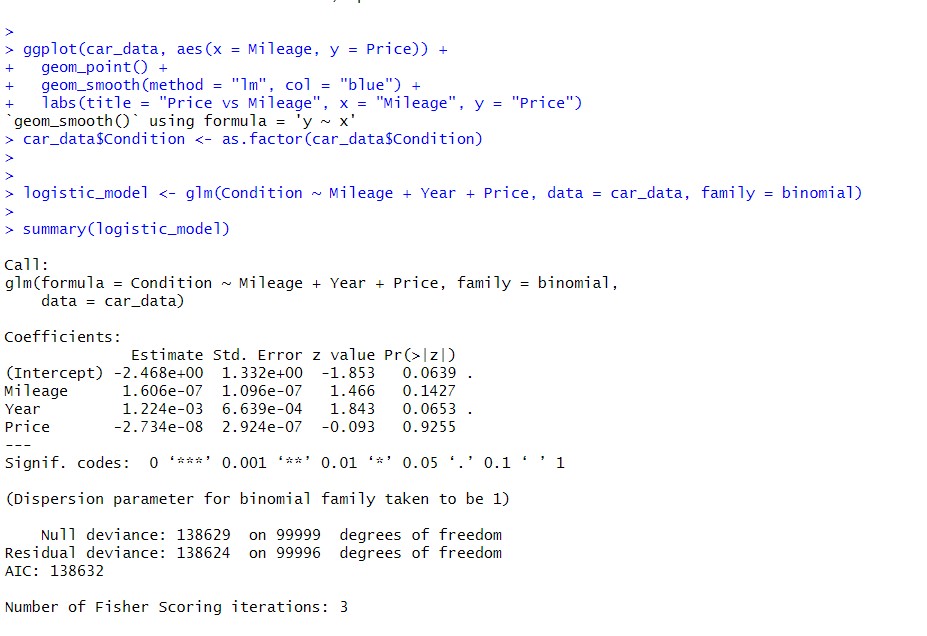
**CODE :**





CODE OUTPUT SNANPSHOTS :





**LINKEDIN:** [https://www.linkedin.com/in/bhumireddy-sudharshan-reddy/recentactivity/all/](https://www.linkedin.com/in/bhumireddy-sudharshan-reddy/recent-activity/all/)

**Conclusion :**

The datasetis a comprehensive collection of vehicle information, including various attributes such as brand, model, year, colour, mileage, price, and condition. It consists of 293 entries and provides insights into the automotive market by detailing both new and used cars across multiple manufacturers. This dataset can be utilized for various analyses, including price prediction models and market trend evaluations. The structured format allows for easy manipulation and exploration of the data, making it suitable for machine learning applications and statistical analysis.

**References :**

**1.**Car Data Report

2.R Documentation for libraries used in this project

3.Online resources on machine learning algorithms

