 **Project: Predictive & Prescriptive Analysis of Vehicle’s Gas Consumption**

**Course Code: DSCI 726**

**Course Title: Operational Analytics**

**Submitted to: Professor Yi Liu**

**Submitted by   
Yogesh Siyyadri  
Mansu Thapa  
Srinivasulu Punugoti**

**Date of Submission: 11/17/2024**

Contents

[Data Acquisition 3](#_Toc182775937)

[Data Exploration 3](#_Toc182775938)

[Data Transformation 5](#_Toc182775939)

[Data Partitioning 6](#_Toc182775940)

[Model Building & Evaluation 6](#_Toc182775941)

[Method 1: Linear Regression 6](#_Toc182775942)

[Method 2: Regression Tree 7](#_Toc182775943)

[Predictive Model 7](#_Toc182775944)

[Prescriptive Analysis 8](#_Toc182775945)

[Decision Variables 8](#_Toc182775946)

[Constraints 8](#_Toc182775947)

[Objective 8](#_Toc182775948)

[Linear Programming 8](#_Toc182775949)

[Conclusion 11](#_Toc182775950)

[Appendix 12](#_Toc182775951)

[Work Plan 12](#_Toc182775952)

**Introduction**

This report delves into both predictive and prescriptive analyses concerning vehicle gas consumption. The predictive aspect forecasts gas consumption based on specific vehicle parameters, while the prescriptive facet focuses on optimizing consumption given certain constraints. Key vehicle specifications analyzed include the number of cylinders, displacement, acceleration, horsepower, and weight.

**Background**

Vehicle gas consumption significantly influences consumer purchasing decisions. Various specifications impact consumption rates and are often inversely related to the vehicle's performance features. This study aims to find an optimal balance that minimizes gas consumption without compromising key specifications.

**Problem Statement**

Following directives from a memorandum to General Motors dated September 04, 2023, this project supports the company’s initiative to develop vehicles that are more fuel-efficient, meeting both market demands and consumer expectations for cost savings.

**Scope and Methodology**

The analysis encapsulates 398 data records, providing a robust basis for both predictive and prescriptive evaluations. The project will:

1. Develop a predictive model to estimate miles per gallon (mpg) based on attributes such as cylinder count, displacement, acceleration, and weight.
2. Determine the optimal mpg achievable under specific vehicle specifications.

In our approach, we will construct and assess various models, selecting the most effective for mpg forecasting. Furthermore, we will use these findings to define specification constraints for optimizing the predictive model.

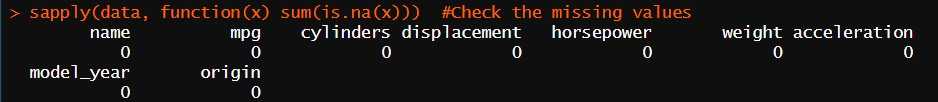
# Data Acquisition

The source of the data is <https://www.kaggle.com/datasets/whenamancodes/automobiles-project-dataset>.

# Data Exploration

1. Missing data

The dataset has 0 missing values.



1. Measure of central tendency and dispersion

Below are the measures of central tendency for all the variables.

Summary statistics

A screen shot of a computer

Description automatically generated

Box plots

A group of black and white lines with text

Description automatically generated

1. Data distributionA group of colored bars

   Description automatically generated with medium confidenceion

Below is the variable distribution, visualized on histograms.

Histograms

1. Correlation

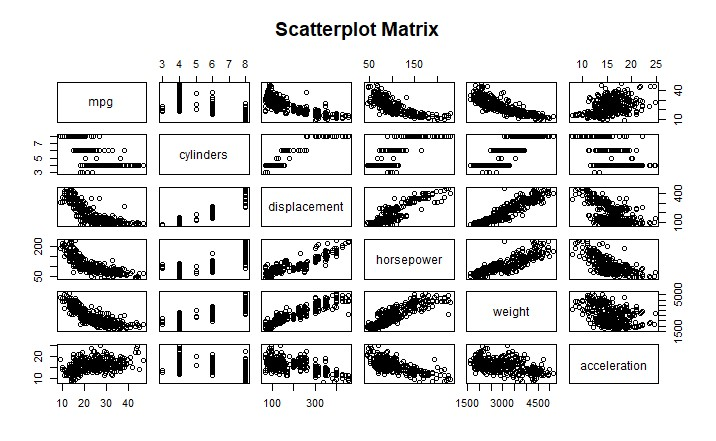
The mpg is positively correlated to and negatively correlated to

Correlation figures

A screen shot of a computer

Description automatically generated

Scatter plot



# Data Transformation

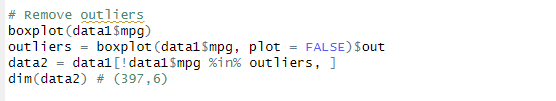
1. Dropping Variables:

To streamline our gas consumption model, we removed several non-impactful variables such as 'name,' 'model year,' and 'origin' from the dataset. This refined the data for more focused analysis.



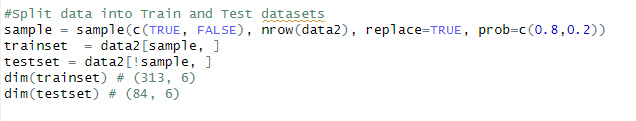
1. Removing outliers:

The dataset contained outliers in the response variable (mpg). To ensure these did not skew our analysis, we removed the records containing these outliers.



# Data Partitioning

1. Training and Testing: To ensure that models built on the dataset are going to perform well on unseen data, the dataset was partitioned into training and test datasets. The different predictive models were trained on one subset of the data (Training set) and its performance tested on the other subset (testing set).
2. Overfitting: To avoid the model from learning the training data too well and performing poorly on new unseen data, the dataset was portioned into 80% training set and 20% testing set.



# Model Building & Evaluation

Two models were developed using the training dataset and subsequently evaluated on the testing set. We assessed their performance using two key metrics: Mean Squared Error (MSE) and R-Squared. MSE quantifies the average squared discrepancies between predicted and actual values, indicating model accuracy. R-Squared measures the proportion of variance in the response variable that can be explained by the predictor variables, reflecting the model's explanatory power.

## Method 1: Linear Regression

|  |  |  |
| --- | --- | --- |
|  | Training set | Testing set |
| MSE | 17.62 | 20.21 |
| R-Squared (%) | 71.7 | 63.40 |

## Method 2: Regression Tree

A graph of a graph with blue rectangular bars

Description automatically generated with medium confidence

|  |  |  |
| --- | --- | --- |
|  | Training set | Testing set |
| MSE | 14.42 | 18.93 |
| R-Squared (%) | 76.8 | 65.7 |

Comparing the two methods, Linear Regression performed better on unseen data (testing set) compared to Regression Tree. Linear Regression is the model proposed for objective 1 (To predict the gas consumption of a vehicle based on its specifications.)

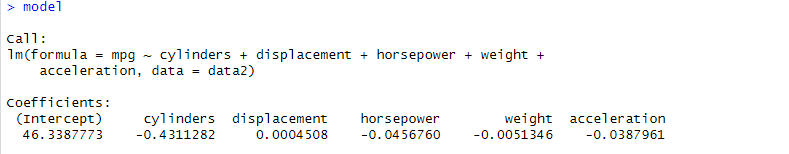
# Predictive Model

The following is the model to predict the gas consumption (mpg) of a vehicle based on its specification (number of cylinders, displacement, horsepower, weight, and acceleration.

A graph of residual plot for regression tree

Description automatically generated

**Mpg = 46.00 - 0.392\*cylinders - 0.0007\*displacement - 0.040\*horsepower - 0.005\*weight - 0.02\*acceleration**



VIF for this model = 3.4. Typically, a VIF < 5 is considered acceptable, since it suggests that the variance of the coefficient estimates is not significantly inflated by the correlation among predictors.

# Prescriptive Analysis

## Decision Variables

From the previous predictive model, the predicting variables for predicting the mpg will be the prescriptive decision variables. i.e., Number of cylinders, displacement, horsepower, weight, and acceleration. All these variables are integers.

## Constraints

To perform the initial mpg maximization, the constraints for the decision variables were set as the respective median (Q2) and maximum (Q4) for the lower and upper bounds, respectively.

Number of cylinders >= 4 but <= 8

Displacement >= 148.5 but <= 455

Horsepower >= 93.5 but <= 230

Weight >= 2804 but <= 5140

Acceleration >= 15.5 but 24.8

## Objective

The objective for this prescriptive analysis is to maximize miles per gallon (mpg) subject to the above decision variables constraints.

Maximize: mpg = 46 – 0.392\*cylinders – 0.0007\*displacement – 0.040\*horsepower – 0.005\*weight – 0.02\*acceleration

## Linear Programming

To optimize (maximize) the mpg linear programming was used as all the constraints and objective functions were linear.

Below is the linear programming model used.

Objective function:

Maximize: 46 – 0.392\*cylinders – 0.0007\*displacement – 0.040\*horsepower – 0.005\*weight – 0.02\*acceleration

Subject to:

Number of cylinders >= 4 but <= 8

Displacement >= 148.5 but <= 455

Horsepower >= 93.5 but <= 230

Weight >= 2804 but <= 5140

Acceleration >= 15.5 but 24.8

**Optimization of Fuel Efficiency for Vehicles Using Premium Gas**

A linear regression model was developed to optimize fuel efficiency, considering variables such as cylinders, displacement, horsepower, weight, and acceleration. The coefficients indicate how each factor influences fuel efficiency. Optimization was performed under bounded constraints for each variable:

* **Optimal Values**:  
  Cylinders: -0.5, Displacement: 0.8001, Horsepower: -0.9947, Weight: 0.9943, Acceleration: 0.8767
* **Optimal Solution**: -23.53 (objective function value)

This result demonstrates the combination of variable values that minimize fuel consumption within realistic constraints.

A screenshot of a graph

Description automatically generated

**Sensitivity Analysis Report**

A screenshot of a computer

Description automatically generated

This section of the report details the findings from the sensitivity analysis of our optimization model, which aimed to refine the fuel efficiency of vehicles using premium gas. The analysis reveals how flexible each variable can be without altering the solution's feasibility, providing insights into the robustness of our optimization results.

**Key Findings**

* **Intercept**: The base fuel efficiency value stands at 48.93. It shows moderate flexibility with an allowable increase of 5.30 and a decrease of 0.48, indicating stable baseline efficiency under minor adjustments.
* **Cylinders**: Shows significant sensitivity with an allowable increase of 5.87 and a substantial allowable decrease of 9.29. This highlights the cylinders' critical role in fuel efficiency, suggesting greater variability and impact on the model's outcomes.
* **Displacement**: Exhibits high sensitivity and flexibility with allowable changes of 6.65 for increases and 7.69 for decreases, implying that adjustments in engine displacement can vary widely without affecting optimal performance.
* **Horsepower**: This parameter can increase by up to 5.30 and decrease by 2.01, indicating its sensitive role in affecting fuel efficiency. The model is somewhat flexible regarding horsepower adjustments.
* **Weight**: Demonstrates considerable elasticity with an allowable increase of 5.09 and decrease of 6.50, affirming weight as a significant but adjustable factor in achieving desired fuel efficiency.
* **Acceleration**: The least flexible, with an allowable increase of only 0.16 and a decrease of 6.54. This suggests that while acceleration has a tight upper limit for changes, there is more leeway to reduce it to enhance efficiency.

The sensitivity analysis provides essential insights into how each vehicle attribute affects the overall optimization of fuel efficiency. The variables show different levels of flexibility, indicating where design modifications can be more aggressive and where they should be more conservative. These findings will guide future adjustments in vehicle specifications to optimize fuel efficiency without compromising the model's integrity and effectiveness.

# Conclusion

Our analysis using Linear Regression and Regression Tree models to predict vehicle gas consumption demonstrated that Linear Regression is the superior method for generalizing on unseen data. It achieved a Mean Squared Error (MSE) of 20.21 and an R-Squared of 63.40% on the testing set, outperforming the Regression Tree model. This method effectively relates vehicle attributes like cylinders and horsepower to fuel efficiency, providing valuable insights for vehicle manufacturers. Linear Regression not only offers a robust framework for predictions but also supports optimal design decisions that align with environmental standards and consumer preferences. This approach equips manufacturers to enhance vehicle efficiency strategically in a competitive market.

# Appendix

## Work Plan

|  |  |  |  |
| --- | --- | --- | --- |
|  | WORK DESCRIPTION (tasks) | PREDICTED TIME (hrs.) | ACTUAL TIME (hrs.) |
| WEEK 3 | Project proposal   1. Collect data for the analysis. 2. Write the Memo to the General Motors | 2 | 3 |
| WEEK 4 | Work plan   1. Develop a timeline for each stage of the project. | 2 | 3 |
| WEEK 5 | Descriptive analysis   1. Perform data cleaning and transformation. 2. Perform exploratory analysis of the dataset using R. | 3 | 5 |
| WEEK 6 | Predictive analysis + R code   1. Build different models to predict the MPG. 2. Evaluate the best-performing model and use it for forecasting the MPG | 9 | 9 |
| WEEK 7 | Prescriptive analysis – Using Excel   1. Evaluate the different constraints. 2. Optimize the MPG based on the constraints. | 6 | 6 |
| WEEK 8 | Executive summary | 3 | 3 |
| WEEK 9 | Final analysis   1. Prepare a detailed technical report on the above project. | 6 | 5 |
| WEEK 10 | Presentation and PowerPoint | 7 | 9 |
| WEEK 11 | Peer review | 5 | 9 |