UNIVERSITY OF
SOUTH DAKOTA

Project Title: Predictive & Prescriptive Analysis of Vehicle's Gas Consumption

Course Code: DSCI 726

Course Title: Operational Analytics

Submitted to: Professor Dr. Yi Liu

Project Team: Yogesh Siyyadri, Mansu Thapa, Srinivasulu Punugoti

Date of Submission: 09/29/2024

Introduction

This document delves into analyzing a vehicle's gas consumption through predictive and prescriptive approaches. Predictive analysis involves estimating fuel consumption based on the vehicle’s specifications, whereas prescriptive analysis focuses on enhancing fuel efficiency while adhering to specification constraints. The study examines factors such as the number of cylinders, engine displacement, acceleration, horsepower, and weight.

Background Research

Fuel consumption is a key factor that customers take into account when choosing a vehicle. Various specifications, including the number of cylinders, engine displacement, acceleration, horsepower, and weight, each have a unique impact on fuel usage. Typically, the higher the quality of these specifications, the more fuel-efficient the vehicle becomes. The prescriptive analysis in this project aims to find an optimal balance that minimizes gas consumption while considering these constraints.

Problem Presentation

As detailed in the memo dispatched to General Motors on September 15, 2024, this project focuses on assisting the company in manufacturing vehicles with improved fuel efficiency to align with market demands. This initiative will also benefit customers by providing fuel-efficient vehicles that offer considerable cost savings.

Specification and Design

This study will encompass both predictive and prescriptive analyses of a vehicle's gas consumption. With 398 data records available, the dataset is expected to be sufficient for conducting both analyses. Upon successful completion, this project will achieve the following:

a. Development of a predictive model to estimate miles per gallon (mpg) based on the vehicle's cylinders, displacement, acceleration, and weight.

b. Determination of the optimal gas consumption in mpg within the vehicle’s specifications.

For the predictive analysis, multiple models will be constructed and assessed, with the most accurate model selected for forecasting. The prescriptive analysis will involve examining specification constraints and using the mpg predictive model as the objective function for optimization.

Data Acquisition

The source of the data is Kaggle Dataset

Data Exploration

a) Missing data:

Our dataset does not have any missing values for any variables.

```
> sapply(data, function(x) sum(is.na(x)))  #Check the missing values
        name          mpg    cylinders displacement   horsepower       weight acceleration
           0            0            0            0            0            0            0
  model_year       origin
           0            0
```

b) Measure of central tendency and dispersion:

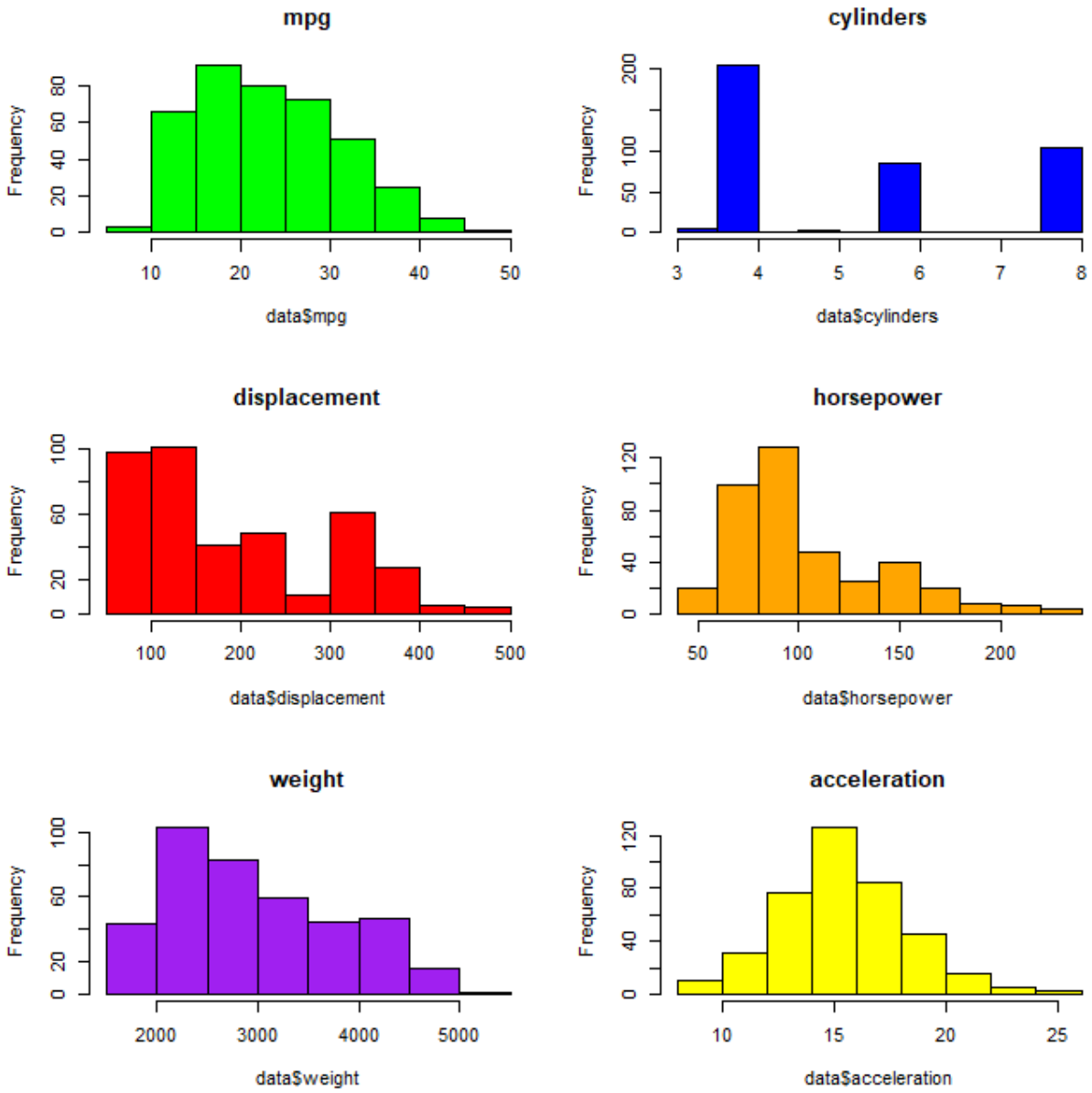
Our results below show the measures of central tendency for all the variables.

Summary statistics

```
          0            0
> summary(data)
     name                mpg          cylinders      displacement     horsepower
 Length:398         Min.   : 9.00   Min.   :3.000   Min.   : 68.0   Min.   : 46.0
 Class :character   1st Qu.:17.50   1st Qu.:4.000   1st Qu.:104.2   1st Qu.: 76.0
 Mode  :character   Median :23.00   Median :4.000   Median :148.5   Median : 95.0
                    Mean   :23.51   Mean   :5.455   Mean   :193.4   Mean   :104.4
                    3rd Qu.:29.00   3rd Qu.:8.000   3rd Qu.:262.0   3rd Qu.:125.0
                    Max.   :46.60   Max.   :8.000   Max.   :455.0   Max.   :230.0
     weight      acceleration     model_year       origin
 Min.   :1613   Min.   : 8.00   Min.   :70.00   Length:398
 1st Qu.:2224   1st Qu.:13.82   1st Qu.:73.00   Class :character
 Median :2804   Median :15.50   Median :76.00   Mode  :character
 Mean   :2970   Mean   :15.57   Mean   :76.01
 3rd Qu.:3608   3rd Qu.:17.18   3rd Qu.:79.00
 Max.   :5140   Max.   :24.80   Max.   :82.00
```

c) Data distribution:

Below is the variable distribution, visualized on histograms.
Histograms

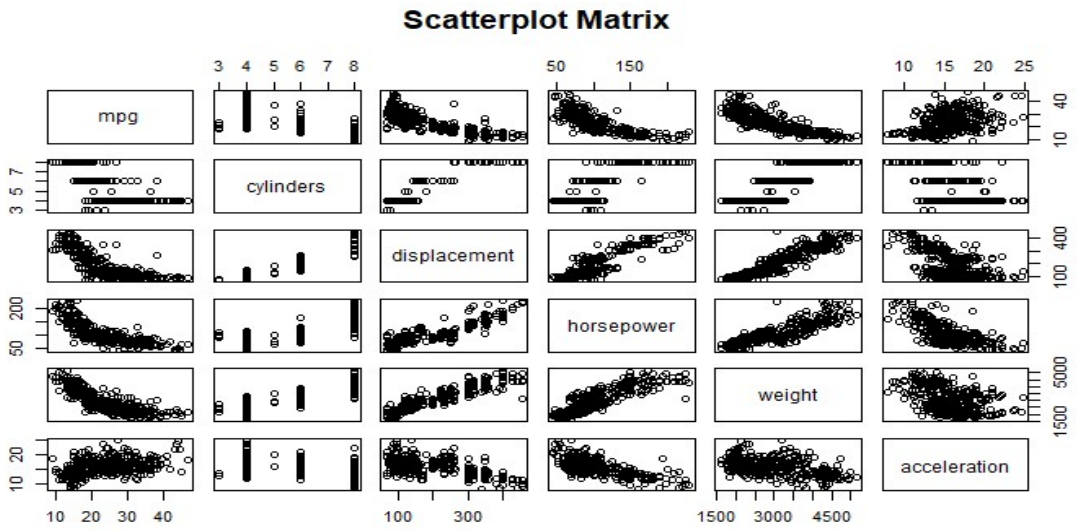


d) <u>Correlation:</u>

The mpg seems to be positively correlated to: …., and negatively correlated to: ….
Correlation figures

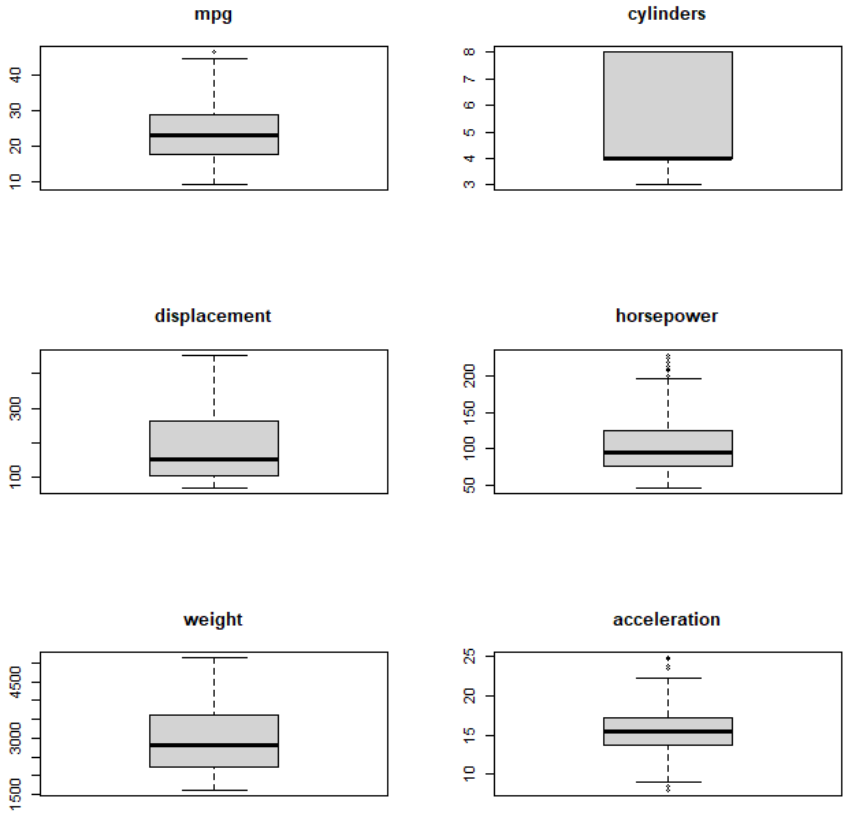
```
> cor(data[, c('mpg' , 'cylinders', 'displacement', 'horsepower', 'weight', 'acceleration')])
                     mpg  cylinders displacement horsepower     weight acceleration
mpg           1.0000000 -0.7753963   -0.8042028 -0.7723732 -0.8317409    0.4202889
cylinders    -0.7753963  1.0000000    0.9507214  0.8400196  0.8960168   -0.5054195
displacement -0.8042028  0.9507214    1.0000000  0.8946477  0.9328241   -0.5436841
horsepower   -0.7723732  0.8400196    0.8946477  1.0000000  0.8614633   -0.6853105
weight       -0.8317409  0.8960168    0.9328241  0.8614633  1.0000000   -0.4174573
acceleration  0.4202889 -0.5054195   -0.5436841 -0.6853105 -0.4174573    1.0000000
```

c. Scatter plot:



d} Outliers:

Mpg, horsepower, and acceleration have outliers.
Boxplots



Our analysis of the dataset has revealed the presence of outliers in three main variables: MPG, Horsepower, and Acceleration. These outliers present an opportunity to gain valuable insights into the dataset's characteristics and enhance our understanding of the vehicles it represents. By conducting a deeper examination of these specific data points and applying domain knowledge, we can identify the underlying factors that cause these deviations and assess their potential influence on the dataset's overall patterns and trends. Investigating outliers is a crucial aspect of data analysis, enabling us to achieve a thorough understanding of the data and make well-informed decisions based on the results.