CHANDIGARH UNIVERSITY
Discover. Learn. Empower.

UNIVERSITY INSTITUTE OF COMPUTING

PROJECT REPORT
ON
Explore the 1985 Cars Dataset

1. Title Page

- Program Name: BCA (Data Science)
- Subject Name/Code: R PROGRAMMING LAB (24CAP-161)

- Submitted by:
  - Name: Mohit Kumar
  - UID: 24BCD10014
  - Section: 24BCD-1A

- Submitted to:
  - Name: Mrs. Ambika
  - Project: 1

CHANDIGARH UNIVERSITY
Discover. Learn. Empower.

College Project Report: Analysis of the 1985 Cars Dataset

Table of Contents

1. Introduction

The purpose of this project is to analyze the 1985 Cars Dataset to help make an informed decision about which car to add to a personal collection. The dataset contains various attributes of cars from 1985, including make, model, engine size, and miles per gallon (mpg). By cleaning and analyzing this data, we aim to identify cars that meet specific criteria, such as fuel efficiency and engine size.

2. Methodology

2.1 Loading Libraries

To begin the analysis, we load the necessary libraries for data manipulation and reading CSV files.

```r
library(readr)
library(dplyr)
```

2.2 Loading the Dataset

We read the dataset from a CSV file into R.

```r
cars ← read_csv("cars85.csv")
```

2.3 Inspecting the Data

We inspect the dataset to understand its structure and contents.

```r
head(cars)
summary(cars)
```

2.4 Cleaning the Data

We remove unnecessary columns and rename columns for clarity.

```r
cars ← select(cars, -normalized_losses)
colnames(cars)[colnames(cars) == "symboling"] ← "risk_factor"
```

2.5 Adding New Columns

We create a new column to measure how each car's highway mpg compares to a defined threshold.

```r
mpg_threshold ← 30
cars ← cars %>%
  mutate(mpg_diff_from_threshold = highway_mpg - mpg_threshold)
```

2.6 Filtering and Arranging Data

We filter the dataset to include only cars that exceed the mpg threshold and arrange them by mpg.

```
mpg_exceeds_threshold <- filter(cars, mpg_diff_from_threshold > 0)
mpg_exceeds_threshold <- arrange(mpg_exceeds_threshold, desc
    (mpg_diff_from_threshold))
```

2.7 Specifying the Make of the Car

We specify a car make to analyze and filter the dataset accordingly.

```r
chosen_make ← "Toyota"
chosen_make_details ← filter(cars, make == chosen_make)
chosen_make_details ← arrange(chosen_make_details, desc(engine_size))
```

3. Code Implementation

Here is the complete code used for the analysis:

```r
library(readr)
library(dplyr)
cars <- read_csv("cars85.csv")
head(cars)
summary(cars)
cars <- select(cars, -normalized_losses)
print(colnames(cars))
colnames(cars)[colnames(cars) == "symboling"] <- "risk_factor"
print(colnames(cars))
mpg_threshold <- 30
cars <- cars %>%
  mutate(mpg_diff_from_threshold = highway_mpg - mpg_threshold)
mpg_exceeds_threshold <- filter(cars, mpg_diff_from_threshold > 0)
mpg_exceeds_threshold <- arrange(mpg_exceeds_threshold, desc
    (mpg_diff_from_threshold))
chosen_make <- "Toyota"
chosen_make_details <- filter(cars, make == chosen_make)
chosen_make_details <- arrange(chosen_make_details, desc(engine_size))
print("Cars exceeding the mpg threshold:")
print(mpg_exceeds_threshold)
print("Ordered cars by engine size:")
print(chosen_make_details)
```

4. Results

The analysis revealed several cars that exceed the 30 mpg threshold, with the highest mpg values identified. Additionally, the details of the chosen make, Toyota, were filtered and arranged by engine size, allowing for a clear comparison of the available options.

5. Conclusion

The analysis of the 1985 Cars Dataset provided valuable insights into fuel-efficient vehicles from that year. By focusing on cars that exceed a specified mpg threshold and filtering by make, we can make informed decisions about potential additions to a personal car collection. The methodology employed in this project can be adapted.