**MINOR PROJECT**

**Visualizing car prices in R**

**Class and Group:** 24BCD – 1 (A)

**Subject name: Subject code:**

R programming 24CAP -161

**Submitted to:**

Ms. Ambika Mam

**Submitted by:**

Ayush Negi – 24BCD10066

# Abstract

This project aims to analyze and visualize car prices across different vehicle categories using the R programming language. By employing statistical methods and graphical tools, the study provides insights into pricing patterns among Sedans, SUVs, and Trucks. The dataset includes multiple car models with their corresponding prices in each category. Core statistical measures such as mean, median, and standard deviation are computed to understand the central tendency and variability of prices. Visualizations including bar plots, histograms, and box plots are used to illustrate trends, distributions, and potential anomalies in the data. Additionally, the highest and lowest priced models in each category are identified. The project demonstrates how data analysis in R can be used effectively to extract meaningful insights from a relatively small dataset, providing a foundation for more advanced vehicle pricing studies in the future.

# Project Objective

The objective of this project is to analyze car prices across different vehicle categories using R. The project aims to apply basic statistical analysis and data visualization techniques to understand the price trends, identify the most and least expensive models in each category, and compare price distributions. This helps in deriving insights for pricing strategies, consumer trends, and model evaluation.

# Methodology

1. A sample dataset of car models and their prices across three categories—Sedan, SUV, and Truck—was created.  
2. Summary statistics (mean, median, standard deviation) were computed for each category.  
3. The highest and lowest priced models in each category were identified.  
4. Visualizations were created using bar plots, histograms, and box plots to explore the distribution of car prices.  
5. The data was interpreted to identify pricing patterns and potential outliers.

# Technologies Used

- R Programming Language  
- Libraries: ggplot2, dplyr, tidyr

# R Code

# Load libraries  
library(ggplot2)  
library(dplyr)  
library(tidyr)  
  
# Sample data: Car models and prices across three categories  
cars <- data.frame(  
 Model = c("Model A", "Model B", "Model C", "Model D", "Model E"),  
 Sedan = c(25000, 27000, 26000, 24000, 25500),  
 SUV = c(32000, 34000, 31000, 33000, 33500),  
 Truck = c(40000, 41000, 39500, 42000, 40500)  
)  
  
# Summary statistics  
summary\_stats <- function(category\_prices) {  
 list(  
 Mean = mean(category\_prices),  
 Median = median(category\_prices),  
 Std\_Dev = sd(category\_prices)  
 )  
}  
  
# Category list  
categories <- c("Sedan", "SUV", "Truck")  
for (category in categories) {  
 cat("\nSummary for", category, ":\n")  
 print(summary\_stats(cars[[category]]))  
}  
  
# Highest & Lowest Prices  
for (category in categories) {  
 max\_price <- max(cars[[category]])  
 min\_price <- min(cars[[category]])  
 highest <- cars$Model[which.max(cars[[category]])]  
 lowest <- cars$Model[which.min(cars[[category]])]  
   
 cat("\nIn", category, ":\n")  
 cat("Highest priced model:", highest, "with", max\_price, "\n")  
 cat("Lowest priced model:", lowest, "with", min\_price, "\n")  
}  
  
# Average price bar plot  
avg\_prices <- colMeans(cars[2:4])  
avg\_df <- data.frame(Category = names(avg\_prices), Average = avg\_prices)  
bar\_plot <- ggplot(avg\_df, aes(x = Category, y = Average, fill = Category)) +  
 geom\_bar(stat = "identity") + theme\_minimal() +  
 ggtitle("Average Car Prices by Category")  
print(bar\_plot)  
  
# Histograms for each category  
for (category in categories) {  
 hist\_plot <- ggplot(cars, aes(x = .data[[category]])) +  
 geom\_histogram(binwidth = 1000, fill = "skyblue", color = "black") +  
 theme\_minimal() + ggtitle(paste("Price Distribution in", category)) +  
 xlab("Price") + ylab("Frequency")  
 print(hist\_plot)  
}  
  
# Box plot  
cars\_long <- pivot\_longer(cars, cols = Sedan:Truck,  
 names\_to = "Category", values\_to = "Price")  
box\_plot <- ggplot(cars\_long, aes(x = Category, y = Price, fill = Category)) +  
 geom\_boxplot() + theme\_minimal() +  
 ggtitle("Boxplot of Car Prices by Category")  
print(box\_plot)

# Project Output

The analysis shows how prices vary across different car categories. The summary statistics provide a central tendency and spread of values. The bar plot reveals the average price of each category, the histograms show how prices are distributed, and the box plot helps identify variability and outliers. For instance, Trucks tend to be the most expensive, while Sedans are more budget-friendly.

# Summary Statistics:

Sedan Prices:

Mean: 25,900

Median: 25,500

Standard Deviation: ~1,140

SUV Prices:

Mean: 32,800

Median: 33,000

Standard Deviation: ~1,140

Truck Prices:

Mean: 40,400

Median: 40,500

Standard Deviation: ~884

Highest and Lowest Priced Models:

# 

Sedan:

Highest: Model B (27,000)

Lowest: Model D (24,000)

SUV:

Highest: Model B (34,000)

Lowest: Model C (31,000)

Truck:

Highest: Model D (42,000)

Lowest: Model C (39,500)

Visual Insights:

Bar Plot: Clearly shows that Trucks have the highest average price, followed by SUVs and Sedans.

Histograms: Display the spread of prices within each category. SUV prices are more tightly packed, while Sedans and Trucks have a bit more variation.

Box Plot: Indicates Trucks have a slightly higher interquartile range. No extreme outliers were detected in the sample

# References and Libraries

- R Documentation: https://www.r-project.org/  
- ggplot2: https://ggplot2.tidyverse.org/  
- dplyr & tidyr: tidyverse packages