Cogs 209 Mini-Project Proposal

Lead Author: Amanda Chau

Co-Authors: Kathryn Oleksak, Stephen Kim, Yvonne Chen

Group Name: INMITD

Title: **Contribution Weighting of Car Features to Predict Retail Price of Cars**

Research Question:

One important factor in considering car purchasing and retail is how a car's features affect its price. Understanding the pricing weighting of said features would help buyers identify important purchasing factors in finding their ideal vehicle. Additionally, understanding pricing relationships for a car's features benefit sellers and manufacturers in maximizing a power vs. price relationship to balance affordability and performance. For example, by analyzing the relationship between the mileage of used cars and how much they are sold for, one could have a way to predict the expected sale price of another vehicle given its specific mileage.

We hypothesize that through regression techniques, we will be able to create a model that will accurately predict the price of a car based on many different descriptive features of the car.

Data/Materials:
We extracted and cleaned our data from Kaggle. The data is available in this Github link. The repository consists of the following files as listed/named below.

File Name	Format	Description
CarPrice_Assignment.csv	CSV file	Data table with 205 rows, 26 columns This data contains various features (columns) including selling price, car dimensions, engine type, etc for different sold cars (rows)
Data Dictionary - carprices.xlsx	XLSX file	This file contains the linking (index to column name) for the data file.
Data Descriptions.csv	CSV file	This file contains a description of the data (explaining the data in each column)

Course Impact/Relevance:
- Prediction (Regression)
  - Feature weighting and analysis using Regression (Ordinary Least Square)
    - Hypothesis Testing
  - Linear Regression, Logistic Regression
- Data Exploration/Processing
  - Categorical feature transformation (training)
  - Feature Importance
- Cross-Validation

This project connects with the topic of regression, where we aim to develop a model that can accurately predict outcomes based on various input features. To achieve this, we will employ techniques such as feature weighing and analysis using ordinary least squares. We will also perform data processing tasks such as applying categorical feature transformation to our data to better organize the data for improved model fitting (e.g., modeling transmission types as binary values).

Furthermore, we will conduct thorough data exploration to identify the most important features and use cross-validation methods to determine the best possible model for our dataset. By doing so, we hope to create a robust and accurate regression model that can provide valuable insights and predictions.

Outcome(s):
For this project, we will create a regression model to accurately predict the price of cars and conduct an analysis of the results. Our analysis will include comparisons of different regression techniques to observe which method performs better in the scenario of car price prediction. We will consider multiple features of the cars not limited to car size, engine size, or color, and how they affect the price. Through analyzing the weights for various features, we can conclude how much each factor contributes to the retail price of a car. In addition, we may analyze how different forms of preprocessing of our data might affect the regressions (e.g., normalization of the data).