Final Project Overview

Estimated time: 5 minutes

In this project, you will use a real-world data set and practice generative AI to create Python codes that can perform data preparation, analysis, and model development.

Project Scenario

You have been employed as a Data Scientist by a consultancy firm. The firm has a client who is a resale car dealer. They have a special feature on Ford cars and they want your firm to design a model that can predict the optimum quotation price for the cars in their lot. They provide you with sales data for the past few years. The dataset contains different features of the cars and the price they were sold at.

The tasks assigned to you are as follows:

1. There might be a few duplicate entries and a few missing values in the dataset. Data cleaning will be a part of the assignment.
2. You have to perform exploratory data analysis to draw keen insights on the data and determine the effect of different features on the price. Some specific requests by the client include:
   a. Identify number of sales for each fuel type
   b. Identify which transmission type has more price outliers
3. Compare the models with linear, polynomial and ridge regressions on single and multiple variables to find the best performing model
4. Perform a Grid Search on the Ridge regression model to identify the optimum hyperparameter of the model for best performance.

About the data set

This dataset contains resale car sale prices for Ford cars. This is a public dataset available on the Kaggle website as Ford Car Pricing Dataset under the CC0: Public Domain license. The dataset has been slightly modified for the purpose of this project.

Attributes of this dataset have been explained below.

Variable	Description
model	Car model name
year	Year of car make
transmission	Type of transmission (Automatic, Manual or Semi-Auto)
mileage	Number of miles traveled
fuelType	The type of fuel the car uses (Petrol, Diesel, Hybrid, Electric, Other)
tax	Annual Tax payable in USD
mpg	Miles per Gallon that the car runs at
engineSize	Engine Size of the car
price	Price of car in USD

Tasks in the project

The project tasks are data preparation, analysis, and model development. Based on the data set, you must write prompts to generate the Python codes for performing specific tasks. You can access a JupyterLite-based testing environment to test the generated codes using the Generative AI classroom prompts.

Author(s)

Abhishek Gagneja