

Hands-on Lab: Final Project: Generative AI for Data Science

Estimated Effort: 60 mins

Project Scenario

You have been employed as a Data Scientist by a consultancy firm. The firm has a client who is a used 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 for the model for best performance.

You decide to use Generative AI to create python codes that can help you analyse the data, determine the best features and create the prediction model as per requirement.

Disclaimer: This is a fictitious scenario created for the purpose of this project. The dataset being used is publicly available.

About the Dataset

This dataset contains used 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

Code execution environment

To test the prompt-generated code, keep the Jupyter Notebook (in the link below) open in a separate tab in your web browser. The notebook has some setup instructions that you should complete now.

[Jupyter-Lite Test Environment](#)

Please note the lab environment above will only work on Windows (Google Chrome or Firefox browser). If you don't have a Windows system with either of these browsers, use the lab environment provided in the next lesson of the module.

The data set for this lab is available in the following URL.

1	<code>URL = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMSkil</code>	
---	--	---

Complete the setup in the Jupyter Notebook and then proceed further.

Important Note: All prompts that are made available have been hidden and the users are encouraged to first try to write their own prompts to create the solutions. Also, the prompts given as solutions have also been maintained as ones which will create generic code structures which you can modify according to the question at hand.