



Laboratório Prático: Projeto Final: IA Generativa para Ciência de Dados

Esforço Estimado: 60 mins

Cenário do Projeto

Você foi contratado como Cientista de Dados por uma empresa de consultoria. A empresa tem um cliente que é um revendedor de carros usados. Eles têm uma característica especial em carros Ford e querem que sua empresa desenhe um modelo que possa prever o preço de cotação ideal para os carros em seu lote. Eles fornecem dados de vendas dos últimos anos. O conjunto de dados contém diferentes características dos carros e o preço pelo qual foram vendidos.

As tarefas atribuídas a você são as seguintes:

1. Pode haver algumas entradas duplicadas e alguns valores ausentes no conjunto de dados. A limpeza de dados será parte da tarefa.
2. Você deve realizar uma análise exploratória de dados para obter insights detalhados sobre os dados e determinar o efeito de diferentes características sobre o preço. Alguns pedidos específicos do cliente incluem:
 - a. Identificar o número de vendas para cada tipo de combustível
 - b. Identificar qual tipo de transmissão tem mais valores discrepantes de preço
3. Comparar os modelos com regressões linear, polinomial e ridge em variáveis únicas e múltiplas para encontrar o modelo com melhor desempenho
4. Realizar uma Busca em Grade no modelo de regressão Ridge para identificar o hiperparâmetro ideal para o modelo com melhor desempenho.

Você decide usar IA Generativa para criar códigos em python que podem ajudá-lo a analisar os dados, determinar as melhores características e criar o modelo de previsão conforme a necessidade.

Aviso: Este é um cenário fictício criado para o propósito deste projeto. O conjunto de dados utilizado está disponível publicamente.

Sobre o laboratório de IA generativa

► [Clique aqui](#)

Notas:

1. Os prompts usados neste laboratório são apenas para sua referência. Você pode criar seus próprios prompts e gerar respostas usando IA generativa.
2. Como as saídas geradas por IA são dinâmicas, você pode receber respostas diferentes, mesmo que tenha usado o mesmo prompt deste laboratório.

Sobre o Conjunto de Dados

Este conjunto de dados contém preços de venda de carros usados para carros Ford. Este é um conjunto de dados público disponível no site do [Kaggle](#) como [Conjunto de Dados de Preços de Carros Ford](#) sob a licença [CC0: Domínio Público](#). O conjunto de dados foi ligeiramente modificado para o propósito deste projeto.

Os atributos deste conjunto de dados foram explicados abaixo.

Variável	Descrição
model	Nome do modelo do carro
year	Ano de fabricação do carro
transmission	Tipo de transmissão (Automática, Manual ou Semi-Automática)
mileage	Número de milhas percorridas
fuelType	O tipo de combustível que o carro utiliza (Gasolina, Diesel, Híbrido, Elétrico, Outro)
tax	Imposto anual a ser pago em USD
mpg	Milhas por Galão que o carro percorre
engineSize	Tamanho do motor do carro
price	Preço do carro em USD

Ambiente de execução de código

Para testar o código gerado pelo prompt, mantenha o Jupyter Notebook (no link abaixo) aberto em uma aba separada no seu navegador. O notebook tem algumas instruções de configuração que você deve completar agora.

[Jupyter-Lite Test Environment](#)

Por favor, note que o ambiente do laboratório acima funcionará apenas no Windows (navegador Google Chrome ou Firefox). Se você não tiver um sistema Windows com um desses navegadores, use o ambiente de laboratório fornecido na próxima lição do módulo.

O conjunto de dados para este laboratório está disponível na seguinte URL.

URL = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMSkillsNetwork-AI0271EN-SkillsNetwork/labs/v1/m3/data/us

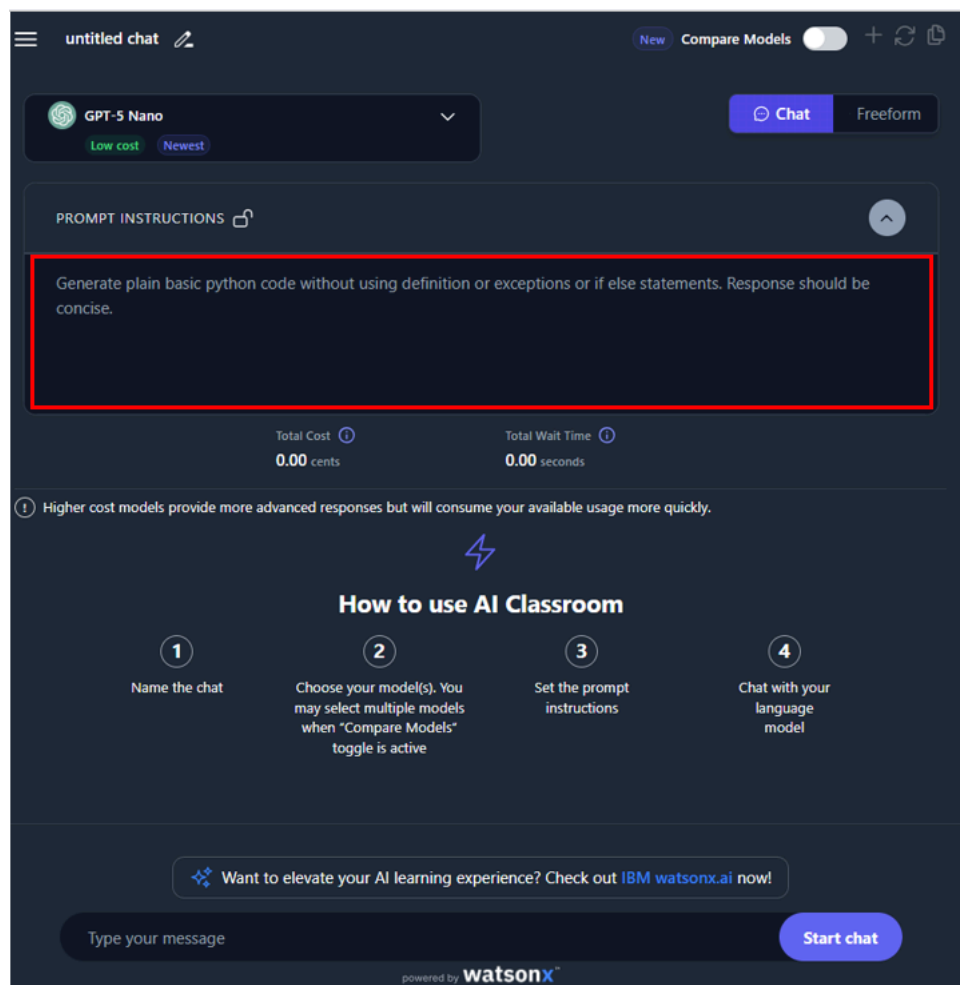
Complete a configuração no Jupyter Notebook e, em seguida, prossiga.

Nota Importante: Você usará o **GPT-5 Nano** para tentar o projeto final. Todos os prompts disponíveis foram ocultados e os usuários são incentivados a primeiro tentar escrever seus próprios prompts para criar as soluções. Além disso, os prompts fornecidos como soluções também foram mantidos como aqueles que criarão estruturas de código genéricas que você pode modificar de acordo com a questão em questão.

Importando o Conjunto de Dados

Você pode começar usando o modelo de IA Generativa para criar um script em python que possa carregar o conjunto de dados em um dataframe do pandas. O arquivo do conjunto de dados já possui os cabeçalhos na primeira linha.

Escreva o prompt para gerar o código mencionado e teste-o no ambiente JupyterLite. Você pode definir o prompt como abaixo para obter uma saída definida:



Para verificar o carregamento apropriado, inclua uma etapa para imprimir os primeiros 5 valores do dataframe carregado.

► [Clique aqui para o prompt](#)