## Bube Ohaneje

## Student Id: 3110527

## Goal setting for project 3830

**Goal Overview**: To clean, preprocess, and analyze the Go Auto dataset in Jupyter Notebook (Python) to ensure it's ready for modeling, and present key insights from EDA for the vehicle pricing model.

**Key Results:**

**Data Cleaning and Preprocessing:**

Action: Explore the dataset to understand its structure. Perform the following tasks:

* Display the count of missing values and identify which columns have missing values.
* Check for and remove duplicates.
* Identify less important features to drop.
* Impute missing values using suitable strategies.
* Remove outliers.

**Measurable:**

* Ensure missing values are handled properly, reducing them to less than 5% of the dataset.
* Document imputation strategies.
* Identify and remove duplicates.
* Apply outlier removal techniques where necessary.

**Complete by: Feb 5, 2025.**

**EDA (Exploratory Data Analysis):**

**Action:** Generate graphs and visualize the relationships between features (e.g., mileage vs. price) to gain insights.

**Measurable:**

* Generate at least 3 visualizations showcasing feature relationships.
* Document 5 key insights from the EDA phase.

Complete by: Feb 13, 2025.

**Data Splitting and Encoding:**

**Action:** Split the dataset into training and test sets. Apply appropriate encoding techniques to the training data.

**Measurable**:

* Train and test data split completed with clear separation.
* Encoding method applied based on feature types (e.g., label encoding for categorical variables)
* Make sure deliverable for Demo 1 is ready with the appropriate presentation document.

**Complete by: Feb 10, 2025.**

**How I’ll Achieve It:**

* Use Jupyter Notebook (Python) with libraries like Pandas, NumPy, Matplotlib, and Seaborn.
* Collaborate with my team, seek feedback, and iterate on my work.
* Complete tasks on time with a focus on quality and accuracy.

Contribution to Project Progress: By cleaning and analyzing the dataset, I will ensure it’s ready for model development and provide valuable insights that will guide the vehicle pricing model. My work will directly support the team's efforts to build an accurate and reliable pricing model.