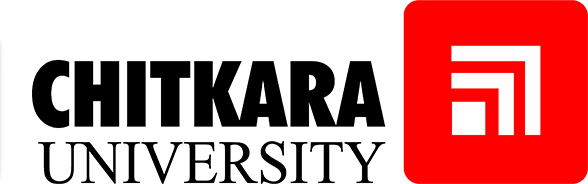
**Capstone Project - 1**

Project Report Semester-IV (Batch-2022)

**Car Price Prediction**



# Supervised By: Submitted By:

Mr. Mohd. Talib Sir

Vrinda Puri 2210992549 (G-28)

Yankit Kumar 2210992552 (G-28)

Akshit Kansal 2210990092 (G-28)

Swastik Garg 2210992432 (G-28)

**Department of Computer Science and Engineering** **Chitkara University Institute of Engineering & Technology,**

**Chitkara University, Punjab**

**Table Of Contents**

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Title** | **Page No.** |
| **1.** | Introduction | 3 |
| **2.** | Problem Statement | 4 |
| **3.** | Software Interaction | 5 |
| **3.** | Exploratory Data Analysis | 6-8 |
| **5.** | Data Preprocessing | 9 |
| **6.** | ML - Models | 9-11 |
| **7.** | Conclusion | 12 |

**Introduction**

Embarking on an innovative venture in the realm of Artificial Intelligence and Machine Learning, our project aims to revolutionize the prediction of car prices. Through meticulous data analysis and advanced algorithmic modeling, we endeavor to offer a sophisticated solution for forecasting the value of automobiles. With a rigorous approach to data collection and model training, our project strives to deliver robust and reliable predictions, empowering stakeholders in the automotive industry to make informed decisions. From manufacturers seeking pricing strategies to consumers navigating the market, our predictive framework promises valuable insights into the dynamic landscape of car valuation. Join us as we merge technology and automotive expertise to redefine the standards of price prediction in the automotive sector.

Through our project, we aspire to bridge the gap between traditional pricing methodologies and the potential offered by cutting-edge AI and ML techniques. By leveraging vast datasets encompassing historical sales data, market trends, and vehicle specifications, we aim to unravel complex patterns and correlations that influence car prices. Our commitment to accuracy and transparency underscores our dedication to delivering a solution that not only meets but exceeds the expectations of industry professionals and enthusiasts alike. Join us as we embark on this journey towards precision and innovation in car price prediction.

**Problem Statement**

In the realm of automotive commerce, accurately forecasting car prices stands as a critical challenge for manufacturers, dealers, and consumers alike. Traditional pricing methods often fall short in capturing the nuanced interplay of factors influencing market value, leading to inefficiencies and missed opportunities. In response, our AIML project seeks to address this pressing issue by developing a robust predictive model capable of reliably estimating car prices.

The primary objective of this project is to leverage Artificial Intelligence and Machine Learning techniques to analyze extensive datasets encompassing historical sales records, vehicle specifications, economic indicators, and market trends. By doing so, we aim to construct a predictive framework that not only accounts for inherent complexities within the automotive market but also adapts to evolving dynamics over time. Through rigorous model training, validation, and optimization, our endeavor is to provide stakeholders with a powerful tool for making informed decisions regarding pricing strategies, inventory management, and purchasing decisions.

By articulating the problem statement in this manner, we underscore the significance of accurate price prediction in the automotive industry while highlighting the innovative approach and objectives of our AIML project.

**Software interaction:**

- **Jupyter Notebook** serves as the primary Integrated Development Environment (IDE) for its interactive and collaborative features.

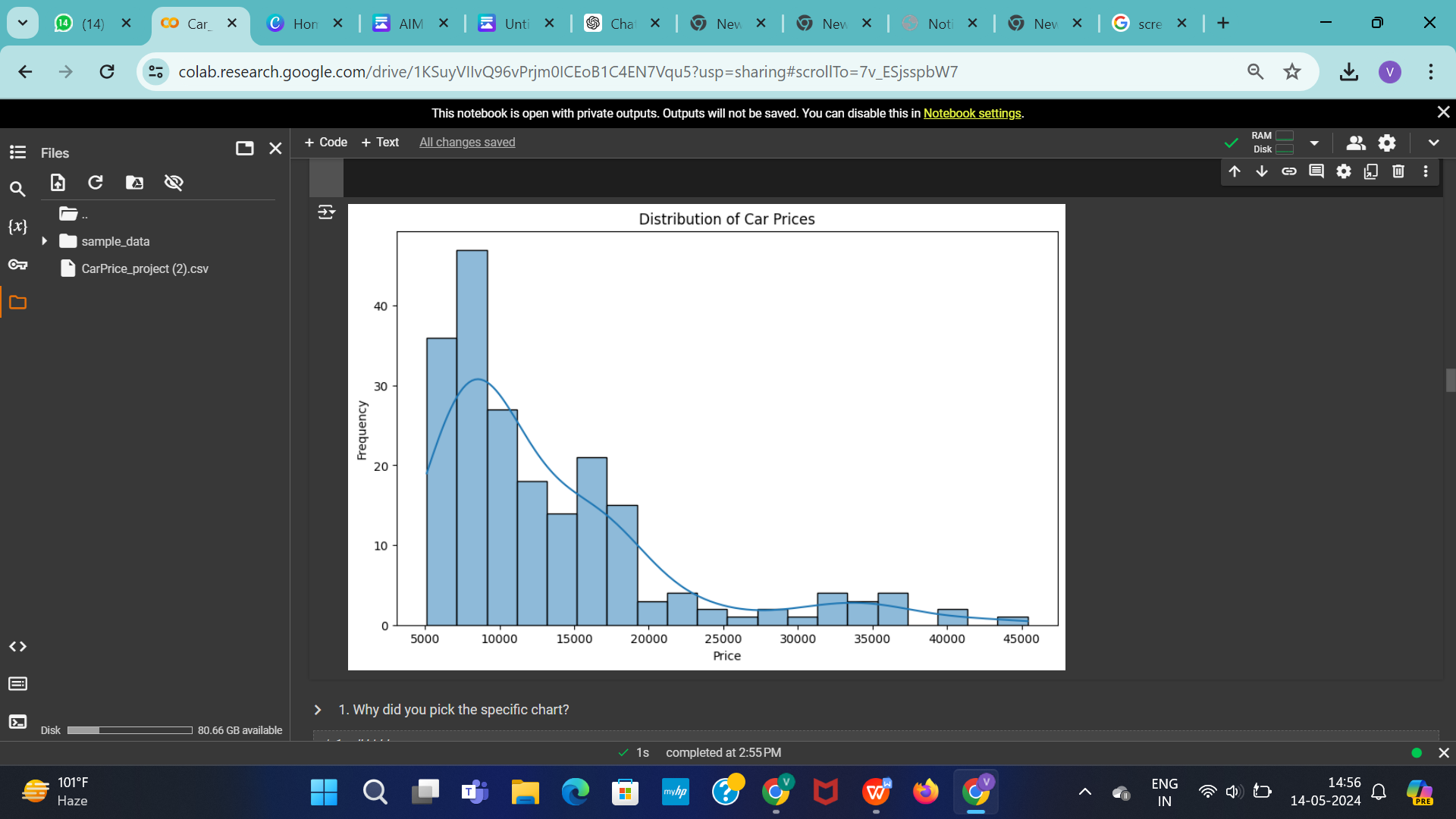
- **Pandas** and **NumPy** play integral roles in data manipulation, pre-processing, and mathematical operations, respectively.

- **Automated Exploratory Data Analysis (EDA**) is facilitated through data prep, streamlining insights discovery.

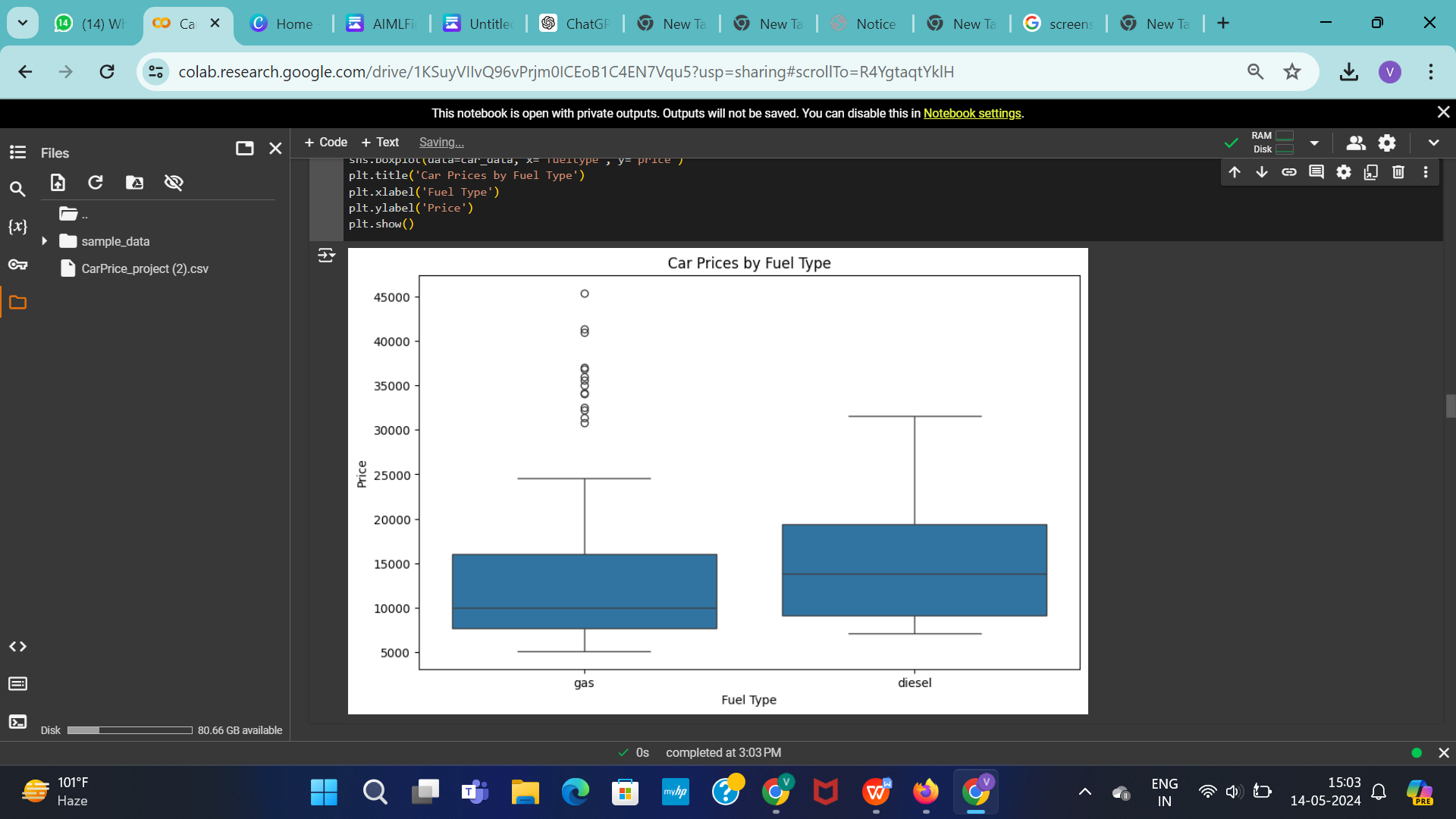
- **Matplotlib**, **Seaborn,** and **Plotly** are utilized for visualization, enabling the creation of compelling graphical representations.

- **GitHub** serves as the cornerstone for version control, ensuring collaboration, change tracking, and repository integrity.

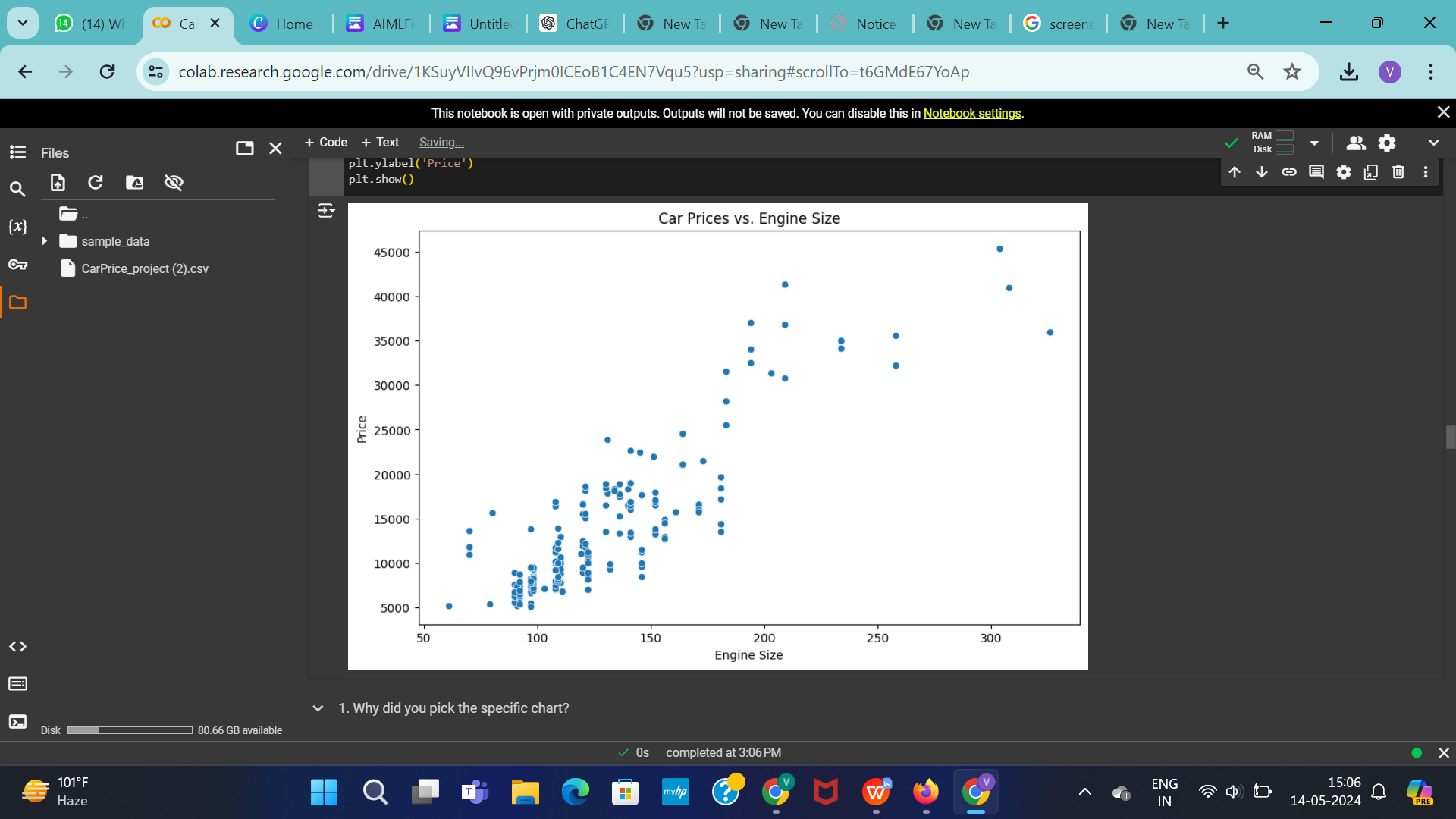
**Exploratory Data Analysis**



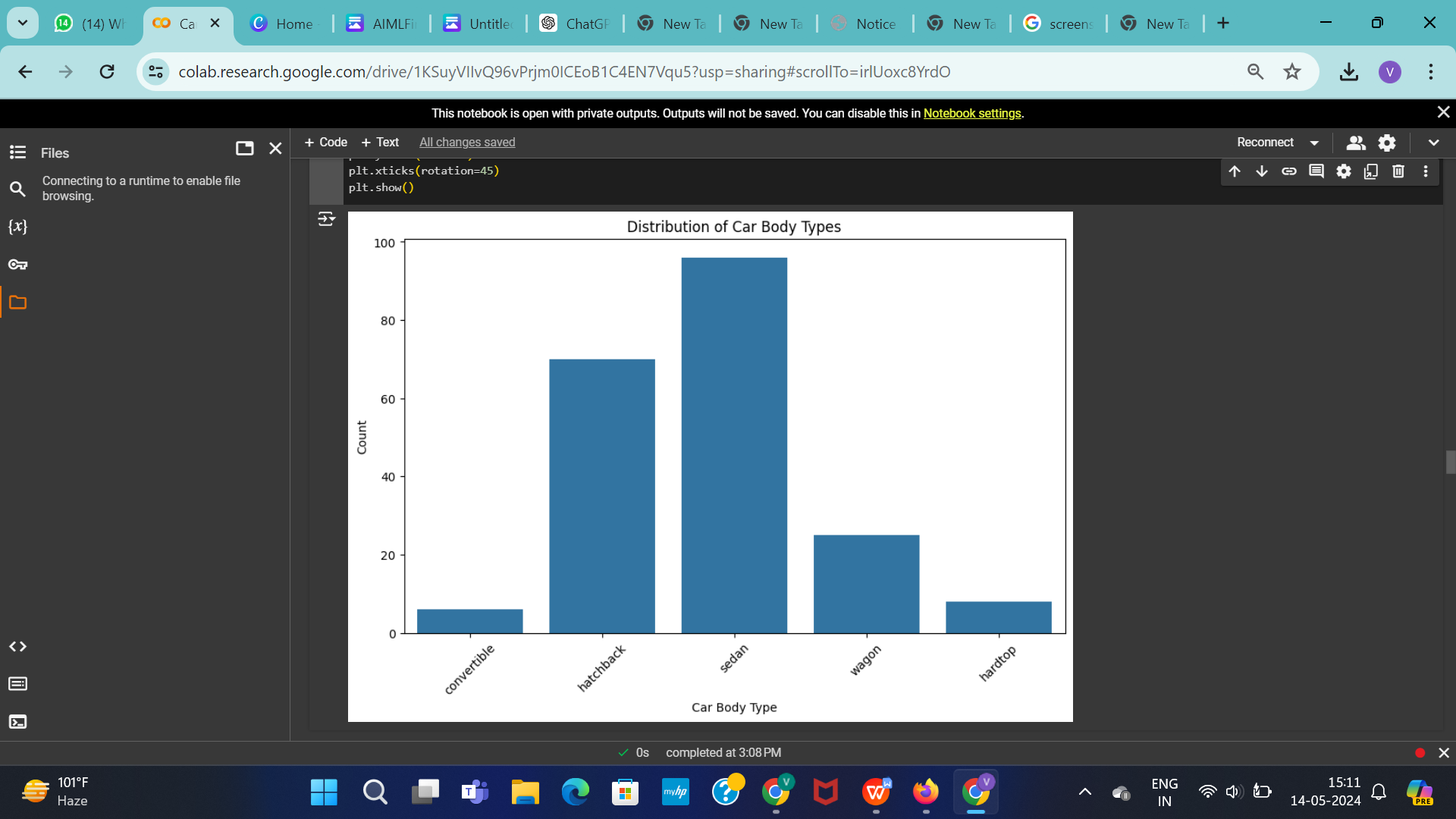
Distribution of Car Prices



Car Prices by Fuel Type



Car Prices vs. Engine Size



Car Body Type vs Count

**Data Preprocessing**

1. **Handling Missing Values:**

The code drops two columns ("Unnamed: 0" and "id") using drop() function. Then, it transforms the target variable "satisfaction" into binary values. Next, it standardizes column names for consistency. However, there's no explicit handling of missing values in the provided snippet. Additional steps such as imputation or deletion might be needed for robust data preprocessing.

1. **Handling Outliers:**

This function identifies outliers within the DataFrame `df` based on the provided `features`. It calculates quartiles and the Interquartile Range (IQR) for each feature, then identifies outlier indices falling beyond 1.5 times the IQR from the quartiles. Outlier indices occurring more than twice across features are stored and returned as `multiple\_outliers`.

**3. Handling Categorical Values:**

This code transforms categorical columns in `train\_df` into binary indicators, assigning 1 to rows matching the first unique value and 0 otherwise. Then, it converts the "satisfaction" column in `test\_df` to binary, representing "satisfied" as 1 and others as 0.

**ML – MODEL 1**

**Linear Regression:**

1.Data Preparation:

Import necessary libraries (e.g., NumPy, pandas).

Load or create your dataset.

Split the dataset into features (X) and target variable (y).

2.Import the linear regression model from a library (e.g., scikit-learn).

Create an instance of the linear regression model.

3.Fit the model to the training data.

Use the .fit() method of the linear regression model, passing the features (X\_train) and target variable (y\_train).

4.Use the trained model to make predictions on test data.

Use the .predict() method of the linear regression model, passing the features (X\_test).

Evaluation:

5.Calculate metrics such as Mean Squared Error (MSE), R-squared, or others depending on the problem

**ML – MODEL 2**

**Decision Tree Classifier:**

1. Import the DecisionTreeClassifier from sklearn.tree.

2. Instantiate a DecisionTreeClassifier object as dt.

3. Define a parameter grid containing potential hyperparameters for tuning.

4. Instantiate a GridSearchCV object (clf) with the DecisionTreeClassifier and the defined parameter grid.

5. Fit the GridSearchCV object to the training data (x\_train, y\_train).

6. Print the best parameters found by the grid search.

7. Make predictions (y\_pred) on the test data using the best model found.

8. Evaluate the model's performance using the accuracy\_score function and print the result.

**ML – MODEL 3**

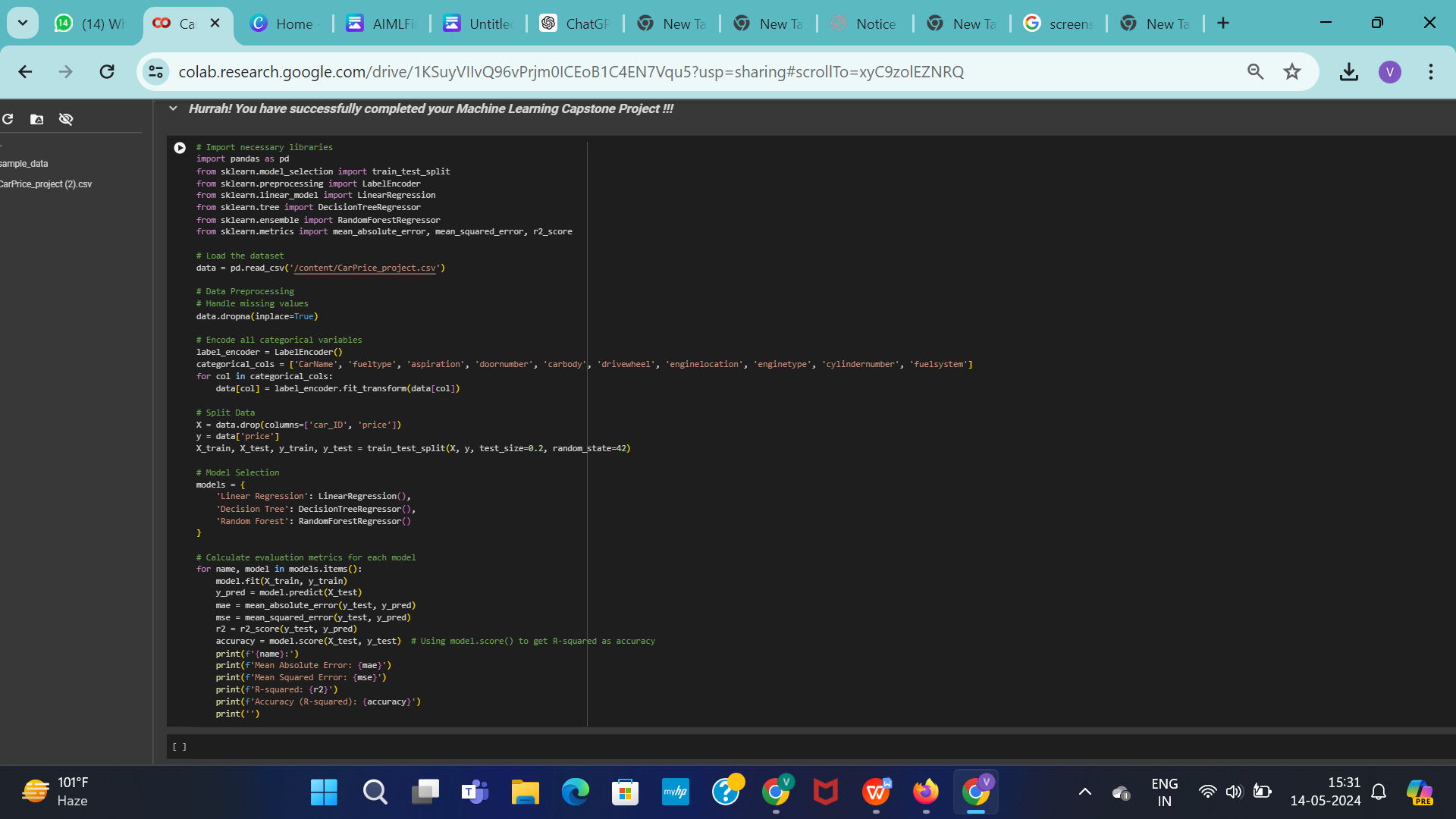
**Random Forest:**

- The code implements a Random Forest Classifier, a model that constructs multiple decision trees and aggregates their predictions for improved accuracy.

- GridSearchCV is employed for hyperparameter tuning, optimizing the model's performance by testing various parameter combinations.

- The best parameters found during the grid search are printed, aiding in understanding the model's configuration.

- Predictions are generated for the test data using the trained model.



**Conclusion**

The implementation of machine learning models such as Logistic Regression, Decision Tree Classifier, Random Forest Classifier holds promise for predicting airline passenger satisfaction.

Through thorough data preprocessing, including feature engineering and handling missing values, and employing hyperparameter tuning techniques like GridSearchCV, these models can effectively learn from the provided data and make accurate predictions.

However, it's crucial to ensure proper evaluation metrics are employed to assess model performance accurately. By integrating these techniques into an ensemble model or deploying the best-performing individual model, airlines can enhance their understanding of passenger satisfaction factors and optimize service delivery.