**Car Price Prediction System**



# 1. Problem Statement:

Issue or problem that you will address:

The project aims to address the challenge of accurately predicting the price of used cars based on various features such as brand, location, year of manufacture, kilometers driven, fuel type, transmission, owner type, mileage, and engine capacity. The primary issue is the lack of an efficient and user-friendly system for predicting the resale value of cars, which is crucial for both buyers and sellers in the second-hand car market.

# 2. Objectives:

What will be the objectives of this work:

* **Develop a Predictive Model:** Create a robust linear regression model using Flutter desktop application that leverages OLS (Ordinary least Squares) techniques to predict the price of used cars based on the provided dataset.
* **User-Friendly Interface:** Design an intuitive and user-friendly interface for the Flutter desktop app, allowing users to input car details effortlessly and receive accurate price predictions in real-time.
* **Visual Representation:** Implement data visualization techniques, including pie charts, bar charts, histograms, and whisker plot to provide users with a clear and comprehensible representation of the dataset, aiding in understanding the factors influencing car prices.
* **Data Accuracy and Reliability:** Ensure that the predictive model is trained on a reliable dataset to maintain accuracy, considering the dynamic nature of the used car market.

# 3. Data Description:

Link of the data:

<https://github.com/sagnikghoshcr7/Car-Price-Prediction/blob/master/data/dataset.csv>

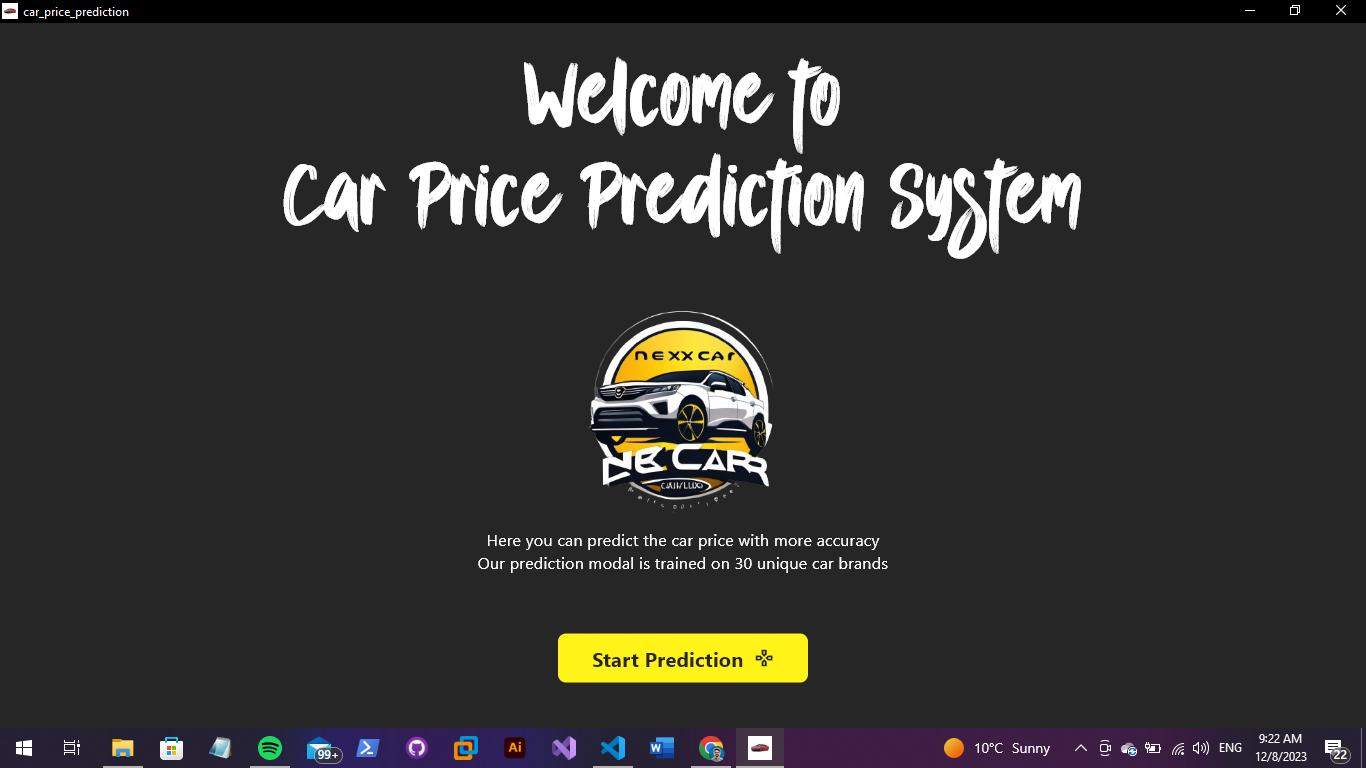
Details description of the data:

The dataset comprises information about used cars, including the following parameters:

* **S.No:** Serial number for identification.
* **Name:** Make and model of the car.
* **Location:** City or region where the car is located.
* **Year:** Year of manufacture.
* **Kilometers Driven:** The total distance the car has been driven.
* **Fuel Type:** Type of fuel the car uses (e.g., Petrol, Diesel, CNG).
* **Transmission:** Type of transmission (e.g., Manual, Automatic).
* **Owner Type:** The number of previous owners.
* **Mileage:** Fuel efficiency of the car in terms of kilometers per unit of fuel.
* **Engine:** Engine capacity in cubic centimeters.

The multiple linear regression model utilizes these parameters to predict the price of a used car accurately. The dataset is diverse and representative of various car models, making the model applicable to a wide range of scenarios in the second-hand car market. The objective is to harness this data to provide users with a reliable tool for estimating the resale value of their vehicles.

# 4. Results:



The front page serves as the gateway to the car price prediction experience. It features a minimalist design with a prominent button as the focal point. This button, when clicked, seamlessly navigates users to the prediction page.

A screenshot of a computer

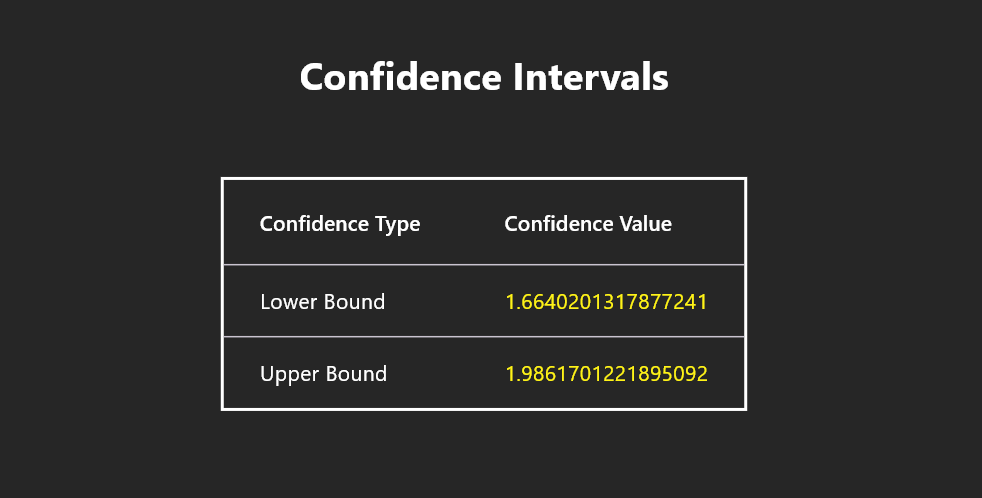
Description automatically generated

* **Dynamic Prediction Interface:**
  + Users can input various parameters including brand, location, year, kilometers driven, fuel type, transmission, owner type, mileage, and engine capacity.
  + The interface is designed to be dynamic and user-friendly.
* **Linear Regression Model:**
  + The system leverages a pre-trained linear regression model to predict the resale price of the specified car.
  + The model has been trained on historical data, enabling it to make predictions based on the input parameters.
* **Predicted Price Display:**
  + Upon submitting the input details, the system promptly generates and displays the predicted price for the specified car.
  + Users gain valuable insights into the estimated resale value, providing a practical tool for decision-making in the second-hand car market.

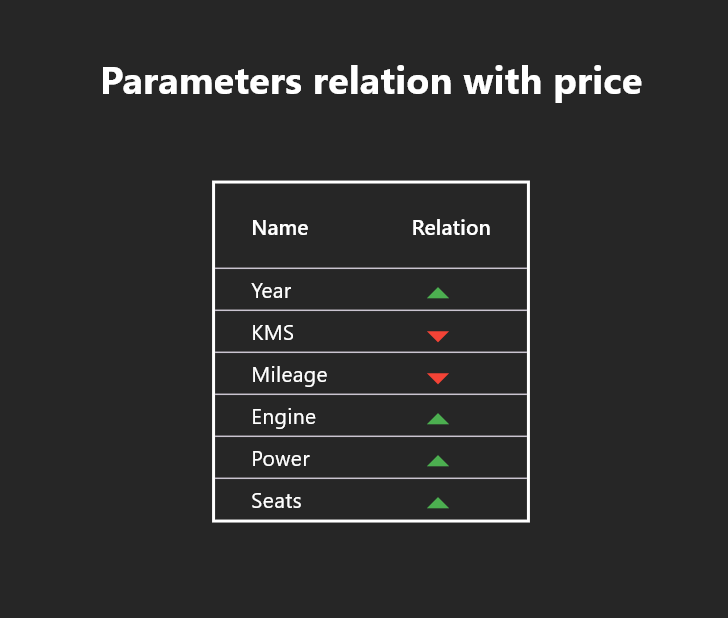
A screenshot of a black screen

Description automatically generated

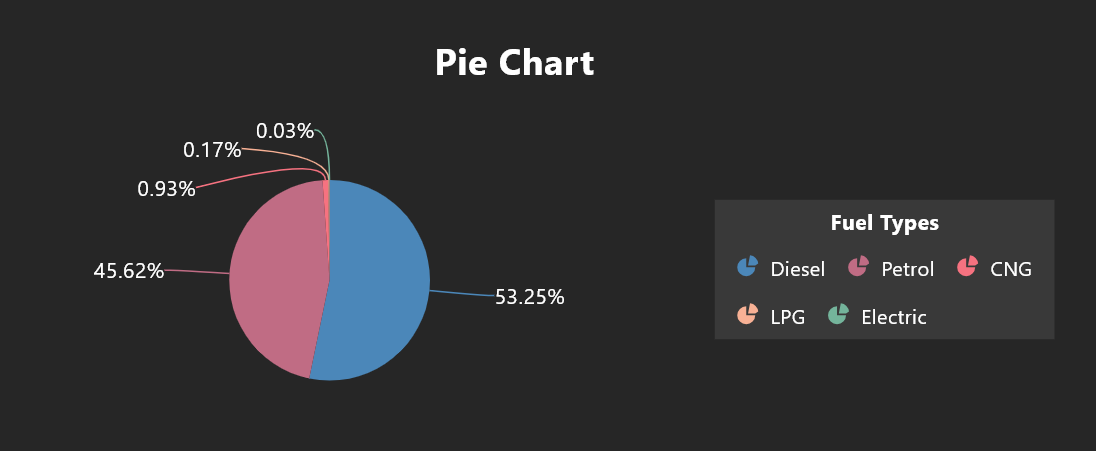
* **Predicted Price Screen Overview:**
  + The screenshot encapsulates the Predicted Price screen, a pivotal component of the application's user interface.
  + The screen exhibits the outcomes of a comprehensive linear regression analysis on a set of cars.
* **Display of Predicted Resale Values:**
  + Users can visually explore the predicted resale values for the specified set of cars.
  + The linear regression model utilizes input parameters to estimate and display the probable resale prices.
* **Insightful Tool for Resale Value Prediction:**
  + The Predicted Price screen serves as a practical and insightful tool for users navigating the second-hand car market.
  + It enables users to understand how the model interprets various features, translating them into estimated resale values.
* **User Observation:**
  + Users have the opportunity to observe and analyze the correlation between input parameters and the predicted resale values.
  + This screen provides a user-friendly interface for gaining actionable insights into the potential resale values of cars based on specified characteristics



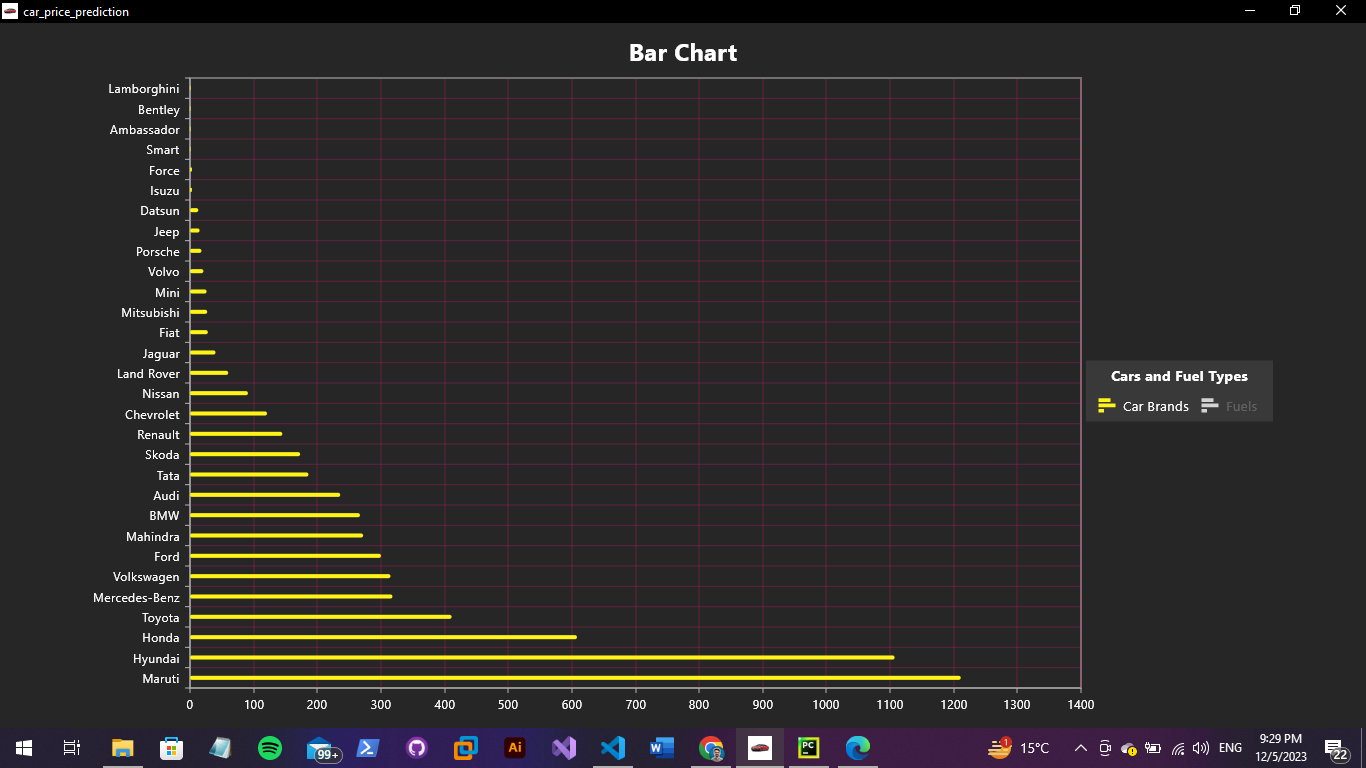
* **Confidence Interval Visualization:**
  + The enclosed screenshot illustrates a graphical representation of confidence intervals for predicted car prices.
* **Range of Likely Values:**
  + Confidence intervals provide a range within which the actual resale values are likely to fall.
  + Users can visually perceive the upper and lower bounds, indicating the model's level of uncertainty.
* **Distribution of Predicted Prices:**
  + The plot displayed in the screenshot demonstrates the distribution of predicted prices.
  + It offers a clear depiction of how the model anticipates the resale values for a range of specified parameters.
* **Insights into Model's Uncertainty:**
  + The visualization serves as a tool for understanding the reliability and variability of the model's predictions.
  + Users can gain valuable insights into the potential range of resale values for second-hand cars, considering the inherent uncertainty in the prediction process.



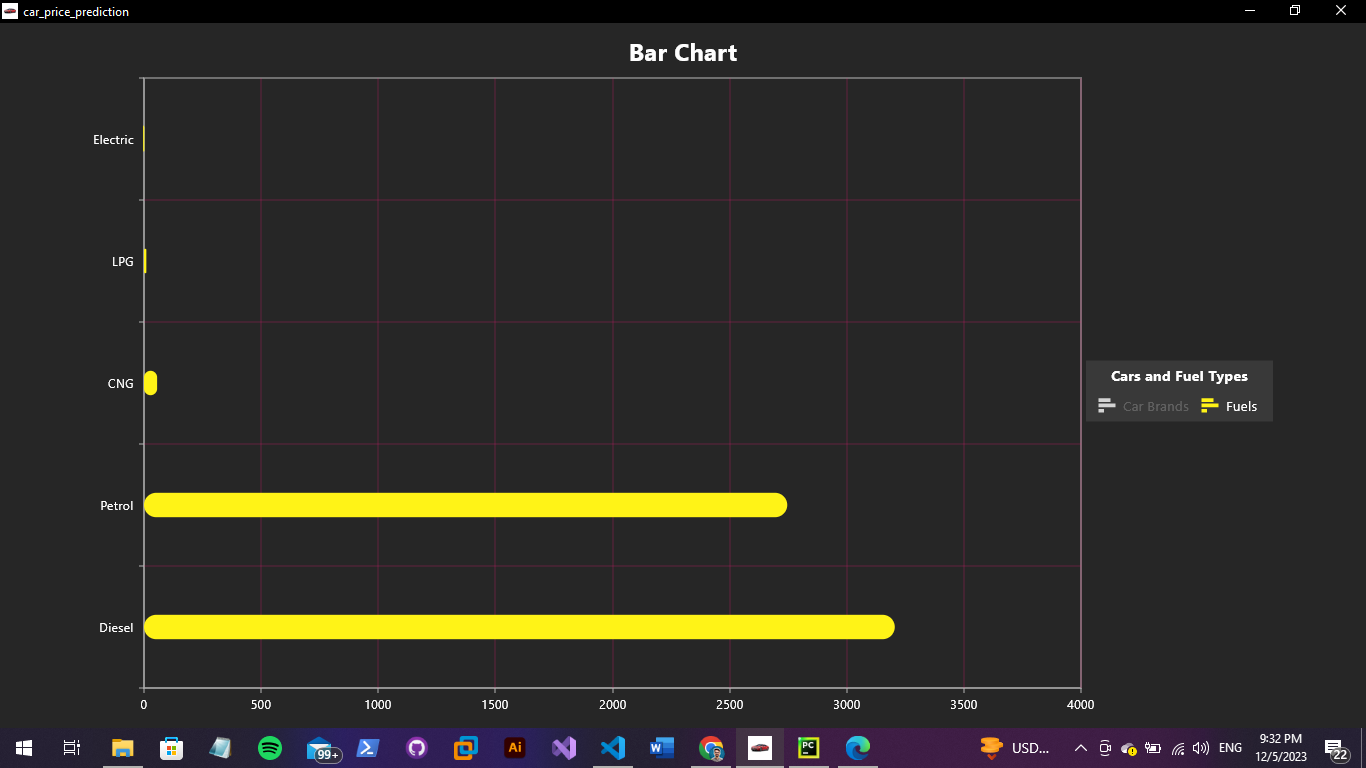
* **Comprehensive Parameter Overview:**
  + The screenshot provides a holistic view of various parameters in relation to car prices.
* **Showcased Parameters:**
  + Parameters such as year, kilometers driven, fuel type, transmission, owner type, mileage, engine capacity, power, number of seats, and additional features are prominently displayed.
* **Corresponding Price Points:**
  + Each parameter is showcased alongside corresponding price points, offering a clear representation of their impact on car prices.
* **Holistic Snapshot:**
  + The consolidated presentation enables users to visually assess the multifaceted relationship between the mentioned parameters and their respective impacts on the pricing of cars.
* **User-Friendly Visual Exploration:**
  + Users can easily explore and interpret the visual representation, gaining a comprehensive understanding of how various features contribute to the overall pricing dynamics in the second-hand car market.



* **Pie Chart Representation:**
  + The enclosed snippet features a visually informative pie chart illustrating the distribution of cars based on different fuel types.
* **Fuel Type Segmentation:**
  + Each segment of the pie corresponds to a specific fuel type, including categories such as Petrol, Diesel, or CNG.
* **Size Indicates Quantity or Percentage:**
  + The size of each pie segment dynamically indicates either the quantity or percentage of cars utilizing a particular fuel type.
* **Selectable Categories:**
  + Users have the flexibility to interact with the pie chart by selecting specific fuel categories they want to display.
  + For instance, users can choose to view a focused pie chart for Diesel and Petrol, tailoring the visual representation to their preferences.
* **Intuitive Visual Exploration:**
  + The pie chart serves as an intuitive tool for users to explore and comprehend the distribution of cars based on diverse fuel types, enhancing their understanding of the fuel dynamics in the second-hand car market.



* **Bar Chart Representation:**
  + This snippet presents a bar chart designed to visually represent the number of cars for each distinct company in the dataset.
* **Manufacturer-Specific Bars:**
  + Each bar within the chart is associated with a specific car manufacturer, creating a visual representation of the dataset's brand distribution.
* **Bar Height Indicates Car Quantity:**
  + The height of each bar dynamically reflects the quantity of cars produced by the corresponding manufacturer.
* **Comparison Across Manufacturers:**
  + The visual arrangement of bars facilitates a quick and straightforward comparison of the car count across different manufacturers.
* **Insights into Brand Distribution:**
  + Users can gain valuable insights into the distribution of car brands within the dataset, making it a useful tool for analyzing the prevalence of different manufacturers in the second-hand car market.

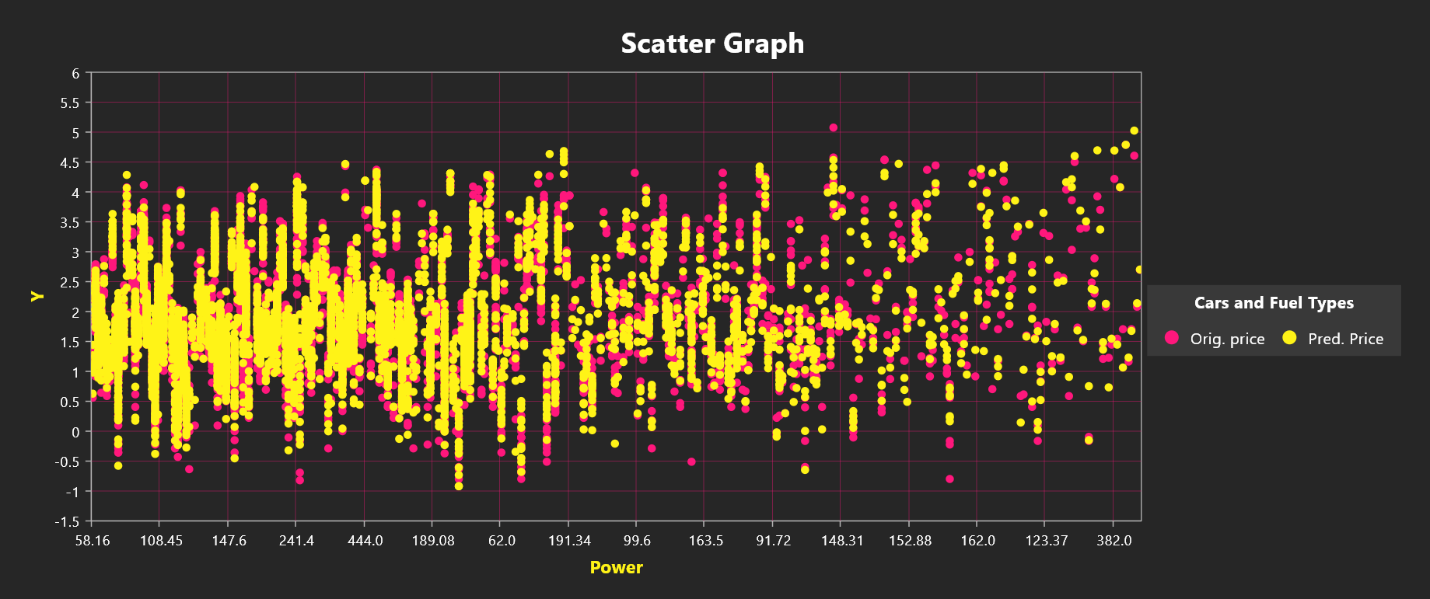


* **Bar Chart Representation:**
  + The provided snippet includes a bar chart designed to visually represent the distribution of cars based on different fuel types.
* **Fuel Category Bars:**
  + Each bar within the chart corresponds to a specific fuel category, such as Petrol, Diesel, or CNG.
* **Bar Height Reflects Car Count:**
  + The height of each bar signifies the count of cars utilizing the respective fuel type.
* **Comparative View of Fuel Types:**
  + The graphical representation offers a comparative view, allowing users to quickly assess the prevalence of each fuel type within the dataset.
* **Facilitates Understanding of Fuel Composition:**
  + The bar chart is a useful tool for understanding the market's fuel composition, providing insights into the distribution of cars based on different fuel categories in the second-hand car market.

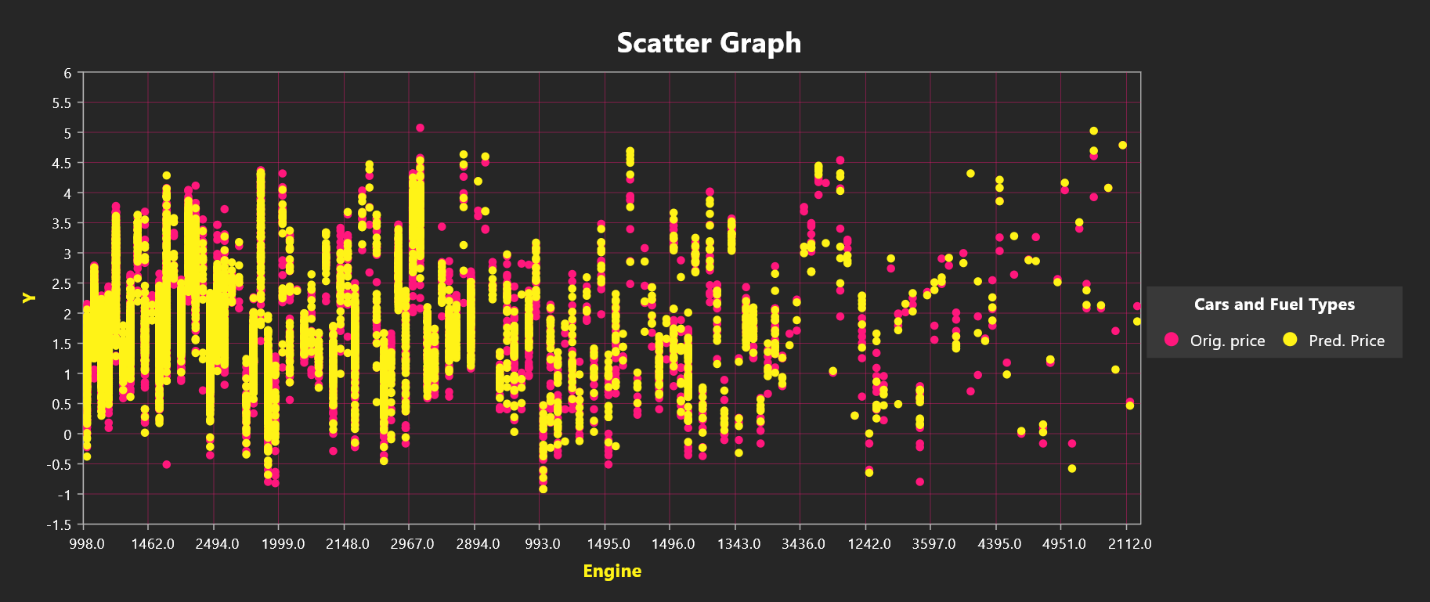
A screen shot of a graph

Description automatically generated

* **Scatter Plot Representation:**
  + This section features a scatter plot that visually maps both the original prices and the predicted prices for cars in the dataset.
* **Individual Data Points:**
  + Each point on the scatter plot represents an individual car within the dataset.
* **X-axis and Y-axis Significance:**
  + The x-axis is indicative of the mileage of the cars, while the y-axis(Price) yhat represents either the original prices or predicted prices.
* **Direct Comparison:**
  + The scatter plot allows for a direct comparison between the actual resale values (original prices) and the values predicted by the model.
* **Insights into Predictive Algorithm Performance:**
  + Users can gain insights into the accuracy and performance of the predictive algorithm by observing how closely the predicted values align with the actual resale values.
* **Visual Assessment of Model Predictions:**
  + The visualization provides a clear and intuitive means for users to assess the model's predictions in relation to the true values, aiding in the evaluation of the model's effectiveness in predicting car resale prices.



* **Scatter Plot Representation:**
  + In this section, a scatter plot is showcased where the x-axis corresponds to the power of cars, and the y-axis represents either the original prices or predicted prices.
* **Individual Data Points:**
  + Each point on the scatter plot corresponds to an individual car within the dataset.
* **X-axis and Y-axis Significance:**
  + The x-axis represents the power of the cars, while the y-axis(Price) yhat showcases either the original prices or predicted prices.
* **Side-by-Side Comparison:**
  + The scatter plot provides a side-by-side comparison, enabling users to visually assess the alignment between the actual resale values and the values predicted by the model.
* **Evaluation of Model Performance:**
  + Users can leverage the visualization to evaluate how well the model predicts resale values based on the power of the cars.
* **Insights into Power's Impact:**
  + The graphical representation offers insights into the relationship between the power of cars and their corresponding resale values, aiding users in understanding the predictive performance of the model for cars with varying power specifications.

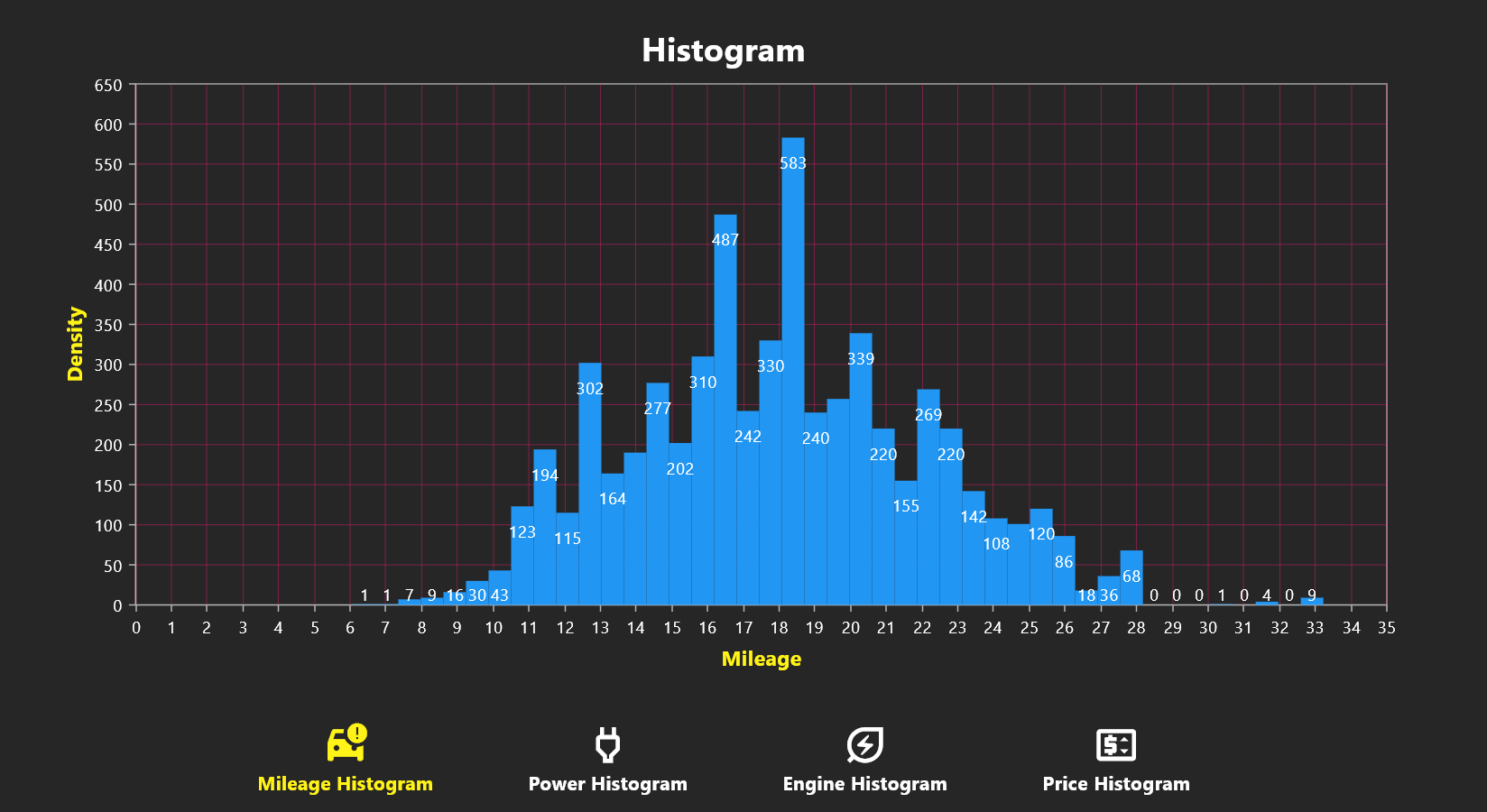


* **Scatter Plot Introduction:**
  + This section introduces a scatter plot featuring the x-axis representing the engine capacity of cars and the y-axis indicating either the original prices or predicted prices.
* **Individual Data Points:**
  + Each point on the scatter plot corresponds to an individual car within the dataset.
* **X-axis and Y-axis Significance:**
  + The x-axis portrays the engine capacity of the cars, while the y-axis continues to feature either the original prices or predicted prices.
* **Comprehensive Comparison:**
  + The scatter plot enables a comprehensive comparison, allowing users to visually assess the relationship between engine capacity and the accuracy of the linear regression model's price predictions.
* **Evaluation of Model Accuracy:**
  + Users can leverage the visualization to evaluate how well the model's predictions align with the actual resale values based on the engine capacity of the cars.
* **Insights into Engine Capacity Impact:**
  + The graphical representation offers insights into the impact of engine capacity on the model's accuracy in predicting car resale prices, providing valuable information for users analyzing cars with varying engine specifications.

A screenshot of a graph

Description automatically generated

* **Scatter Plot Iteration:**
  + In this iteration, the scatter plot introduces the number of seats in a car as a factor, plotted on the x-axis, with the y-axis representing either the original prices or predicted prices.
* **Individual Data Points:**
  + Each point on the scatter plot corresponds to an individual car within the dataset.
* **X-axis Represents Number of Seats:**
  + The x-axis portrays the number of seats in a car, offering a new perspective on how this feature influences the accuracy of price predictions.
* **Side-by-Side Comparison:**
  + The scatter plot maintains a side-by-side comparison, enabling users to visually assess the relationship between the number of seats and the accuracy of price predictions from the linear regression model.
* **Enhanced Perspective:**
  + The inclusion of the number of seats as a factor provides an enhanced perspective on how this particular feature impacts the accuracy of price predictions, contributing to a more comprehensive understanding of the model's performance.
* **Insights into Seat Count Impact:**
  + Users can leverage the visualization to gain insights into how the number of seats influences the accuracy of the linear regression model's price predictions for individual cars in the dataset.

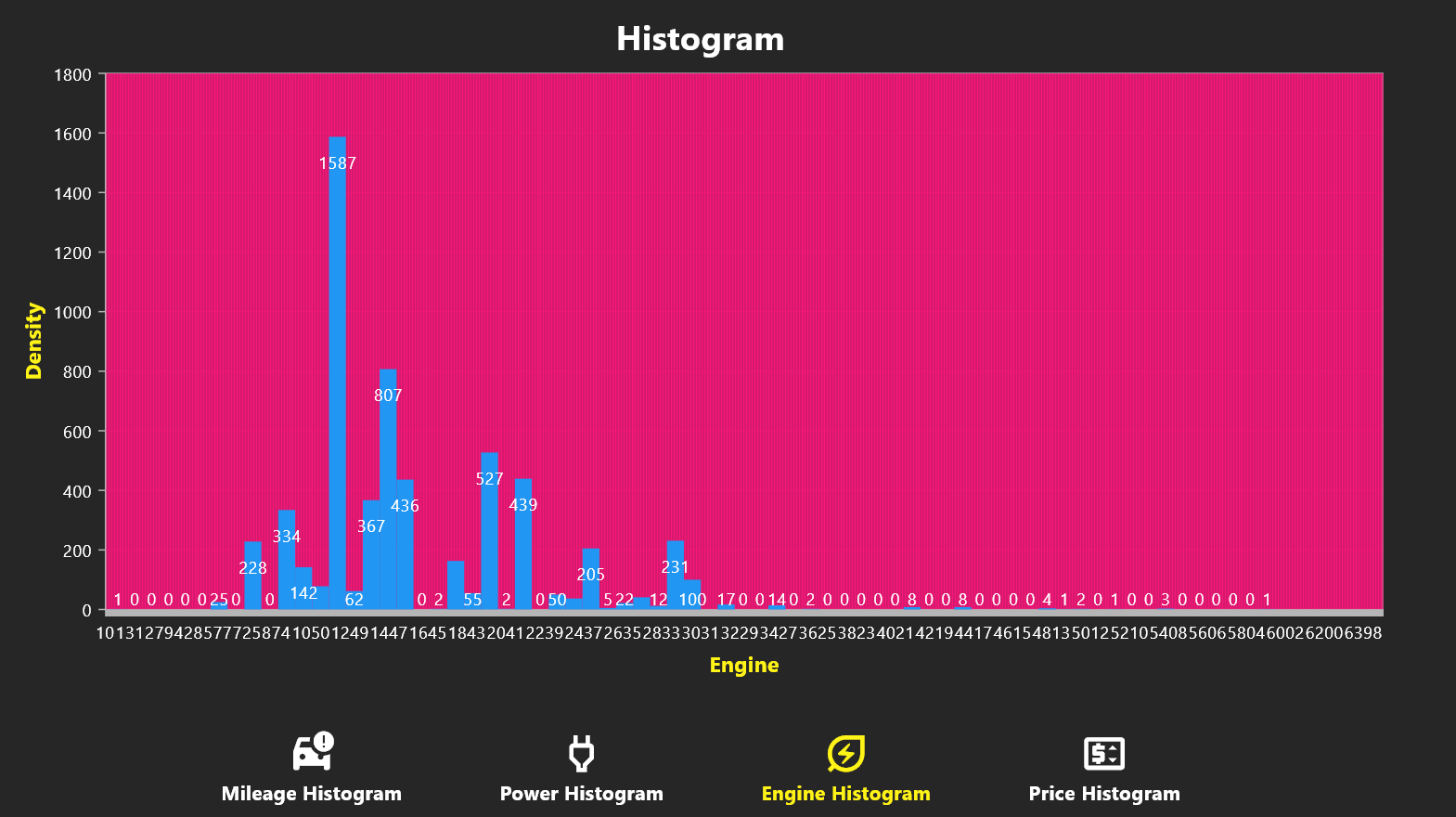


* **Histogram Screen - Mileage Distribution:**
  + The enclosed screenshot features a histogram providing insights into the distribution of mileage among a set of cars within the dataset.
* **X-axis Represents Mileage Intervals:**
  + The x-axis is dedicated to mileage intervals, delineating different ranges of mileage values.
* **Y-axis Indicates Car Count:**
  + The y-axis quantifies the number of cars falling within each mileage category.
* **Visual Spread and Concentration:**
  + The histogram visually represents the spread and concentration of mileage values, offering an overview of how mileage is distributed across the dataset.
* **Identification of Common Ranges and Outliers:**
  + Users can identify common mileage ranges and potential outliers through the graphical presentation, aiding in a nuanced understanding of the dataset's mileage characteristics.
* **Enhanced Understanding of Mileage Distribution:**
  + The graphical presentation enhances understanding by providing insights into patterns and trends related to mileage within the second-hand car market, contributing to informed decision-making for potential buyers and sellers.

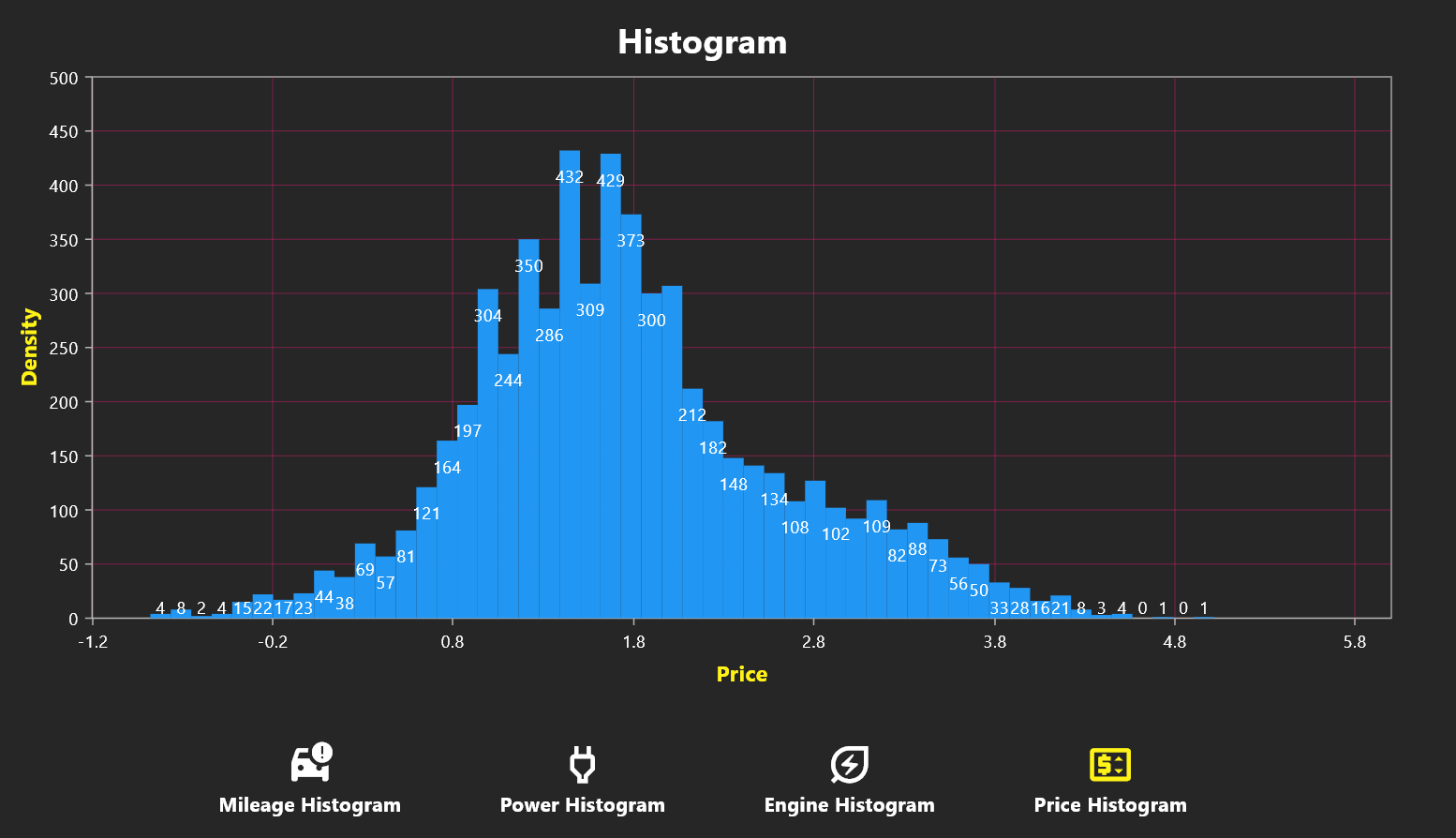
A screenshot of a computer screen

Description automatically generated

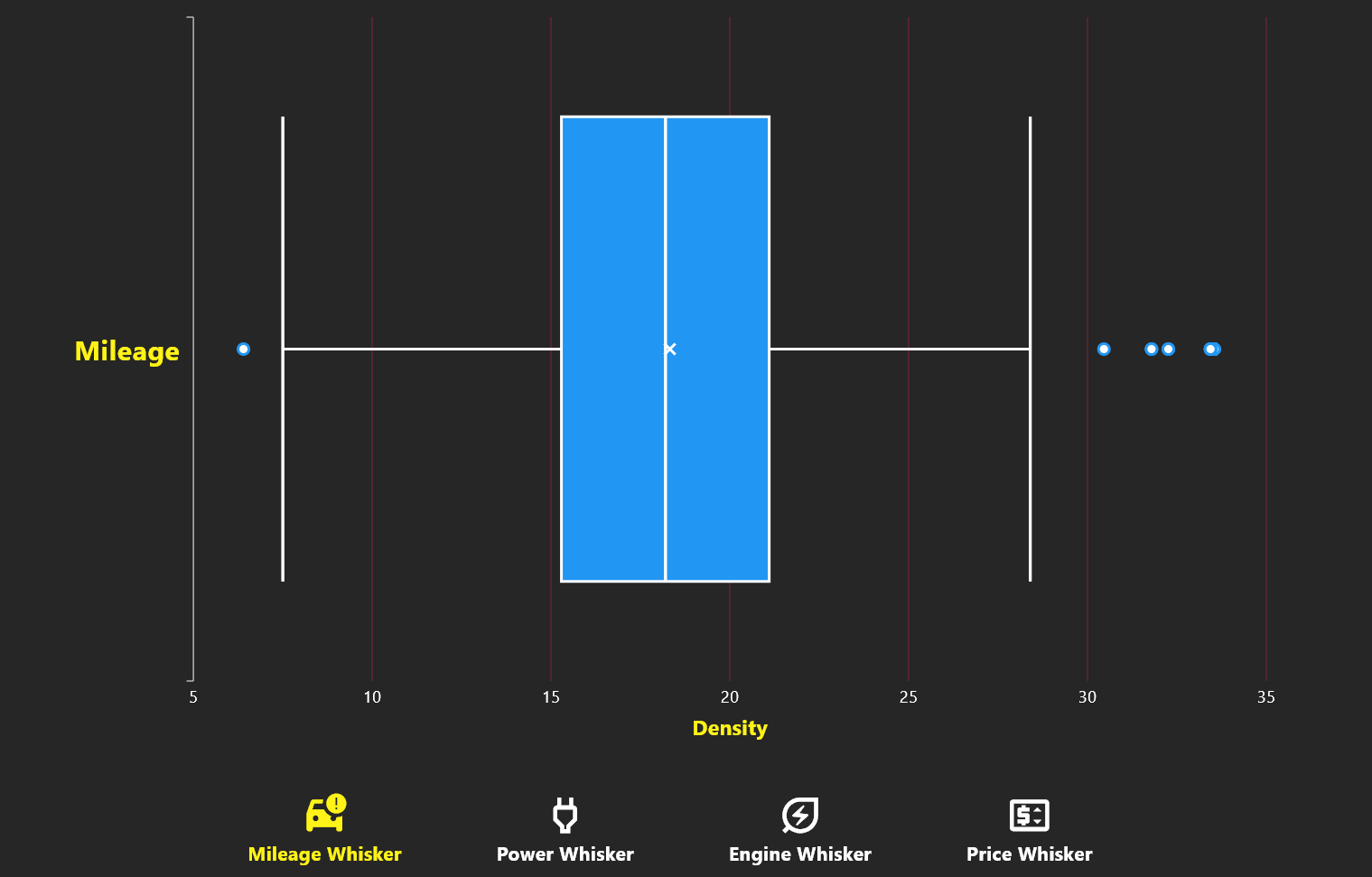
* **Histogram Screen - Engine Power Distribution:**
  + The enclosed screenshot features a histogram that visually portrays the distribution of engine power across a collection of cars within the dataset.
* **X-axis Represents Power Intervals:**
  + Power intervals are delineated on the x-axis, providing a breakdown of different power categories.
* **Y-axis Indicates Car Count:**
  + The y-axis quantifies the number of cars falling within each power category.
* **Visual Tool for Comprehending Diversity:**
  + The histogram serves as a visual tool, aiding users in comprehending the diversity and concentration of engine power values present in the dataset.
* **Identification of Prevalent Ranges and Outliers:**
  + Users can identify prevalent power ranges and potential outliers through the graphical representation, contributing to a nuanced understanding of the distribution of engine power in the second-hand car market.
* **Enhanced Understanding of Engine Power Distribution:**
  + The graphical presentation enhances understanding by offering a visual overview of how engine power is distributed across the dataset, providing valuable insights for market analysis and decision-making.



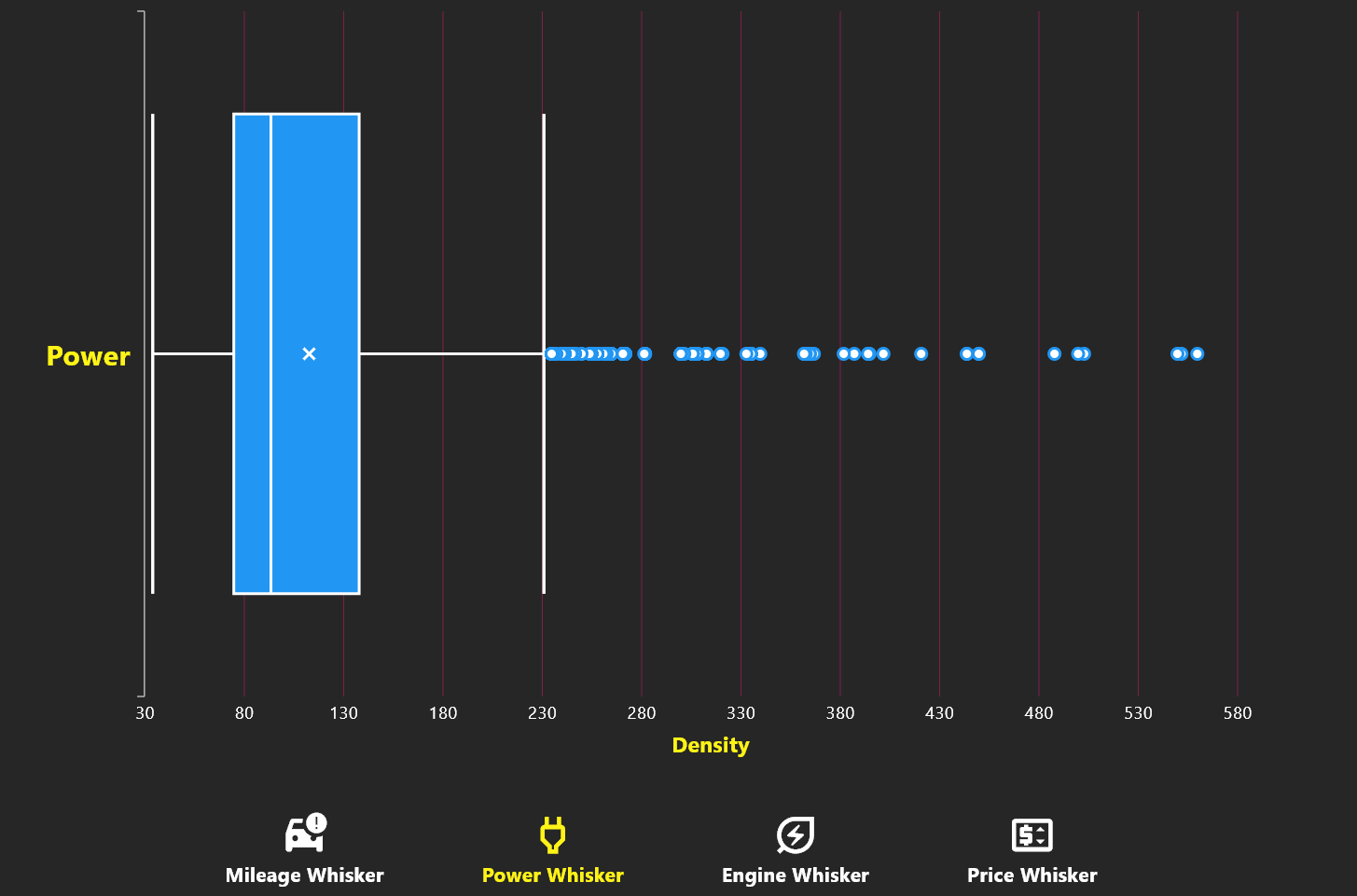
* **Histogram Screen - Engine Capacity Distribution:**
  + The enclosed screenshot features a histogram illustrating the distribution of engine capacities across a set of cars within the dataset.
* **X-axis Categorizes Engine Capacities:**
  + Engine capacities are categorized on the x-axis, providing a breakdown of different capacity ranges.
* **Y-axis Quantifies Car Count:**
  + The y-axis quantifies the number of cars falling within each engine capacity range.
* **Visual Representation of Prevalence and Spread:**
  + The histogram serves as a visual representation, offering insights into the prevalence and spread of engine capacities within the dataset.
* **Identification of Common Sizes and Outliers:**
  + Users can identify common engine sizes and potential outliers through the graphical representation, contributing to a nuanced understanding of how engine capacity is distributed in the second-hand car market.
* **Enhanced Understanding of Engine Capacity Distribution:**
  + By observing the distribution, users can enhance their understanding of the prevailing engine sizes in the dataset, aiding in market analysis and decision-making related to engine capacity considerations



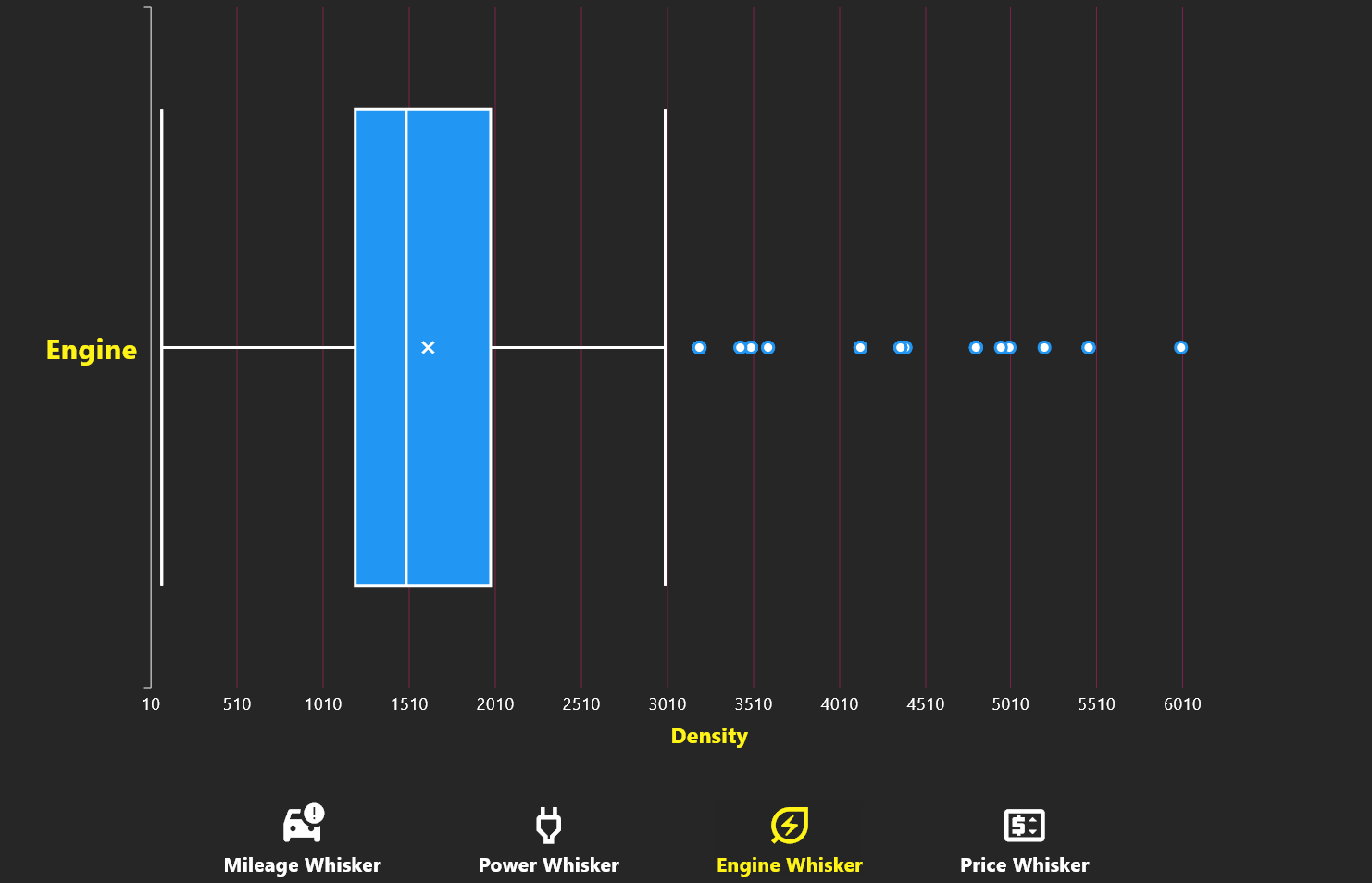
* **Histogram Screen - Car Price Distribution:**
  + The provided screenshot features a histogram that illustrates the distribution of car prices within specific price ranges.
* **X-axis Depicts Price Intervals:**
  + Price intervals are delineated on the x-axis, providing a segmentation of different price categories.
* **Y-axis Enumerates Car Count:**
  + The y-axis enumerates the number of cars falling within each price category.
* **Visual Depiction of Prevalence and Spread:**
  + The histogram offers a visual depiction, providing insights into the prevalence and spread of car prices within the dataset.
* **Identification of Common Ranges and Outliers:**
  + Users can identify common price ranges and potential outliers through the graphical representation, contributing to a nuanced understanding of how car prices are distributed in the second-hand car market.
* **Enhanced Insights into Price Distribution:**
  + The graphical presentation enhances insights into the distribution of prices, offering valuable information for market analysis and decision-making in the context of car pricing dynamics.



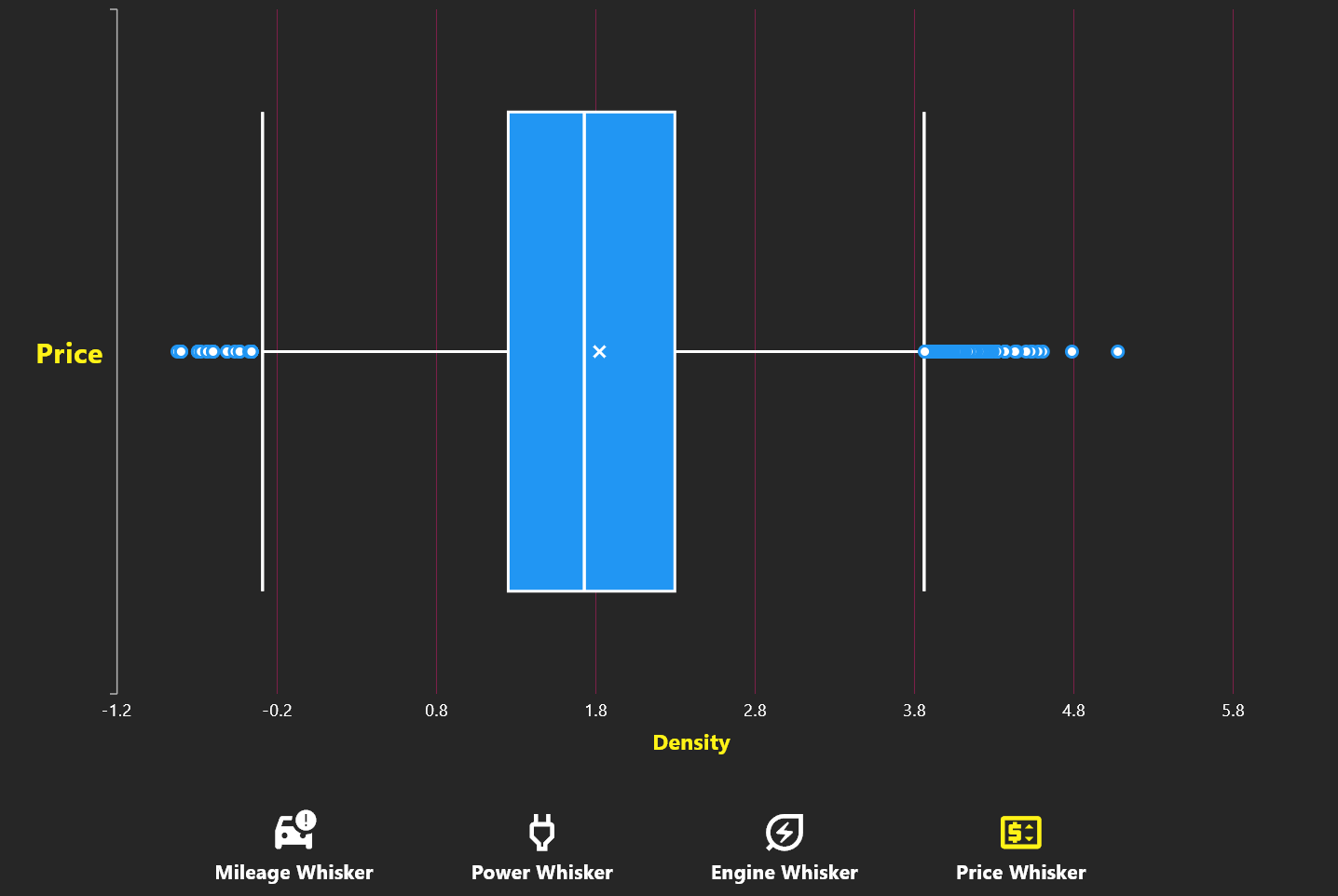
* **Whisker Chart Screen - Mileage Distribution by Price Range:**
  + The enclosed screenshot features a whisker chart illustrating the distribution of mileage values within different price ranges for cars.
* **Visual Representation of Central Tendencies and Spread:**
  + The whisker chart provides a visual representation of central tendencies, spread, and potential outliers of mileage data grouped by specific price intervals.
* **Insights into Mileage Variation Across Price Categories:**
  + The chart aids in understanding how mileage varies across different price categories, offering insights into the relationship between car prices and their corresponding mileage.
* **Enhanced Analysis of Mileage Patterns:**
  + This graphical presentation enhances the analysis of mileage patterns and trends within distinct price ranges in the second-hand car market.
* **Facilitates Identification of Outliers:**
  + Users can identify potential outliers and variations in mileage more easily through the graphical depiction, contributing to a more comprehensive understanding of the relationship between mileage and car prices.



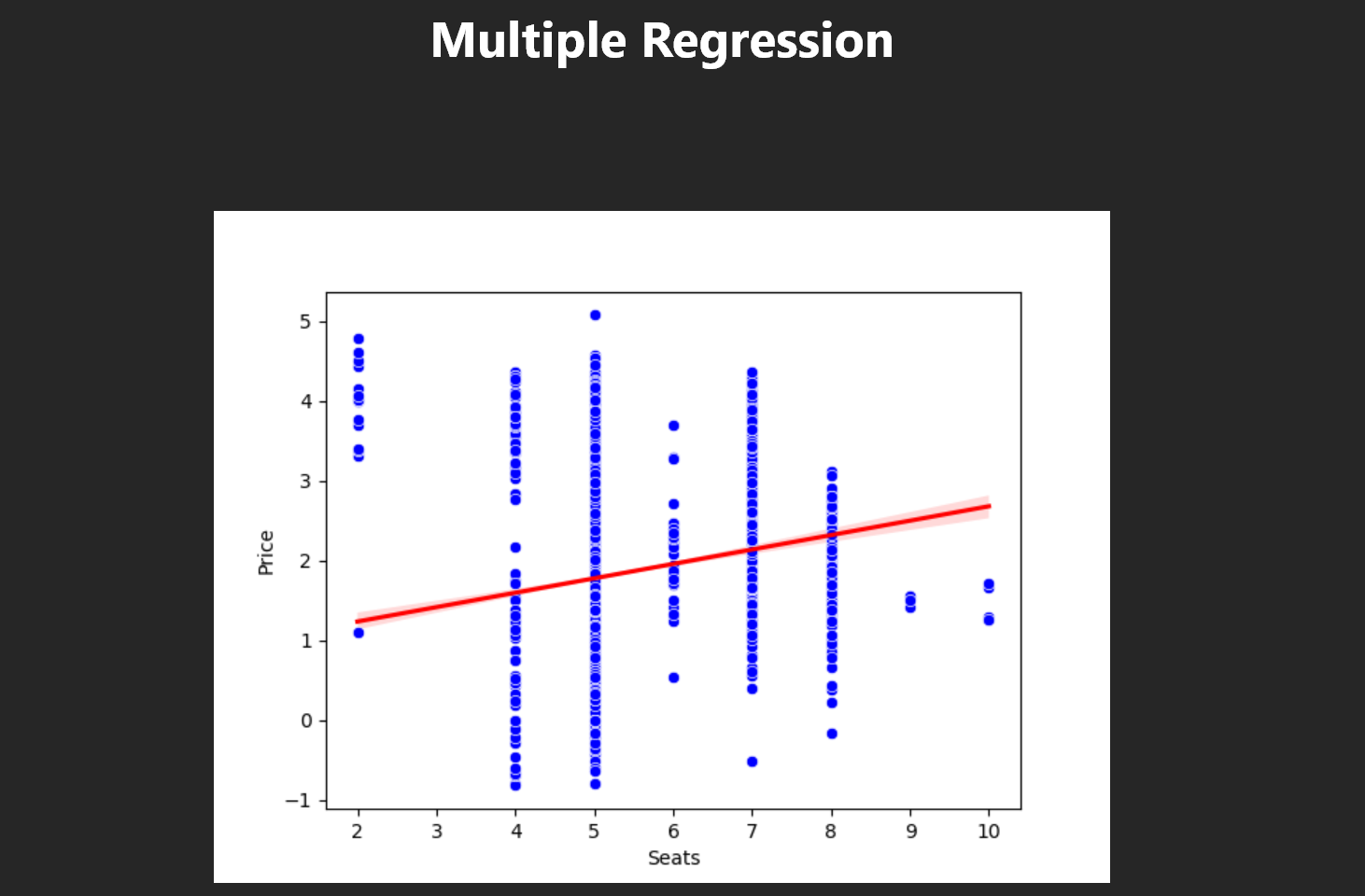
* **Whisker Chart Screen - Power Distribution by Price Range:**
  + The enclosed screenshot presents a whisker chart illustrating the distribution of power values within distinct price ranges for cars.
* **Visual Representation of Central Tendencies and Variability:**
  + The whisker chart provides a visual representation of central tendencies, variability, and potential outliers of power data grouped by specific price intervals.
* **Insights into Power Variation Across Price Categories:**
  + The chart aids in understanding how power varies across different price categories, offering insights into the relationship between car prices and their corresponding power specifications.
* **Enhanced Analysis of Power Patterns:**
  + This graphical presentation enhances the analysis of power patterns and trends within distinct price ranges in the second-hand car market.
* **Facilitates Identification of Outliers:**
  + Users can identify potential outliers and variations in power more easily through the graphical depiction, contributing to a more comprehensive understanding of the relationship between power and car prices.



* **Whisker Chart Screen - Engine Capacity Distribution by Price Range:**
  + The provided screenshot showcases a whisker chart illustrating the distribution of engine capacities within different price ranges for cars.
* **Visual Representation of Central Tendencies and Spread:**
  + The whisker chart offers a visual depiction of central tendencies, spread, and potential outliers of engine capacity data grouped by specific price intervals.
* **Insights into Engine Capacity Variation Across Price Categories:**
  + The chart aids in understanding how engine capacities vary across different price categories, providing insights into the relationship between car prices and their corresponding engine sizes.
* **Enhanced Analysis of Engine Capacity Patterns:**
  + This graphical presentation enhances the analysis of engine capacity patterns and trends within distinct price ranges in the second-hand car market.
* **Facilitates Identification of Outliers:**
  + Users can identify potential outliers and variations in engine capacity more effectively through the graphical depiction, contributing to a more comprehensive understanding of the relationship between engine capacity and car prices.



* **Whisker Chart Screen - Car Price Distribution by Price Range:**
  + The enclosed screenshot showcases a whisker chart illustrating the distribution of car prices within distinct price ranges.
* **Visual Representation of Central Tendencies and Spread:**
  + The whisker chart provides a visual representation of central tendencies, spread, and potential outliers of car prices grouped by specific price intervals.
* **Insights into Car Price Variation Across Price Categories:**
  + The chart aids in understanding how car prices vary across different price categories, offering insights into the pricing dynamics within the second-hand car market.
* **Enhanced Analysis of Price Patterns:**
  + This graphical presentation enhances the analysis of price patterns and trends within distinct price ranges, providing valuable information for market evaluation and decision-making.
* **Facilitates Identification of Outliers:**
  + Users can identify potential outliers and variations in car prices more effectively through the graphical depiction, contributing to a more comprehensive understanding of the pricing landscape in the second-hand car market.



* **Linear Regression with Average Line - Scatter Plot Iteration:**
  + In this iteration, the scatter plot introduces the number of seats in a car as a factor, plotted on the x-axis, with the y-axis representing either the original prices or predicted prices.
* **Individual Data Points:**
  + Each point on the scatter plot corresponds to an individual car within the dataset.
* **X-axis Represents Number of Seats:**
  + The x-axis portrays the number of seats in a car, offering a new perspective on how this feature influences the accuracy of price predictions.
* **Side-by-Side Comparison with Average Line:**
  + The scatter plot maintains a side-by-side comparison, enabling users to visually assess the relationship between the number of seats and the accuracy of price predictions from the linear regression model. Additionally, an average line is included for reference.
* **Enhanced Perspective:**
  + The inclusion of the number of seats as a factor, along with the average line, provides an enhanced perspective on how this particular feature impacts the accuracy of price predictions. Users can reference the average line to evaluate the model's performance relative to the dataset's overall trend.
* **Insights into Seat Count Impact:**
  + Users can leverage the visualization to gain insights into how the number of seats influences the accuracy of the linear regression model's price predictions for individual cars in the dataset. The average line offers a benchmark for comparison.

A screen shot of a graph

Description automatically generated

* **Linear Regression with Average Line - Scatter Plot Iteration: Power and Price Comparison:**
  + In this iteration, the scatter plot introduces the power of cars as a variable, plotted on the x-axis, with the y-axis representing either the original prices or predicted prices.
* **Individual Data Points:**
  + Each point on the scatter plot corresponds to an individual car within the dataset.
* **X-axis Represents Power of Cars:**
  + The x-axis illustrates the power of the cars, offering insights into the relationship between power specifications and price predictions.
* **Side-by-Side Comparison with Average Line:**
  + The scatter plot provides a side-by-side comparison, enabling users to visually assess the alignment between the actual resale values and the values predicted by the model. An average line is included for reference.
* **Evaluation of Model Performance:**
  + Users can leverage the visualization to evaluate how well the model predicts resale values based on the power of the cars.
* **Insights into Power's Impact:**
  + The graphical representation offers insights into the relationship between the power of cars and their corresponding resale values, aiding users in understanding the predictive performance of the model for cars with varying power specifications. The average line serves as a benchmark for comparison.

A screen shot of a graph

Description automatically generated

* **Linear Regression with Average Line - Scatter Plot Iteration: Engine Capacity and Price Comparison:**
  + In this iteration, the scatter plot introduces the engine capacity of cars as a variable, plotted on the x-axis, with the y-axis representing either the original prices or predicted prices.
* **Individual Data Points:**
  + Each point on the scatter plot corresponds to an individual car within the dataset.
* **X-axis Represents Engine Capacity:**
  + The x-axis illustrates the engine capacity of the cars, offering insights into the relationship between engine specifications and price predictions.
* **Comprehensive Comparison with Average Line:**
  + The scatter plot facilitates a comprehensive comparison, allowing users to visually assess the alignment between the actual resale values and the values predicted by the model. An average line is included for reference.
* **Evaluation of Model Accuracy:**
  + Users can leverage the visualization to evaluate how well the model predicts resale values based on the engine capacity of the cars.
* **Insights into Engine Capacity Impact:**
  + The graphical representation offers insights into the impact of engine capacity on the model's accuracy in predicting car resale prices, providing valuable information for users analyzing cars with varying engine specifications. The average line serves as a benchmark for comparison.

A screen shot of a graph

Description automatically generated

* **Regression Model with Linear Line - Scatter Plot Representation:**
  + This section introduces a scatter plot featuring both the original prices and the predicted prices for cars in the dataset, with a linear line representing the regression model.
* **Individual Data Points:**
  + Each point on the scatter plot corresponds to an individual car within the dataset.
* **X-axis and Y-axis Significance:**
  + The x-axis indicates the mileage of the cars, while the y-axis (Price) yhat represents either the original prices or predicted prices.
* **Direct Comparison:**
  + The scatter plot enables a direct comparison between the actual resale values (original prices) and the values predicted by the model, facilitated by the linear line.
* **Insights into Predictive Algorithm Performance:**
  + Users can derive insights into the accuracy and performance of the predictive algorithm by observing how closely the predicted values align with the actual resale values.
* **Visual Assessment of Model Predictions:**
  + The visualization offers a clear and intuitive means for users to visually assess the model's predictions in relation to the true values, aiding in the evaluation of the model's effectiveness in predicting car resale prices.

**SUMMARY OF MODEL:**

Predicted Price\_Log (statsmodels): OLS Regression Results

Dep. Variable: Price\_Log R-squared: 0.926

Model: OLS Adj. R-squared: 0.925

Method: Least Squares F-statistic: 1410.

Date: Tue, 05 Dec 2023 Prob (F-statistic): 0.00

Time: 19:33:13 Log-Likelihood: 108.23

No. Observations: 6019 AIC: -108.5

Df Residuals: 5965 BIC: 253.5

Df Model: 53

Covariance Type: nonrobust

coef std err t P>|t| [0.025 0.975]

const -123.2312 1.187 -103.814 0.000 -125.558 -120.904

Year 0.1276 0.001 103.656 0.000 0.125 0.130

Kilometers\_Driven -8.247e-08 3.49e-08 -2.362 0.018 -1.51e-07 -1.4e-08

Mileage -0.0188 0.001 -12.986 0.000 -0.022 -0.016

Engine 0.0001 1.69e-05 8.071 0.000 0.000 0.000

Power 0.0052 0.000 29.447 0.000 0.005 0.006

Seats 0.0429 0.006 6.985 0.000 0.031 0.055

Encoded\_Ahmedabad -11.1822 0.109 -102.930 0.000 -11.395 -10.969

Encoded\_Bangalore -11.0415 0.109 -101.737 0.000 -11.254 -10.829

Encoded\_Chennai -11.1663 0.107 -103.906 0.000 -11.377 -10.956

Encoded\_Coimbatore -11.1088 0.110 -100.673 0.000 -11.325 -10.892

Encoded\_Delhi -11.2635 0.108 -104.019 0.000 -11.476 -11.051

Encoded\_Hyderabad -11.0911 0.107 -103.347 0.000 -11.301 -10.881

Encoded\_Jaipur -11.2629 0.108 -104.358 0.000 -11.474 -11.051

Encoded\_Kochi -11.2334 0.110 -101.806 0.000 -11.450 -11.017

Encoded\_Kolkata -11.4085 0.108 -105.954 0.000 -11.620 -11.197

Encoded\_Mumbai -11.2319 0.108 -103.720 0.000 -11.444 -11.020

Encoded\_Pune -11.2412 0.108 -104.259 0.000 -11.453 -11.030

Encoded\_CNG -24.8921 0.241 -103.381 0.000 -25.364 -24.420

Encoded\_Diesel -24.6631 0.240 -102.728 0.000 -25.134 -24.192

Encoded\_Electric -23.7918 0.277 -86.044 0.000 -24.334 -23.250

Encoded\_LPG -24.9540 0.246 -101.291 0.000 -25.437 -24.471

Encoded\_Petrol -24.9302 0.240 -103.752 0.000 -25.401 -24.459

Encoded\_Automatic -61.5550 0.594 -103.559 0.000 -62.720 -60.390

Encoded\_Manual -61.6762 0.593 -104.062 0.000 -62.838 -60.514

Encoded\_First -30.7702 0.301 -102.277 0.000 -31.360 -30.180

Encoded\_Fourth & Above -30.7221 0.301 -102.222 0.000 -31.311 -30.133

Encoded\_Second -30.8407 0.298 -103.324 0.000 -31.426 -30.256

Encoded\_Third -30.8982 0.296 -104.398 0.000 -31.478 -30.318

Encoded\_Ambassador -4.0823 0.236 -17.278 0.000 -4.546 -3.619

Encoded\_Audi -3.7073 0.046 -81.396 0.000 -3.797 -3.618

Encoded\_BMW -3.7243 0.044 -83.737 0.000 -3.811 -3.637

Encoded\_Bentley -4.0960 0.238 -17.214 0.000 -4.562 -3.630

Encoded\_Chevrolet -4.6102 0.048 -96.403 0.000 -4.704 -4.516

Encoded\_Datsun -4.7179 0.080 -59.046 0.000 -4.875 -4.561

Encoded\_Fiat -4.5526 0.062 -73.595 0.000 -4.674 -4.431

Encoded\_Force -4.1833 0.141 -29.668 0.000 -4.460 -3.907

Encoded\_Ford -4.3567 0.047 -93.546 0.000 -4.448 -4.265

Encoded\_Honda -4.2333 0.045 -94.869 0.000 -4.321 -4.146

Encoded\_Hyundai -4.2824 0.045 -95.596 0.000 -4.370 -4.195

Encoded\_Isuzu -4.4944 0.143 -31.485 0.000 -4.774 -4.215

Encoded\_Jaguar -3.6721 0.057 -64.382 0.000 -3.784 -3.560

Encoded\_Jeep -4.0677 0.077 -53.152 0.000 -4.218 -3.918

Encoded\_Lamborghini -3.6167 0.241 -15.010 0.000 -4.089 -3.144

Encoded\_Land Rover -3.4325 0.054 -63.634 0.000 -3.538 -3.327

Encoded\_Mahindra -4.4440 0.048 -92.430 0.000 -4.538 -4.350

Encoded\_Maruti -4.2747 0.045 -95.789 0.000 -4.362 -4.187

Encoded\_Mercedes-Benz -3.6701 0.045 -81.008 0.000 -3.759 -3.581

Encoded\_Mini -3.2088 0.063 -50.586 0.000 -3.333 -3.084

Encoded\_Mitsubishi -4.0568 0.063 -64.022 0.000 -4.181 -3.933

Encoded\_Nissan -4.3276 0.051 -85.426 0.000 -4.427 -4.228

Encoded\_Porsche -3.8131 0.071 -53.529 0.000 -3.953 -3.673

Encoded\_Renault -4.3292 0.049 -88.075 0.000 -4.426 -4.233

Encoded\_Skoda -4.2151 0.046 -91.172 0.000 -4.306 -4.124

Encoded\_Smart -4.0038 0.236 -16.982 0.000 -4.466 -3.542

Encoded\_Tata -4.7641 0.048 -99.074 0.000 -4.858 -4.670

Encoded\_Toyota -4.0987 0.047 -87.399 0.000 -4.191 -4.007

Encoded\_Volkswagen -4.3056 0.047 -92.336 0.000 -4.397 -4.214

Encoded\_Volvo -3.8899 0.066 -58.592 0.000 -4.020 -3.760

==============================================================================

Omnibus: 1688.491 Durbin-Watson: 1.989

Prob(Omnibus): 0.000 Jarque-Bera (JB): 31617.206

Skew: -0.863 Prob(JB): 0.00

Kurtosis: 14.095 Cond. No. 1.19e+16

==============================================================================

# 5. Codes

Link of the complete front end (Flutter) and backend (Python) code is:

<https://github.com/Programming-Phantoms/Car_Prediction_System.git>

The provided backend Python code is a Flask application that serves as a backend for a linear regression model for predicting car prices. Here's a brief description of the code:

1. **Imports:**
   * numpy and pandas for numerical operations and data manipulation.
   * Flask for creating a web application.
   * matplotlib**.**pyplot for data visualization.
   * train\_test\_split and LinearRegression from scikit**-**learn for machine learning tasks.
   * statsmodels**.**api for statistical analysis and building an Ordinary Least Squares (OLS) regression model.
2. **Flask App Initialization:**
   * The Flask application is initialized, creating an instance of the Flask class.
3. **Dataset Loading:**
   * The code loads a dataset containing car information (carsdf.csv).
4. **Route Definition:**
   * A single route **'/'** is defined, which triggers the train\_linear\_regression function when the root URL is accessed.
5. **Linear Regression Model Training:**
   * Inside train\_linear\_regression, a preprocessed and encoded dataset is loaded (encoded.csv).
   * The script assumes 'Price\_Log' as the target variable and performs OLS regression using statsmodels.
6. **Data Visualization:**
   * The script generates visualizations such as bar charts and pie charts to show the number of cars per company and the distribution of cars by fuel type, respectively.
   * Histograms are created for various features like price log, year, kilometers driven, engine, mileage, seats, and power.
7. **Confidence Interval Calculation:**
   * A function **CI** is defined to calculate the confidence interval for predicted prices.
8. **JSON Response Preparation:**
   * The script prepares a JSON response containing various information for front end implementation in FLUTTER:
     + Linear Regression Model Results (LRM): Actual prices (y), predicted prices (yhat), model coefficients, confidence intervals, and summary information.
     + Scatter plot data (scatter\_data): Year, Kilometers Driven, Mileage, Engine, Power, and Seats.
     + Histogram data (histogram\_of\_price\_Log): Histogram values and bin edges for various features.
     + Bar chart data (bar\_chart\_data): Number of cars per company.
     + Pie chart data (pie\_chart\_data): Distribution of cars by fuel type.
9. **Flask App Execution:**
   * The Flask application is run, and the server starts to listen for incoming requests.

**Note:**

* + Ensure that the paths to CSV files are correct.
  + The application runs in debug mode (debug=True) for development purposes. Change it to debug=False in production.

This Flask app serves as an API, providing various endpoints for accessing predictions, visualizations, and model results related to second-hand car prices.

Code:

import numpy as np

from flask import Flask, jsonify

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

# Initialize the Flask app

app = Flask(\_\_name\_\_)

# Load the raw dataset containing car information

car = pd.read\_csv('carsdf.csv')

# Define a route for the root URL

@app.route('/')

def train\_linear\_regression():

import pandas as pd

from sklearn.model\_selection import train\_test\_split

import statsmodels.api as sm

# Load the preprocessed and encoded dataset

df = pd.read\_csv('encoded.csv')

# Assuming 'Price\_Log' is the target variable

X = df.drop(['Price\_Log', 'Price'], axis=1)

y = df['Price\_Log']

# Add a constant term for the intercept

X\_train\_statsmodels = sm.add\_constant(X)

model\_statsmodels = sm.OLS(y, X\_train\_statsmodels).fit()

# Ensure X\_test has the same columns as X\_train\_statsmodels

X\_test\_statsmodels = sm.add\_constant(X)

X\_test\_statsmodels = X\_test\_statsmodels[X\_train\_statsmodels.columns] # Align columns

# Predict with the statsmodels model

y\_pre = model\_statsmodels.predict(X\_test\_statsmodels)

# Count the number of cars for each company

company\_counts = car['Brand'].value\_counts()

# Plot the bar chart

plt.bar(company\_counts.index, company\_counts.values, color='skyblue')

plt.xlabel('Company')

plt.ylabel('Number of Cars')

plt.title('Number of Cars per Company')

# Save the bar chart image

plt.savefig('bar\_chart.png')

# Count the number of cars for each fuel type

fuel\_counts = car['Fuel\_Type'].value\_counts()

# Plot the pie chart

plt.pie(fuel\_counts, labels=fuel\_counts.index, autopct='%1.1f%%', startangle=90)

plt.title('Distribution of Cars by Fuel Type')

# Save the pie chart image

plt.savefig('pie\_chart.png')

plt.close()

# Load parameter data

pre\_data = pd.read\_csv('Parameters.csv')

# Confidence interval function

def CI(X, confidence\_level=0.95):

meann = np.mean(X)

stdd = np.std(X, ddof=1)

lower = meann - (0.1915 \* stdd)

upper = meann + (0.1915 \* stdd)

return lower, upper

# Calculate confidence interval for predicted prices

lower, upper = CI(y\_pre)

# Create histograms for various features

hist\_values\_Price\_Log, bin\_edges\_Price\_Log, \_ = plt.hist(df['Price\_Log'], bins='auto')

hist\_values\_Year, bin\_edges\_Year, \_ = plt.hist(df['Year'], bins='auto')

hist\_values\_Kilometers\_Driven, bin\_edges\_Kilometers\_Driven, \_ = plt.hist(df['Kilometers\_Driven'], bins='auto')

hist\_values\_Engine, bin\_edges\_Engine, \_ = plt.hist(df['Engine'], bins='auto')

hist\_values\_Mileage, bin\_edges\_Mileage, \_ = plt.hist(df['Mileage'], bins='auto')

hist\_values\_Seats, bin\_edges\_Seats, \_ = plt.hist(df['Seats'], bins='auto')

hist\_values\_Power, bin\_edges\_Power, \_ = plt.hist(df['Power'], bins='auto')

# Prepare JSON response

histogram\_data = {

'values\_Price\_Log': hist\_values\_Price\_Log.tolist(),

'bin\_edges\_Price\_Log': bin\_edges\_Price\_Log.tolist(),

'hist\_values\_Year': hist\_values\_Year.tolist(),

'bin\_edges\_Year': bin\_edges\_Year.tolist(),

'hist\_values\_Kilometers\_Driven': hist\_values\_Kilometers\_Driven.tolist(),

'bin\_edges\_Kilometers\_Driven': bin\_edges\_Kilometers\_Driven.tolist(),

'hist\_values\_Mileage': hist\_values\_Mileage.tolist(),

'bin\_edges\_Mileage': bin\_edges\_Mileage.tolist(),

'hist\_values\_Engine': hist\_values\_Engine.tolist(),

'bin\_edges\_Engine': bin\_edges\_Engine.tolist(),

'hist\_values\_Power': hist\_values\_Power.tolist(),

'bin\_edges\_Power': bin\_edges\_Power.tolist(),

'hist\_values\_Seats': hist\_values\_Seats.tolist(),

'bin\_edges\_Seats': bin\_edges\_Seats.tolist()

}

# Extract summary information from the statsmodels model

summary\_info = {

'f\_statistic': dict(

zip(model\_statsmodels.summary().tables[0].data[0], model\_statsmodels.summary().tables[0].data[1])),

'rsquared': float(model\_statsmodels.summary().tables[0].data[6][1]),

'adj\_rsquared': float(model\_statsmodels.summary().tables[0].data[7][1]),

}

# Create a dictionary with all the results

results = {

'LRM': {'y': y.tolist(), 'yhat': y\_pre.tolist(), 'coefficients\_bs': model\_statsmodels.params[1:].tolist(),

'constant\_as': model\_statsmodels.params[0].tolist(), 'summary\_of\_model': summary\_info,

'Variables': pre\_data['Variables'].tolist(), 'Parameters': pre\_data['Parameters'].tolist(),

'CI\_L': lower, 'CI\_U': upper},

'scatter\_data': {'Year': df['Year'].tolist(), 'Kilometers\_Driven': df['Kilometers\_Driven'].tolist(),

'Mileage': df['Mileage'].tolist(), 'Engine': df['Engine'].tolist(),

'Power': df['Power'].tolist(), 'Seats': df['Seats'].tolist()},

'histogram\_of\_price\_Log': {'hist': histogram\_data},

'bar\_chart\_data': {'labels': company\_counts.index.tolist(), 'values': company\_counts.values.tolist()},

'pie\_chart\_data': {'labels': fuel\_counts.index.tolist(), 'values': fuel\_counts.values.tolist()}

}

# Return results as JSON

return jsonify({'result': results})

# Run the Flask application

if \_\_name\_\_ == '\_\_main\_\_':

app.run(debug=True)

# 6. Conclusion

In conclusion, the project successfully addresses the challenge of predicting used car prices through the implementation of a robust linear regression model within a user-friendly Flutter desktop application. By leveraging OLS (Ordinary least Squares) techniques and visual representations such as pie charts, bar charts, histograms, and whisker plot. The model provides accurate and real-time price predictions based on a diverse dataset encompassing critical parameters like brand, location, year of manufacture, kilometers driven, fuel type, transmission, owner type, mileage, and engine capacity. The user interface enhances accessibility, allowing users to effortlessly input car details and receive reliable estimates. The project not only fulfills its objectives of developing an accurate predictive model and creating an intuitive interface but also contributes as a valuable tool for both buyers and sellers in the dynamic second-hand car market.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*END\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*