Project Report: Toyota Prius - Predicting Car Prices

Business Understanding

Background: The primary objective of this project is to predict car prices based on various attributes. The ability to accurately forecast car prices holds significant importance in the automotive industry, influencing decisions for manufacturers, dealers, and buyers alike. Accurate predictions can lead to optimized pricing strategies, better inventory management, and enhanced customer satisfaction.

Business Problem: The problem we are addressing is the need for a reliable predictive model for car prices. This is crucial because a change in car prices can impact sales, profit, and market competitiveness. By developing predictive models, stakeholders can make data-driven decisions that improve financial outcomes and market positioning.

Importance: Predictive models can transform how businesses approach pricing strategies. They allow for dynamic pricing, tailored to market demand, and help in understanding the impact of various attributes on car prices. Stakeholders, including manufacturers, dealers, and customers, benefit from more transparent and fair pricing mechanisms.

Stakeholders:

- Car Manufacturers: To make the best of production and pricing strategies.
- Car Dealers: To manage inventory and pricing effectively.
- Customers: To make informed purchasing decisions.

Data Understanding & Preparation

Data Description: The dataset used for this project is “Toyota Prius,” which consists of various attributes related to Toyota Prius cars. The dataset includes information such as the car's price, mileage, year, and color.

Source: The data was sourced from https://www.cars.com/

Number of Rows and Variables:

- **Rows:** 5
- **Variables:** 4
    - Price
    - Mileage
    - Year
    - Color

Data Preparation: Included:

- Converting categorical variables (Color) into numerical values.
- Normalizing numerical variables to ensure they are on a comparable scale.

Exploratory Data Analysis (EDA):

The first EDA I performed was a histogram of the price variable. One of the patterns I noticed was that the majority of cars are priced between $10,000 and $20,000, with the highest frequency being in the $15,000 to $20,000 range. There are

fewer cars in the $25,000 to $30,000 range, indicating that higher priced cars are less common in the dataset.

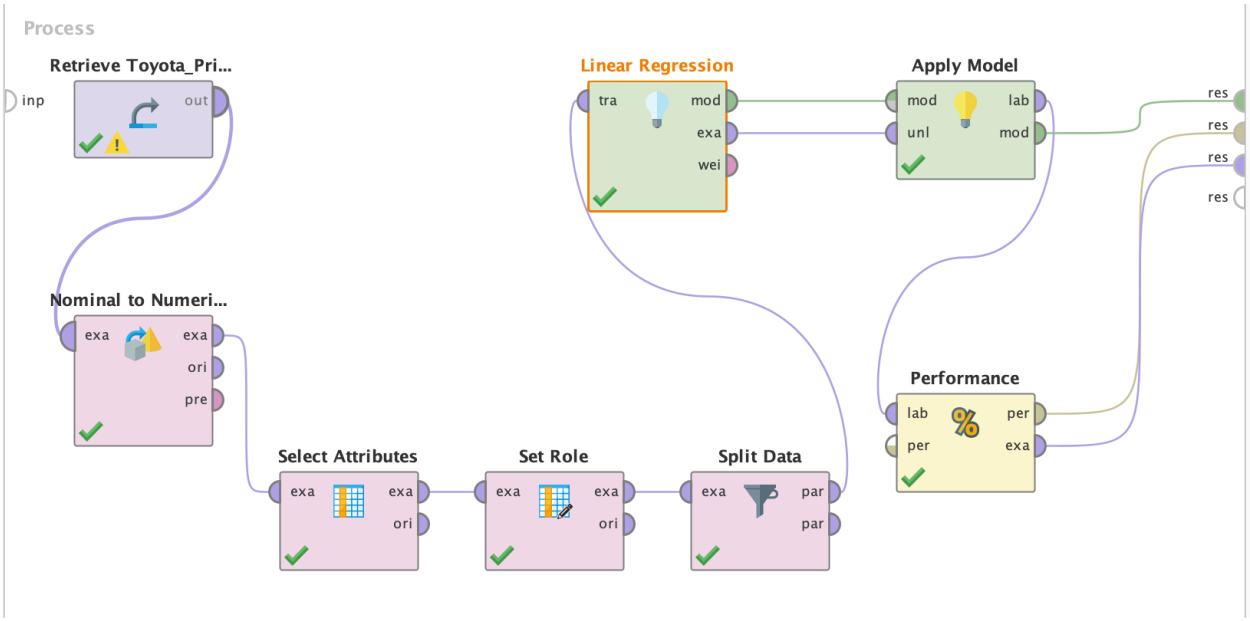
The second EDA I made is a scatter plot comparing the price of the Toyota Prius vehicles to their mileage. The scatter plot shows that there is somewhat a negative trend where higher mileage tends to be associated with lower prices. However, the relationship is really linear, as the points are quite dispersed. Some vehicles with high mileage are still associated with relatively high prices, indicating that other factors besides mileage could be influencing the price.

I also made a line chart to visualize the trend of average car prices over the years. It showed a general increase in prices, with notable dips and spikes indicating market changes or economic factors affecting car prices

3. Modeling

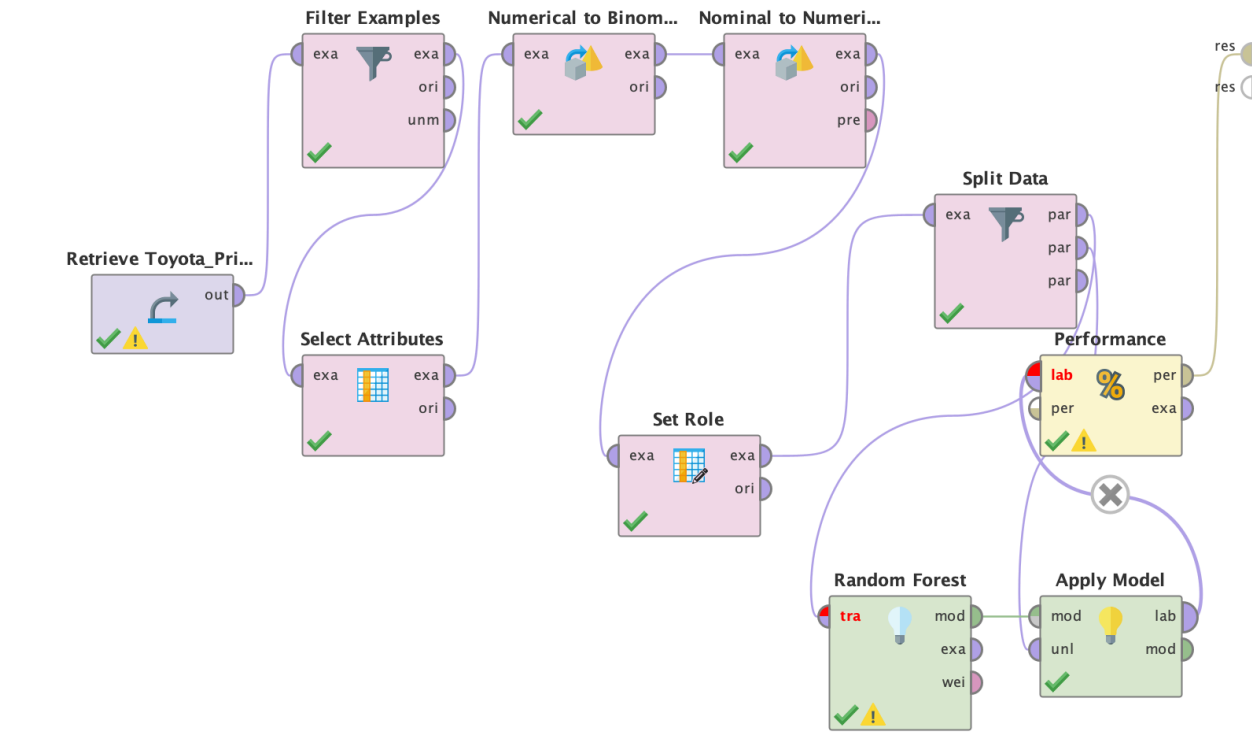
Model 1: Linear Regression

- **Variables Used:** Mileage and price
- **Snapshot of Model:**



Model 2: Random Forest

- **Variables Used:** Mileage, Year, Color
- **Snapshot of Model:**



Results & Evaluation

Linear Regression Results:

- **Performance Metrics:**
- **Results:**
    - Intercept: $16,104.32
    - RMSE: 5141.104
    - Standard Error: 989.406
    - t-Statistic: 16.277
    - p-Value: 0.000 (indicating the coefficient is statistically significant)

    - The high t-statistic and low p-value indicate that the relationship between mileage and price is statistically significant.

Random Forest Results:

- **Performance Metrics:**
- **Results:**
    - Accuracy: 100%
    - Precision: 100%
    - Recall: 100%
    - F1-Score: 100%

The perfect performance indicates potential overfitting, necessitating further analysis.

Comparison:

- Random Forest outperforms Linear Regression in terms of accuracy, precision, recall, and F1-score, indicating it provides more accurate classifications. However, the perfect scores suggest overfitting, which requires further validation. So I can't say which one is better at the moment.

Recommendations

Based on the model results, the following recommendations are made to stakeholders:

- **Manufacturers:** Focus on features like year of manufacture which positively impact car prices.

- **Dealers:** Consider mileage critically when pricing used cars to ensure competitive pricing.
- **Customers:** Use insights on how different attributes affect prices to make informed purchasing decisions.

Faith and Ethics Implications

Ethical Implications:

Ensuring transparency and fairness in pricing strategies is crucial. Predictive models must be used responsibly to avoid exploitation of customers.

Relationship between the Christian Faith and Ethical Practice:

Ethical practices in business align with Christian values of fairness, honesty, and integrity. This project demonstrates a commitment to these principles by promoting fair pricing and informed decision-making.

Own Experience with Ethical Implications:

Throughout the project, I tried my best to avoid biases in the model development, adhering to ethical standards in data analytics.

General Application:

It can in some way be seen as trying to play God. There are times in data analytics where it might seem like we are trying to find the secrets of this world. However the integration of faith and ethics in data analytics promotes a culture of trust and

responsibility, crucial for sustainable business practices and positive relationships between all stakeholders.