

Kadıköy – Tavşantepe

MACHINE LEARNING MODEL

SAUDI ARABIA USED CARS



BY SHAFA SALZABILA MEIDITA

CONTENT

01

Business
Understanding

02

Data
Understanding

03

Machine Learning
Regression Model

04

Explainable
AI

05

Summary and
Recommendation

06

Predict Using GCP



BUSINESS UNDERSTANDING

01

CONTEXT



consumers in Saudi Arabia tend to favour used cars over new ones.

Market Overview

Online Platform



Consumer Behavior



Market Dynamic

develop a predictive model based on multiple factors that can accurately anticipate used automobile prices in Saudi Arabia.

GOALS

ANALYTICAL APPROACH

Attributes:

Type, Region, Make, Gear Type, Origin, Options, Year, Engine Size, Mileage, Negotiable, and Price.

Data Collection

Data Preprocessing

Exploratory Data Analysis

Feature Engineering

Modeling

Model Deployment

- Data Cleaning:**
Handle missing values, outliers, and inconsistencies in the dataset.
- Data Transformation:**
Scaling and Encoding

- Descriptive Stats:**
Generate summary statistics.
- Visualizations:**
Use visualizations (e.g., histograms, scatter plots, and etc.)

- Feature Selection:**
Identify and select relevant features that have a significant impact on the target variable (Price).
- Feature Importance:**
Utilize techniques like correlation analysis,

1. Algorithm Choice:

Consider various machine learning algorithms

2. Model Evaluation :

Evaluate model performance using metrics & Perform validation on unseen data

- Integration:**
Integrate the predictive model into the Google Cloud Platform (GCP)
- Monitoring and Maintenance**

OUTCOME

Stakeholder

Stakeholders can improve their ability to make decisions by enabling sellers to set competitive rates

Customers

Enhance the customer experience by giving customers the information they need

Dealers

Enhance Inventory Control by Dealers and online platforms can optimize their pricing and inventory strategies by utilizing predictive analytics.



DATA UNDERSTANDING

02

DATA UNDERSTANDING

Data Dictionary

Total Data : 5624 Records

Features : Type, Region, Make, Gear
Type, Origin, Options, Year, Engine
Size, Mileage, Negotiable, Price



Data Cleaning

- Unknown Value : 61 rows
- Duplicated Data : 4 rows
- Drop Data (based on Threshold) :
2754 rows
- Rename Columns
- Switch Columns Order



Data Preparation

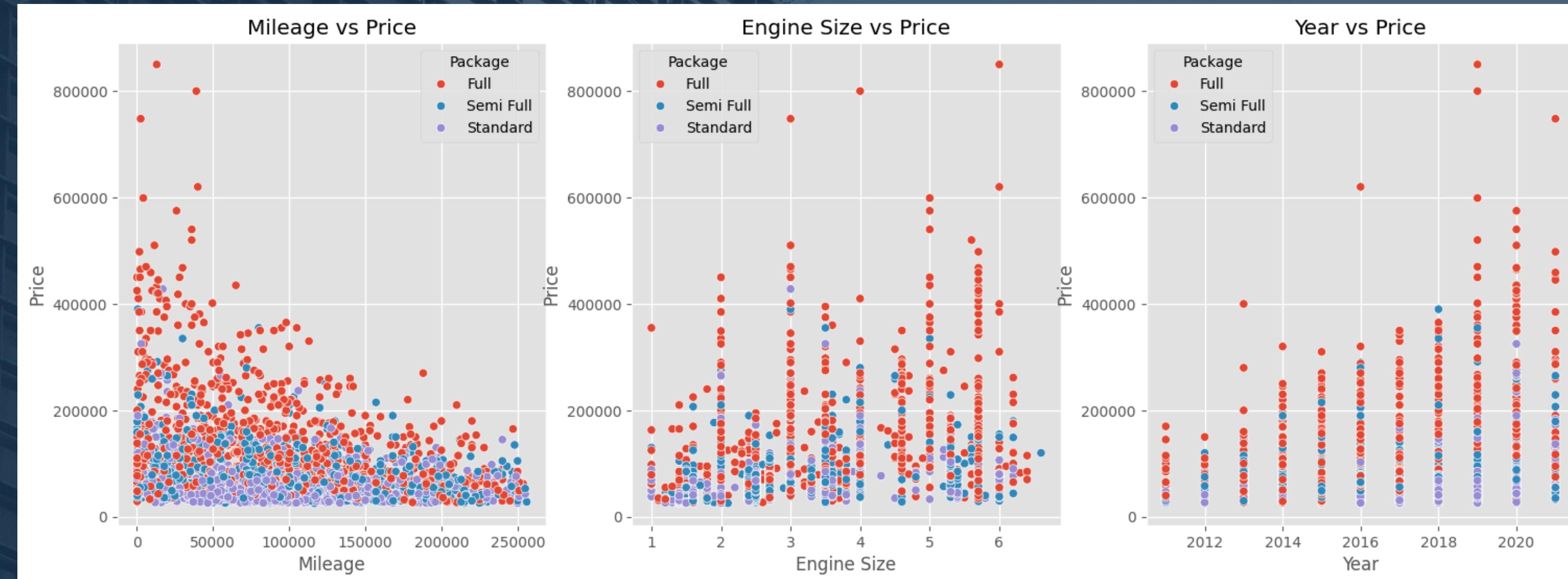
- Importing all the libraries
- Read the dataset
- Showing all informations



Exploratory Data Analysis

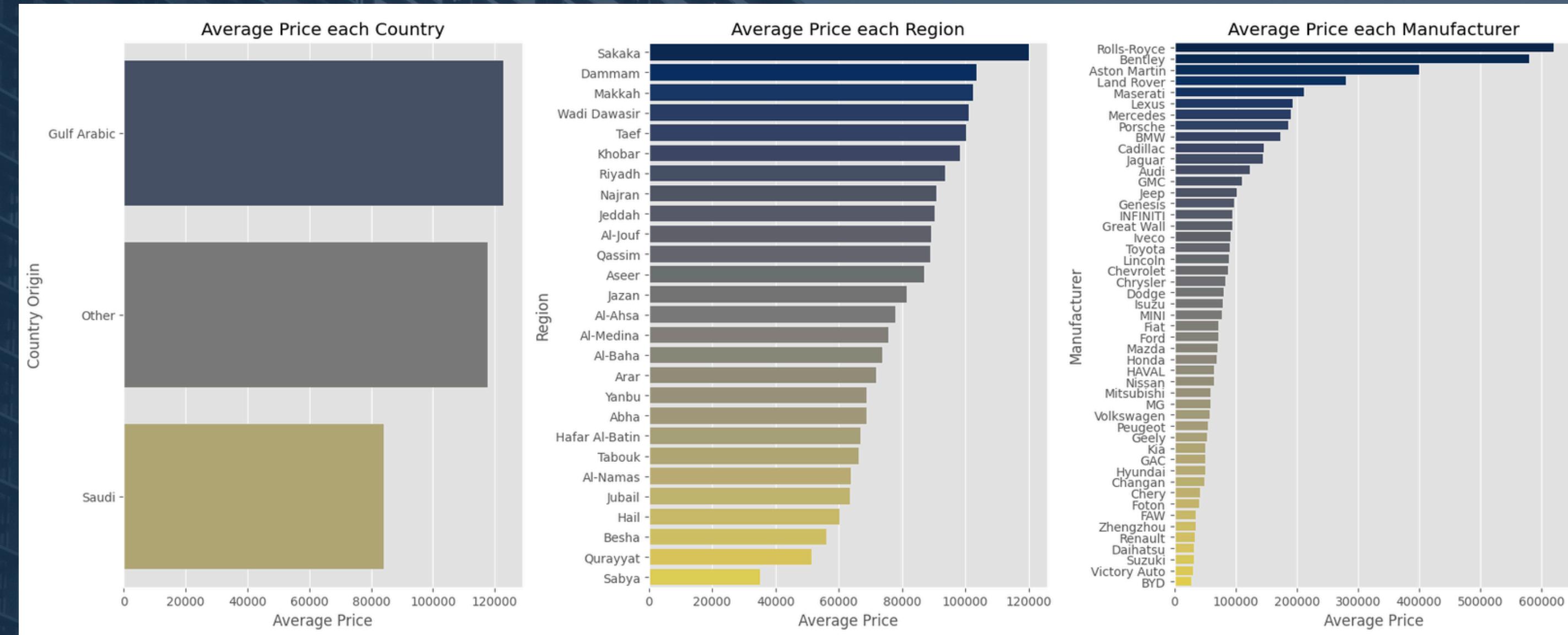
Next Slide 

PRICE DISTRIBUTION



The scatter plots provided illustrate the relationship between car prices and three variables: mileage, engine size, and year of manufacture

AVERAGE PRICE COMPARISON



These insights highlight the diversity in pricing across different origins, regions, and manufacturers, reflecting varying market dynamics and consumer preferences in the Saudi Arabian used car market.

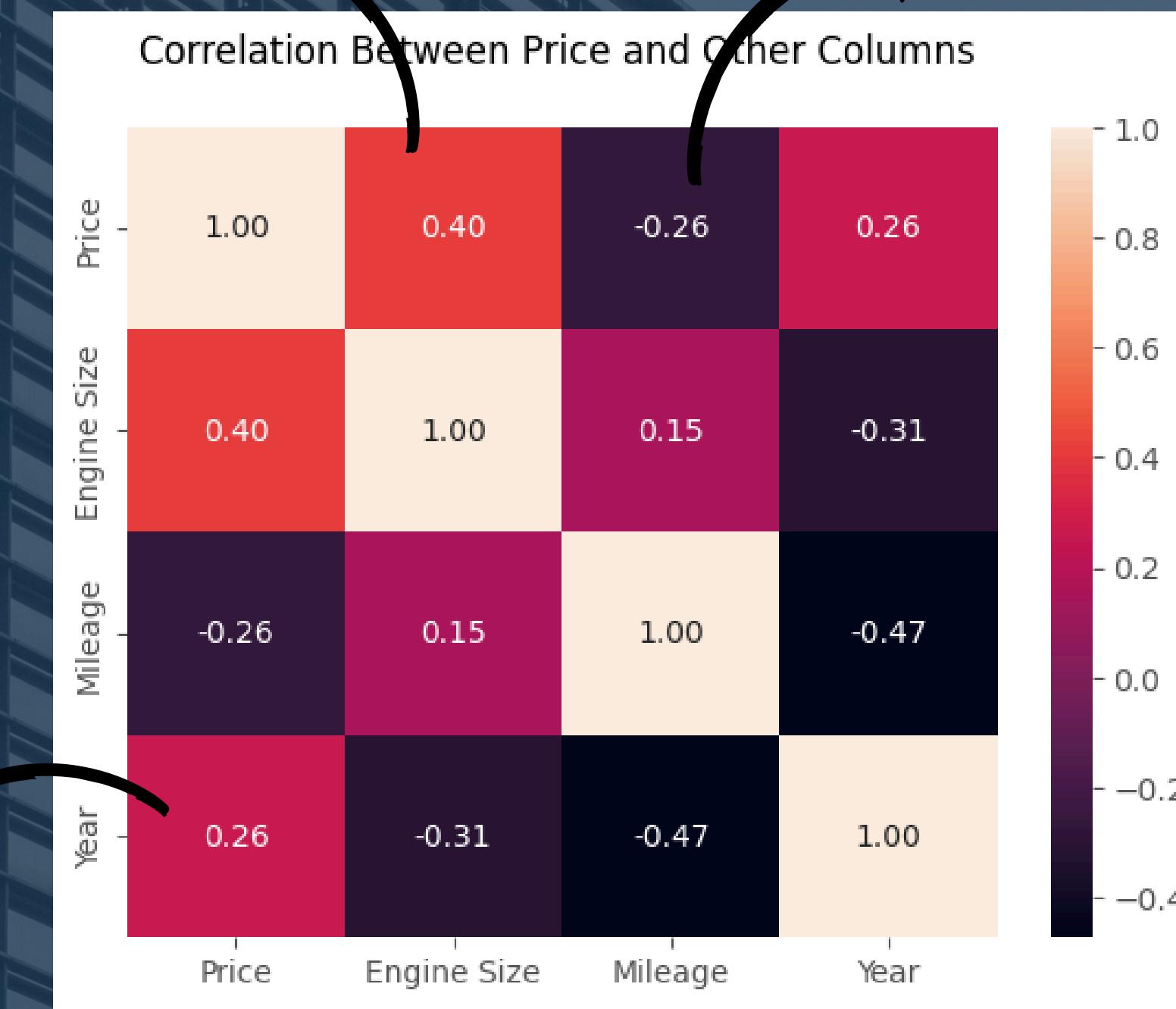
HEATMAP

Correlation
between
Price and
other
numerical
columns

moderate positive

weak negative

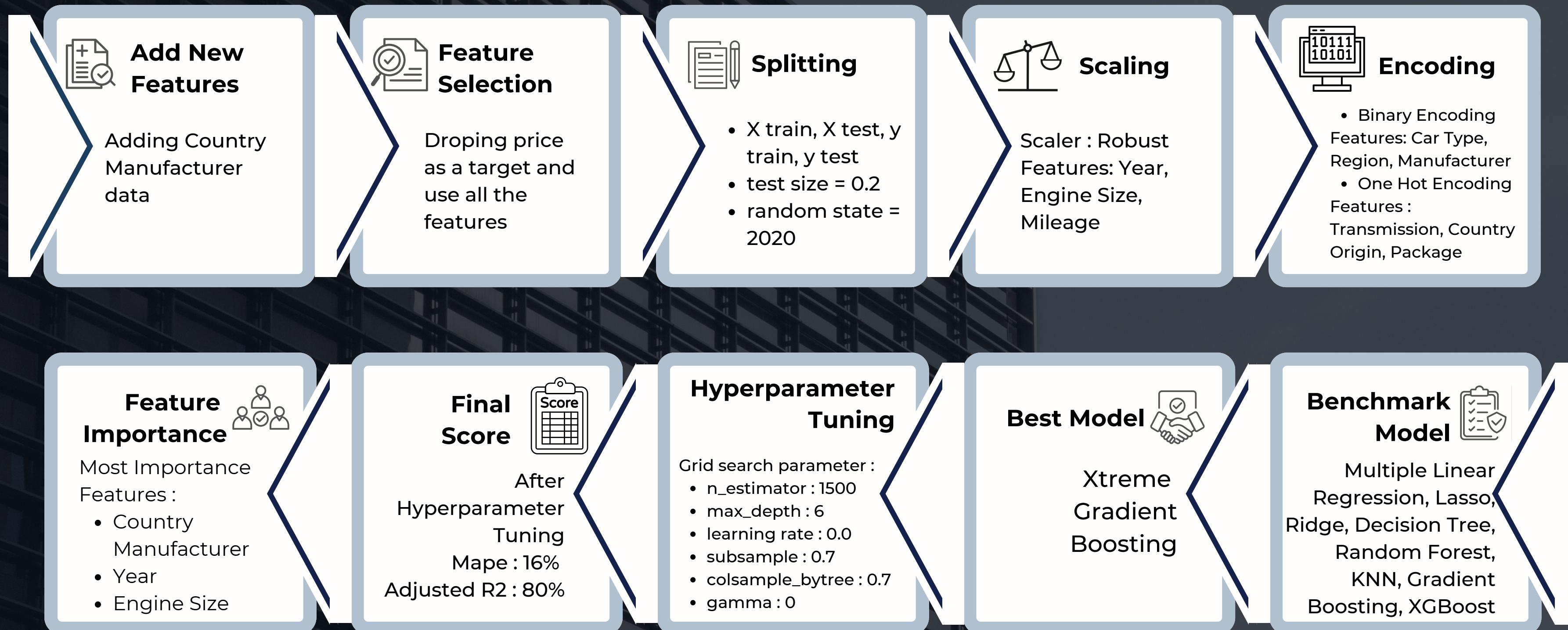
weak positive





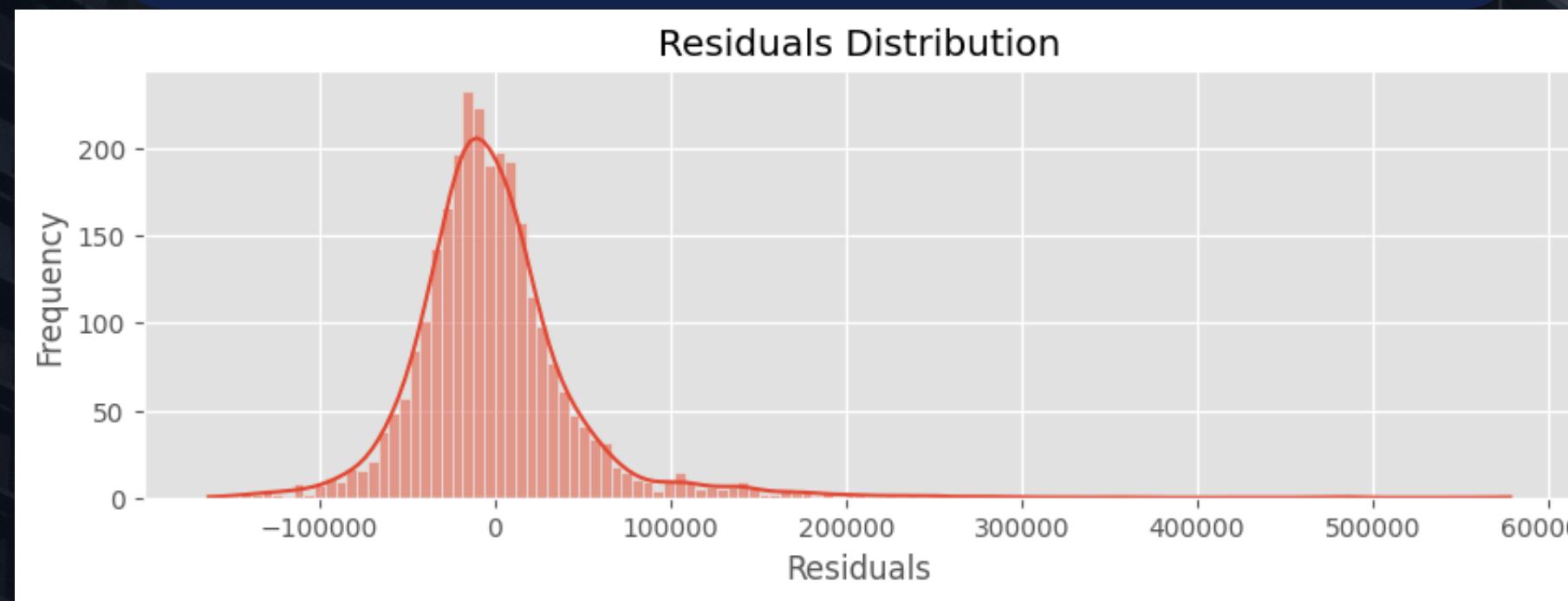
MACHINE LEARNING MODEL

MACHINE LEARNING PROCESSES



MULTIPLE LINEAR REGRESSION

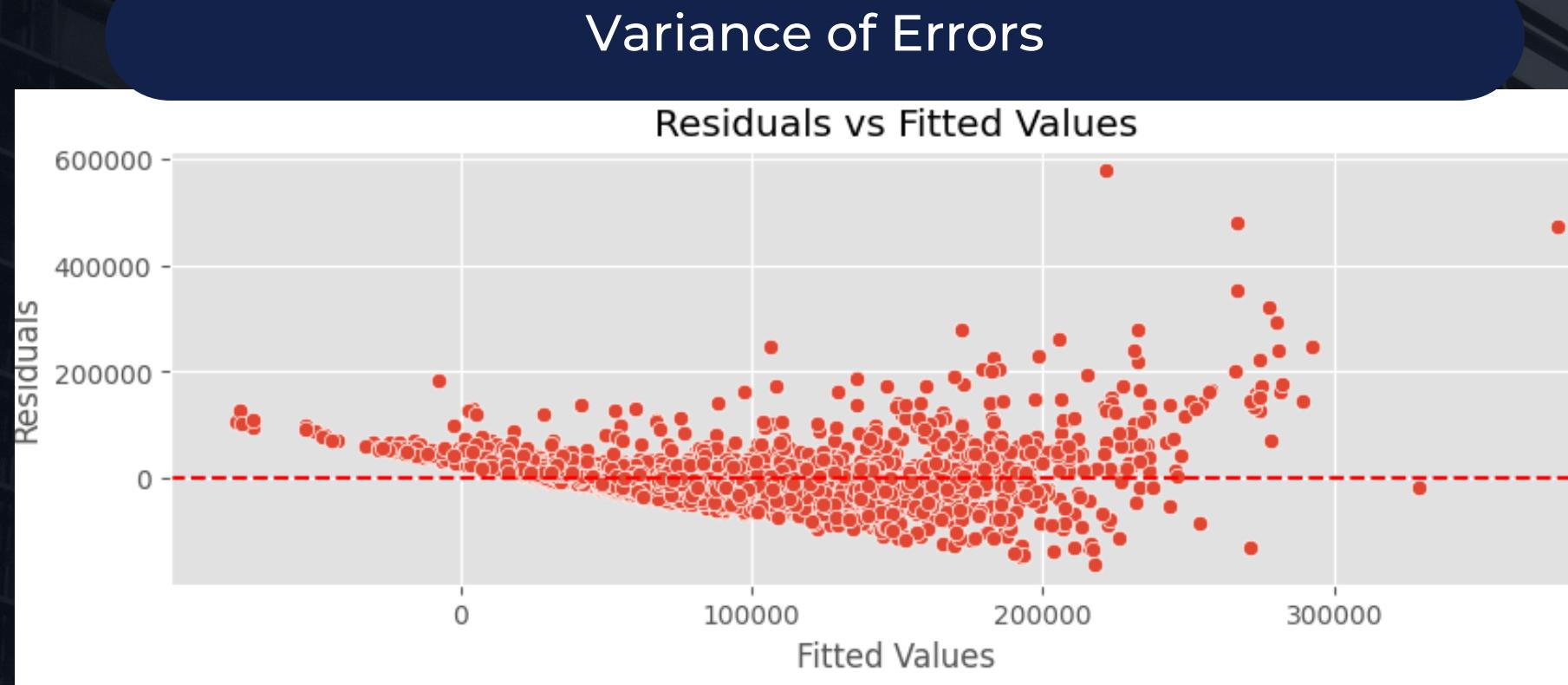
Normality Assumptions



Evaluation Metrics Score

MAPE
42 %

Adjusted R2
57%



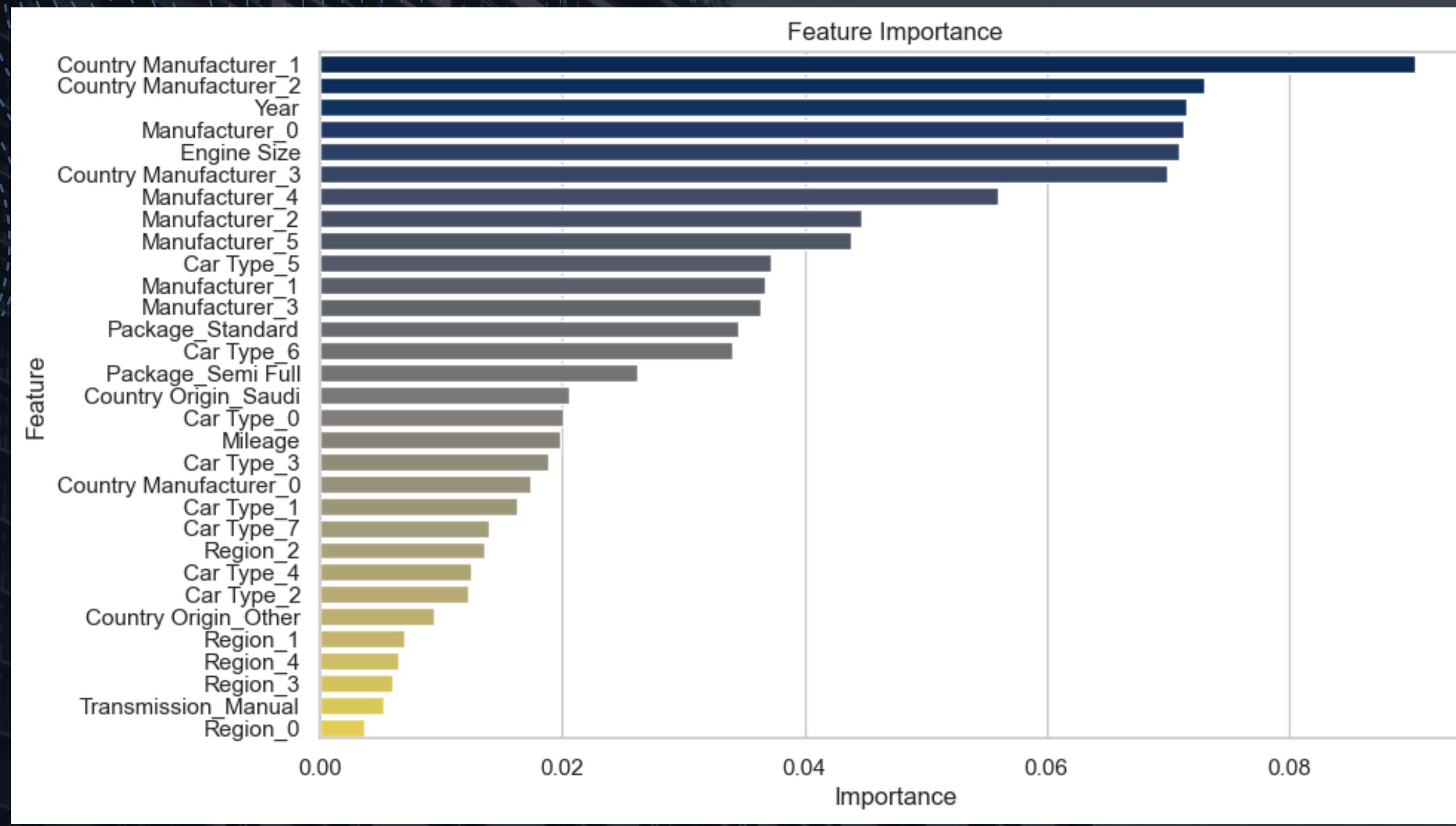
From that analysis, we don't use Multiple Linear Regression as our model, because:

1. Right Skewed
2. Heteroscedasticity
3. Suboptimal Metrics Performance

OTHER MODEL



FEATURE IMPORTANCE

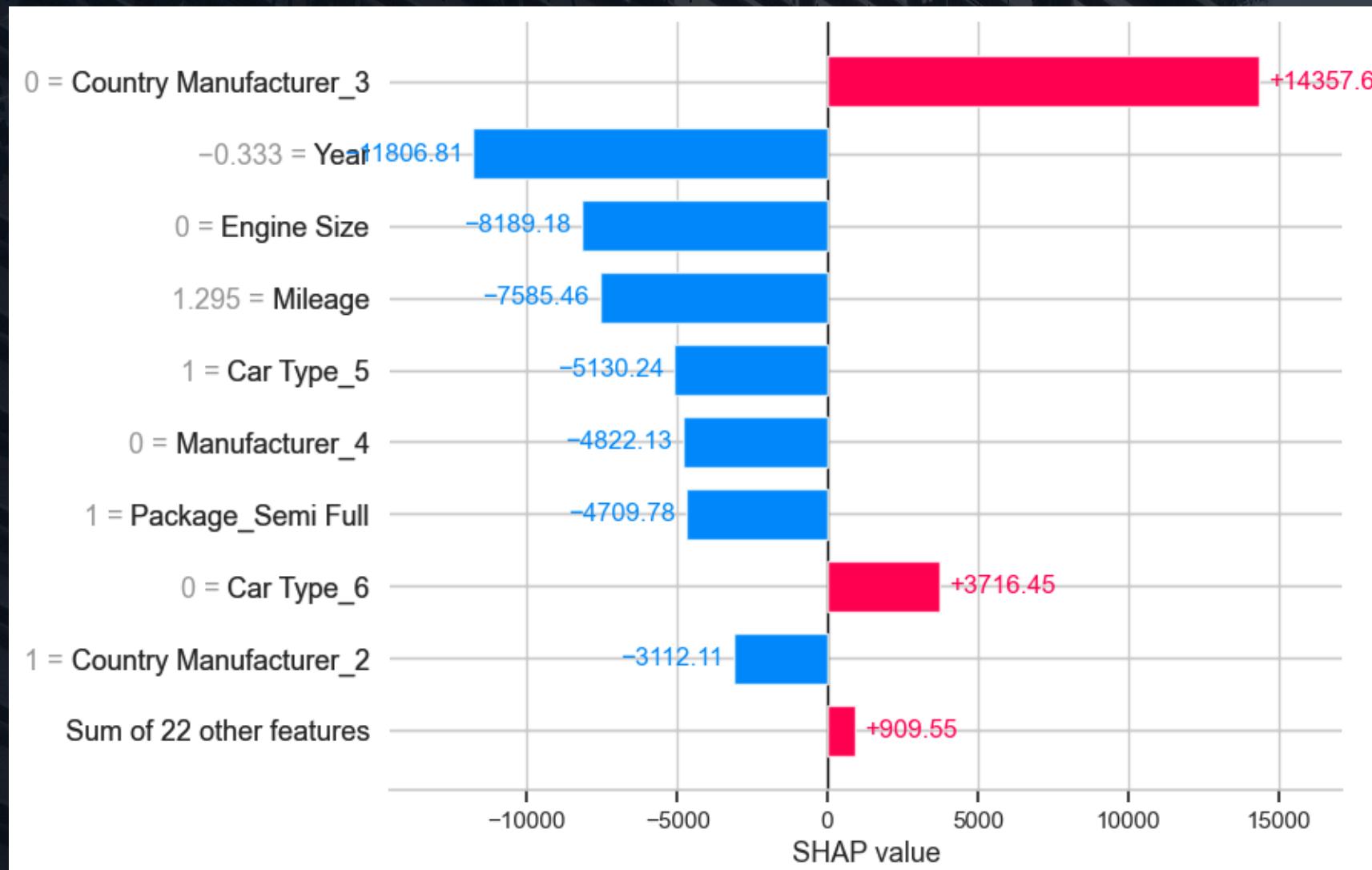


The feature importance graph for predicting used car prices in Saudi Arabia highlights that the most crucial factors are engine size, year, and manufacturer country. This indicates that a vehicle's engine size, manufacture year, and the country of the manufacturer significantly impact its price.



EXPLAINABLE AI

SHAPLEY ADDITIVE EXPLANATIONS



The SHAP graph provides a clear view of how feature contributes to the model's prediction, allowing for an understanding of feature importance and their respective influence on the target variable (car price).





SUMMARY & RECOMMENDATION

SUMMARY



Best Model

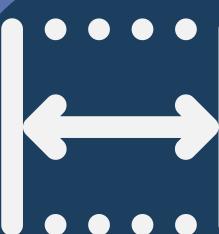
Xtreme Gradient Boosting



Best Score

Mape : 16%

Adjusted R2 : 80%



Range Best MAPE

Price : 100K - 300K SAR

MAPE : 12%

1. Enhanced Pricing Strategies
2. Market Transparency Initiatives
3. User Experience Optimization
4. Continuous Model Improvement
5. Data Governance and Compliance
6. Market Expansion and User Engagement
7. Integration and Deployment

RECOMMENDATION



PREDICT USING GCP