DATA SCIENCE
PORTFOLIO

Reza Fathurahman Sihab

About Me

Hello, my name is Reza Fathurahman Sihab. I hold a bachelor's degree in Education and have experience as a Team Core Database National at NGO World Cleanup Day Indonesia alongside engaging in various volunteer activities.

Currently, I'm eagerly and dedicatedly learning data science to enrich my knowledge and elevate my skill set. Actively participating in webinars and data science bootcamps, I am committed to advancing my expertise in this field.

Education and Training

- The Islamic State University of Jakarta Departemen of Physics Education
- Dibimbing Data Science Bootcamp

Certification

- Data Science Bootcamp
- Potential Academic Test BAPPENAS
- TOEFL ITP - 530

Working Experiences

- ***Data Science Internship***

Bukit Vista (June - July 2023)
Collecting, organizing, and optimizing data from website sources to construct a valuable structured knowledge base, involving web scraping and preprocessing for reliable data input to machine learning algorithms

- ***Project Admin Freelance***

CV. Karya Mandiri Contractor - (July 2022 – Present)
Managed project documents, resource procurement, equipment acquisition, data storage, and organization to ensure the availability of essential project information.

- ***Team Core Database National***

NGO World Cleanup Day Indonesia - (April 2020 - February 2021)
Became a coordinator and responsible for handling the database team in analyzing, checking, inputting, and controlling data nationally

Data Science Project

Classification

A Survei on The Effects of COVID-19 on The Education, Social Life and Mental Health of Students

rezafsihab1/mini_final_projects (github.com)

Clustering

Customer Segmentation

rezafsihab1/Customer-Segmentation (github.com)

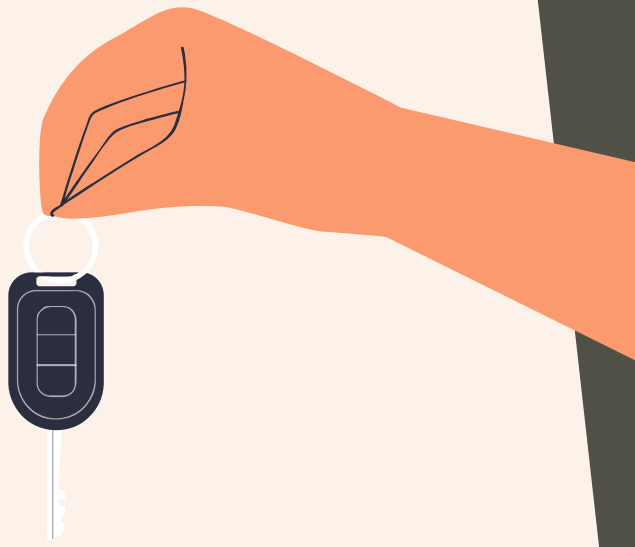
Regression

Used Car Price Prediction

rezafsihab1/car-price-prediction (github.com)

Used Car Price Prediction

By Reza Fathurahman Sihab



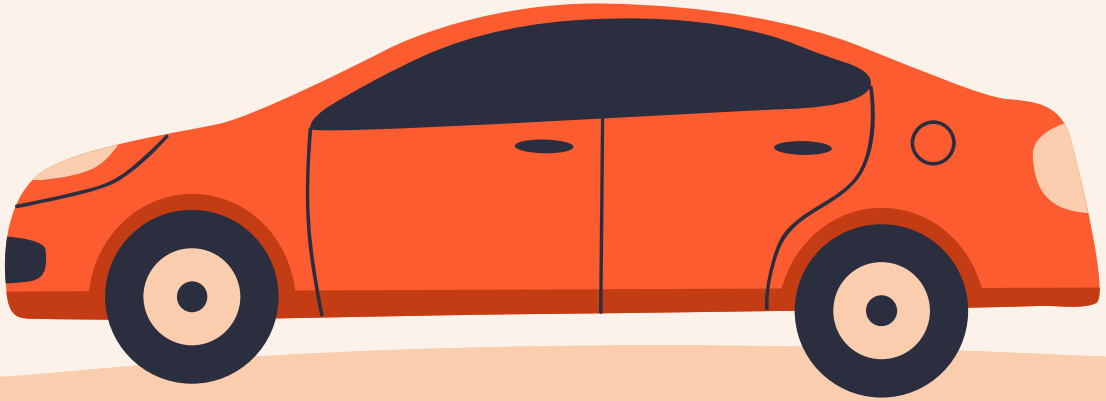
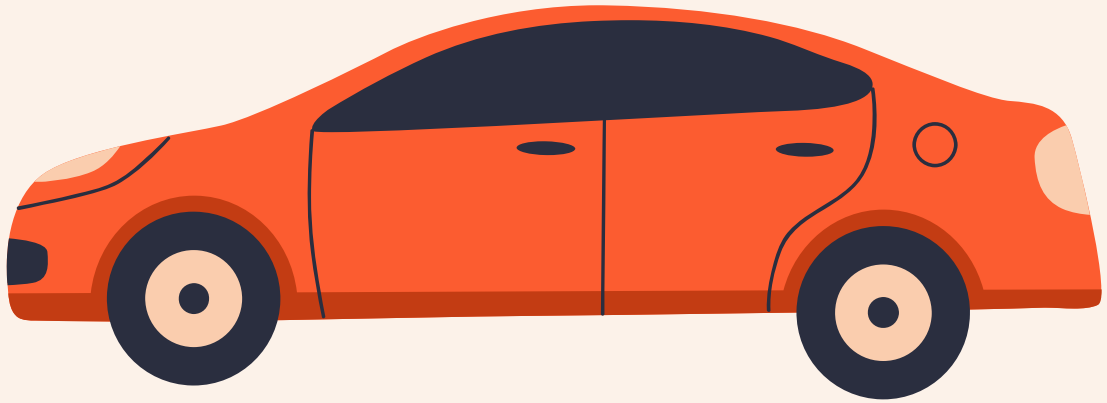


Table of Contents



Project Background

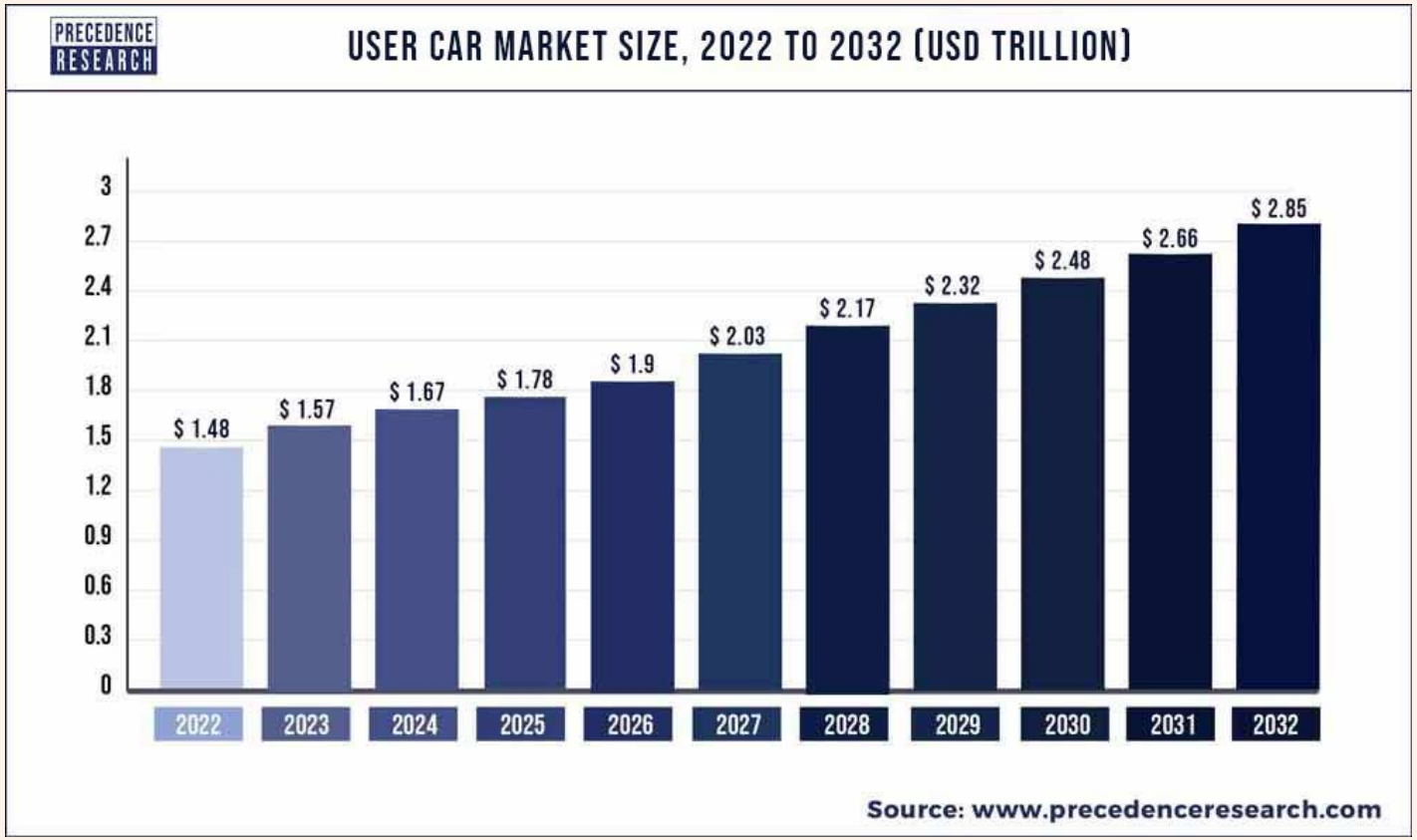
Factors driving the growth of the used car market:

01 **Rising New Car Prices**

02 **Budget Constraints for New Cars**

03 **Surge in Used Car Sales**

04 **Affordability as a Key Driver**



The total value of all used cars sold and purchased worldwide is anticipated to experience a remarkable average annual growth rate of 6.80% from the year 2022 to 2032.

Business Problem and Objectives

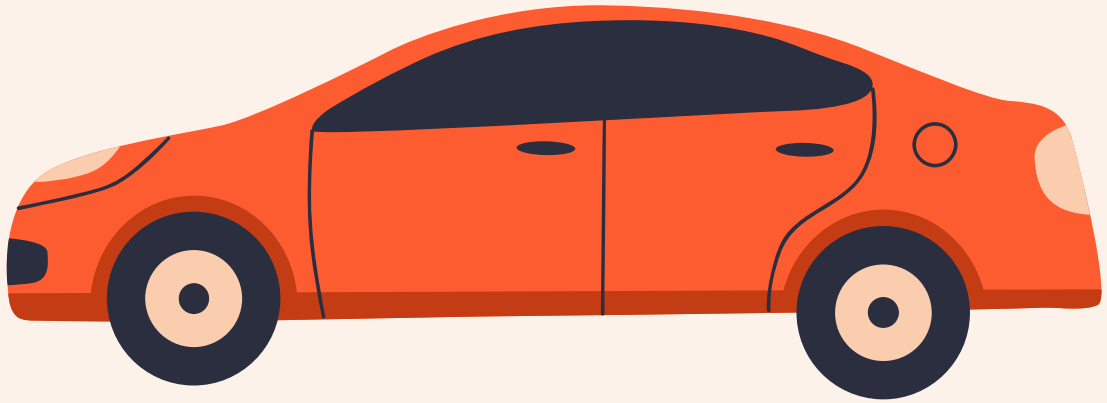
Business Problem

Uncertainty in Used Car Values

Objectives

To develop an advanced predictive model for used car prices, for a wide range of stakeholders within the used car ecosystem, including:

1. Used car sellers (dealers)
2. Online pricing services
3. Individuals



Data Information

Data about used car from 1962 - 2024

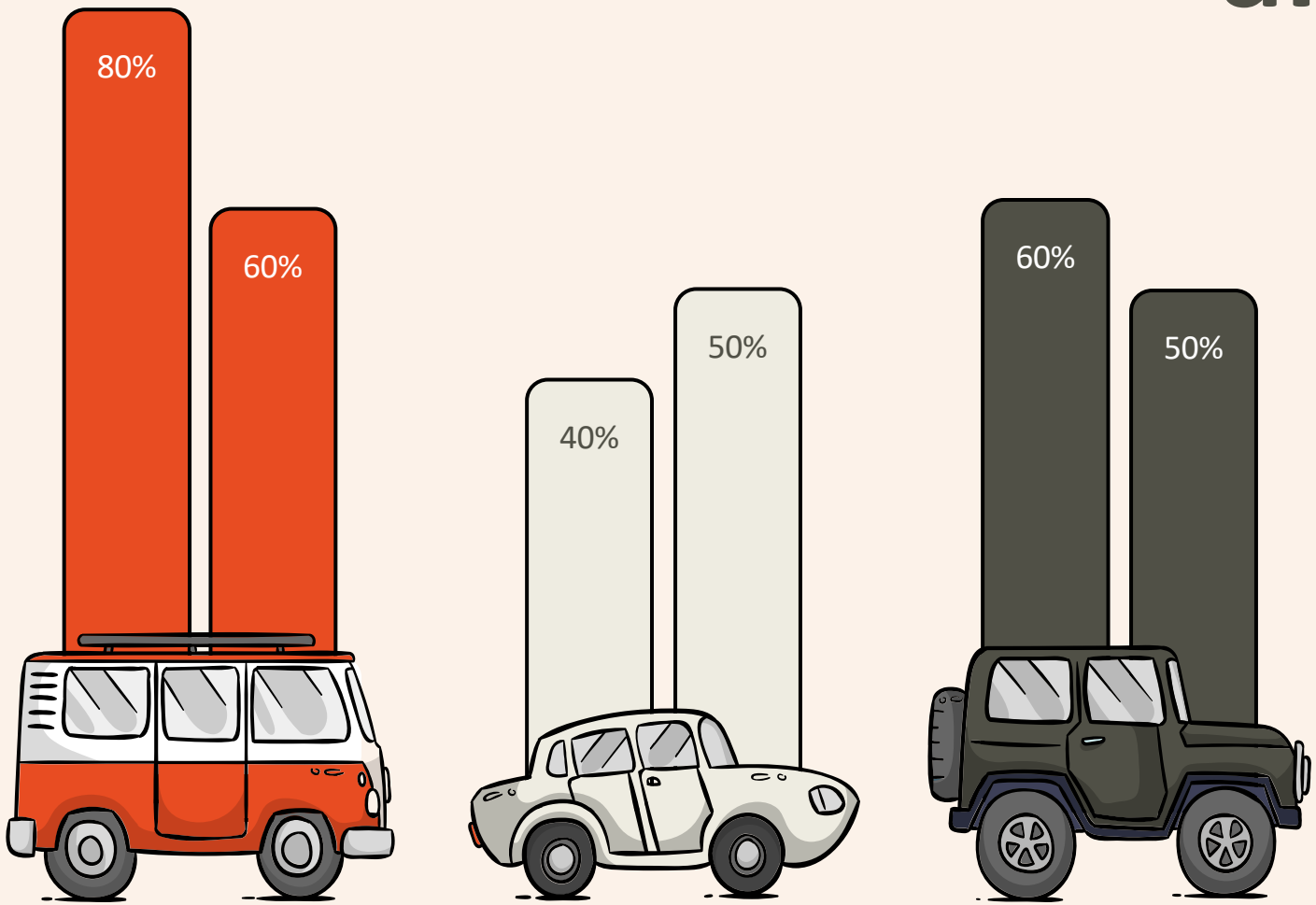
Cleaned dataset:
19109 rows
39 columns

Categorical values
brand
model
engine
transmission
fuel_type
drivetrain
interior_color
exterior_color

Numerical values
year
mileage
engine_size
min_mpg
max_mpg
price

Binary values	
automatic_transmission	remote_start
damaged	sunroof/moonroof
first_owner	automatic_emergency_braking
personal_using	stability_control
turbo	leather_seats
alloy_wheels	memory_seat
adaptive_cruise_control	third_row_seating
navigation_system	apple_car_play/android_auto
power_liftgate	bluetooth
backup_camera	usb_port
keyless_start	

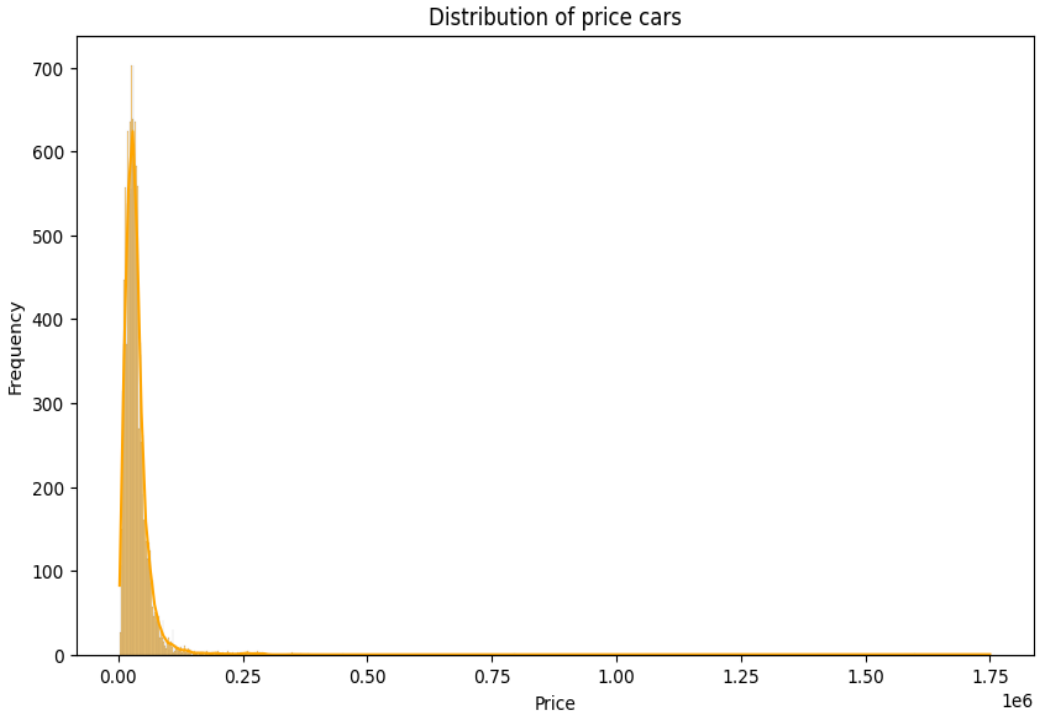
Explanatory Data Analysis and Visualisation

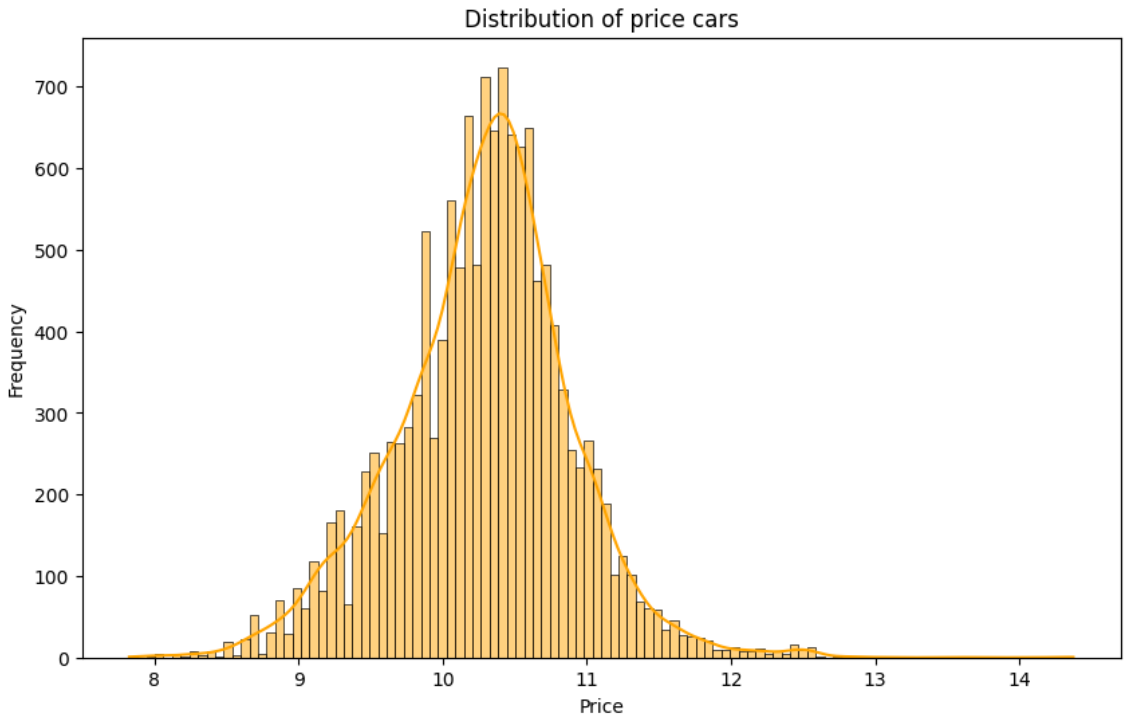


* While exploring the data, we'll look at the different combinations of features with the help of visuals.

An Examination of Price Trend

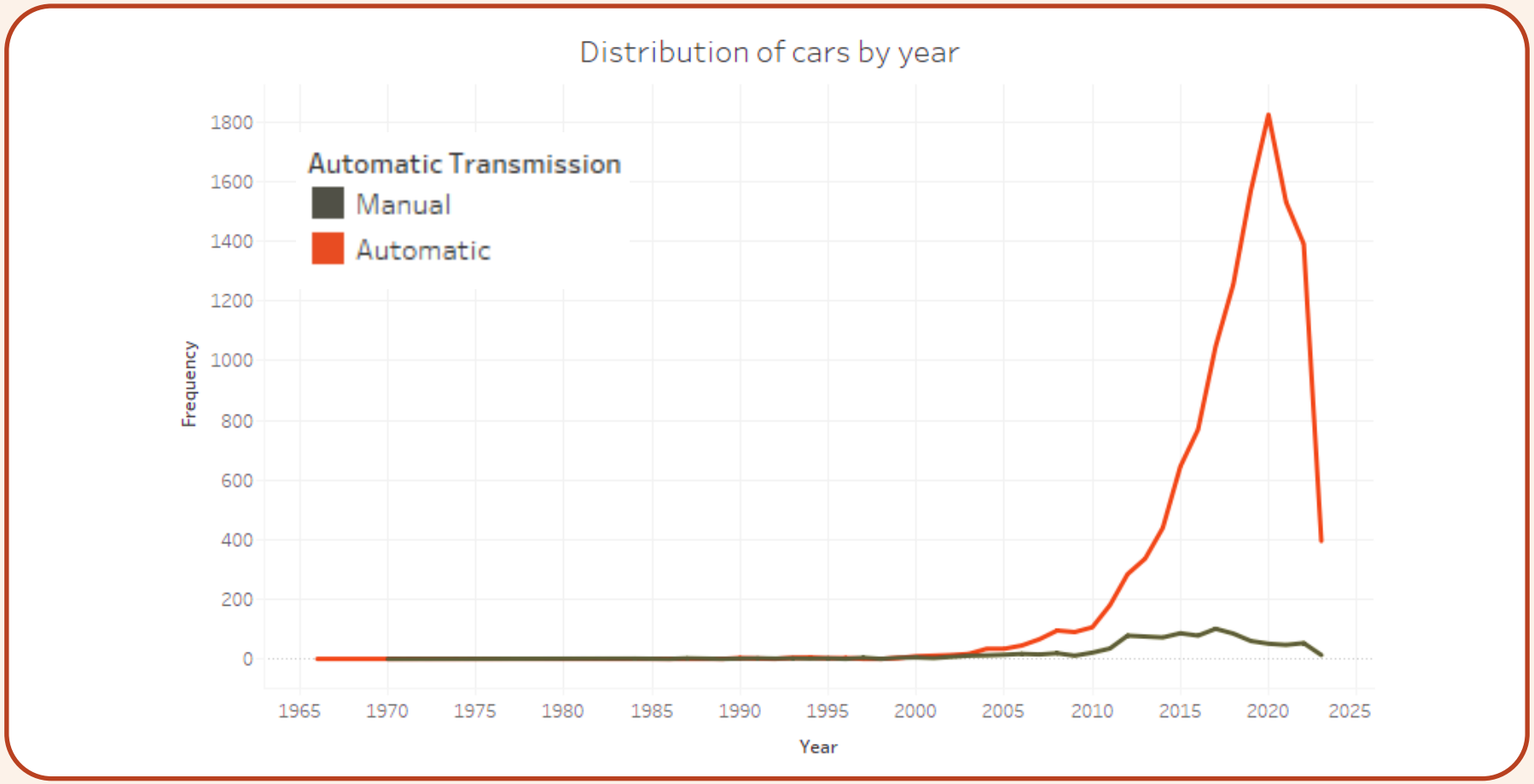
* Before applying any models, taking a look at price data may give us some ideas.





Most of the used cars are less than $20,000. In addition, we see that there are still considerable number of cars that is over $20k price.

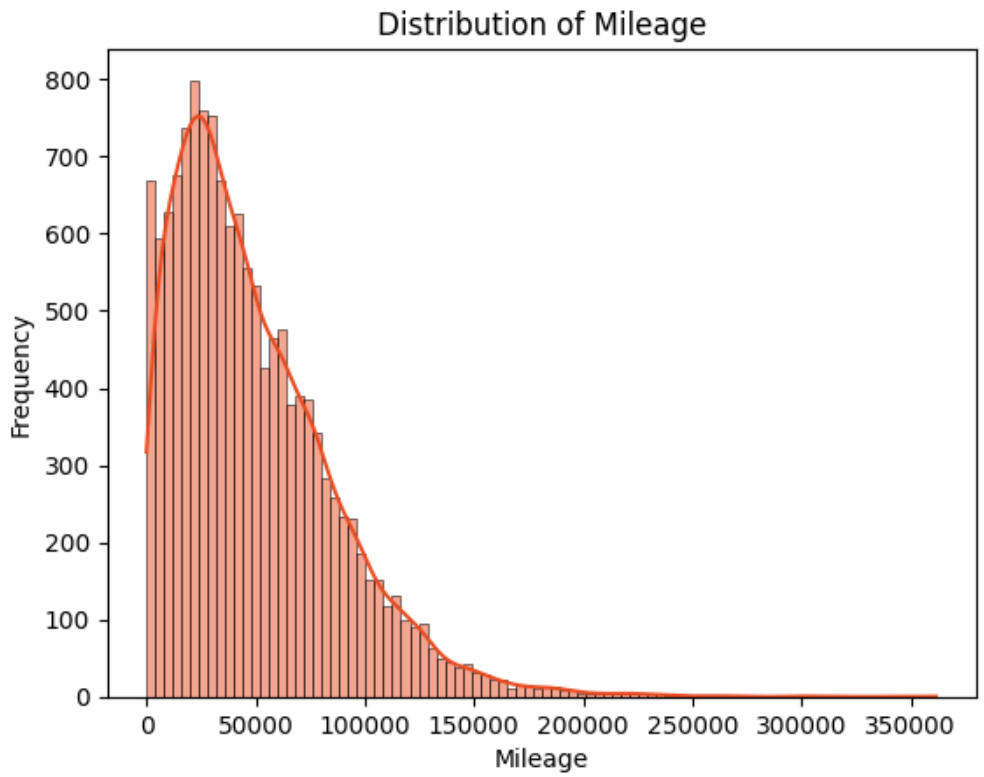
Used Car Quantity Trend by Production Year

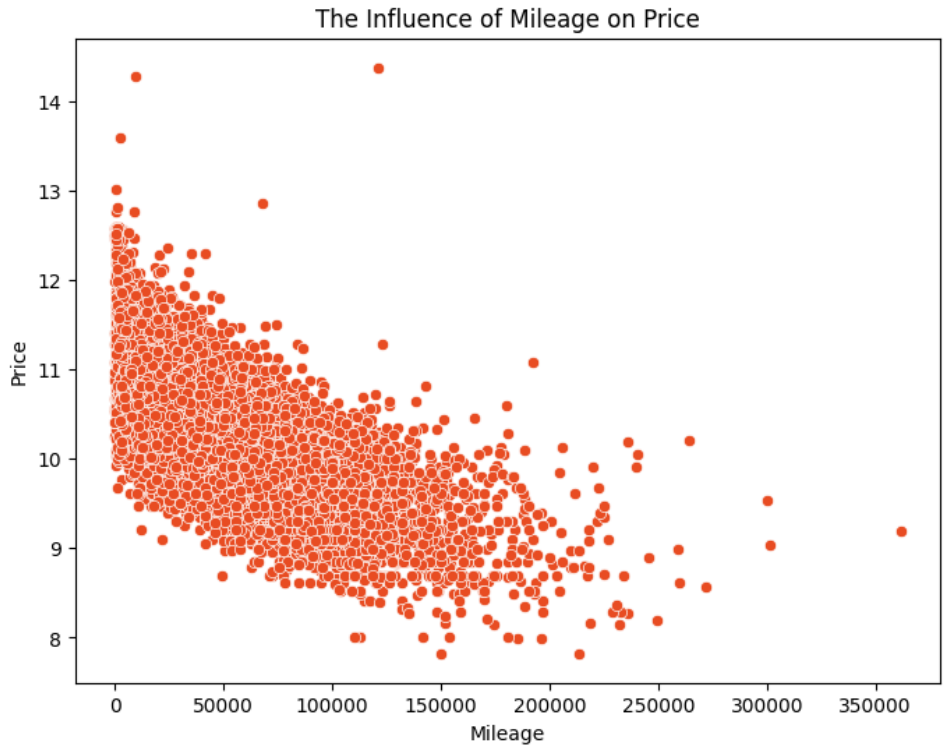


Based on the trend of car production years, online sales are predominantly dominated by automatic transmission vehicles, with the majority of cars falling within the production years of 2010 to 2023.

Other Popular Features of Used Cars

* Therefore, it is hard to make a strong estimate of a price of a car just by considering the type or condition of a car. But we can tell it certain condition cars are popular and higher chance to be sold.

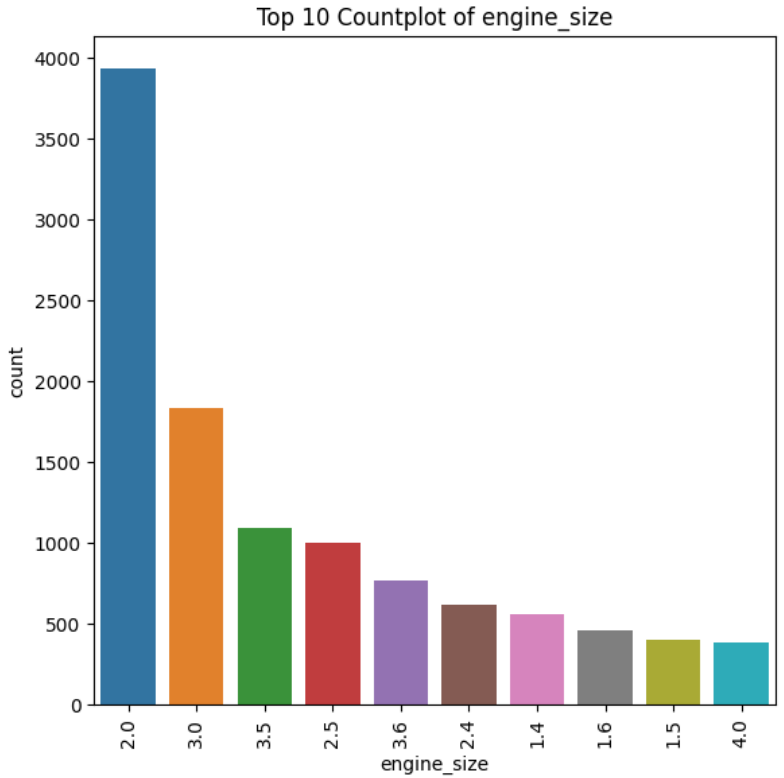


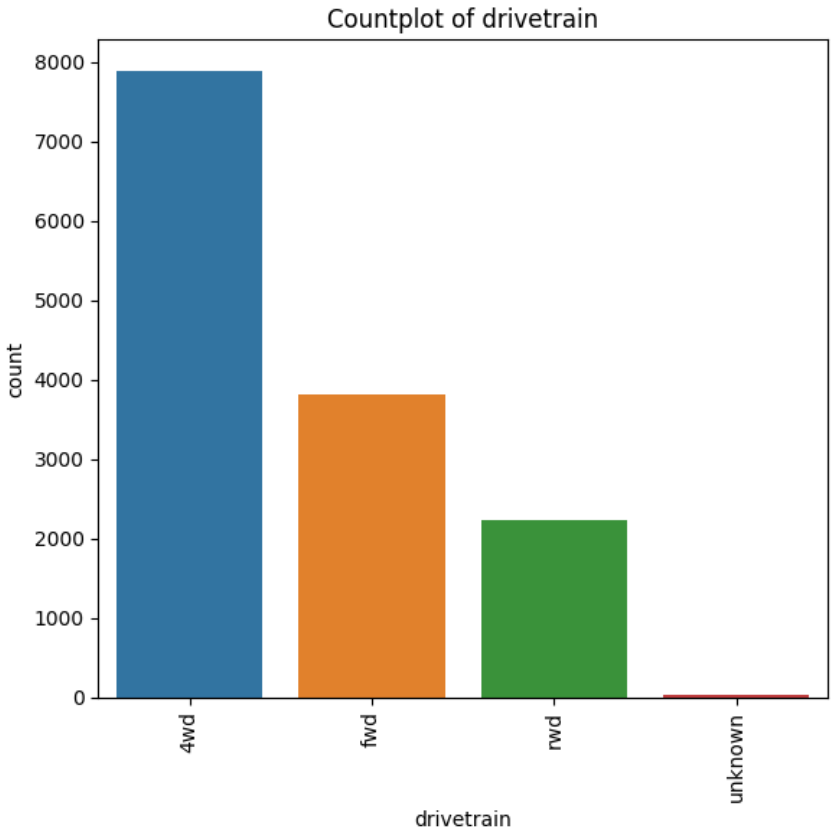


Most Popular used cars are the ones that has mileage around 25k

Vehicles with low mileage often command higher prices

Other Popular Features of Used Cars





The most prevalent vehicles in the market are those with engine sizes ranging from 2 to 3.5 liters, falling into the sedan, SUV, or crossover categories.

4wd cars are the most popular in terms of numbers. In the long run, they can keep their ability to rum better compared to rwd and fwd drive train.

Machine Learning

Data Preprocessing

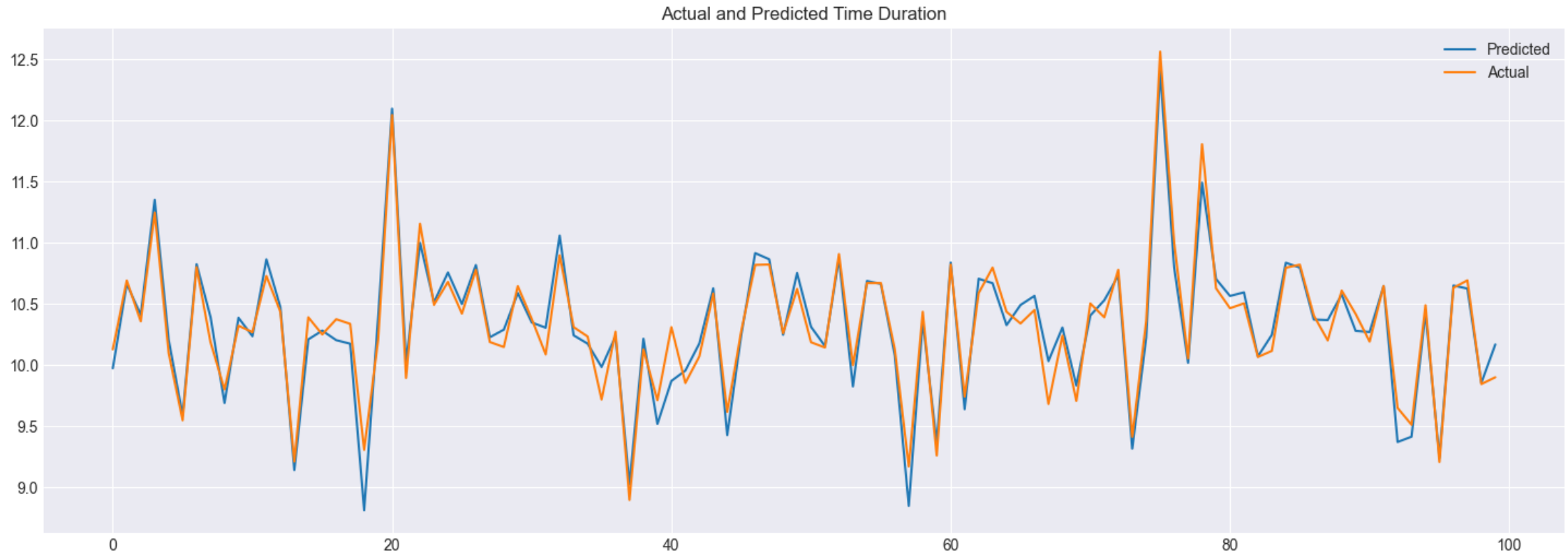
1. **Handling missing values**
2. **Removing duplicated values**
3. **Data cleaning and feature engineering**
    a. Change string to lowercase
    b. Convert target to numerical value
    c. Classify the brands based on their production countries
    d. Cleaning columns with similar meanings
    e. Grouping colors
4. **Handling outliers**
5. **Encoding**

    One-hot encoding

Regression Model Result

Model		R2 Score		RMSE		MAE	
		Train	Test	Train	Test	Train	Test
Baseline	Linear Regressor	0.22	0.40	31436	18704	12034	11174
	Gradient Boosting	0.77	0.61	16784	15047	7193	7399
	XGB Regressor	0.97	0.69	5131	13462	3428	6071
	LGBM Regressor	0.71	0.73	19134	12426	5575	6381
Model Improvements	Linear Regressor	0.84	0.84	0.24	0.23	0.16	0.16
	Gradient Boosting	0.90	0.88	0.18	0.20	0.13	0.14
	XGB Regressor	0.97	0.92	0.09	0.16	0.07	0.11
	LGBM Regressor	0.94	0.91	0.14	0.17	0.10	0.12
Hyperparameter Tuning	**LGBM Regressor**	**0.96**	**0.92**	**0.11**	**0.16**	**0.08**	**0.11**

An Evaluation of LGBM Regression Model Results on Test Data



Model	R2 Score		RMSE		MAE	
	Train	Test	Train	Test	Train	Test
LGBM Regressor + Hyperparameter tuning	0.96	0.92	0.11	0.16	0.08	0.11

Feature Importance

Color

Bold red exterior matches stylish interior palette.

50%

Features

Car showcases advanced features for enhanced driving experience

70%

Mileage

Low mileage signals well-maintained condition and longevity.

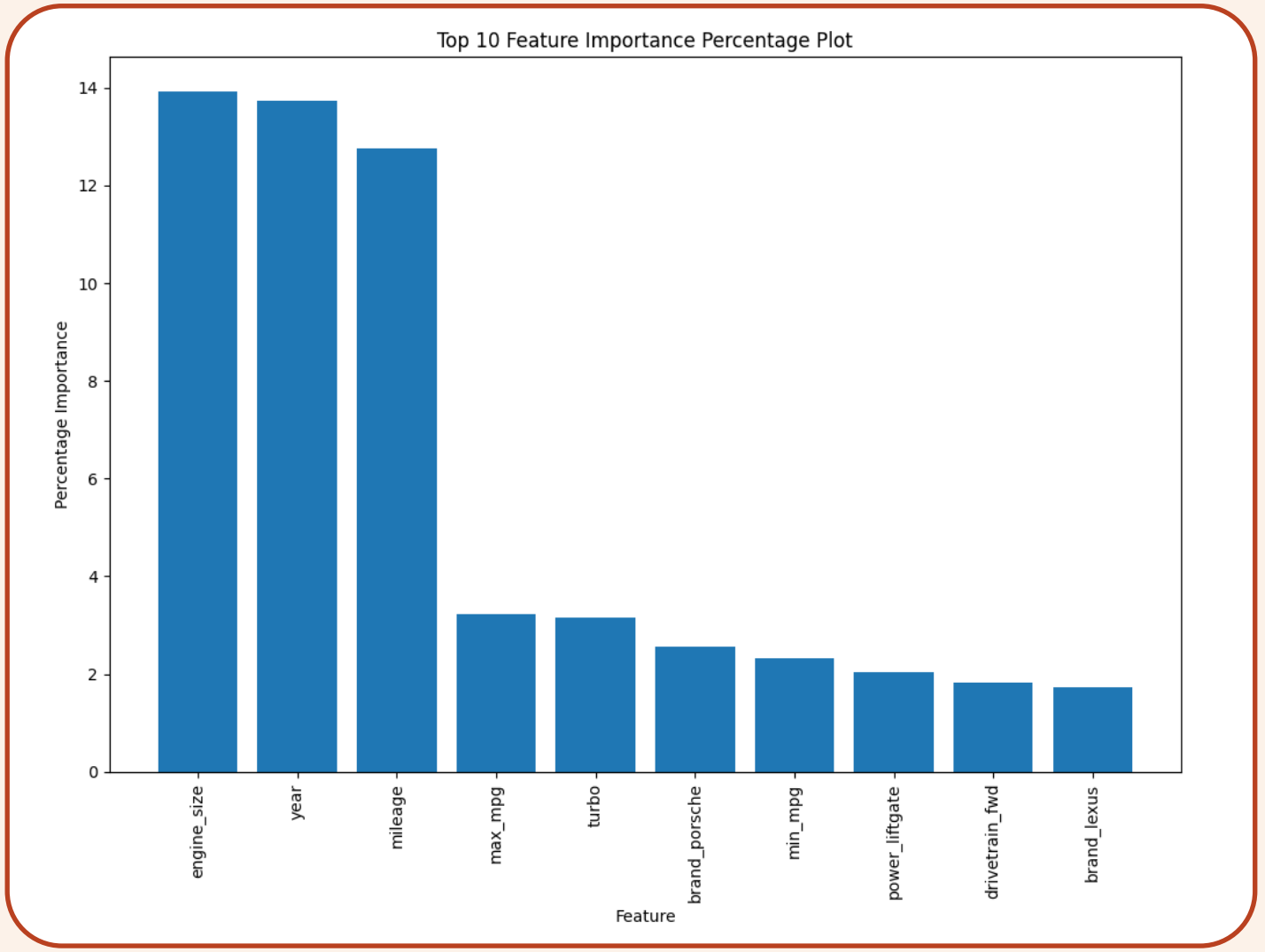
60%

Engine

Robust engine size ensures impressive performance capability.

80%

Feature Importance



Color

Bold red exterior matches stylish interior palette.

50%

Mileage

Low mileage signals well-maintained condition and longevity.

60%

Features

Car showcases advanced features for enhanced driving experience

70%

Engine

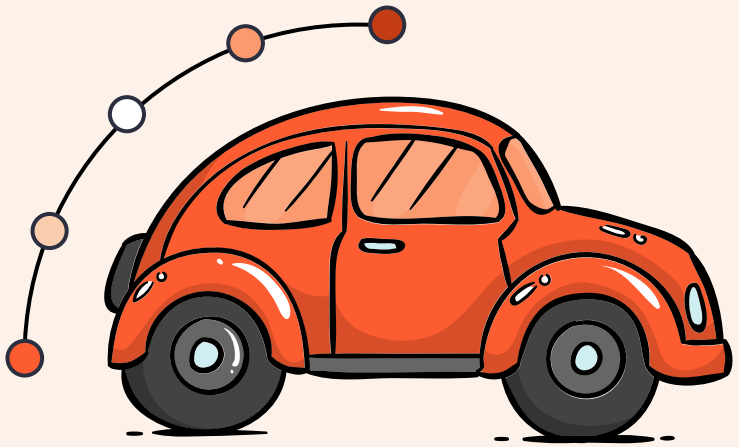
Robust engine size ensures impressive performance capability.

80%

Conclusion and Recommendation

Conclusion

- Engine size, production year, and mileage are the most influential factors in Used Car Price Prediction.
- The XGB Regressor model excels in predicting used car prices with excellent performance, with the highest R2 scores and the lowest RMSE and MAE values.
- The LGBM Regressor model is also a strong alternative with a balanced performance and execution time.



"What is the most effective strategy to determine the accurate potential pricing for used cars?“

Recommendations:

1. Enhance Price Prediction Accuracy:

Utilize the XGB Regressor/ LGBM model for more accurate used car price predictions due to its superior R2 score and lower RMSE and MAE values.

2. Leverage Key Features:

Emphasize pivotal features—year, engine size, and mileage—in marketing efforts to capitalize on their significant influence on pricing.

3. Optimize Product Portfolio:

Tailor your inventory based on insights from influential features, targeting vehicles with characteristics favored by the market.

Thank You

Credits:

- Kaggle
- Slidesgo
- Freepik

Source:

Dataset:

https://www.kaggle.com/datasets/tugberkkaran/used-car-listings-features-and-prices-carscom

Contact Me:

- Whatsapp: https://wa.me/6289627904468
- E-mail: Rezafsihab98@gmail.com
- LinkedIn: https://www.linkedin.com/in/rezafsihab/