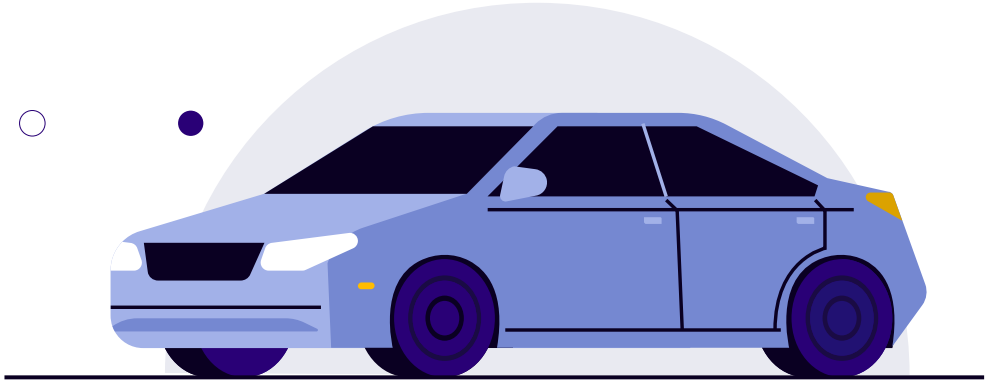
Predicting Used Car Prices in Kenya

Leveraging Machine Learning for accurate price estimation

- Munene Jackson
- Kelvin Mwaura
- Marion Achieng
- Gerald Mwangi
- Benedict Kuloba



Overview

- The used car market in Kenya is dynamic and rapidly evolving.
- Determining fair prices for used cars is challenging due to factors like engine, mileage, transmission and many more.
- This project aims to create a machine learning model to predict used car prices accurately.

Objectives

01. Develop a predictive model for estimating used car prices in Kenya.

02. Identify key factors influencing car prices.

03. Create a user-friendly tool for buyers and sellers to make informed decisions.

Data Overview

The dataset for this project was collected from various car bazaars and showrooms in Nairobi County and nearby areas, including car listings up to 2021.

Column Name	Description	Data Type
No	Unique identifier for each vehicle listing.	Integer (int64)
Name	Model name of the vehicle.	String (object)
Year	Year the vehicle was manufactured.	Integer (int64)
Kilometers_Driven	Total distance covered by the vehicle in kilometers.	Integer (int64)
Fuel_Type	Type of fuel used by the vehicle (e.g., petrol, diesel).	String (object)
Transmission	Transmission type of the vehicle (e.g., manual, automatic).	String (object)
Use	where the ehicle was previously used (e.g foreign,local)	String (object)
Engine	Engine capacity or specification of the vehicle.	String (object)
Power	Power output of the vehicle's engine, typically in brakehorsepower.	String (object)
Seats	Number of seats available in the vehicle.	Integer (int64)
Car_price	Price of the vehicle listed in Kenyan Shillings.	Float (float64)

For more info:
mwangigerald436@gmail.com

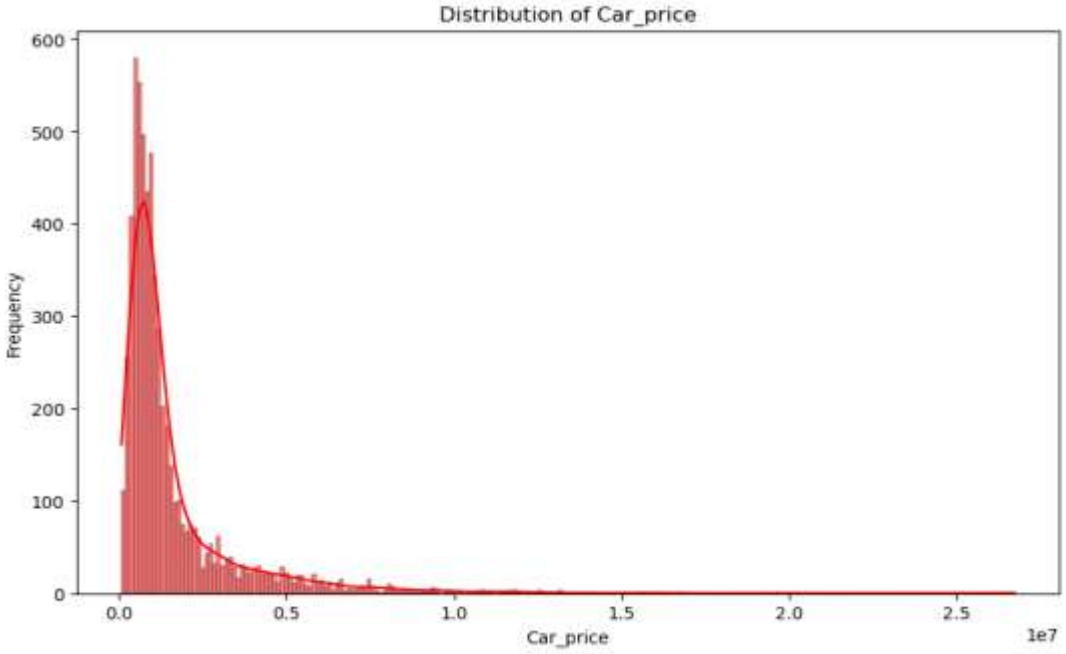
Data Cleaning & Preparation

Ensuring data quality is crucial for building an accurate predictive model. The dataset required extensive cleaning and transformation to make it suitable for analysis and modeling.

- Removed non-numeric characters and converted necessary columns to numeric types.
- Filled missing values with median/mode.
- Standardized text data and created a new column for car age.

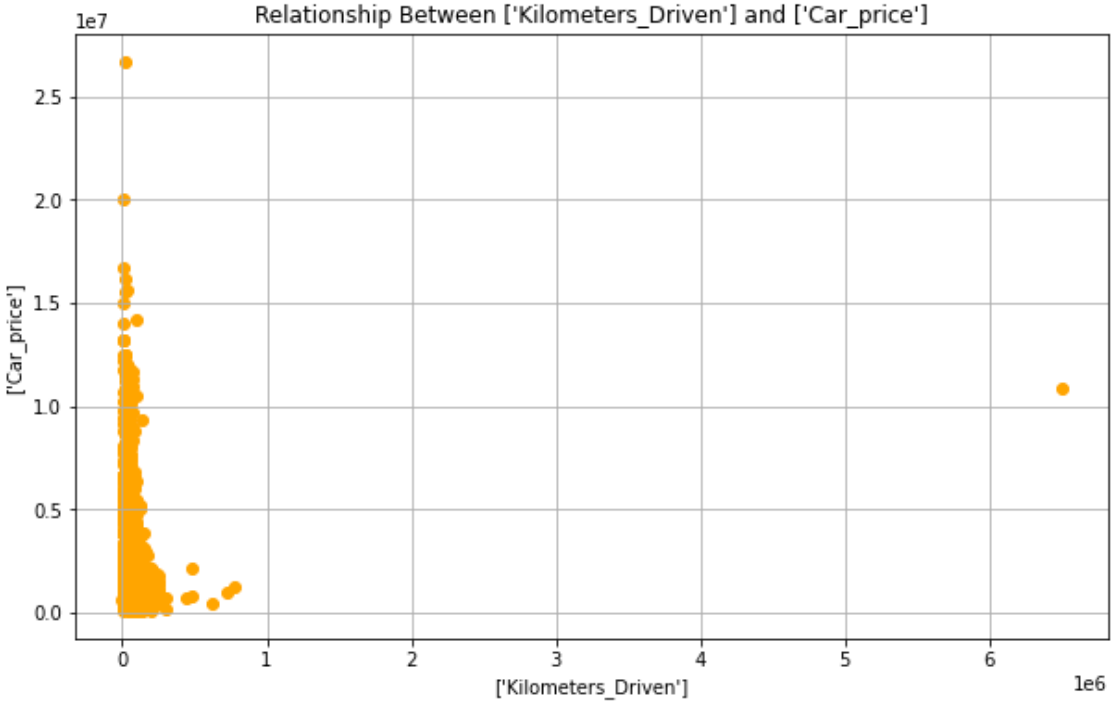
EDA

Performed univariate and bivariate analyses to understand data distribution and relationships.



Univariate Analysis

This shows that a lot of used cars are priced lower, according to this dataset.



Bivariate Analysis

We looked at a lot of relationships between different features like understanding how mileage affects the price of cars.

Key Findings

Age

- Observed a strong negative correlation between car age and price.
- Older cars tend to have lower prices due to depreciation.

Mileage

- Strong negative correlation between mileage and price.
- Cars with higher mileage are generally priced lower due to wear and tear.

Engine and Power

- Positive correlation between engine size, power, and price.
- Cars with larger engines and higher power output tend to be priced higher.

Brand

- Certain brands and models command higher prices due to their reputation, reliability, and demand.
- Luxury brands and popular models tend to retain their value better.

Modeling

Model Selection

Tested various regression models: Linear, Ridge, Lasso, ElasticNet, Random Forest, KNN, XGBoost, CatBoost, SVR, and Polynomial Regression.

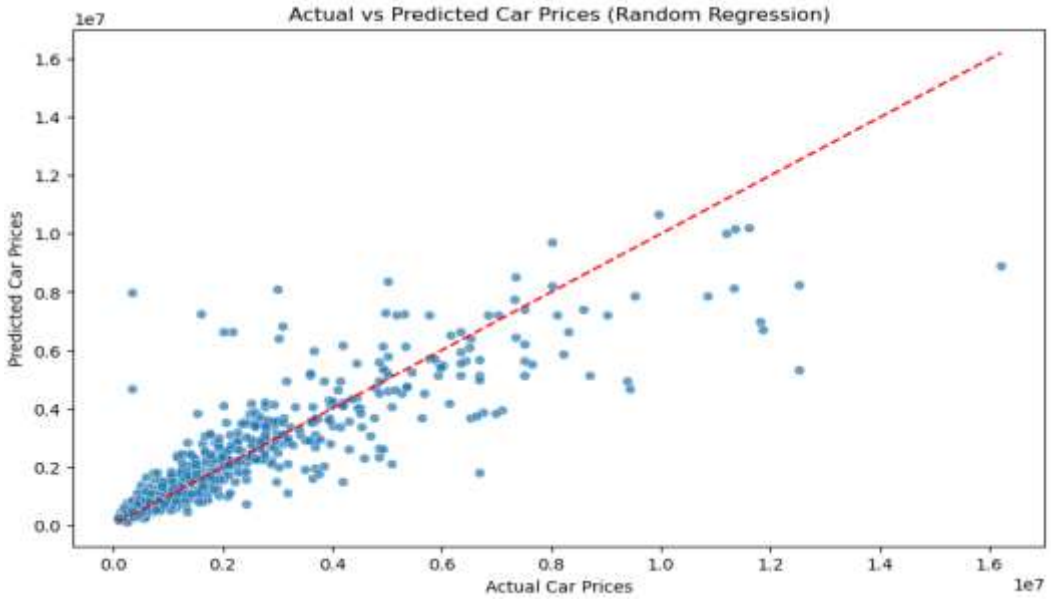
Evaluated models using cross-validation to determine performance.

Best Models

Below are the best performing models in order.

- Random Forest Regressor
- XGBoost Regressor
- Ridge Regressor
- CatBoost Regressor
- KNN Regressor

Random Forest Regressor



Random Forest Regressor achieved the best performance.

- MAE: Ksh. 393,696.89
- MSE: Ksh. 696,940,315,948.30
- R^2: 0.7969

Recommendations

For Buyers

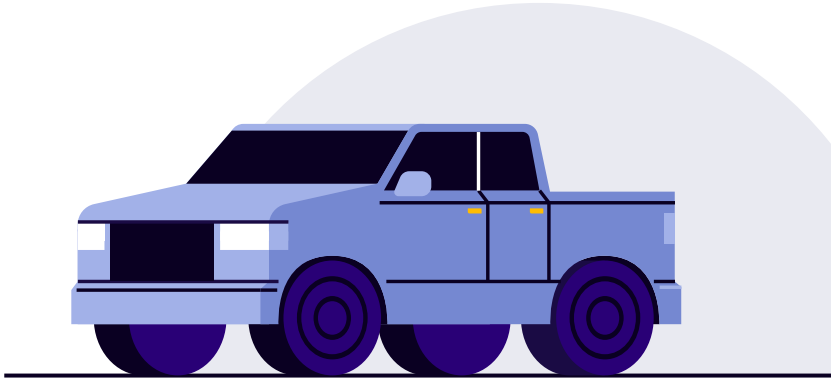
Leverage Data Insights: Use data-driven insights to make informed purchasing decisions, ensuring a fair price for the car's condition and specifications.

For Sellers

Competitive Pricing: Set competitive prices based on the car's age, condition, and market demand to attract potential buyers.
Market Trends: Stay informed about market trends and preferences, such as the growing demand for eco-friendly vehicles, to adjust pricing strategies accordingly.

For Car Dealers

Inventory Management: Optimize inventory management by understanding the demand for different car types and models.
Customer Education: Educate customers about the key factors influencing car prices and provide transparent information to build trust.



Conclusion

- The developed machine learning model offers a reliable tool for predicting used car prices in Kenya.
- The model empowers users to make informed decisions based on data-driven insights.
- Future work can focus on refining the model and incorporating additional features for even greater accuracy.

Thanks!

Do you have any questions?

- https://www.linkedin.com/in/jackson-mwaniki/
- https://www.linkedin.com/in/marion-achieng-0747712a3/
- https://www.linkedin.com/in/benedict-kuloba-76954512a/
- https://www.linkedin.com/in/kelvin-mwaura-a642351a2/