

Regression Project

Used Car Price Prediction

**Hrishinandan N
BCB 304**



Problem Statement

Build a model to predict used car prices (USD) using historical data and vehicle specifications.

Source: Kaggle ([click here](#))

Data Size: 10000 records, 12 features

Target Variable: price_usd

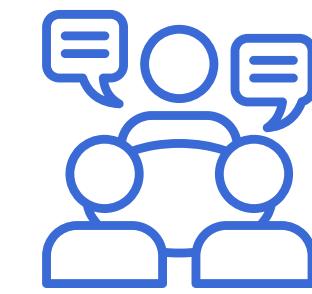
Numerical

- make_year
- engine_cc
- owner_count
- accidents_reported
- mileage_kmpl
- price_usd

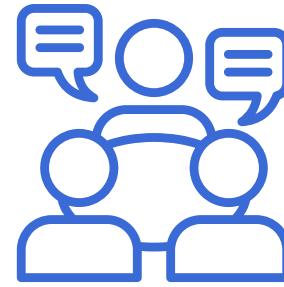
Categorical

- brand
- color
- transmission
- fuel_type
- service_history
- insurance_valid

Key Features



Dataset Overview



Data Preprocessing & Cleaning

Imputation using Mode

**Outlier Management
using IQR**

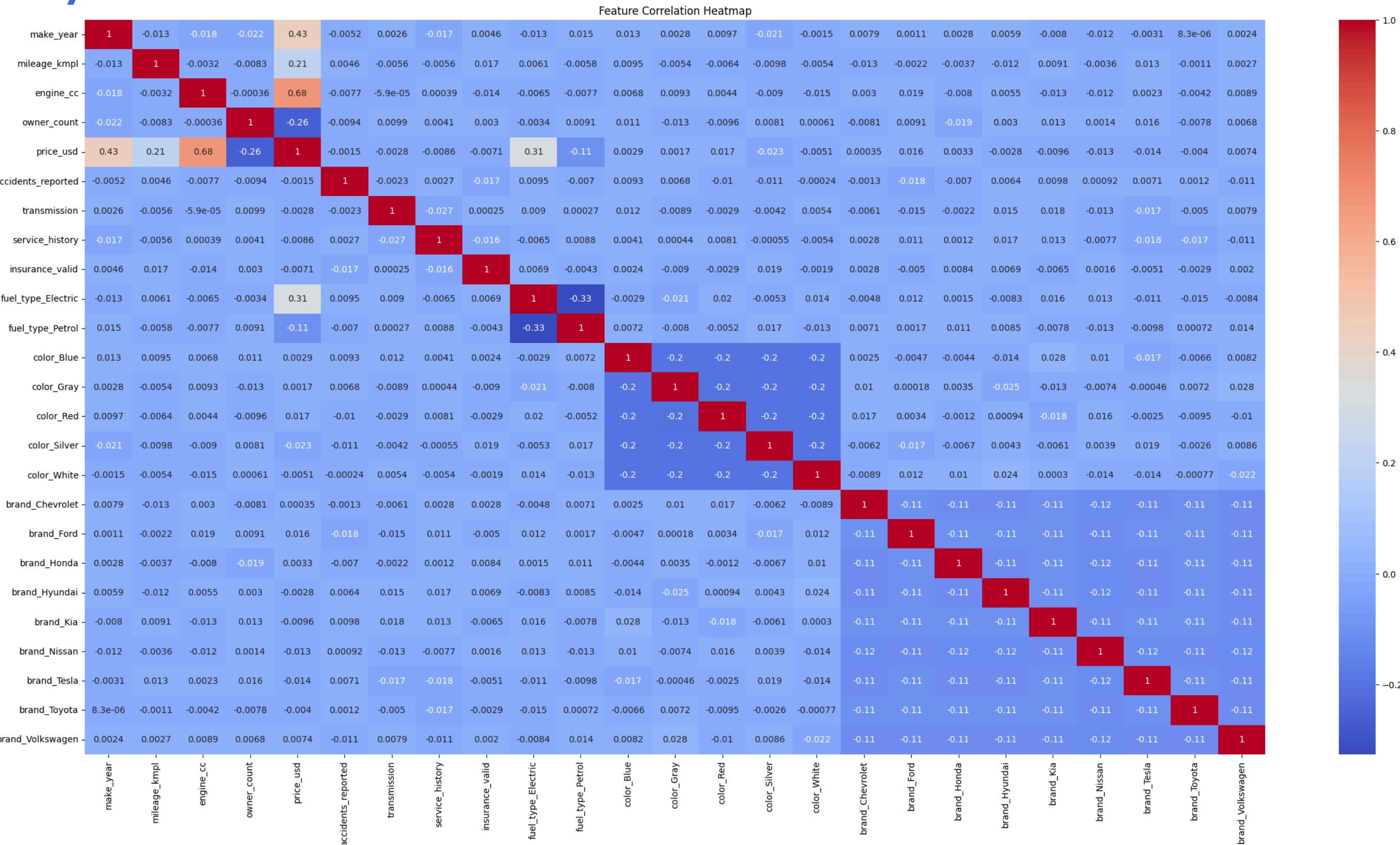
**Standardization for better
model convergence**

Feature Transformation

- Label Encoding for binary features
- One-Hot Encoding for multi-level categorical features

Exploratory Data Analysis

Correlation Analysis



Model Performance & Comparison

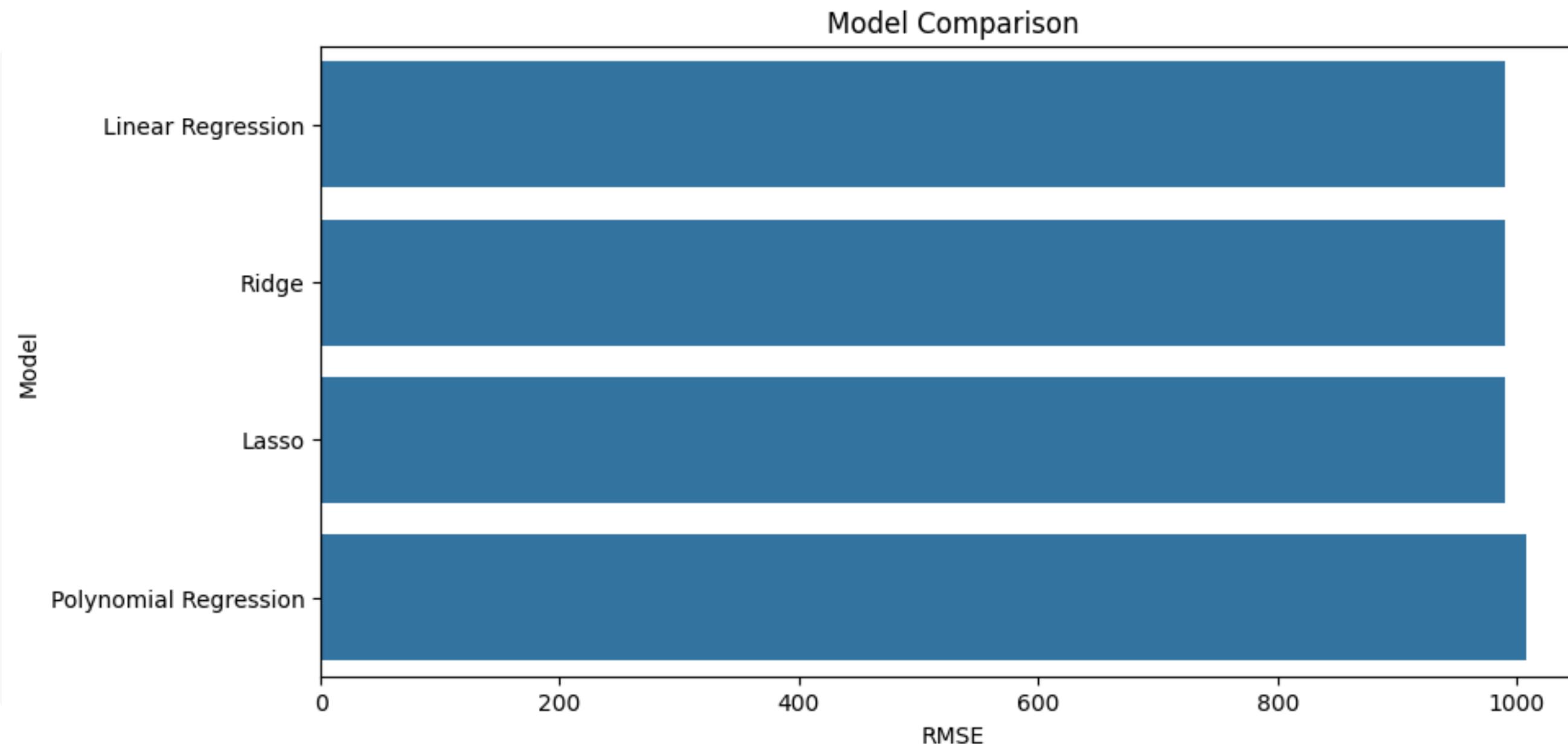
Evaluation Metrics Used

- R2 Score
- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE).

	R2 Score	MAE	MSE	RMSE
Linear Regression	0.876648	790.800540	9.813285e+05	990.620251
Ridge	0.876648	790.811695	9.813273e+05	990.619645
Lasso	0.876628	791.160844	9.814909e+05	990.702238
Polynomial Regression	0.873007	800.210429	1.010298e+06	1005.135824

Ridge Regression was identified as the most stable model with the lowest RMSE.

Results



Thank
You