The Goal:
Predicting Depreciation of Used Vehicles

By: Temesgen Tesfay

Mentor: Ben Bell

Problem Statment:

Car dealerships and customers predict used car prices more often based on:

- Odometer
- Vehicles age
- Others features such as model, manufacturer, etc..

However, predicting Used cars prices based on the above parameters alone is not enough to get accurate pricing.

Success criteria

- Getting depreciation per year value to specific car gives clear information for clients to evaluate the car price , whereas dealerships also set the price more accurately.

Scope of Solution Space

- The depreciation per year rate of used cars is constant over the entire age of the car.
- Total loss and great loss by accident are exclusive in the analysis.
- Only extreme outliers from price and odometer removed.

Limitation

- Limited Dataset and Features.
- Data Imbalance
- Antique cars behave strangely

Data sources:

1. Kaggle Dataset
2. Kelly Blue Book(KBB)

❖ **Data Wrangling:**

Raw Dataset Dimension:

- 18 variables
- 423857 rows

Cleaned Dataset Dimension:

- 73 features
- 7488 rows

Merging dataset

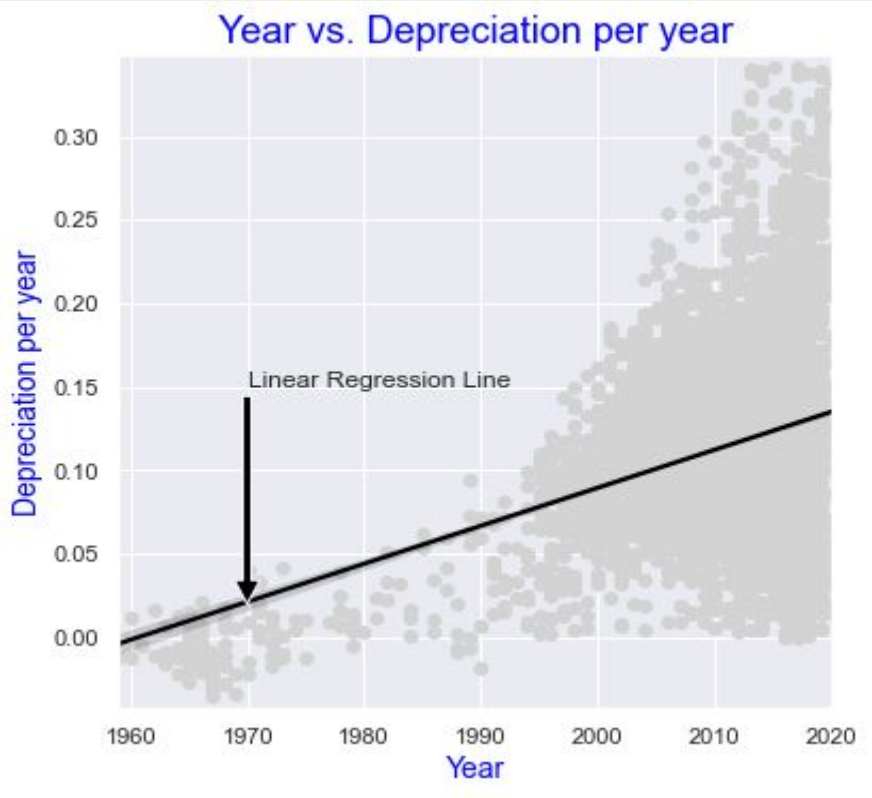
- Two datasets merged on manufacturer and model features

Exploratory Data Analysis (EDA)

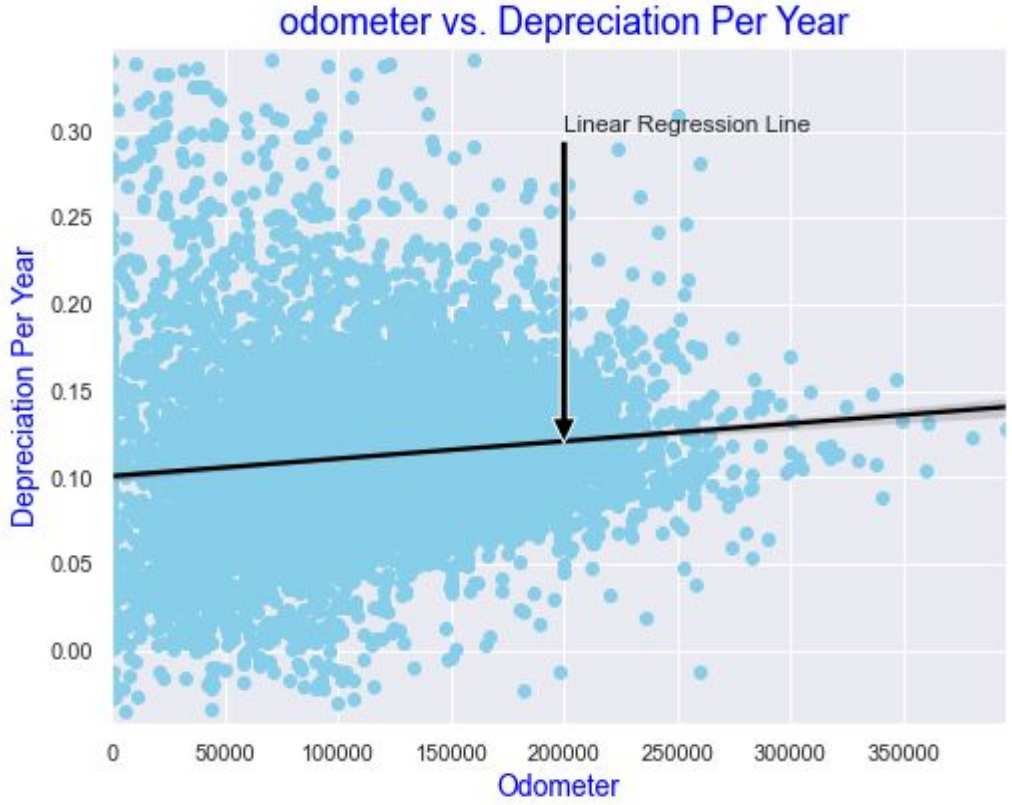
- Cleaning dataset and data type conversion.
- Name difference addressed manually.
- Removed duplicates and Merging data frames
- The features such as depreciation, depreciation per year, new_price, year_difference are derived from the datasets .

Year vs Depreciation per year

- New cars depreciate faster than the old cars- period between 2015 -2020
- In 1960 the depreciation per year is below zero, implies the value of antique cars sold more than the cost of cars when new.

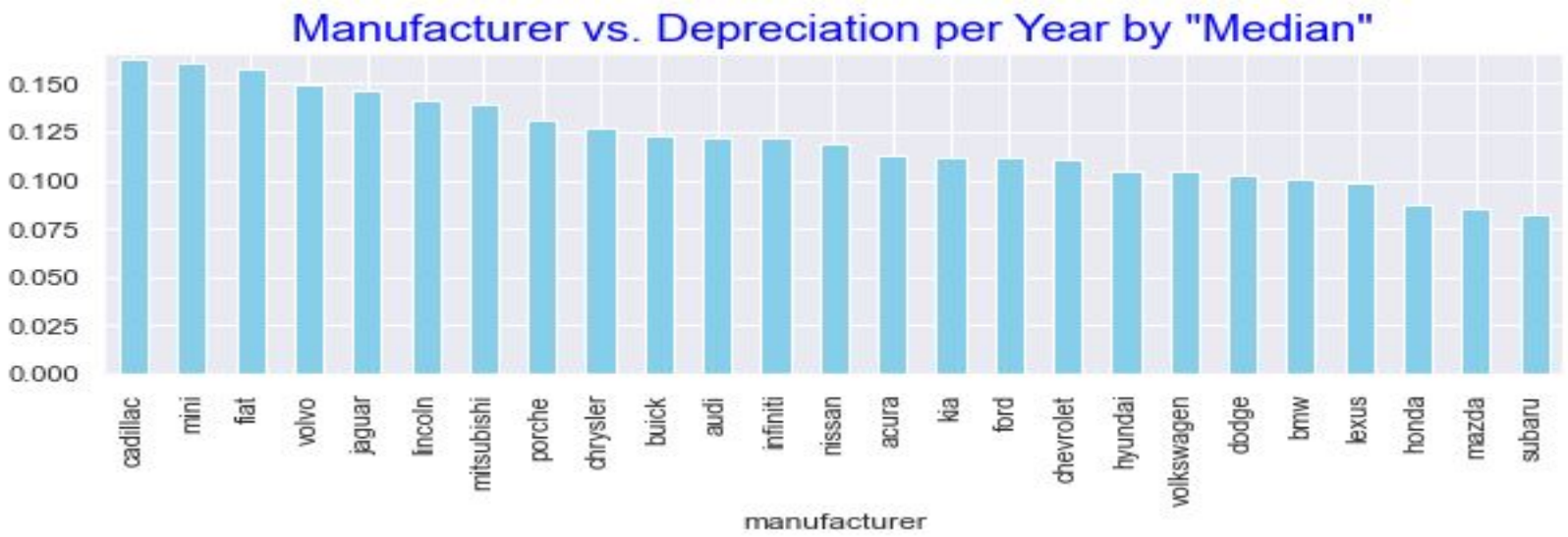


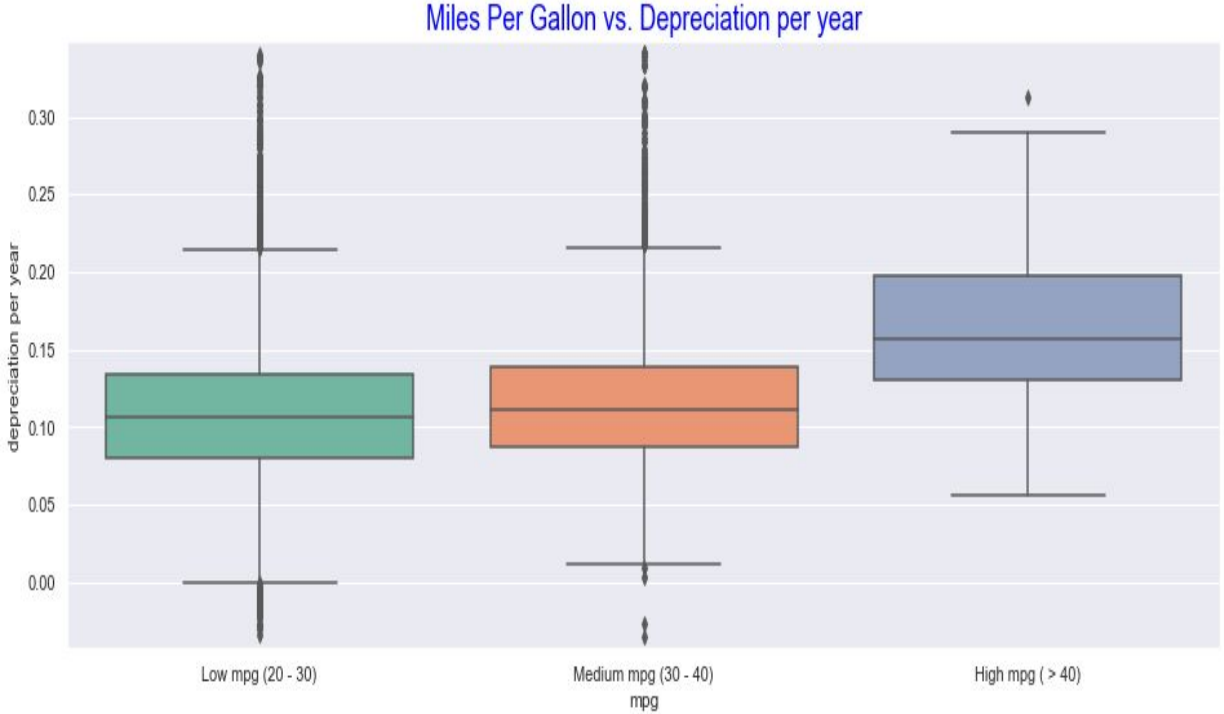
Outlier depreciation rates are most common in cars with low mileage



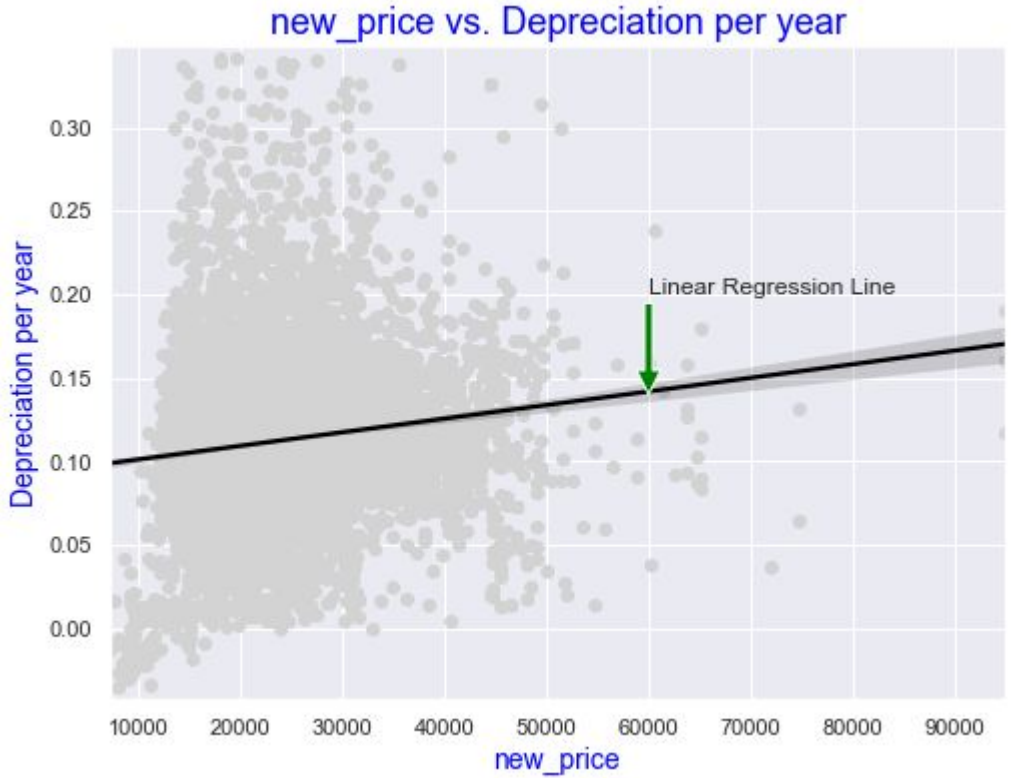
The median bar graph of depreciation per year versus manufacturer .

- Cadillac , Mini and Fiat are the highest depreciation per year over 15%.
- Honda, Mazda, Subaru has low depreciation per year less 10% .



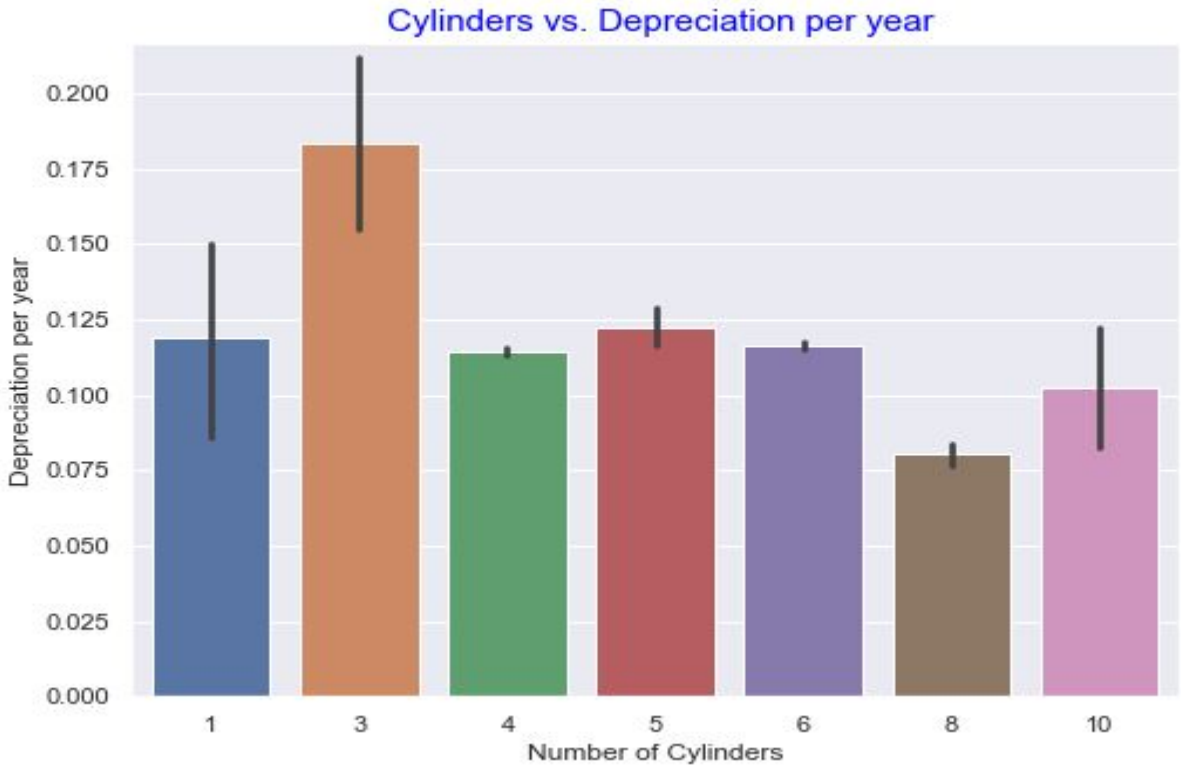


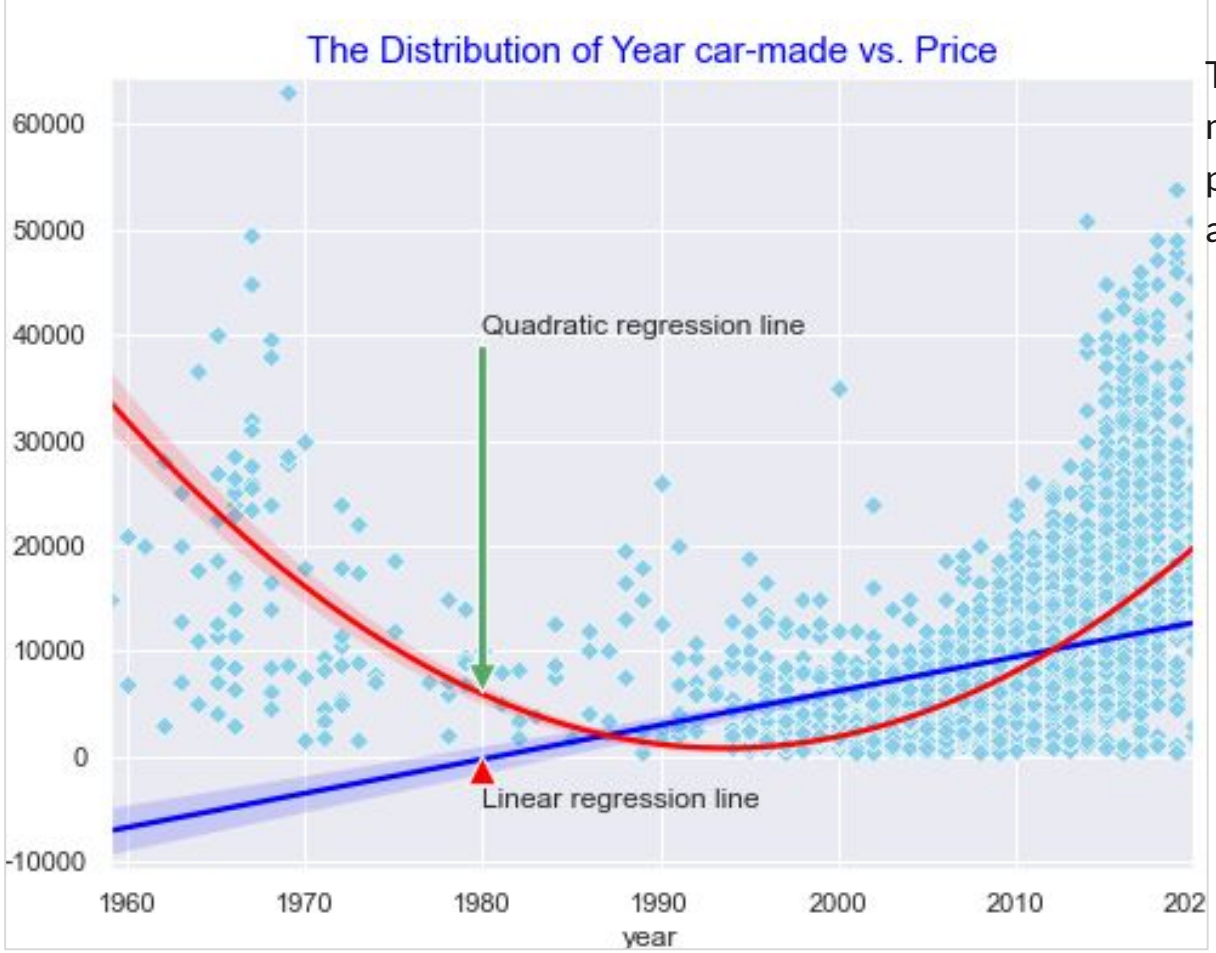
The depreciation per year rate is higher in used cars run over 40 miles per gallon.



- Depreciation per year is increases as the cars price getting more expensive.
- Clients decide to buy vehicles on the range of the depreciation per year rate.

The number of cylinders a car has greatly affects the overall power of the engine. Depreciation per year rate significantly varies from 8% to 18%.Cars with 8-cylinders maintain their value better than most others.





The Dataset includes cars year model from 1960 - to date. The price is high in new cars and antique cars.

- Cars prices dropped until 1985, but those cars made before 1985, regain their value exponentially.

The Models Algorithm Evaluators:

- LinearRegression
- Ridge
- XGBoost
- DecisionTreeRegressor
- RandomForestRegressor

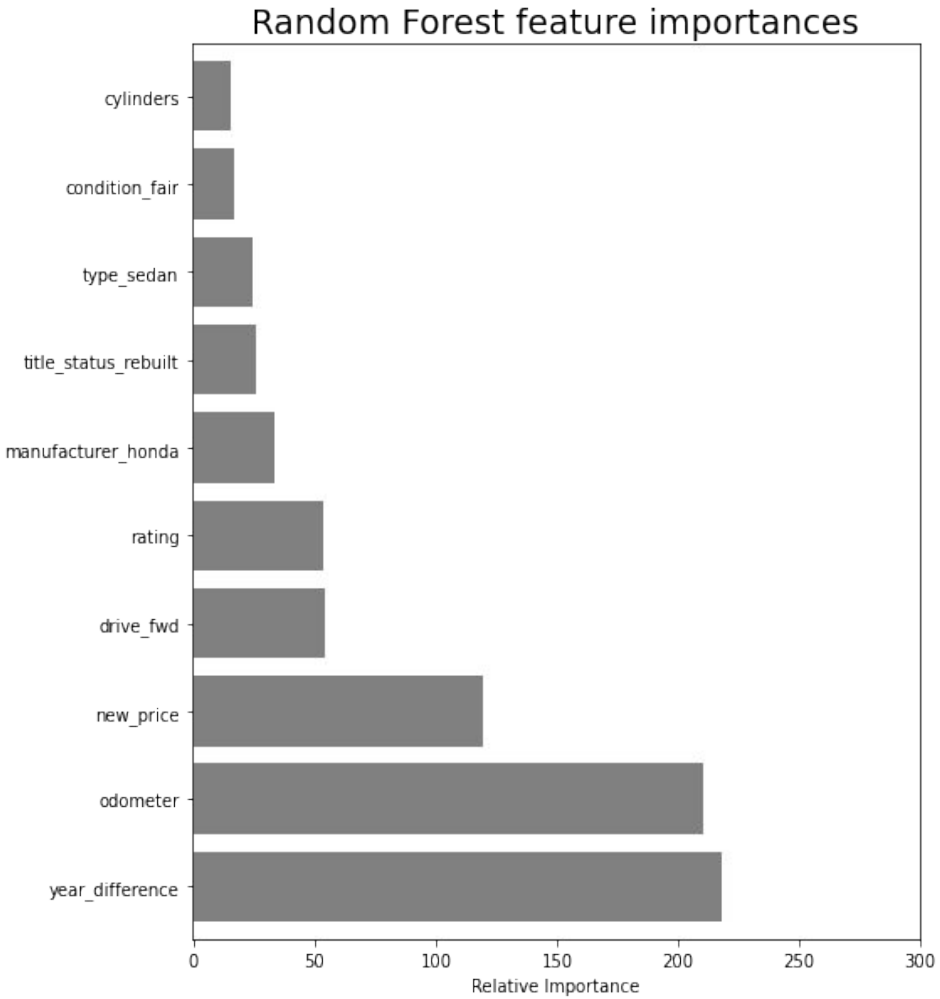
Model Metrics

The R-squared score and others parameters mentioned below is a result of the model GridSearchCV.

Model	Parameter	Metrics	Metric
		R-Square	RMST
Linear Regression	-	0.4278	0.03823
Ridge Regression	Alpha=0.1	0.4238	0.03829
XGBoost Regression	Alpha=0.1 n_estimators=200	0.7759	0.02966
Decision Tree Regression	max_depth=11 min_samples_split=18	0.3672	0.04578
Random Forest Regression	max_depth=40 n_estimatores=200	0.5724	0.0344

Random Forest Regressor

- The most important features are year_difference and Odometer, and new_price are top three features based on variance explained to predict depreciation per year.



Conclusion:

Vehicle features are relevant in price prediction, however, they are not enough to get accurate price because vehicles are differ in depreciation rate. Depreciation per year is a specific value to each car that minimize the pricing error significantly besides, it gives clear information to clients in used vehicles transaction.

I will work to improve the prediction power of the model especially by adding more features and data and will try to collect the exact value of used vehicles price when new.