Vehicle Price Prediction

Presented by:
Praveen Kumar Gonuguntla
Pavan Kumar Boddu
Hemanth Varma Dantuluri
Koushik Mannaru
Sreeja Komma Reddy

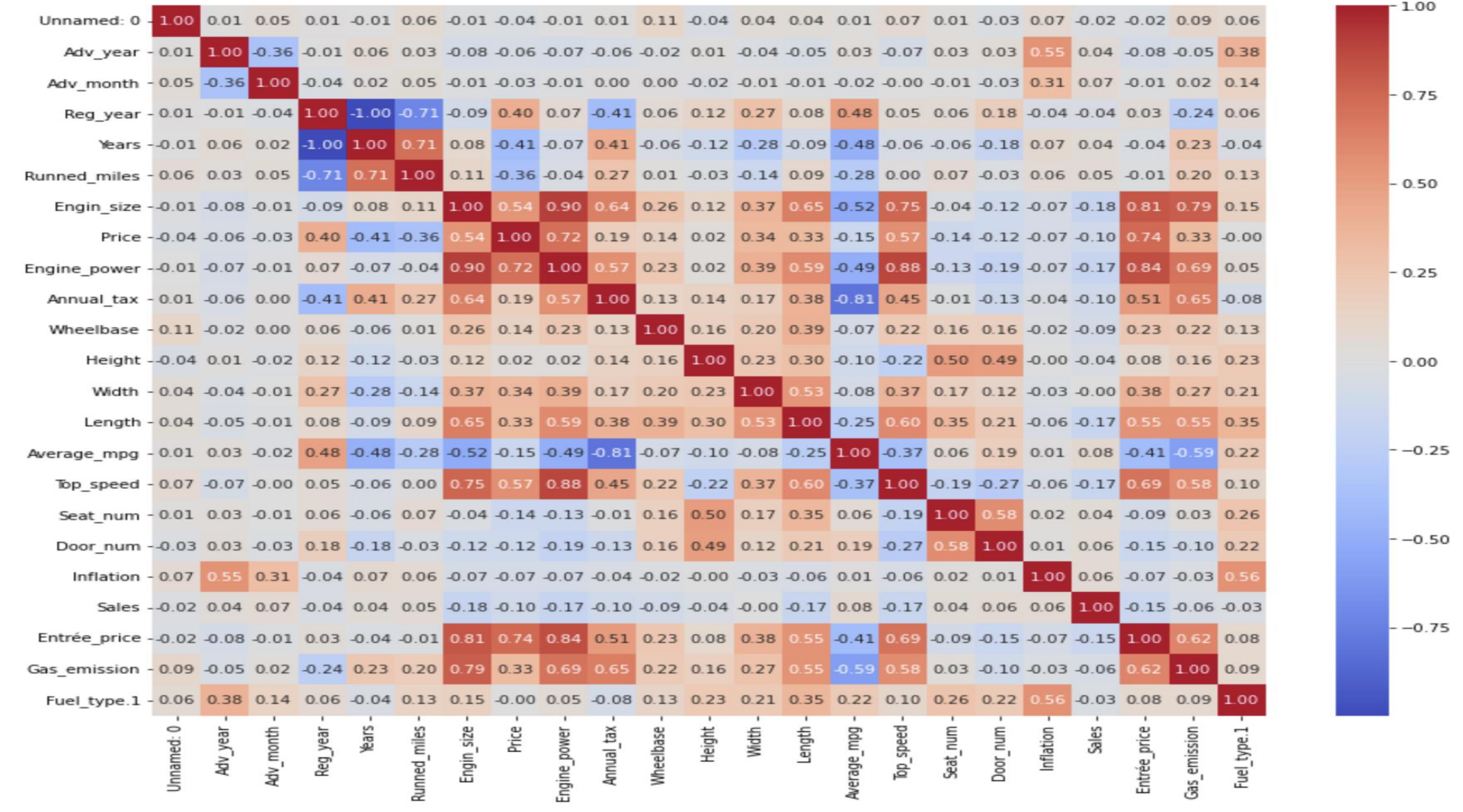
Introduction

- Vehicle price prediction is a task that has become increasingly relevant in recent years, as the automotive industry has seen a surge in demand for personalized vehicles.
- One of the main challenges in this field is predicting the price of a vehicle accurately, which involves analyzing numerous factors such as the make and model, mileage, year, and condition of the vehicle.
- It is possible to leverage historical data to predict the price of a vehicle with a high degree of accuracy.
- Supervised learning algorithms can be trained on past data to identify patterns and relationships between the variables, allowing for the creation of a predictive model that can estimate the price of a vehicle based on its features.
- This project explores the use of various supervised learning algorithms for vehicle price prediction.

Dataset

- *DVM Car Dataset: A large-scale dataset for Automotive Applications*. DVM Car Dataset: A Large-Scale Dataset for Automotive Applications. https://deepvisualmarketing.github.io/
- *Historical inflation rates: 1914-2023*. US Inflation Calculator | Easily calculate how the buying power of the U.S. dollar has changed from 1913 to 2023. Get inflation rates and U.S. inflation news. (2023, April 12). https://www.usinflationcalculator.com/inflation/historical-inflation-rates/

Column Definitions


Unnamed: 0 1.00 0.01 0.05 0.01 -0.01 0.06 -0.01 -0.04 -0.01 0.01 0.11 -0.04 0.04 0.04 0.01 0.07 0.01 -0.03 0.07 -0.02 -0.02 0.09 0.06
Adv_year 0.01 1.00 -0.36 -0.01 0.06 0.03 -0.08 -0.06 -0.07 -0.06 -0.02 0.01 -0.04 -0.05 0.03 -0.07 0.03 0.03 0.55 0.04 -0.08 -0.05 0.38
Adv_month 0.05 -0.36 1.00 -0.04 0.02 0.05 -0.01 -0.03 -0.01 0.00 0.00 -0.02 -0.01 -0.01 -0.02 -0.00 -0.01 -0.03 0.31 0.07 -0.01 0.02 0.14
Reg_year 0.01 -0.01 -0.04 1.00 -1.00 -0.71 -0.09 0.40 0.07 -0.41 0.06 0.12 0.27 0.08 0.48 0.05 0.06 0.18 -0.04 -0.04 0.03 -0.24 0.06
Years -0.01 0.06 0.02 -1.00 1.00 0.71 0.08 -0.41 -0.07 0.41 -0.06 -0.12 -0.28 -0.09 -0.48 -0.06 -0.06 -0.18 0.07 0.04 -0.04 0.23 -0.04
Runned_miles 0.06 0.03 0.05 -0.71 0.71 1.00 0.11 -0.36 -0.04 0.27 0.01 -0.03 -0.14 0.09 -0.28 0.00 0.07 -0.03 0.06 0.05 -0.01 0.20 0.13
Engin_size -0.01 -0.08 -0.01 -0.09 0.08 0.11 1.00 0.54 0.90 0.64 0.26 0.12 0.37 0.65 -0.52 0.75 -0.04 -0.12 -0.07 -0.18 0.81 0.79 0.15
Price -0.04 -0.06 -0.03 0.40 -0.41 -0.36 0.54 1.00 0.72 0.19 0.14 0.02 0.34 0.33 -0.15 0.57 -0.14 -0.12 -0.07 -0.10 0.74 0.33 -0.00
Engine_power -0.01 -0.07 -0.01 0.07 -0.07 -0.04 0.90 0.72 1.00 0.57 0.23 0.02 0.39 0.59 -0.49 0.88 -0.13 -0.19 -0.07 -0.17 0.84 0.69 0.05
Annual_tax 0.01 -0.06 0.00 -0.41 0.41 0.27 0.64 0.19 0.57 1.00 0.13 0.14 0.17 0.38 -0.81 0.45 -0.01 -0.13 -0.04 -0.10 0.51 0.65 -0.08
Wheelbase 0.11 -0.02 0.00 0.06 -0.06 0.01 0.26 0.14 0.23 0.13 1.00 0.16 0.20 0.39 -0.07 0.22 0.16 0.16 -0.02 -0.09 0.23 0.22 0.13
Height -0.04 0.01 -0.02 0.12 -0.12 -0.03 0.12 0.02 0.02 0.14 0.16 1.00 0.23 0.30 -0.10 -0.22 0.50 0.49 -0.00 -0.04 0.08 0.16 0.23
Width 0.04 -0.04 -0.01 0.27 -0.28 -0.14 0.37 0.34 0.39 0.17 0.20 0.23 1.00 0.53 -0.08 0.37 0.17 0.12 -0.03 -0.00 0.38 0.27 0.21
Length 0.04 -0.05 -0.01 0.08 -0.09 0.09 0.65 0.33 0.59 0.38 0.39 0.30 0.53 1.00 -0.25 0.60 0.35 0.21 -0.06 -0.17 0.55 0.55 0.35
Average_mpg 0.01 0.03 -0.02 0.48 -0.48 -0.28 -0.52 -0.15 -0.49 -0.81 -0.07 -0.10 -0.08 -0.25 1.00 -0.37 0.06 0.19 0.01 0.08 -0.41 -0.59 0.22
Top_speed 0.07 -0.07 -0.00 0.05 -0.06 0.00 0.75 0.57 0.88 0.45 0.22 -0.22 0.37 0.60 -0.37 1.00 -0.19 -0.27 -0.06 -0.17 0.69 0.58 0.10
Seat_num 0.01 0.03 -0.01 0.06 -0.06 0.07 -0.04 -0.14 -0.13 -0.01 0.16 0.50 0.17 0.35 0.06 -0.19 1.00 0.58 0.02 0.04 -0.09 0.03 0.26
Door_num -0.03 0.03 -0.03 0.18 -0.18 -0.03 -0.12 -0.12 -0.19 -0.13 0.16 0.49 0.12 0.21 0.19 -0.27 0.58 1.00 0.01 0.06 -0.15 -0.10 0.22
Inflation 0.07 0.55 0.31 -0.04 0.07 0.06 -0.07 -0.07 -0.07 -0.04 -0.02 -0.00 -0.03 -0.06 0.01 -0.06 0.02 0.01 1.00 0.06 -0.07 -0.03 0.56
Sales -0.02 0.04 0.07 -0.04 0.04 0.05 -0.18 -0.10 -0.17 -0.10 -0.09 -0.04 -0.00 -0.17 0.08 -0.17 0.04 0.06 0.06 1.00 -0.15 -0.06 -0.03
Entrée_price -0.02 -0.08 -0.01 0.03 -0.04 -0.01 0.81 0.74 0.84 0.51 0.23 0.08 0.38 0.55 -0.41 0.69 -0.09 -0.15 -0.07 -0.15 1.00 0.62 0.08
Gas_emission 0.09 -0.05 0.02 -0.24 0.23 0.20 0.79 0.33 0.69 0.65 0.22 0.16 0.27 0.55 -0.59 0.58 0.03 -0.10 -0.03 -0.06 0.62 1.00 0.09
Fuel_type.1 0.06 0.38 0.14 0.06 -0.04 0.13 0.15 -0.00 0.05 -0.08 0.13 0.23 0.21 0.35 0.22 0.10 0.26 0.22 0.56 -0.03 0.08 0.09 1.00
Unnamed: 0
Adv_year
Adv_month
Reg_year
Years
Runned_miles
Engin_size
Price
Engine_power
Annual_tax
Wheelbase
Height
Width
Length
Average_mpg
Top_speed
Seat_num
Door_num
Inflation
Sales
Entrée_price
Gas_emission
Fuel_type.1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

Before >>

```
O_df.dtypes
```

```
Unnamed: 0          int64
Maker              object
Genmodel           object
Genmodel_id        object
Adv_id             object
Adv_year            int64
Adv_month           int64
Reg_year            int64
Years               int64
Bodytype           object
Runned_miles        int64
Engin_size        float64
Gearbox            object
Fuel_type          object
Price               int64
Engine_power      float64
Annual_tax        float64
Wheelbase           int64
Height              int64
Width               int64
Length              int64
Average_mpg         int64
Top_speed           int64
Seat_num            int64
Door_num            int64
Inflation         float64
Sales               int64
Entrée_price        int64
Gas_emission        int64
Fuel_type.1       float64
dtype: object
```

After >>

```
df.dtypes
```

```
Unnamed: 0          int64
Adv_year            int64
Adv_month           int64
Reg_year            int64
Years               int64
Bodytype           object
Runned_miles        int64
Engin_size        float64
Gearbox            object
Fuel_type          object
Price               int64
Engine_power      float64
Annual_tax        float64
Wheelbase           int64
Height              int64
Width               int64
Length              int64
Average_mpg         int64
Top_speed           int64
Seat_num            int64
Door_num            int64
Inflation         float64
Sales               int64
Entrée_price        int64
Gas_emission        int64
Fuel_type.1       float64
dtype: object
```

Final >>

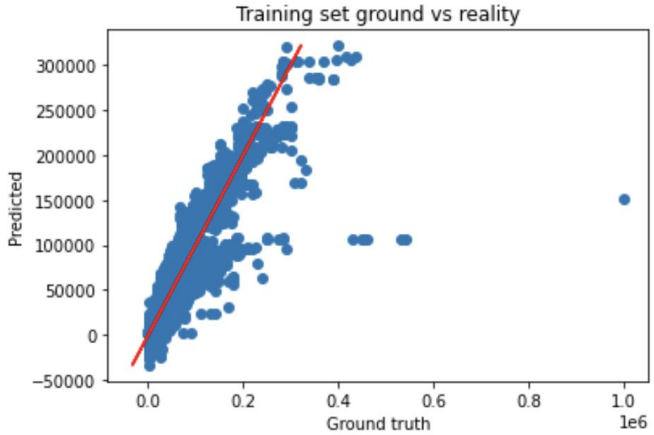
```
df_encoded.dtypes
```

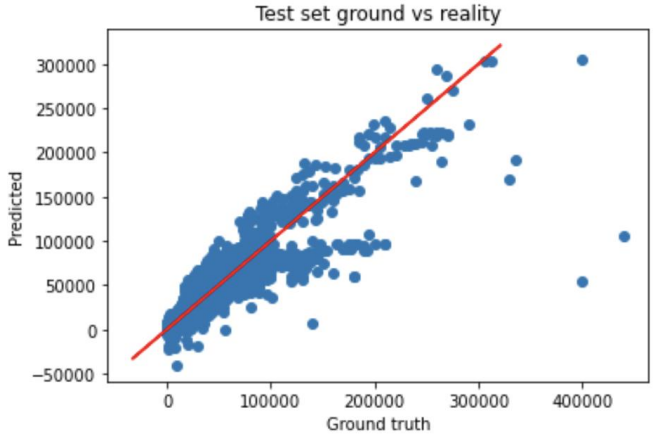
```
Unnamed: 0                                  int64
Adv_year                                    int64
Adv_month                                   int64
Reg_year                                    int64
Years                                       int64
Runned_miles                                int64
Engin_size                                float64
Price                                       int64
Engine_power                              float64
Annual_tax                                float64
Wheelbase                                   int64
Height                                      int64
Width                                       int64
Length                                      int64
Average_mpg                                 int64
Top_speed                                   int64
Seat_num                                    int64
Door_num                                    int64
Inflation                                 float64
Sales                                       int64
Entrée_price                                int64
Gas_emission                                int64
Fuel_type.1                               float64
Bodytype_Convertible                        uint8
Bodytype_Coupe                              uint8
Bodytype_Estate                             uint8
Bodytype_Hatchback                          uint8
Bodytype_Limousine                          uint8
Bodytype_MPV                                uint8
Bodytype_SUV                                uint8
Bodytype_Saloon                             uint8
Bodytype_Window Van                         uint8
Gearbox_Manual                              uint8
Fuel_type_Diesel                            uint8
Fuel_type_Hybrid  Diesel/Electric           uint8
Fuel_type_Hybrid  Diesel/Electric Plug-in   uint8
Fuel_type_Hybrid  Petrol/Electric           uint8
Fuel_type_Hybrid  Petrol/Electric Plug-in   uint8
Fuel_type_Petrol                            uint8
Fuel_type_Petrol Ethanol                    uint8
dtype: object
```

Polynomial Regression

```
Training set Mean Squared Error (MSE): 39898541.767
Training set R^2 score: 0.871

Training set Mean Squared Error (MSE): 39898541.767
Test set R^2 score: 0.876
```





OLS Regression Results

Dep. Variable:	Price_x	R-squared:	0.719
Model:	OLS	Adj. R-squared:	0.719
Method:	Least Squares	F-statistic:	2.564e+04
Date:	Tue, 09 May 2023	Prob (F-statistic):	0.00
Time:	23:25:35	Log-Likelihood:	-1.5915e+06
No. Observations:	150686	AIC:	3.183e+06
Df Residuals:	150670	BIC:	3.183e+06
Df Model:	15		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.887e+04	4690.171	4.023	0.000	9676.551	2.81e+04
Years_x	-1205.5911	8.570	-140.678	0.000	-1222.388	-1188.794
Entrée_price_x	0.7699	0.002	380.510	0.000	0.766	0.774
Runned_miles_x	-0.0497	0.001	-61.877	0.000	-0.051	-0.048
Length_x	-3.0603	0.098	-31.234	0.000	-3.252	-2.868
Seat_num_x	-666.6300	44.947	-14.831	0.000	-754.726	-578.534
Bodytype_Convertible	1843.8668	4676.919	0.394	0.693	-7322.799	1.1e+04
Bodytype_Coupe	4485.2271	4675.823	0.959	0.337	-4679.290	1.36e+04
Bodytype_Estate	2944.2580	4674.976	0.630	0.529	-6218.601	1.21e+04
Bodytype_Hatchback	2160.1706	4674.156	0.462	0.644	-7001.080	1.13e+04
Bodytype_Limousine	-5.747e+04	4828.462	-11.902	0.000	-6.69e+04	-4.8e+04
Bodytype_MPV	1446.2631	4674.483	0.309	0.757	-7715.629	1.06e+04
Bodytype_SUV	1621.9588	4674.067	0.347	0.729	-7539.118	1.08e+04
Bodytype_Saloon	-933.1512	4674.604	-0.200	0.842	-1.01e+04	8228.978
Bodytype_Window Van	1.42e+04	5167.320	2.748	0.006	4073.899	2.43e+04
Gas_emission_x	-3.4861	0.676	-5.159	0.000	-4.810	-2.162

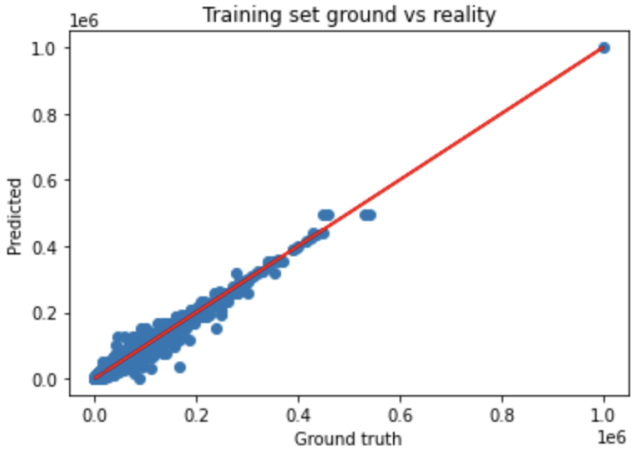
Omnibus:	282776.067	Durbin-Watson:	2.000
Prob(Omnibus):	0.000	Jarque-Bera (JB):	4080192985.611
Skew:	13.447	Prob(JB):	0.00
Kurtosis:	808.689	Cond. No.	4.28e+07

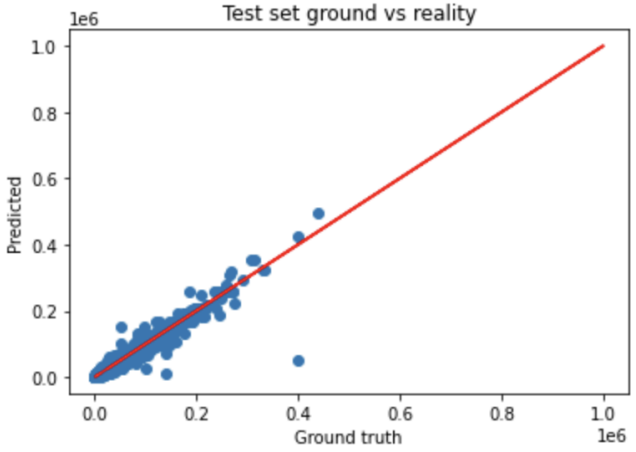
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.28e+07. This might indicate that there are
strong multicollinearity or other numerical problems.
Training set R^2 score: 0.719
Test set R^2 score: 0.722

Decision Tree

```
Training set Mean Squared Error (MSE): 8099126.783
Training set R^2 score: 0.974

Test set Mean Squared Error (MSE): 13731375.131
Test set R^2 score: 0.953
```

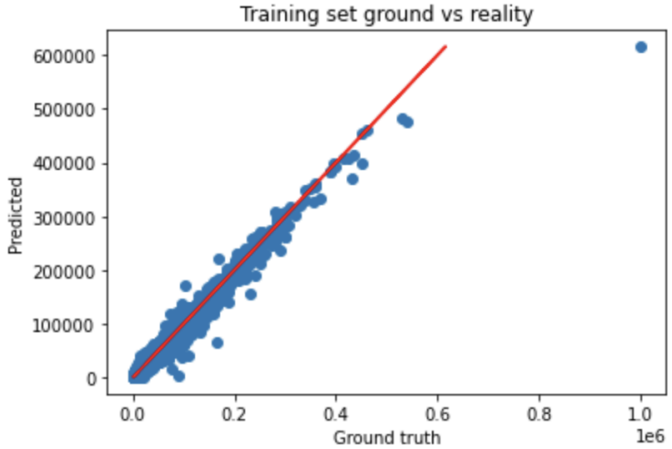


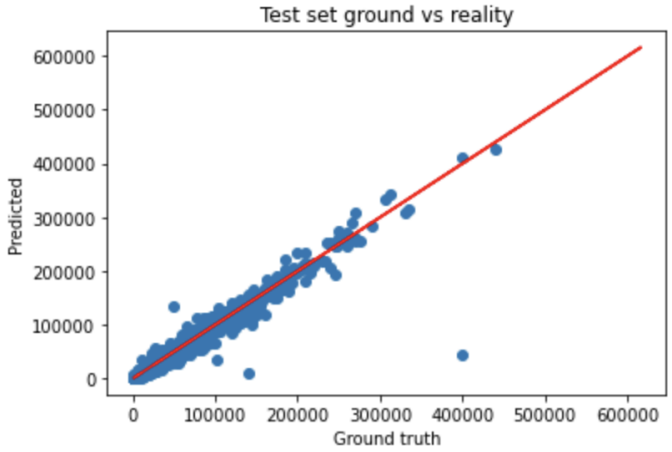


Random Forest

```
Training set Mean Squared Error (MSE): 6754562.675
Training set R^2 score: 0.978

Test set Mean Squared Error (MSE): 10515721.002
Test set R^2 score: 0.964
```

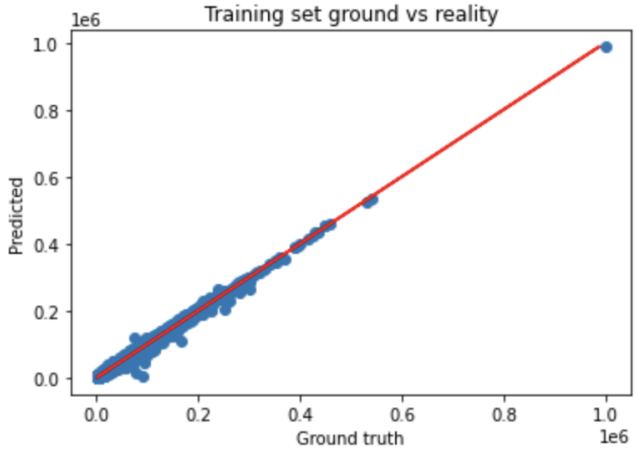


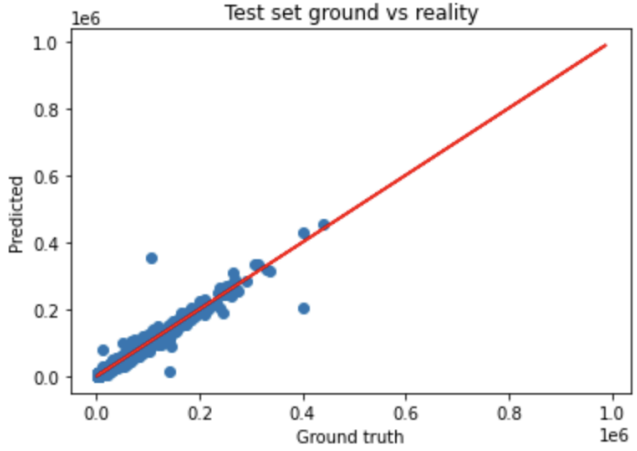


Gradient Boosting

```
Training set Mean Squared Error (MSE): 2643583.247
Training set R^2 score: 0.991

Test set Mean Squared Error (MSE): 7291046.840
Test set R^2 score: 0.975
```





Bagging Regressor

```
R^2 score for train dataset =  0.9914

R^2 score for test dataset =  0.9738
```