

Regression

Used Porsche 911 prices



Introduction

Central Question:

How predictable are used Porsche listing prices?

Approach:

Accuracy over interpretability

Source of data:

Scraped from autotrader.co.uk

AutoTrader Used cars New cars Sell your car Value your car Car reviews Car leasing Electric cars Finance

[Back to results](#)

Get an insurance quote

1 of 4

Porsche 911
1977 (S reg)
3.0 Classic SC Targa 2dr

£34,911

[Home Delivery](#) [COVID-19 safety measures](#)

[Message seller](#)

Car price
Seller's admin fee [i](#)

Total price

++ NUMBERS MATCHING ++

HUDSON CARS
CARDIFF

Hudson Cars Cardiff

★★★★★ 5.0 (93 reviews)

[Visit website](#)

[\(07537\) 125492](#)

[Location map](#) [Get directions](#)

[Home Delivery](#)

Free delivery up to 50 miles, £1 per mile thereafter

Overview

203,000 miles
647,873 miles below average [i](#)

Convertible 3.0L Manual Petrol

2 doors

Methodology

Web-scraping : Selenium and BeautifulSoup

Data storage: Python Pickle

EDA and data-readying: Python Pandas

Regression: Statsmodels and sklearn

Visualizations: matplotlib and Seaborn



Data

Of 1742 Porsche 911's on [autotrader.co.uk](https://www.autotrader.co.uk), 1560 cars with 'full' data:

- Mileage
- Engine size
- Number of doors
- Number of cylinders
- Top speed
- 0-60mph in seconds
- Engine power in bhp
- Years old
- Gearbox (e.g. 'Manual')
- Fuel type (e.g. 'Petrol')
- Body type (e.g. 'Coupe')



Data – continued

Target is listing price:

- Most expensive car: £495,990
- Least expensive: £8,950
- Mean price: £79,177

Some feature value ranges:

- Engine power: 204 bhp – 690 bhp
- Top speed: 118mph – 211mph
- Year registered: 1973 – 2021
- Mileage: 10 miles - 286,000 miles



Results

- Of 27 models, best predictor is:
 - OLS model
 - 45 features - all but 8 auto-generated
 - Power transformation of target (Box-Cox)
- Mean Absolute Error on test data of just under £8000



Conclusion

- Of all the models the better predictors are the more complex ones
- The best predictor is not that great



Future work

- There is a data gap, for example:
 - Condition of car
 - Type of seller
 - Number of owners
 - Quality of documentation related to car
 - Special edition information
- Regularization techniques may yield prediction improvements
- However maybe inherently unpredictable





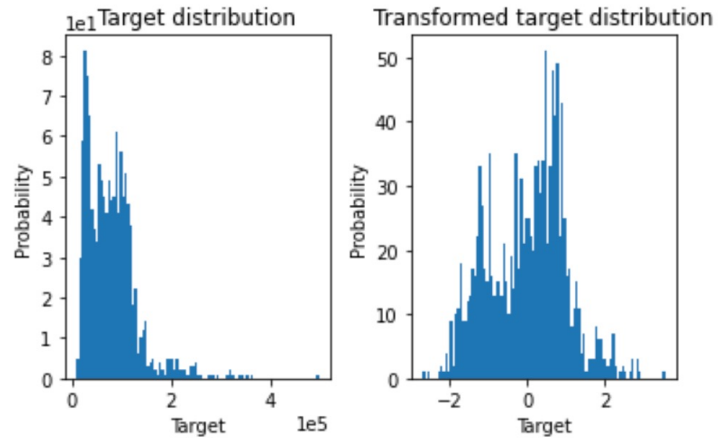
APPENDIX

Regressions performed:

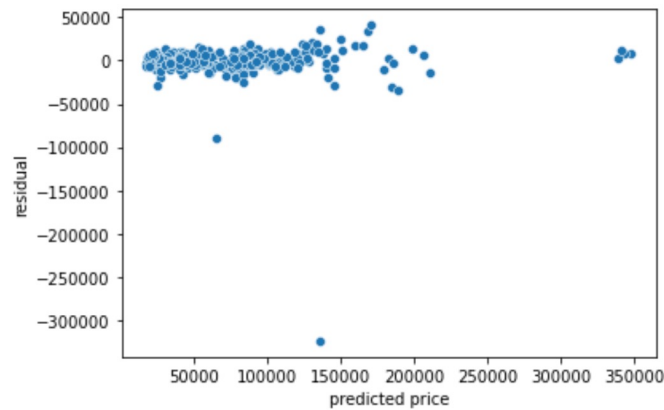
| Name of X df | Type of regre: | No. features | Polynomial and Interaction | VIFs >5 | Features Transformed | Target Transformed | Cross-validated | r-squ validation | MAE validation | r-sq train | MAE train | r-sq test | MAE test |
|--------------------|----------------|--------------|----------------------------|---------|----------------------|--------------------|-----------------|------------------|----------------|------------|-----------|-----------|----------|
| X_non_test | OLS | 8 | no | yes | no | no | yes | 0.665 | 17705.84 | | | | |
| X_non_test_2 | OLS | 6 | no | yes | no | no | yes | 0.674 | 18342.57 | | | | |
| X_non_test_3 | OLS | 4 | no | no | no | no | yes | 0.649 | 18847.14 | | | | |
| X_non_test_4 | OLS | 5 | no | no | no | no | yes | 0.677 | 17824.94 | | | | |
| X_non_test_5 | OLS | 4 | no | no | no | no | yes | 0.680 | 17836.38 | | | | |
| X_non_test_6 | OLS | 5 | no | yes | no | no | yes | 0.688 | 17159.58 | | | | |
| poly_non_test_5_df | OLS | 15 | yes - degree 2 | yes | no | no | yes | 0.817 | 12092.47 | | | | |
| poly_non_test_6_df | OLS | 21 | yes - degree 2 | yes | no | no | yes | 0.796 | 12119.95 | | | | |
| poly_3_df | OLS | 35 | yes - degree 3 | yes | no | no | yes | 0.803 | 11073.62 | | | | |
| poly_non_test_6_df | OLS | 21 | yes - degree 2 | yes | no | log | no | 0.904 | 12349.53 | | | | |
| poly_non_test_5_df | OLS | 15 | yes - degree 2 | yes | no | log | no | 0.755 | 12349.50 | | | | |
| poly_non_test | OLS | 45 | yes - degree 2 | yes | no | box-cox | yes | 0.875 | 8748.95 | 0.887 | 8258.01 | 0.853 | 7813.63 |
| poly_non_test_5_df | OLS | 15 | yes - degree 2 | yes | no | box-cox | yes | 0.829 | 11515.84 | | | | |
| poly_non_test_6_df | OLS | 21 | yes - degree 2 | yes | no | box-cox | yes | 0.811 | 11545.78 | | | | |
| X_non_test_5 | OLS | 4 | no | no | no | box-cox | yes | 0.755 | 13536.04 | | | | |
| X_non_test | OLS | 8 | no | yes | no | box-cox | yes | 0.618 | 13931.11 | | | | |
| X_non_test | OLS | 8 | no | yes | no | 900 quantiles | yes | 0.689 | 14265.40 | | | | |
| X_non_test_5 | OLS | 4 | no | no | no | 900 quantiles | yes | 0.725 | 14020.86 | | | | |
| poly_non_test_5_df | OLS | 15 | yes - degree 2 | yes | no | 900 quantiles | yes | 0.815 | 11434.73 | | | | |
| poly_non_test_6_df | OLS | 21 | yes - degree 2 | yes | no | 900 quantiles | yes | 0.728 | 11694.08 | | | | |
| poly_non_test | OLS | 45 | yes - degree 2 | yes | no | 900 quantiles | yes | 0.867 | 9445.24 | 0.881 | 8703.11 | 0.854 | 8543.75 |
| poly_non_test_6_df | LassoCV | 21 | yes - degree 2 | yes | standard scaler | no | yes | 0.788 | 12786.83 | | | | |
| poly_non_test_6_df | RidgeCV | 21 | yes - degree 2 | yes | standard scaler | no | yes | 0.801 | 12546.25 | | | | |
| poly_non_test_6_df | ElasticNetCV | 21 | yes - degree 2 | yes | standard scaler | no | yes | 0.775 | 13098.56 | | | | |
| poly_non_test_6_df | ElasticNetCV | 21 | yes - degree 2 | yes | standard scaler | 900 quantiles | yes | 0.742 | 11793.69 | | | | |
| poly_non_test | ElasticNetCV | 45 | yes - degree 2 | yes | standard scaler | 900 quantiles | yes | 0.750 | 12104.47 | | | | |
| poly_non_test | ElasticNetCV | 45 | yes - degree 2 | yes | standard scaler | box-cox | yes | 0.753 | 11600.61 | | | | |

Visualizations relating to most accurate predictive model:

Box-Cox transformation of target variable:



Plot of predicted price versus residual on held out data:



Normal Q-Q plot of residuals for predictions on held out data:

