

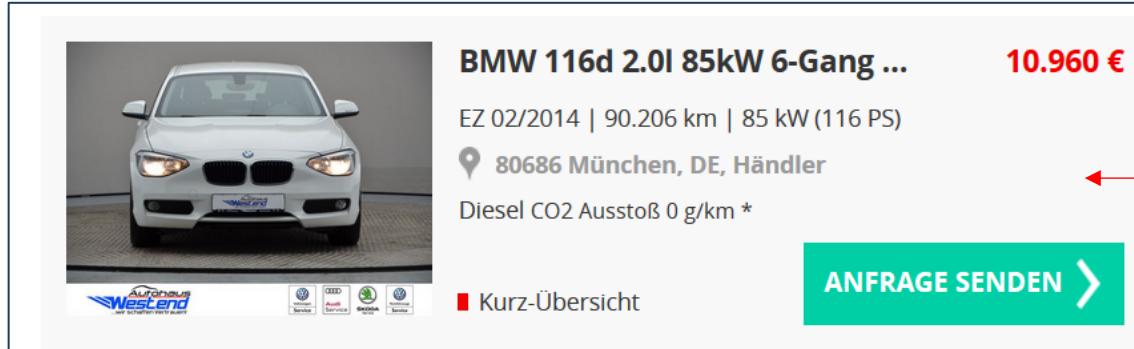


Finding a Good Value BMW 116

METIS DATA SCIENCE
BOOTCAMP

FEBRUARY 4, 2020

The challenge: finding good value offers amongst hundreds of BMW 116 ads



BMW 116d 2.0l 85kW 6-Gang ... **10.960 €**

EZ 02/2014 | 90.206 km | 85 kW (116 PS)

80686 München, DE, Händler

Diesel CO2 Ausstoß 0 g/km *

Kurz-Übersicht **ANFRAGE SENDEN >**



BMW 116d Klimaaut. LM PDC ... **10.980 €**

EZ 08/2013 | 79.003 km | 85 kW (116 PS)

59348 Lüdinghausen, DE, Händler

Diesel 4.1 l/100km (komb.) CO2 Ausstoß 109 g/km *

Kurz-Übersicht **ANFRAGE SENDEN >**



BMW 116d EfficientDynamics... **10.981 €**

EZ 02/2014 | 93.101 km | 85 kW (116 PS)

67227 Frankenthal, DE, Händler

Diesel 3.8 l/100km (komb.) CO2 Ausstoß 0 g/km *

Kurz-Übersicht **ANFRAGE SENDEN >**

BMW116

500+ ads

20+ configurable features

WHAT IS GOOD VALUE FOR MONEY?

Linear regression model predicting ‘market conform’ prices of used cars

A linear regression model was developed through an iterative approach

Approach

- 1) Web Scrapping (517 ads)
- 2) Data Cleansing and Exploratory Data Analysis
- 3) Modelling (train – validate – test)
 - Feature engineering
(adj. R-squared 0.74 → 0.84)
 - Model evaluation
 - Testing of assumptions

BeautifulSoup



Python
TimeDelta
Operation | Object | Example



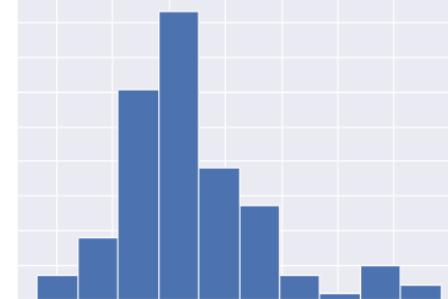
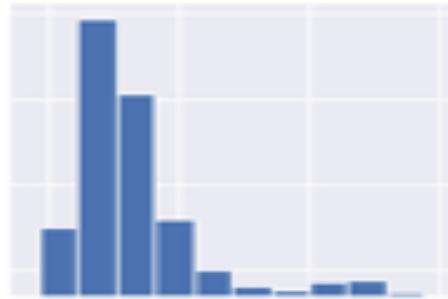
Seaborn

matplotlib

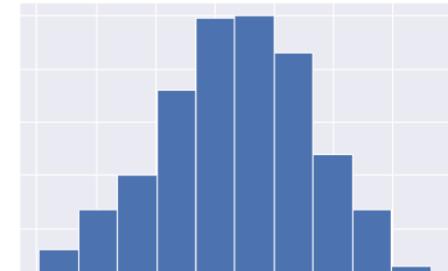
scikit
learn

Feature Engineering increased adjusted R-squared from 0.74 to 0.84

Log-transformation of target variable ‘price’



Box-Cox transformation of ‘emissions’



Adj. R-squared
0.74 → 0.84

Combining and dropping features

- Aircon type (dummy)
- Emissions
- Environmental certificate
- Prior owners
- Fuel type
- Cylinder capacity

The final model is statistically significant and yields an adjusted R-squared of 0.84

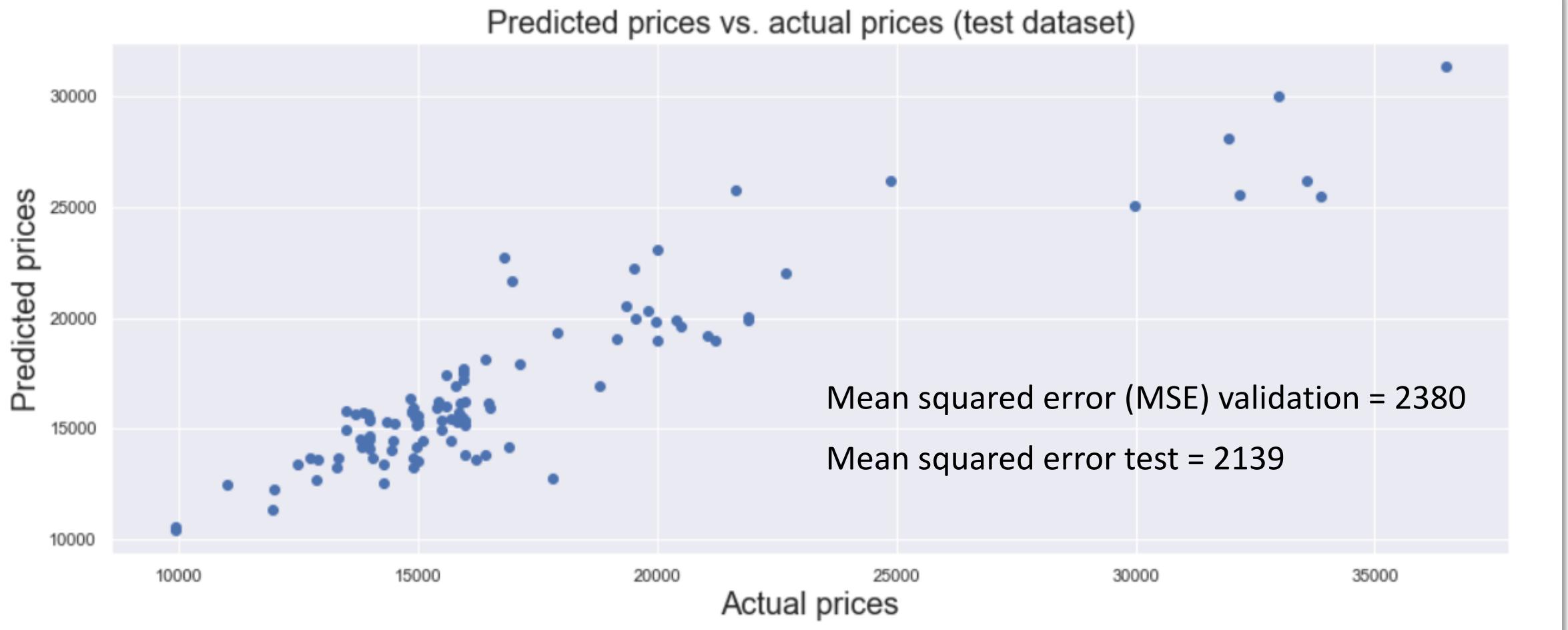
() sign of coefficient

$\ln(\text{Price}) = \text{Intercept}$
+ Horse power (+)
+ Dummy: 5 Doors (+)
+ Fuel consumption (-)
+ Dummy: Manual gears (-)
+ Box-Cox: Mileage (-)
+ Age (-)
+ aircon¹

- ✓ Adj. R-squared = **0.84**
- ✓ Prob(F-statistic) = **1.89e- 117**
- ✓ Skew = **0.606**
- ✓ Kurtosis = **3.651**
- ✓ Omnibus **21.381**

¹ Not statistically significant

The model works well for most test data, except highly expensive cars

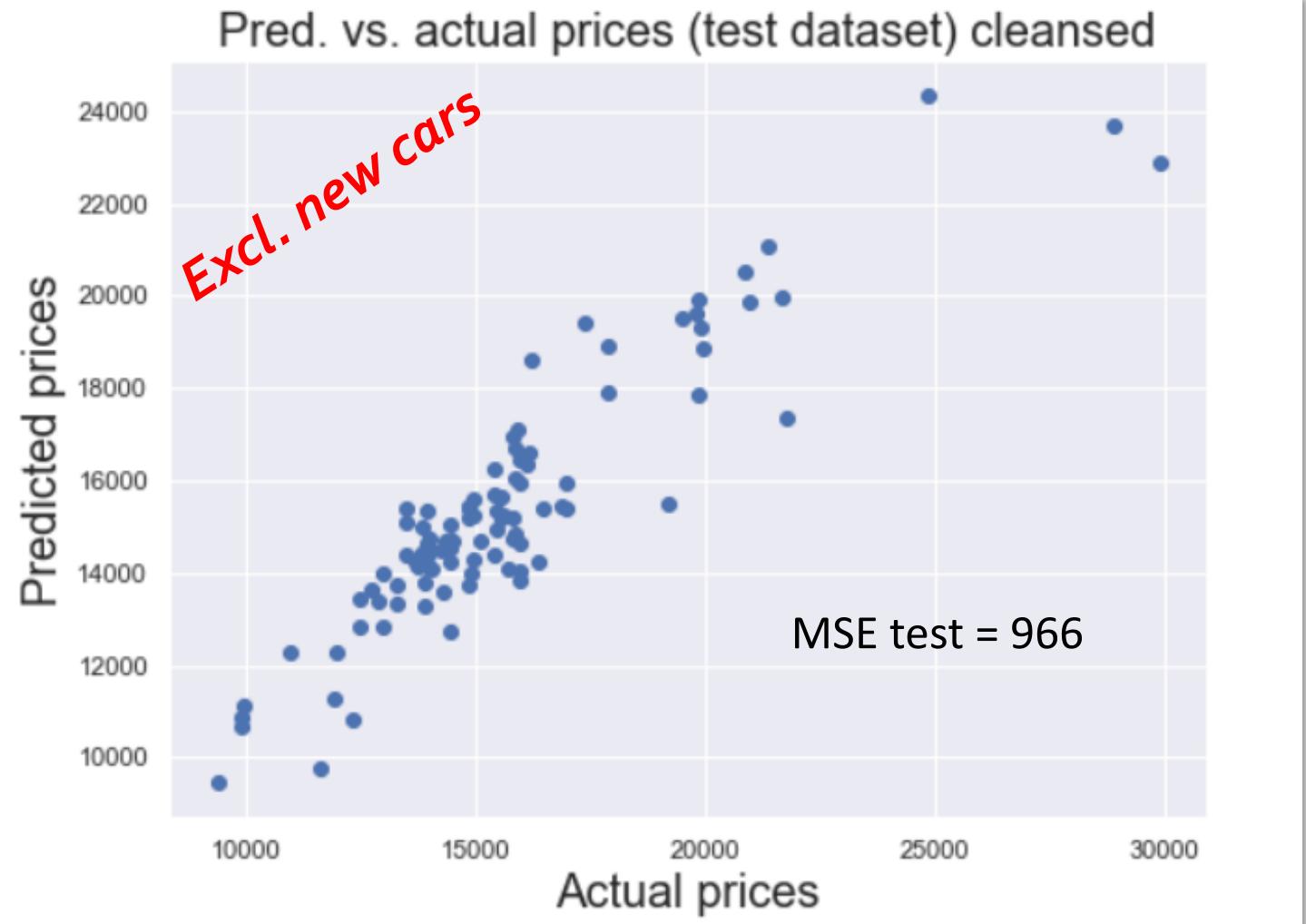


When excluding cars that are almost new, the model has a very high predictive power

Descriptive statistics of expensive cars (price >30k EUR)

- n = 25
- Avg. mileage = 4,139 km
- Avg. horse power = 116 PS
- Age = 0
- ...

All very expensive cars are almost new!



Use Cases of model for buyers and sellers of used cars

USED CARS

- › **Buyers:** Avoid overpaying
- › **Sellers:** Find appropriate selling price



NEW CARS

- › Select features that impact resell-value
- › Identify good time to sell



THANK YOU!



Diagnostic plots illustrate that regression is linear in parameters and correctly specified

