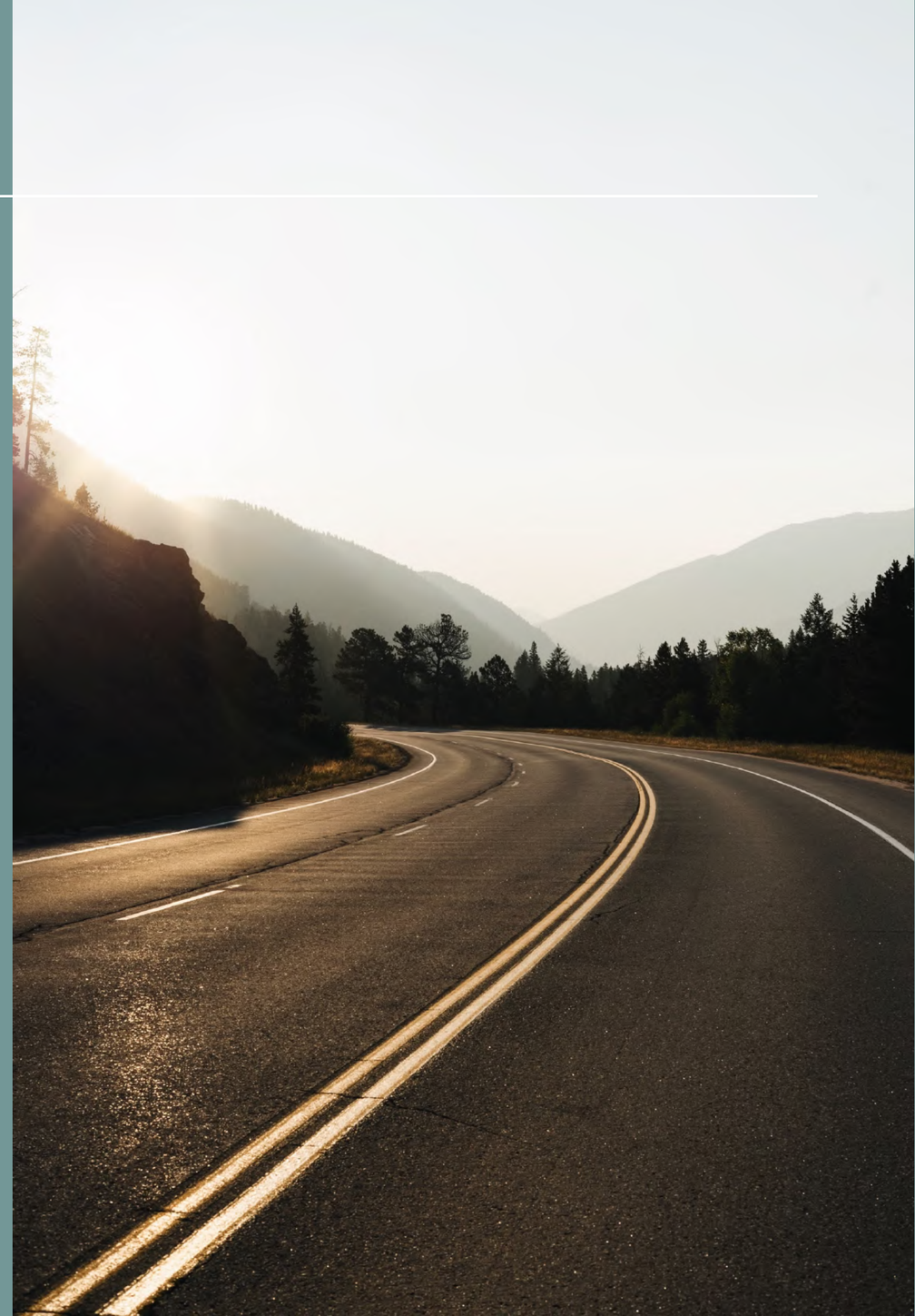
Maschinelles Lernen
Prüfungsstudienarbeit
2022

Daria Likhacheva
Olha Solodovnyk
Sofia Gutoranska

MOTORCYCLE PRICE PREDICTION

INHALT



Daria Likhacheva
Olha Solodovnyk
Sofia Gutoranska

DESKRIPTIVE ANALYSE

→ Summary der Daten

name	selling_price	year	seller_type	owner	km_driven	ex_showroom_price
Bajaj Pulsar 150 : 41	Min. : 5000	Min. :1988	Dealer : 6	1st owner:924	Min. : 350	Min. : 30490
Royal Enfield Classic 350 : 27	1st Qu.: 28000	1st Qu.:2011	Individual:1055	2nd owner:123	1st Qu.: 13500	1st Qu.: 54605
Honda Activa [2000-2015] : 23	Median : 45000	Median :2015		3rd owner: 11	Median : 25000	Median : 60122
Honda CB Hornet 160R : 22	Mean : 59638	Mean :2014		4th owner: 3	Mean : 34360	Mean : 79346
Bajaj Pulsar 180 : 20	3rd Qu.: 70000	3rd Qu.:2017			3rd Qu.: 43000	3rd Qu.: 80821
Royal Enfield Thunderbird 350: 19	Max. :760000	Max. :2020			Max. :880000	Max. :1278000
(Other) :909						

→ Boxplots und Histogramme für die metrischen Variablen ohne Ausreißer

→ Streudiagramme zwischen metrische Input Variablen und Output Variable

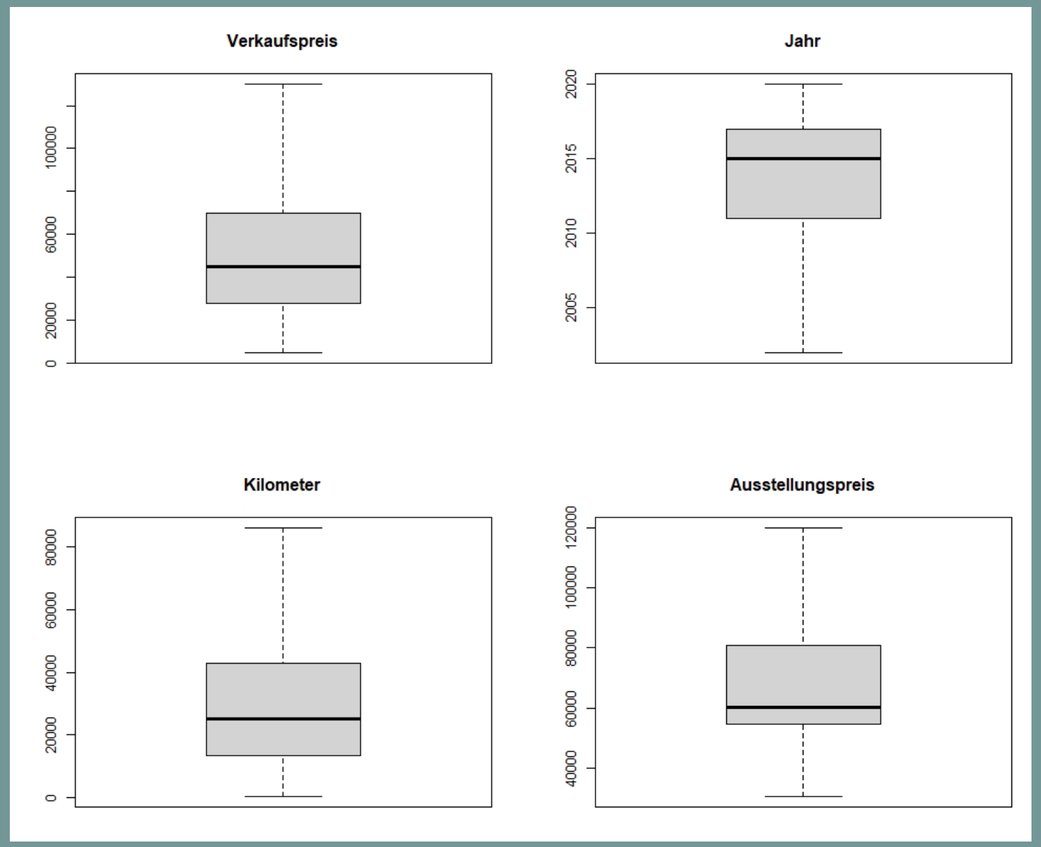
→ Boxplot für Verkaufspreis in Abhängigkeit von Besitzer

→ Boxplots für die Abhängigkeit zwischen Verkaufspreis und entsprechend Ausstellungspreis, Jahr und Kilometer

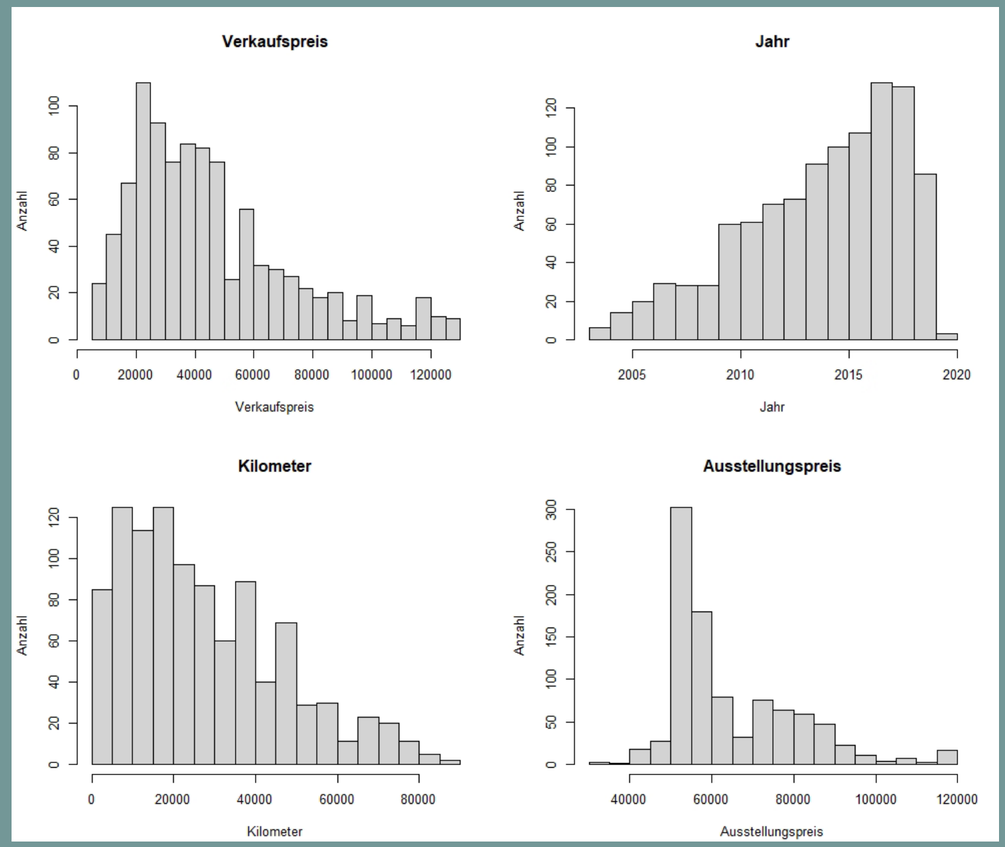
→ Berechnung der Korrelationen

- Verkaufspreis und Ausstellungspreis
0.7629427
- Verkaufspreis und Jahr
0.4021884

DESKRIPTIVE ANALYSE

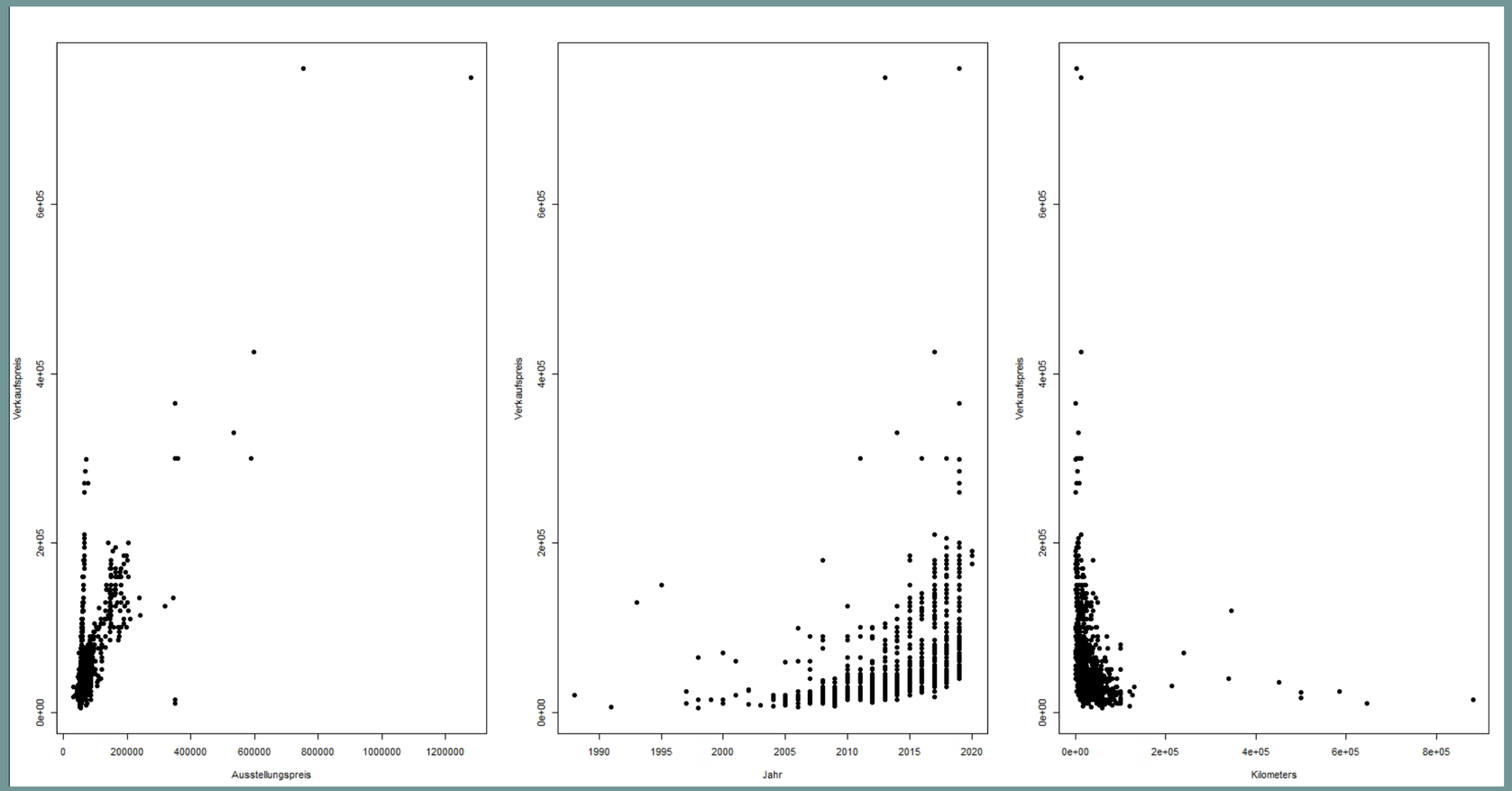


Boxplots für die metrischen Variablen ohne Ausreißer

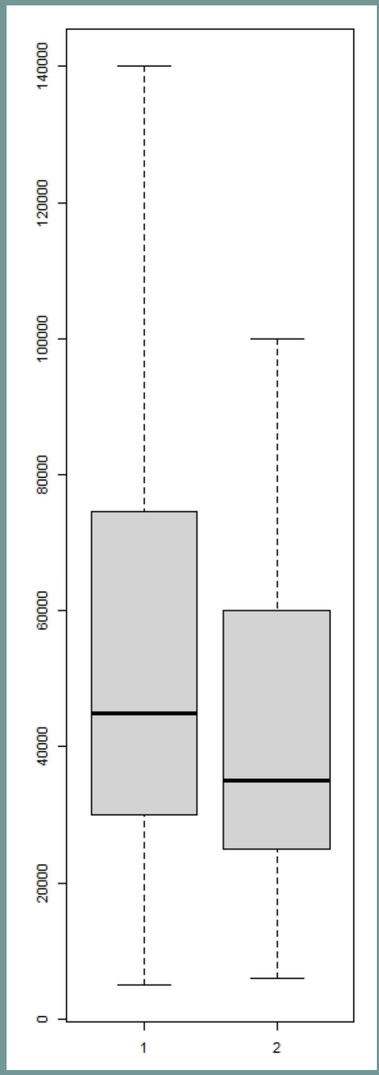


Histogramme für die metrischen Variablen ohne Ausreißer

DESKRIPTIVE ANALYSE

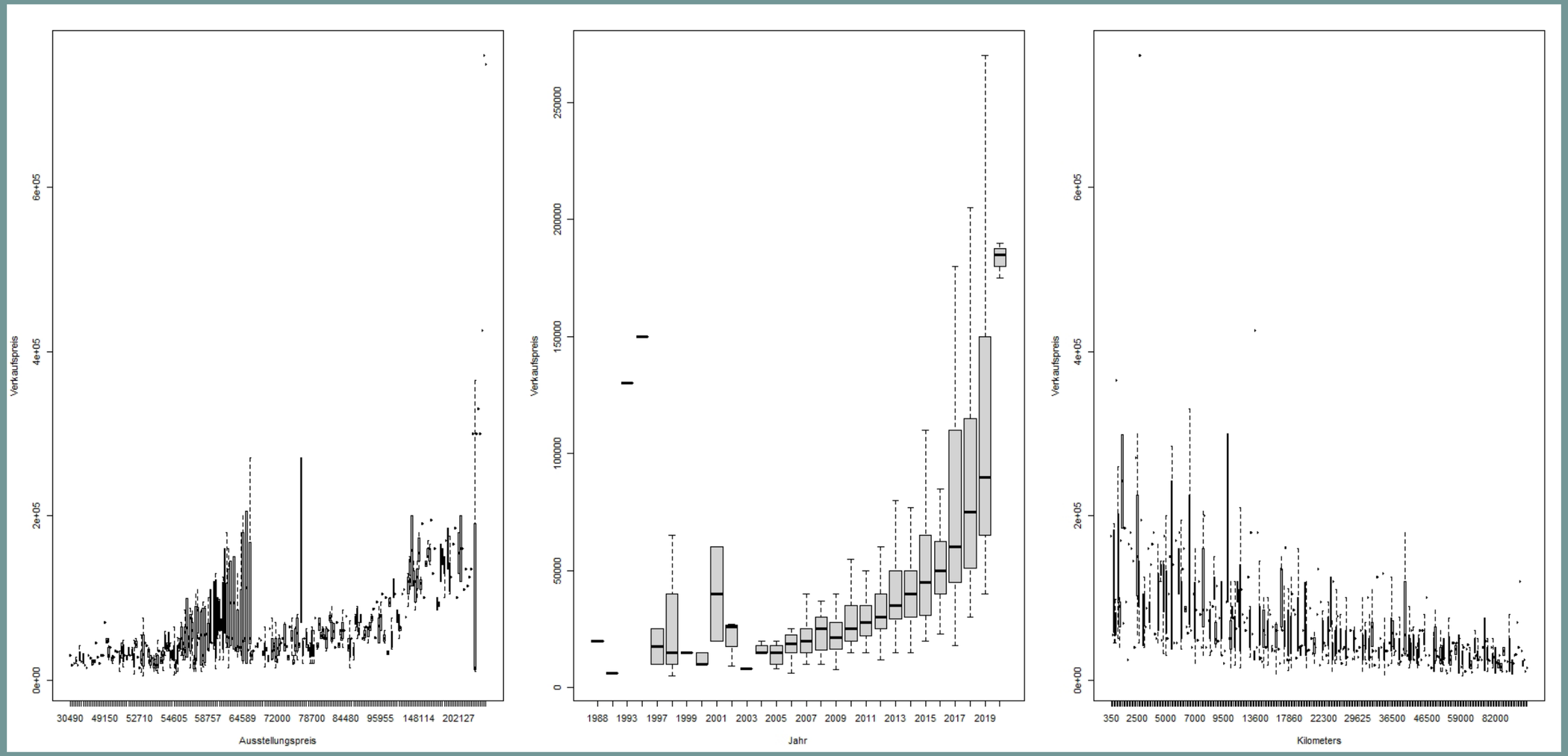


Streudiagramme zwischen metrische Input Variablen und Output Variable



Boxplot für "selling_price" in Abhängigkeit von "owner"

DESKRIPTIVE ANALYSE



Boxplots für die Abhängigkeit zwischen Verkaufspreis und entsprechend Ausstellungspreis, Jahr und Kilometer

Daria Likhacheva
Olha Solodovnyk
Sofia Gutoranska

LINEARE REGRESSION

Lineare Regression mit verschiedenen Einflussvariablen durchführen

→ Jahr + Km

MAPE = 37.5%

→ Jahr + Ausstellungspreis

MAPE = 38.4%

→ Km + Ausstellungspreis

MAPE = 50.6%

→ Km + Jahr + Ausstellungspreis

MAPE = 35.8%

* Mean absolute percentage error

Maschinelles Lernen
Prüfungsstudienarbeit
2022

Daria Likhacheva
Olha Solodovnyk
Sofia Gutoranska

K-NEAREST NEIGHBORS

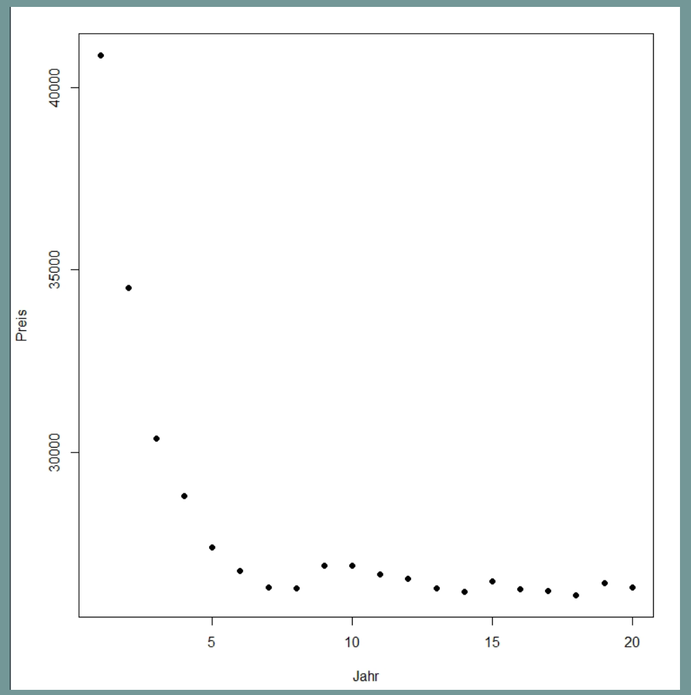
→ Prognostiziert wird der Verkaufspreis in Abhängigkeit von dem Jahr

→ Ausgabe der Güten für die verschiedenen Werte von k

```
> guete
 [1] 27566.50 36331.64 33724.05 33108.55 32255.67 32250.70 33146.77 33239.36 32191.26 32127.76 32066.86 32017.51
[13] 31436.82 31052.73 30924.87 30921.53 30365.95 29904.90 29483.50 29645.97
```

→ Ausgabe des optimalen k mit minimalen Prognosefehler → 1

→ Berechnung der Prognosen mit k-Nearest Neighbors für k=1 und des Prognosefehlers → 26189.31



Ausgabe der Gueten für die verschiedenen Werte von k

Maschinelles Lernen
Prüfungsstudienarbeit
2022

Daria Likhacheva
Olha Solodovnyk
Sofia Gutoranska

SVM

ZIELVARIABLE - selling_price

EINFLUSSVARIABLEN - name, year, km_driven, ex_showroom_price

ANZAHL DER DATEN - 910 (70% zu 30% Aufteilung)

```
> summary(Daten)
                              name     selling_price          year
Bajaj Pulsar 150                : 41   Min.   :  5000   Min.   :1988
Royal Enfield Classic 350       : 27   1st Qu.: 28000   1st Qu.:2011
Honda Activa [2000-2015]        : 23   Median : 45000   Median :2015
Honda CB Hornet 160R            : 22   Mean   : 59638   Mean   :2014
Bajaj Pulsar 180                : 20   3rd Qu.: 70000   3rd Qu.:2017
Royal Enfield Thunderbird 350   : 19   Max.   :760000   Max.   :2020
(Other)                         :909
    seller_type          owner       km_driven      ex_showroom_price
Dealer     :   6   1st owner:924   Min.   :   350   Min.   :  30490
Individual:1055   2nd owner:123   1st Qu.: 13500   1st Qu.:  54605
                  3rd owner: 11   Median : 25000   Median :  60122
                  4th owner:  3   Mean   : 34360   Mean   :  79346
                                  3rd Qu.: 43000   3rd Qu.:  80821
                                  Max.   :880000   Max.   :1278000
```

Zusammenfassung von Datensatz

```
> summary(Daten)
                              name     selling_price          year
Bajaj Pulsar 150                : 39   Min.   :  5000   Min.   :1988
Bajaj Pulsar 180                : 20   1st Qu.: 26625   1st Qu.:2011
Honda CB Hornet 160R            : 20   Median : 40000   Median :2015
Royal Enfield Thunderbird 350   : 19   Mean   : 48945   Mean   :2014
Honda Activa [2000-2015]        : 18   3rd Qu.: 60000   3rd Qu.:2017
Bajaj Discover 125              : 16   Max.   :299000   Max.   :2020
(Other)                         :778
    seller_type          owner       km_driven      ex_showroom_price
Dealer     :   5   1st owner:795   Min.   :   380   Min.   : 30490
Individual:905   2nd owner:104   1st Qu.: 15000   1st Qu.: 54586
                 3rd owner: 11   Median : 26000   Median : 58000
                 4th owner:  0   Mean   : 30021   Mean   : 65040
                                 3rd Qu.: 42000   3rd Qu.: 74295
                                 Max.   : 89000   Max.   :120000
```

Zusammenfassung von Datensatz ohne Ausreißer

Maschinelles Lernen
Prüfungsstudienarbeit
2022

Daria Likhacheva
Olha Solodovnyk
Sofia Gutoranska

SVM

Die ersten 10 Motorräder aus dem Datensatz (Datensatz ohne Ausreißer)

	name	selling_price	year	seller_type	owner	km_driven	ex_showroom_price
2	Honda Dio	45000	2017	Individual	1st owner	5650	54605
4	Yamaha Fazer FI V 2.0 [2016-2018]	65000	2015	Individual	1st owner	23000	89643
5	Yamaha SZ [2013-2014]	20000	2011	Individual	2nd owner	21000	54760
6	Honda CB Twister	18000	2010	Individual	1st owner	60000	53857
7	Honda CB Hornet 160R	78500	2018	Individual	1st owner	17000	87719
8	Royal Enfield Bullet 350 [2007-2011]	180000	2008	Individual	2nd owner	39000	64071
9	Hero Honda CBZ extreme	30000	2010	Individual	1st owner	32000	75502
10	Bajaj Discover 125	50000	2016	Individual	1st owner	42000	60122
11	Yamaha FZ16	35000	2015	Individual	1st owner	32000	78712
12	Honda Navi	28000	2016	Individual	2nd owner	10000	47255

selling_price Prognosen für die ersten 10 Motorräder

2	4	5	6	7	8	9	10	11	12
50018.807	61547.421	23439.479	21445.804	69923.547	108760.692	33436.681	31823.830	42131.089	31450.680

selling_price Prognosen für neues Motorrad

name	year	km_driven	ex_showroom_price	selling_price Prognosen
Honda CB Hornet 160R	2018	10 000	95 000	**72554.028**

Mittlere absolute Abweichung (MAD)

12.89267 %

Maschinelles Lernen
Prüfungsstudienarbeit
2022

Daria Likhacheva
Olha Solodovnyk
Sofia Gutoranska

L2-BOOSTING

ZIELVARIABLE -

- selling_price

EINFLUSSVARIABLEN -

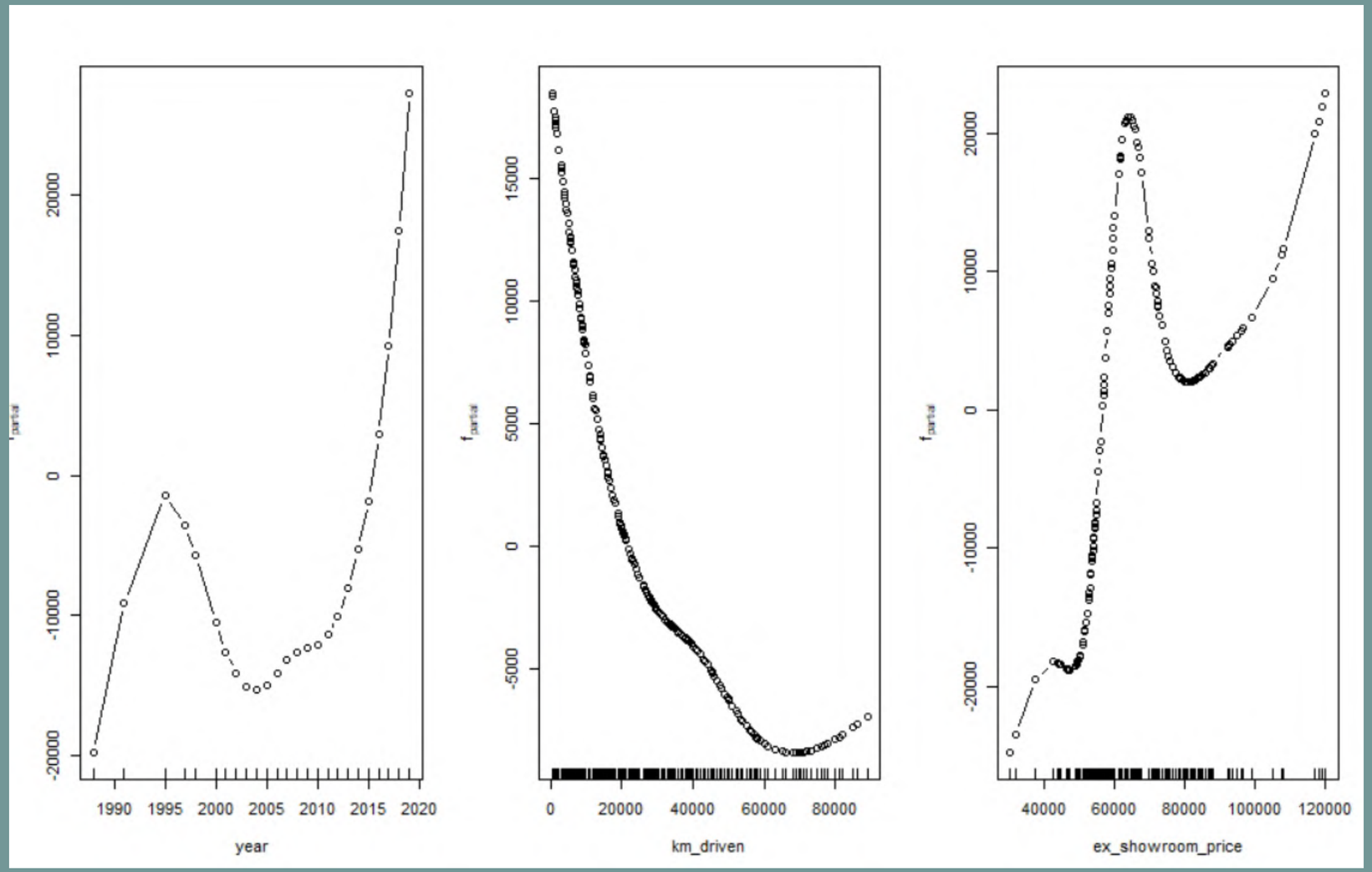
- year
- km_driven
- ex_showroom_price

ANZAHL DER DATEN -

- 910 (70% zu 30% Aufteilung)

Mittlere absolute Abweichung (MAD)

35.87828 %



Grafische Darstellung des Modells nach der Kreuzvalidierung

Daria Likhacheva
Olha Solodovnyk
Sofia Gutoranska

LASSO

Vorbereitung der Daten:
- Irrelevante Spalten entfernen
- Ausreißer entfernen
- Datensatz in Trainings- und Testdaten aufteilen

Berechnung von LASSO mit der restriktiveren Einstellung s="lambda.1se":
- Jahr und Km können als Einflussvariablen verwendet werden

Berechnung von LASSO mit der Standardeinstellung s="lambda.min" :
- Jahr, Km und Ausstellungspreis können als Einflussvariablen verwendet werden

Berechnung 100 mal wiederholen:
- Jahr, Km und Ausstellungspreis können als Einflussvariablen verwendet werden

Ergebnisse interpretieren:
- Für weitere Analyse werden die Variablen "Jahr", "Km" und "Ausstellungspreis"

Die Variablen, die häufig ausgewählt wurden:

year	100
km_driven	89
ex_showroom_price	14

Maschinelles Lernen
Prüfungsstudienarbeit
2022

Daria Likhacheva
Olha Solodovnyk
Sofia Gutoranska

DANKE FÜR IHRE AUFMERKSAMKEIT!