



# Praxis-Projekt

Faktoren der Co<sub>2</sub>-Emission von  
Autos

Niko Gaiser (804968)

Tim Jüstel (804123)

# Gliederung

1. Hintergrund
2. Leitfragen
3. Systematische Literaturrecherche
4. Daten
5. Explorativen Datenanalyse
6. Lineare Regression
7. Multiple Regression
8. KNN
9. Decision Tree

# 1. Hintergrund

- Naturkatastrophen, wie 2021 in NRW und RLP sind auf den menschengemachten Klimawandel zurückzuführen (Europäische Union 2021)
- Verkehrssektor ist für 20% der Treibhausgasemissionen verantwortlich (Umweltbundesamt 2021)

## 2. Leitfrage

**RQ1:** Inwiefern beeinflussen welche Merkmale die CO<sub>2</sub>-Emissionen von Kraftfahrzeugen?

**RQ2:** Wie lassen sich die CO<sub>2</sub>-Emissionen durch einen Machinelearning-Ansatz vorhersagen?

### 3. Systematische Literaturrecherche

ID	Suchterm	Suchmaschine	Ergebnisse
1	COMPONENTS AND AFFECTING AND CO2 EMISSIONS AND CARS	RG , WOS 89,	[Shi 20] [Fon 17]
2	CAR AND EMISSIONS AND MACHINE AND LEARNING	ACM 1412, WOS 35,	[Che 18] [Kar 10] [Kam 19] [Lee 21]

# 3. Systematische Literaturrecherche

**1.**

[Shi 20] – Faktoren, die Spritverbrauch und CO<sub>2</sub>-Emissionen bei Kraftfahrzeugen beeinflussen

[Fon 17] – Spritverbrauch und CO<sub>2</sub> Emissionen von Kraftfahrzeugen Labor vs. Echte Welt

**2.**

[Che 18] - Anwendung von Machinelearning-Algorithmen auf mit Sensor gemessene CO<sub>2</sub>-Werte

[Kar 10] - Die Rolle von Datamining bei der CO<sub>2</sub> Reduktion

[Lee 21] - Vorhersage der Emissionen von Dieselfahrzeugen mit Machinelearning-Ansatz

[Kam 19] – Vorhersage von Emissionsdaten anhand eines Machinelearning-Ansatzes

# 4. Daten

Verwendeter Datensatz: [CO2 Emission by Vehicles | Kaggle](https://www.kaggle.com/debajyotipodder/co2-emission-by-vehicles)

- Datensatz beschreibt, wie die Co2-Emissionen bei Autos von verschiedenen Features abhängen
- Datensatz ist von der Kanadischen Regierung
- Beinhaltet Daten in einem Zeitraum von 7 Jahren
- Insgesamt **7385 Zeilen** und **12 Spalten**

Quelle: (<https://www.kaggle.com/debajyotipodder/co2-emission-by-vehicles>)

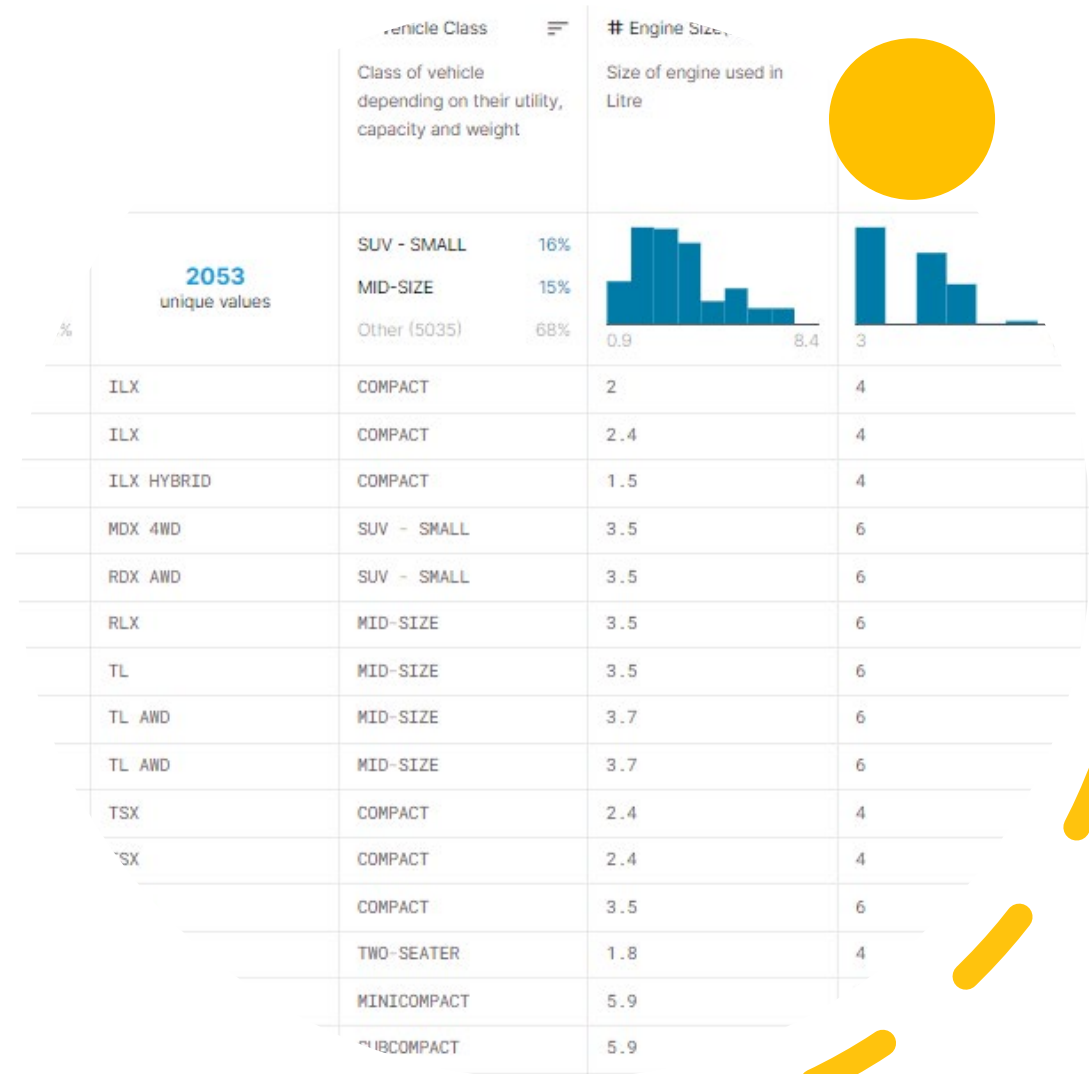


<https://media.canada.travel/de-DE/resources/canada-in-brief>

# 4. Daten

Spalten in Datensatz:

- Auto Hersteller
- Auto Modell
- Fahrzeugklasse
- Motorengröße (Hubraum)
- Anzahl an Zylindern
- Getriebe
- Kraftstoffart
- Kraftstoffverbrauch Innerorts (L/100 km)
- Kraftstoffverbrauch Außerorts (L/100 km)
- Kraftstoffverbrauch Kombiniert (L/100 km)
- Kraftstoffverbrauch Kombiniert (mpg)
- Co2-Emission (g/ 1km)



# 5. Explorativen Datenanalyse

Verwendete Software:



# 5. Explorativen Datenanalyse

```
import numpy as np # linear algebra

import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

import warnings

import seaborn as sns

import matplotlib.pyplot as plt

import plotly.express as px
import plotly.graph_objects as go

from plotly.subplots import make_subplots
```

Import der benötigten Bibliotheken

# 5. Explorativen Datenanalyse

- Erste Reduzierung der Merkmale
  1. Kraftstoffverbrauch Innerorts (L/100 km)
  2. Kraftstoffverbrauch Außerorts (L/100 km)
  3. Kraftstoffverbrauch Kombiniert (mpg)

# 5. Explorativen Datenanalyse

```
emissions = pd.read_csv(r'C:\Users\timju\OneDrive\Documents\Hochschule  
Reutlingen\Formale Methoden\Praxisprojekt\CO2 Emissions_Canada.csv' ,  
usecols = [ 'Make', 'Model', 'Vehicle Class', 'Engine  
Size(L)', 'Cylinders', 'Transmission', 'Fuel Type', 'Fuel Consumption Comb  
(L/100 km)', 'CO2 Emissions(g/km)' ])
```

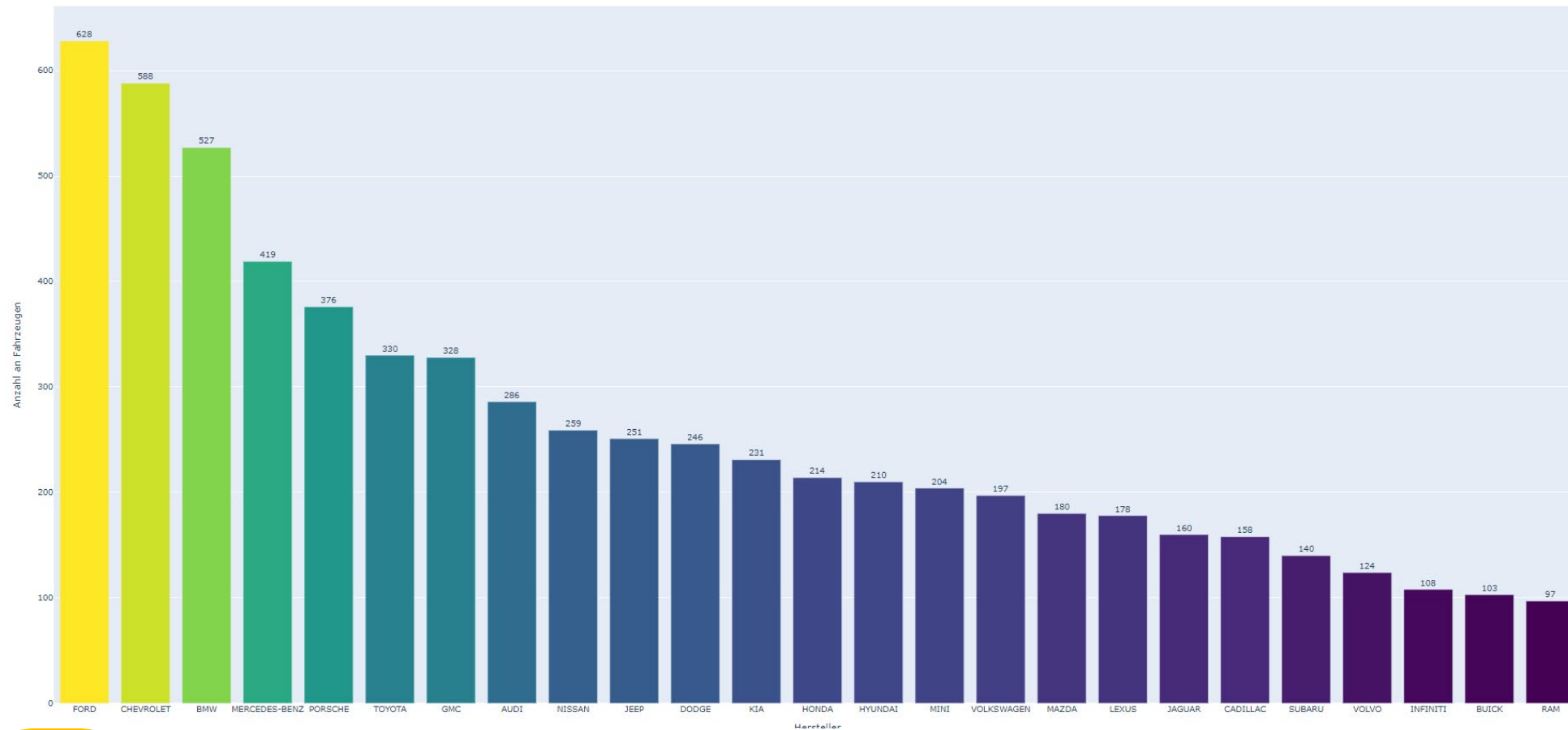
Import des Datensatzes

# 5. Explorativen Datenanalyse

```
renamed_col = {  
    'Make': 'Hersteller',  
    'Vehicle Class': 'Fahrzeug_Klasse',  
    'Engine Size(L)': 'Hubraum',  
    'Fuel Type': 'Kraftstofftyp',  
    'Fuel Consumption Comb (L/100 km)': 'Spritverbrauch_Kombiniert',  
    'CO2 Emissions(g/km)': 'co2(g/km)',  
    'Cylinders': "Zylinder"  
}
```

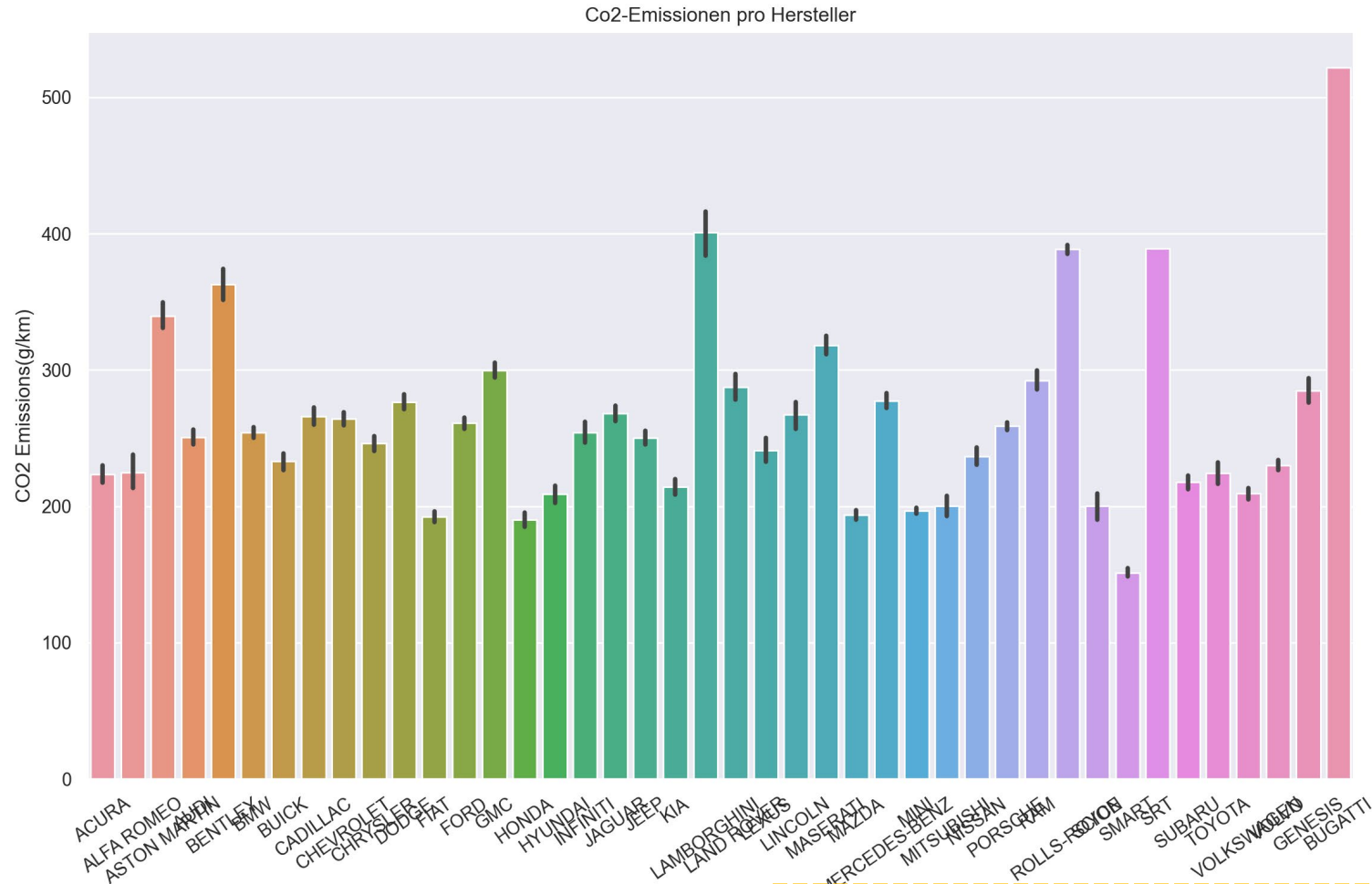
Umbenennung der Spalten

# 5. Explorativen Datenanalyse



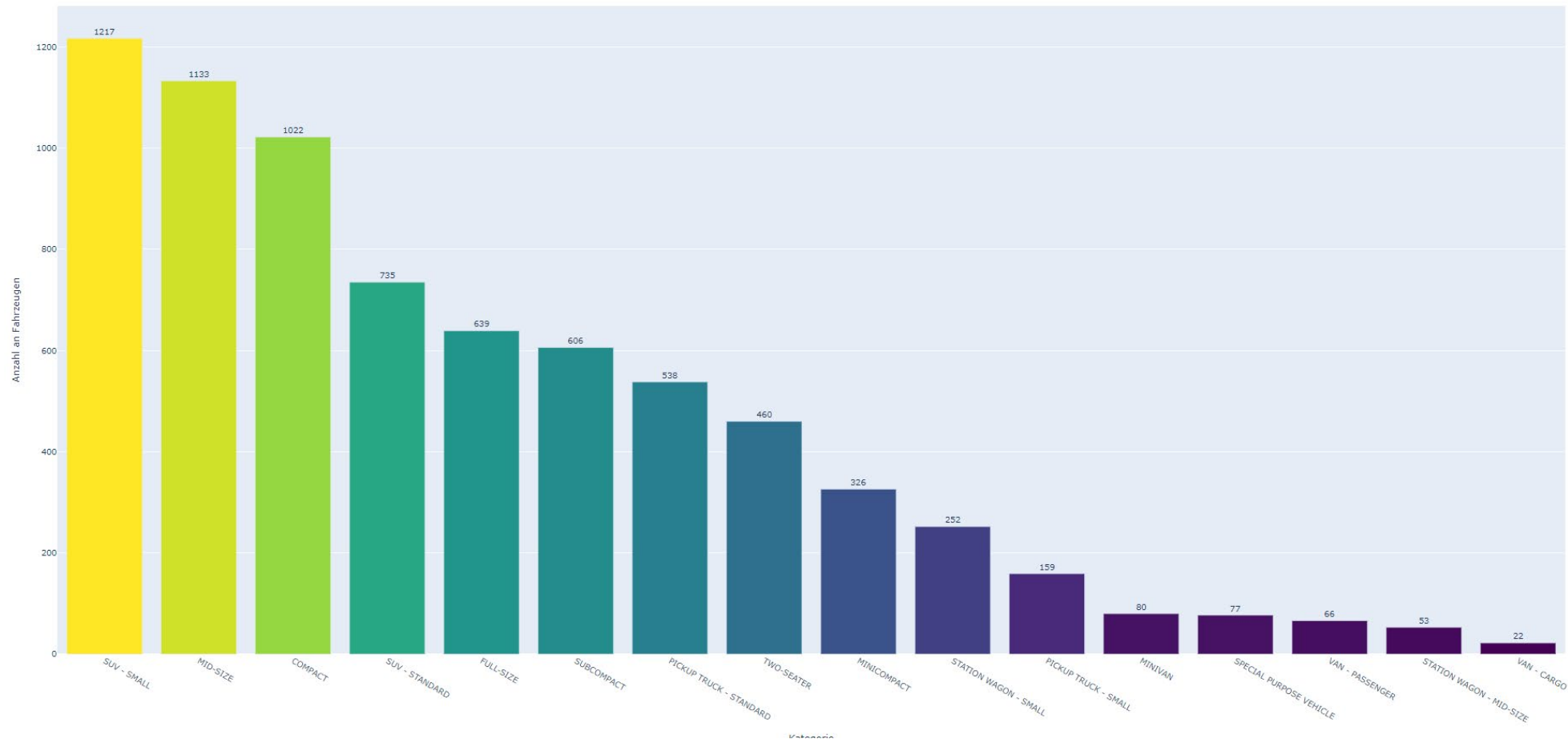
Verteilung Hersteller (TOP 25)

# 5. Explorativen Datenanalyse



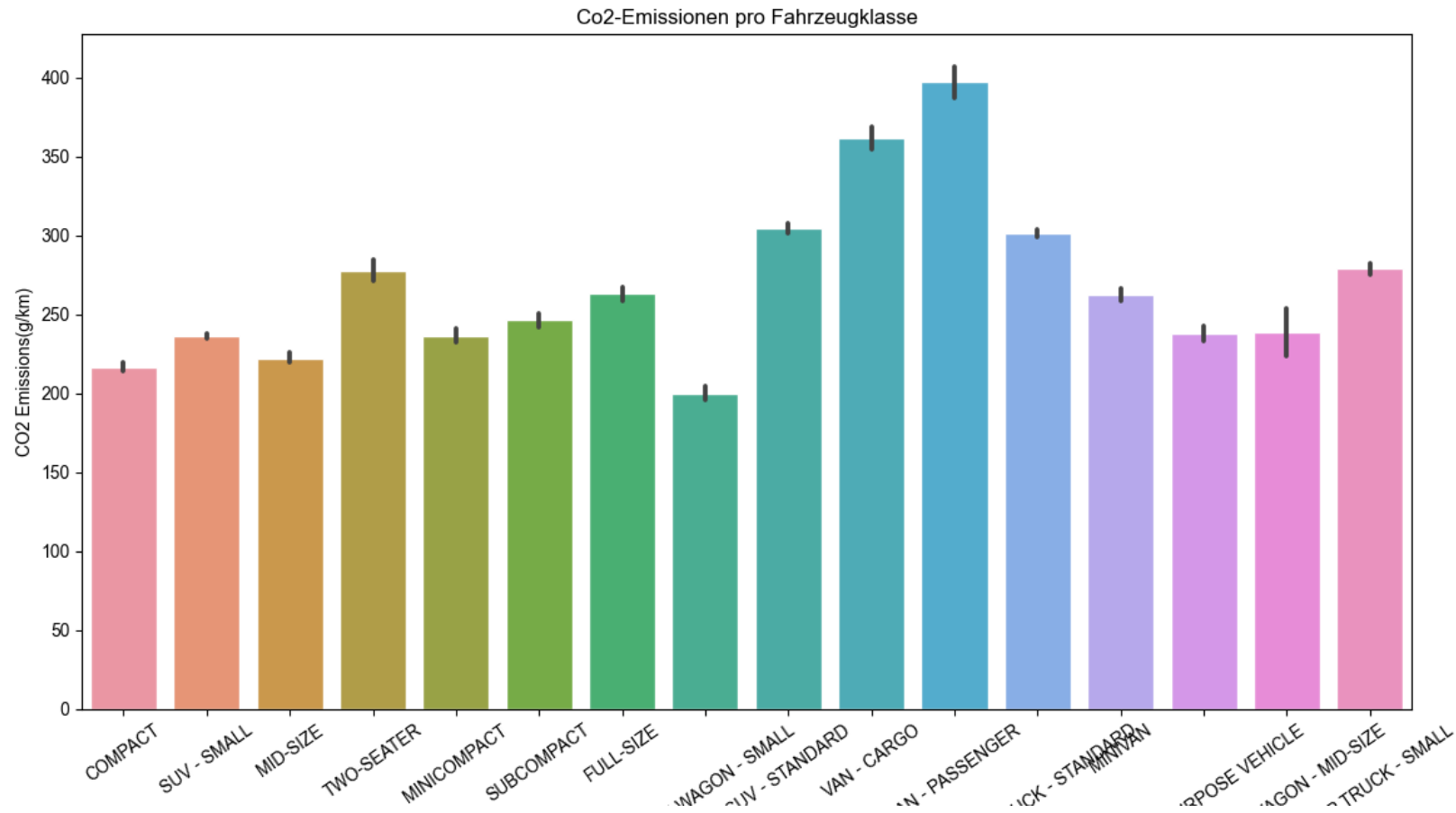
Co2-Emissionen Hersteller

# 5. Explorativen Datenanalyse



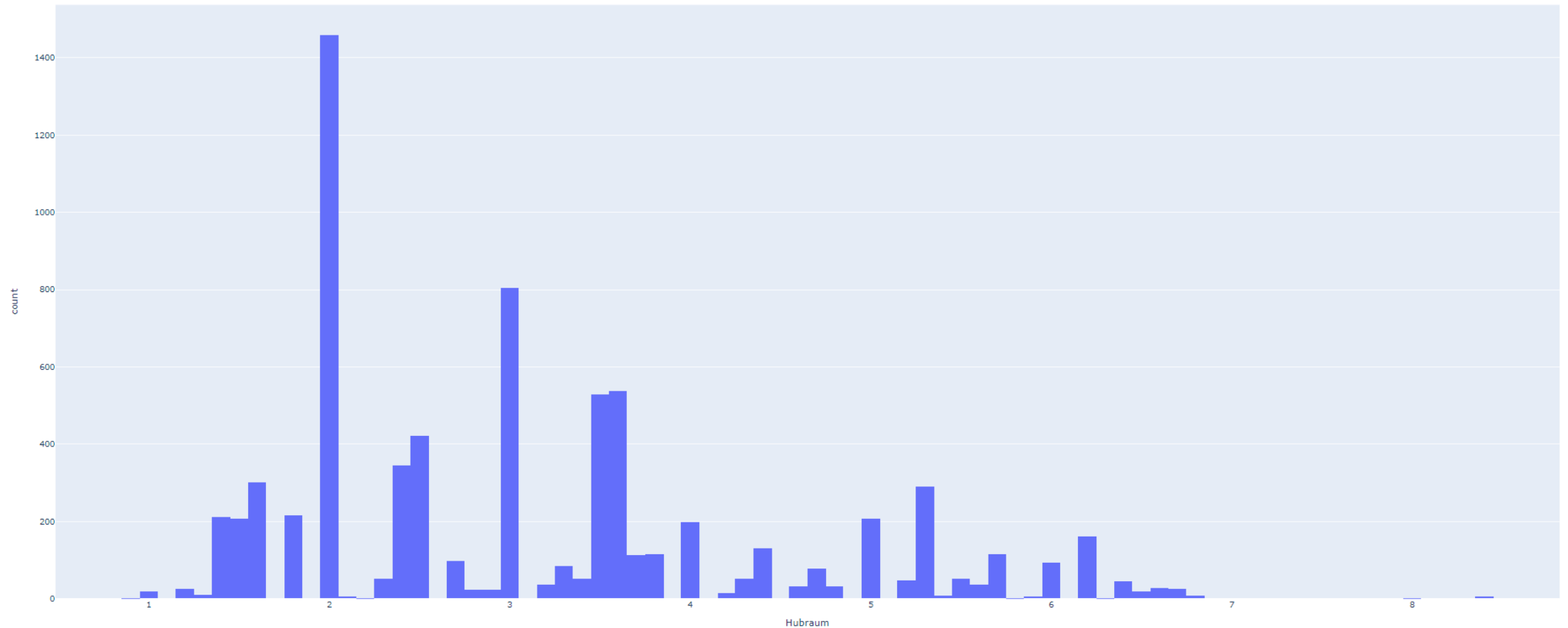
Verteilung Fahrzeugkategorie

# 5. Explorativen Datenanalyse



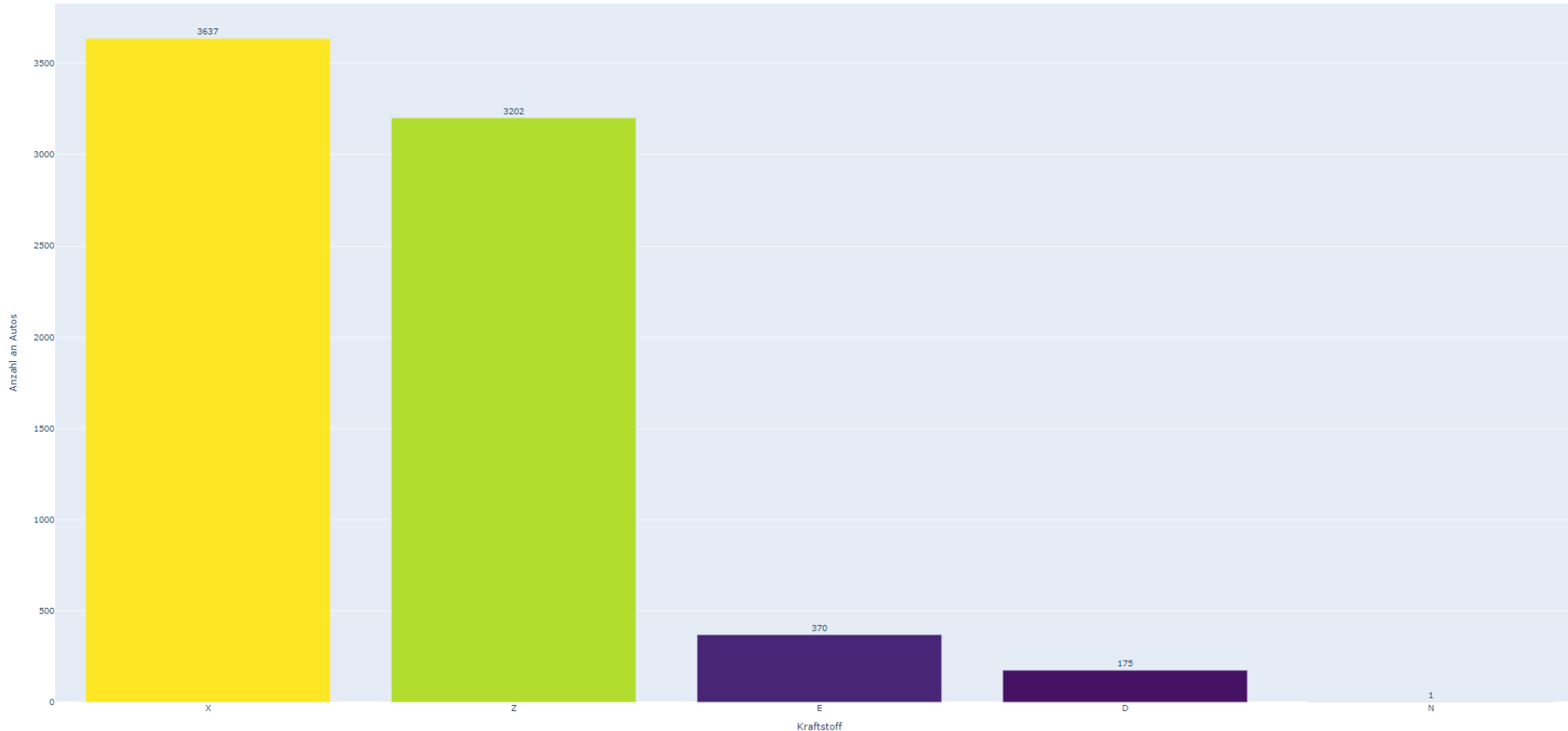
Co2-Emissionen pro Fahrzeugkategorie

# 5. Explorativen Datenanalyse



Verteilung Hubraum

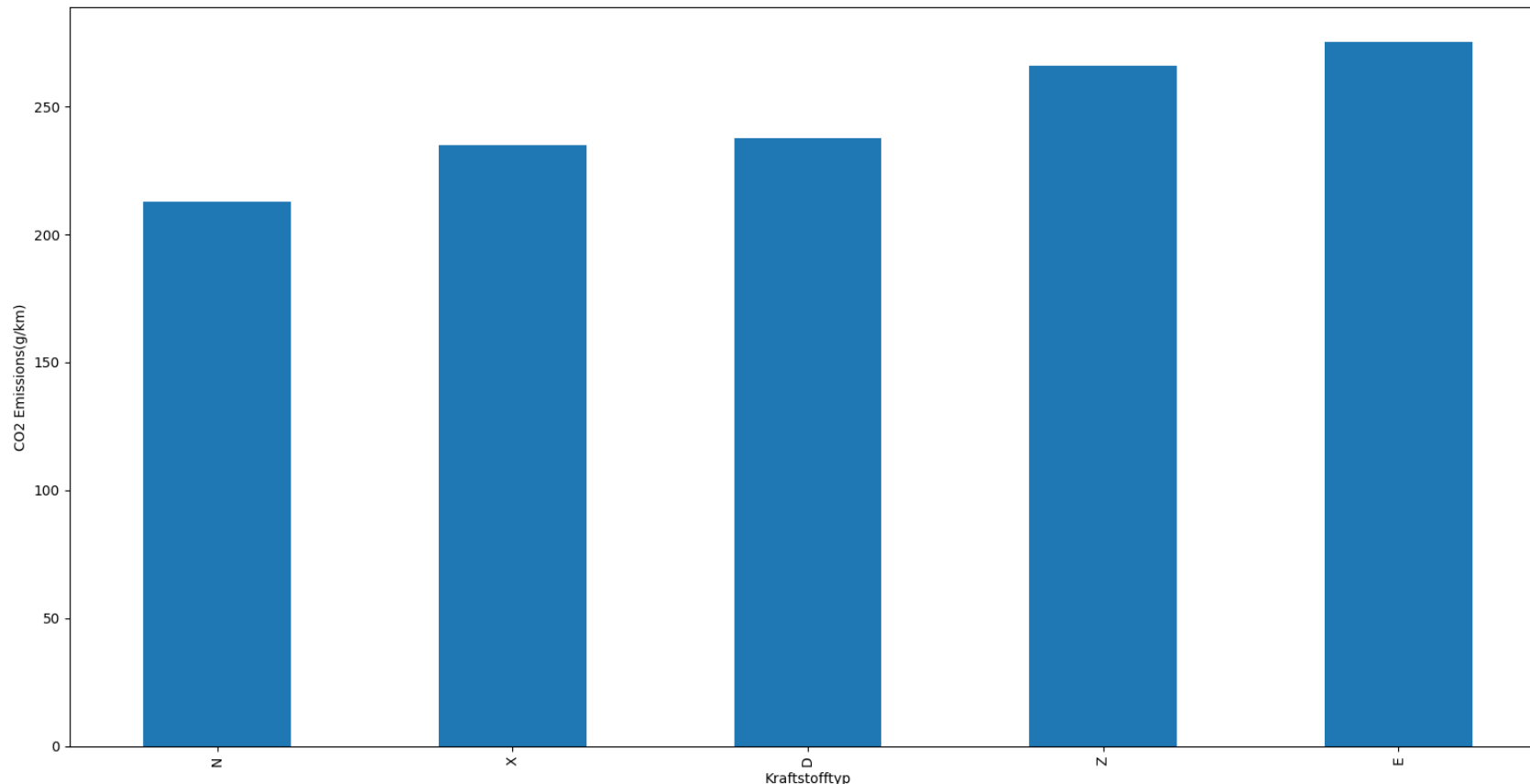
# 5. Explorativen Datenanalyse



X = Normales Benzin  
Z = Premium Benzin  
D = Diesel  
E = Ethanol (E85)  
N = Gas

Verteilung Kraftstoffe

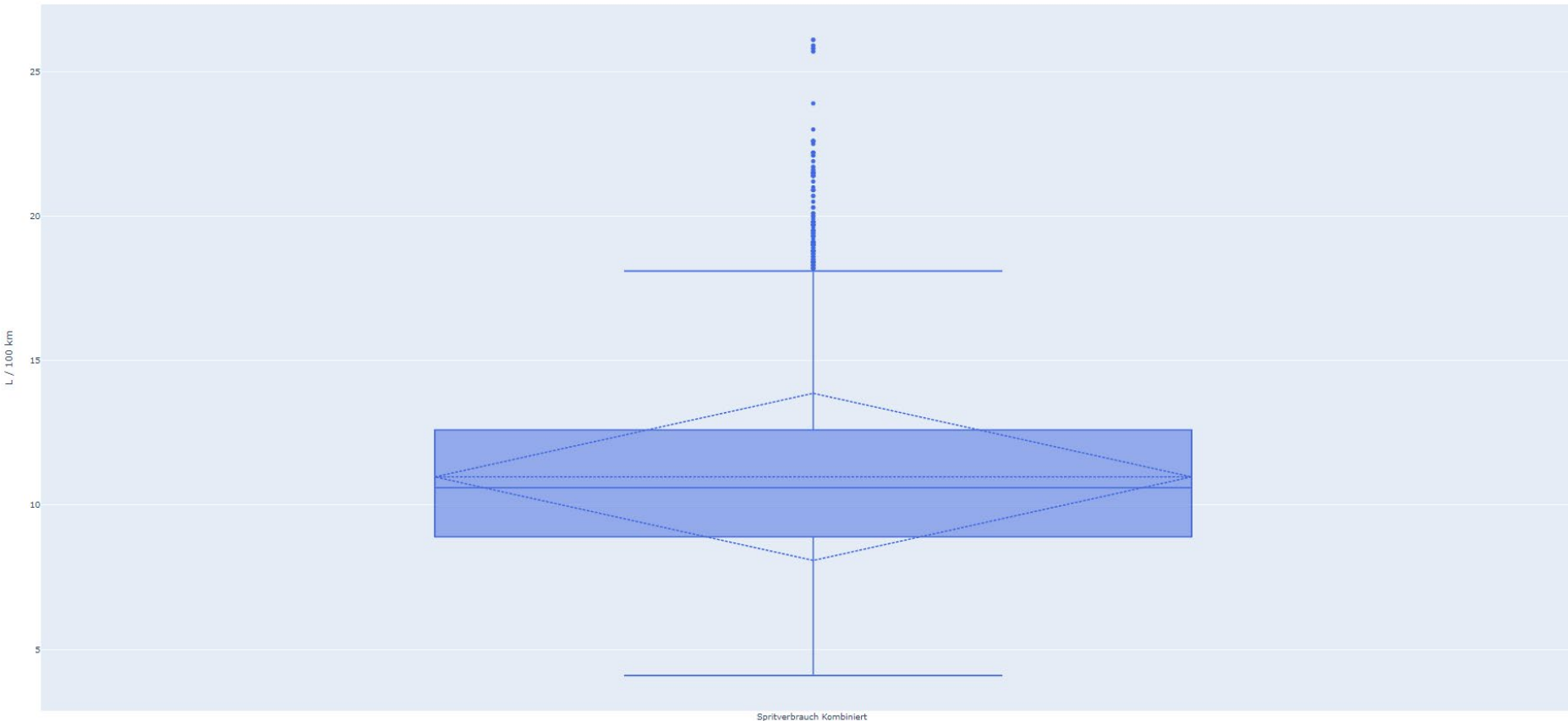
# 5. Explorativen Datenanalyse



X = Normales Benzin  
Z = Premium Benzin  
D = Diesel  
E = Ethanol (E85)  
N = Gas

Verteilung CO2-Ausstoß Kraftstoffe

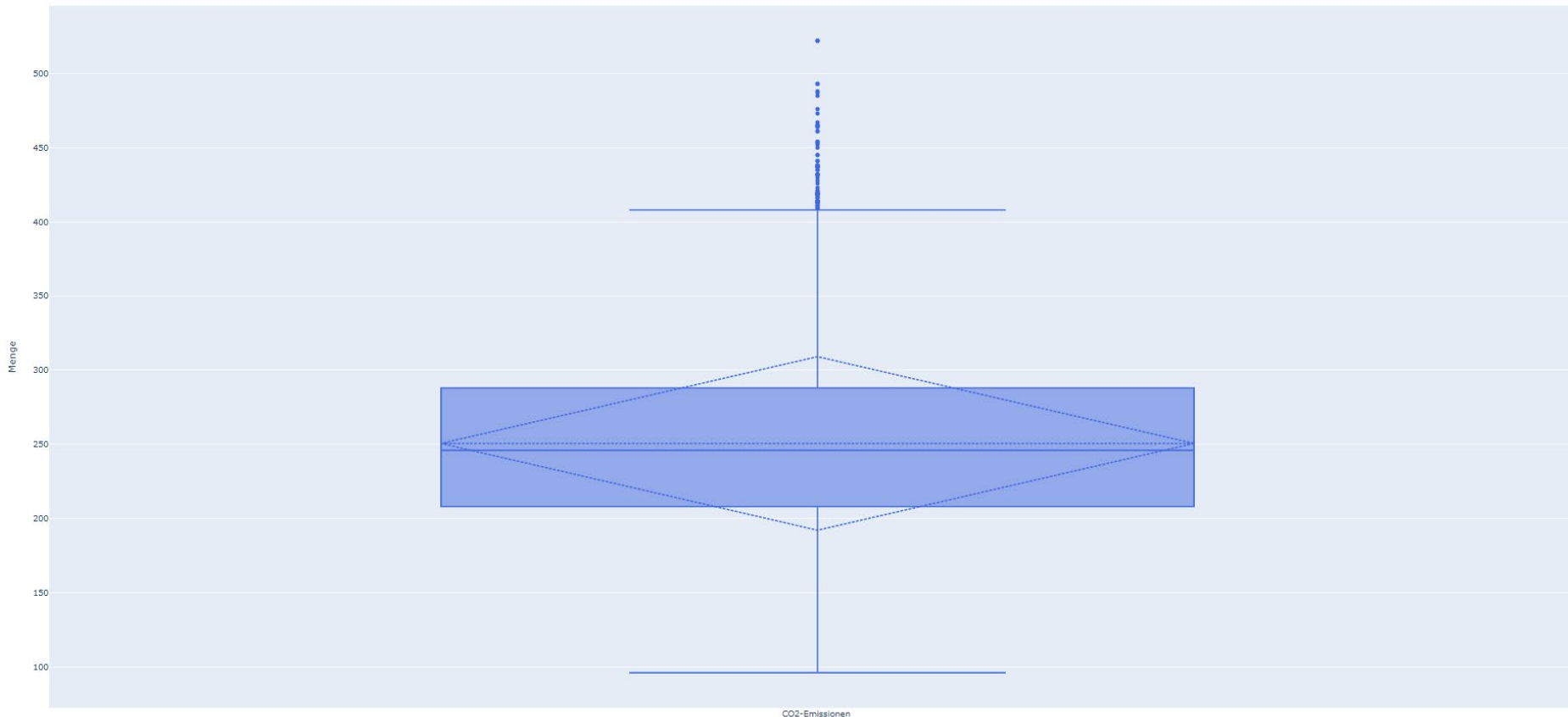
# 5. Explorativen Datenanalyse



Median: 10,6 L/km  
Q1: 8,9 L/km  
Q3: 12,6 L/km  
SD: 2,89 L/km  
Max = 26,1 L/km

Verteilung Benzinverbrauch kombiniert

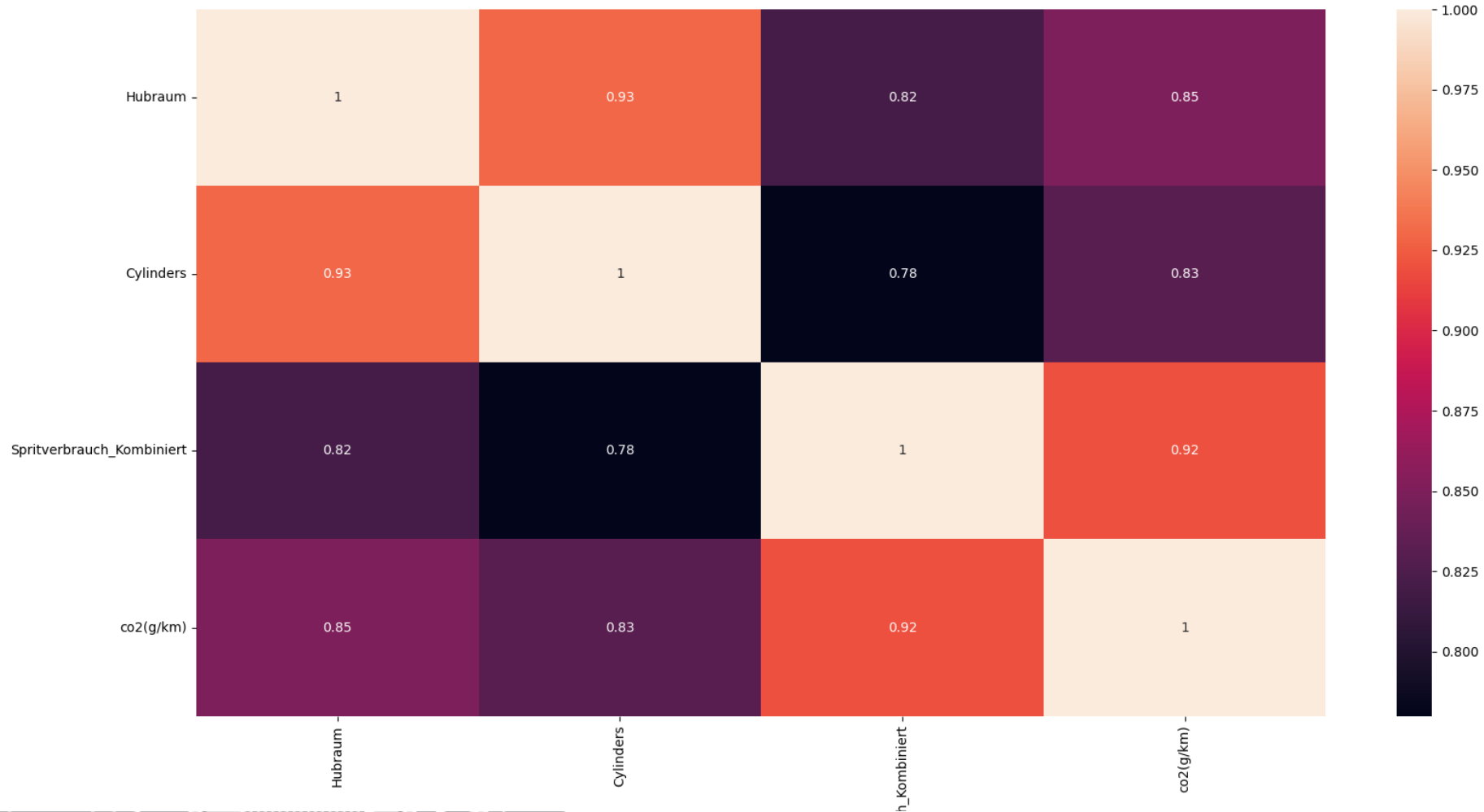
# 5. Explorativen Datenanalyse



Median: 246 g/km  
Q1: 208 g/km  
Q3: 288 g/km  
SD: 58,5 g/km  
Max = 522 g/km

Verteilung CO2-Emissionen

# 5. Explorativen Datenanalyse



# 6. Implementierung Maschinelle Lernmethoden

## Datenaufbereitung:

- Plausibilisierung
- Prüfen ob Daten Vollständig
- Transformation der Daten

# 6. Implementierung Maschinelle Lernmethoden

- Zweite Reduzierung der Merkmale

1. Auto Hersteller
2. Auto Modell
3. Fahrzeugklasse
4. Getriebe
5. Kraftstoffart

→ Reduzierung auf numerische Werte

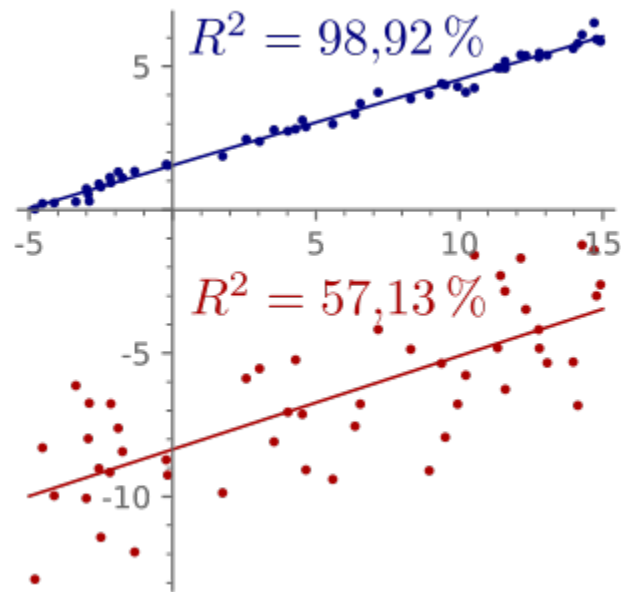
# 6. Implementierung Maschinelle Lernmethoden

- Einfache Regression
- Multiple Regression
- K-Nearest Neighbor (KNN)
- Decision Tree

scikit-learn

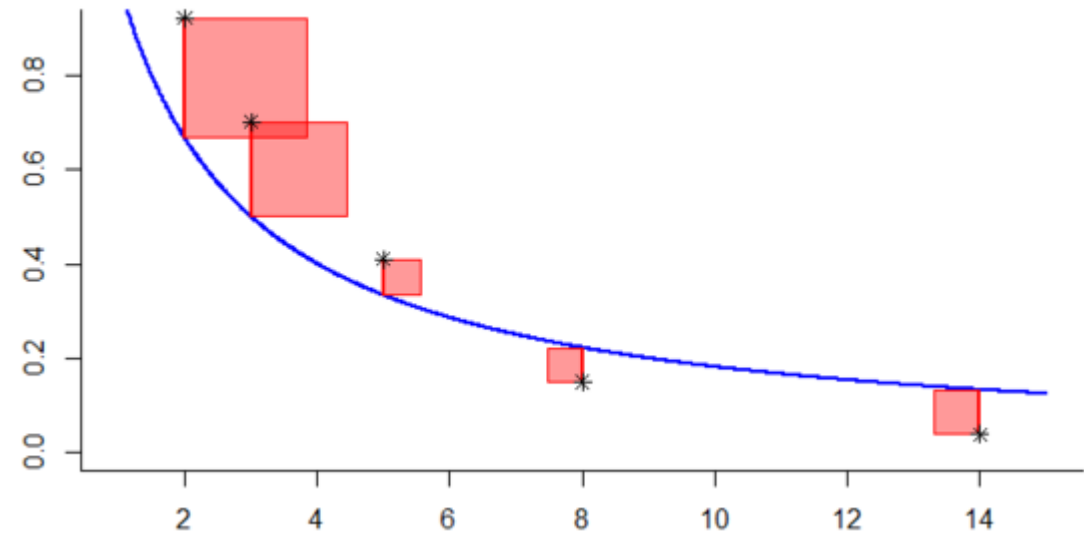
# 6. Implementierung Maschinelle Lernmethoden

Bestimmtheitsmaß



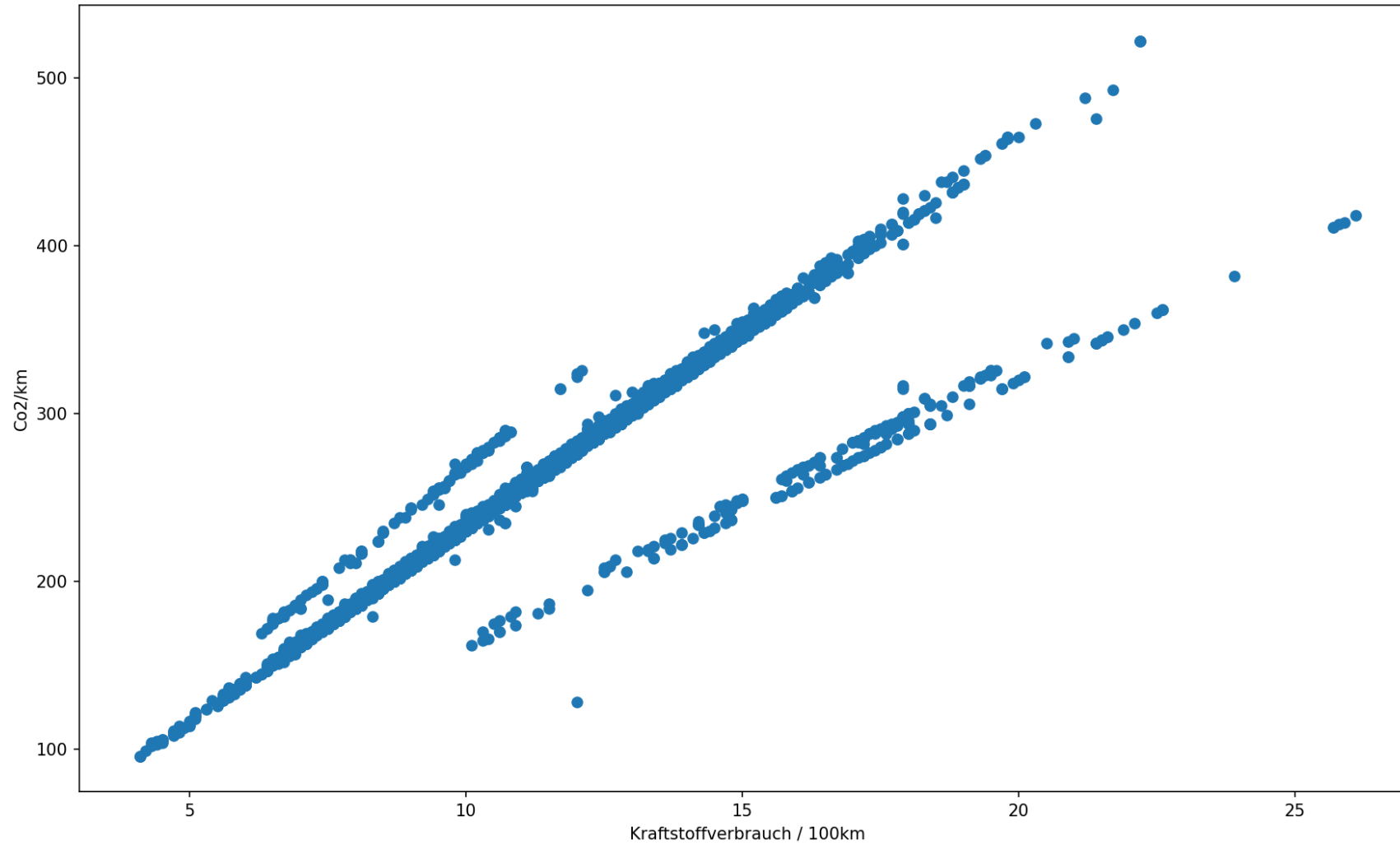
<https://de.wikipedia.org/wiki/Bestimmtheitsmaß>

Residuenquadratsumme (MSE)



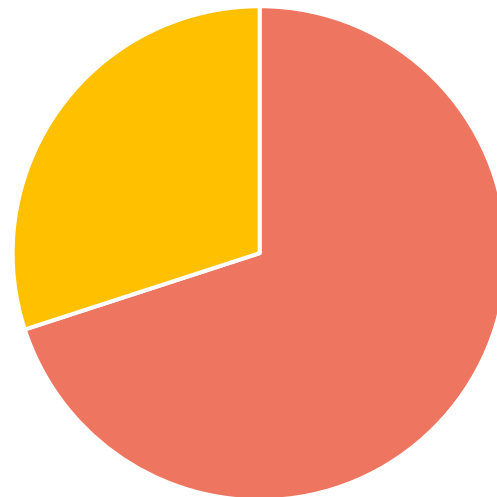
<https://de.wikipedia.org/wiki/Bestimmtheitsmaß>

# 6. Lineare Regression



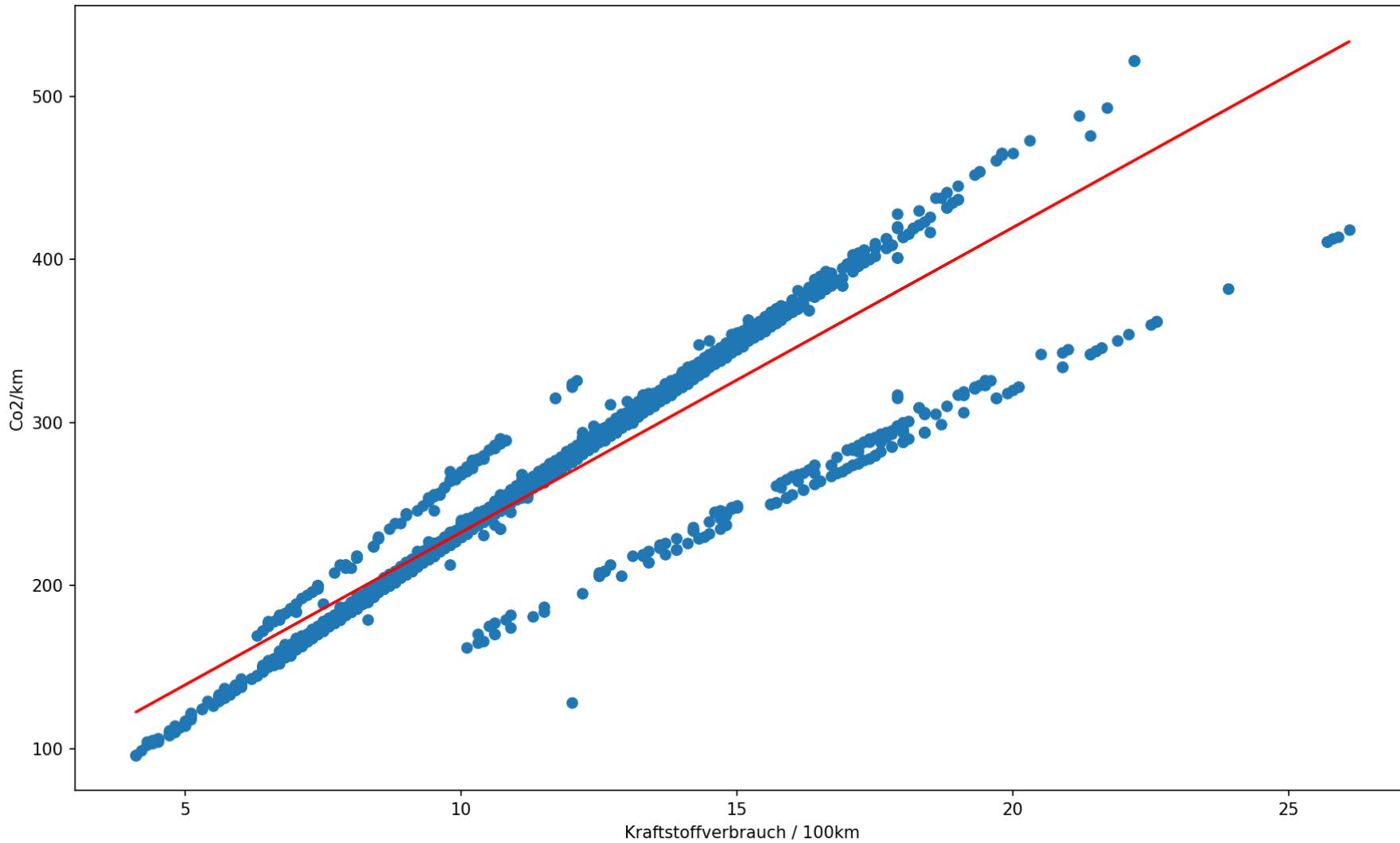
# 6. Lineare Regression

- Erklärende Variable = **Spritverbrauch Kombiniert | Hubraum**
- Zu erklärende Variable = **CO2 /km**



■ 1. Training ■ 2. Test

# 6. Lineare Regression Spritverbrauch Kombiniert



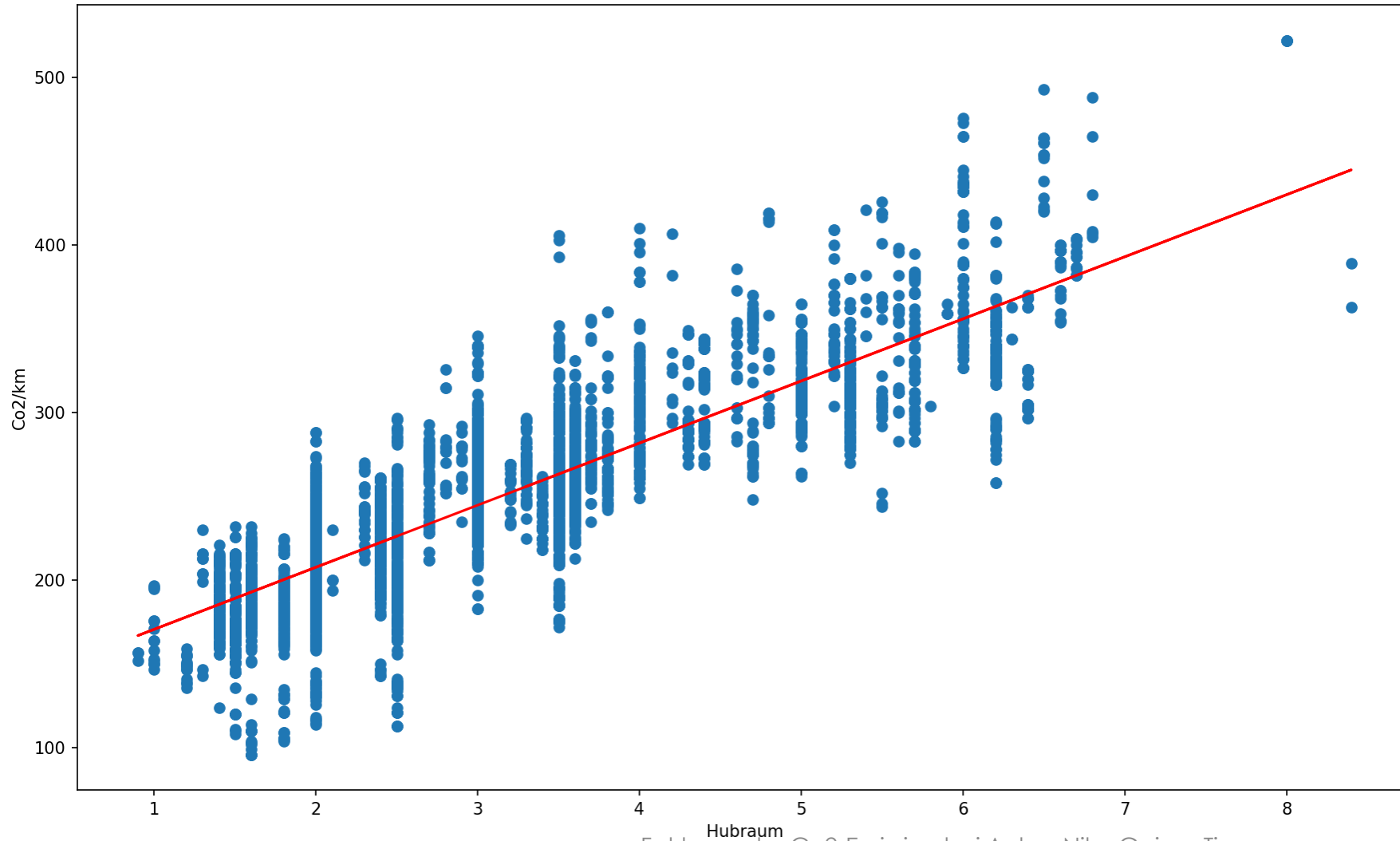
R2 Wert: 0.8047  
MSE: 558.79

## 6. Lineare Regression

```
slr = LinearRegression()
y_col = "co2(g/km)"
X = emissions.drop(y_col, axis=1)
y = emissions[y_col]
X = emissions[['Hubraum']]
X_training, X_test, y_training, y_test = train_test_split(X, y, test_size=0.3, random_state=72018) #split into train and test parts
|
slr.fit(X_training,y_training)
plt.scatter(X_training,y_training)
plt.xlabel('Hubraum' )
plt.ylabel('Co2/km')
plt.plot(X_training, slr.predict(X_training), color='red')
y_pred = slr.predict(X_test)
y_pred = slr.predict(X_test)

print(f'R2 Wert: {r2_score(y_pred,y_test)}')
print("Residuenquadratsumme (MSE): %.2f" % np.mean((y_pred - y_test) ** 2))
```

# 6. Lineare Regression Hubraum



R2 Wert: 0.644  
MSE: 895.56

# 7. Multiple Regression

- Erklärende Variable = Spritverbrauch Kombiniert & Hubraum & Zylinder
- Zu erklärende Variable = CO2 /km

```
slr = LinearRegression()
y_col = "co2(g/km)"
#X = emissions.drop(y_col, axis=1)
y = emissions[y_col]
X = emissions[['Hubraum', 'Spritverbrauch_Kombiniert', 'Zylinder']]
X_training, X_test, y_training, y_test = train_test_split(X, y, test_size=0.3, random_state=72018) #split into train and test parts

y_pred = slr.predict(X_test)

print(f'R2 Wert: {r2_score(y_pred,y_test)}')
print("Residuenquadratsumme (MSE): %.2f" % np.mean((y_pred - y_test) ** 2))
```

R2 Wert: 0.8589

MSE: 421.15

# 8. KNN

- Erklärende Variable = **Spritverbrauch Kombiniert & Hubraum & Zylinder**
- Zu erklärende Variable = CO2 /km

```
knn_model = KNeighborsRegressor().fit(X_training, y_training)
y_pred = knn_model.predict(X_test)
print(knn_model.score(X_test, y_test))
print("Residuenquadratsumme KNN (MSE): %.2f" % np.mean((y_pred - y_test) ** 2))
```

R2 Wert: 0.9746

MSE: 83.32

# 9. Decision Three

- Erklärende Variable = **Spritverbrauch Kombiniert & Hubraum & Zylinder**
- Zu erklärende Variable = CO2 /km

```
from sklearn.tree import DecisionTreeRegressor
DTmodel = DecisionTreeRegressor(random_state = 0)
DTmodel.fit(X_training, y_training)
DTmodelYpred = DTmodel.predict(X_test)
DTmodelScore = r2_score(y_test, DTmodelYpred)
print('DT R2: %.3f' % DTmodelScore)
print(["Residuenquadratsumme DT (MSE): %.2f" % np.mean((DTmodelYpred - y_test) ** 2)])
```

R2 Wert: 0.986

MSE: 45.03

# Literatur

- **Umweltbundesamt (2021):** Emissionsquellen. Online verfügbar unter <https://www.umweltbundesamt.de/themen/klima-energie/treibhausgas-emissionen/emissionsquellen#energie-verkehr>
- **Europäische Union (2021):** Übereinkommen von Paris. Online verfügbar unter [https://ec.europa.eu/clima/eu-action/international-action-climate-change/climate-negotiations/paris-agreement\\_de](https://ec.europa.eu/clima/eu-action/international-action-climate-change/climate-negotiations/paris-agreement_de), zuletzt geprüft am 16.01.2022.