

In [1]:

```
import pandas as pd
import os
```

## Bronze Schicht

"Automobile Data Set" unter folgendem Link: [https://archive.ics.uci.edu/ml/machine-learning-databases/autos/imports-85.data](https://archive.ics.uci.edu/ml/machine-learning-databases autos/imports-85.data).

Der Datensatz listet Fahrzeuge ab dem Jahr 1985 auf, der Inhalt ist in englischer Sprache.

Die Daten zeigen die normalisierten Nutzungsschäden im Vergleich zu anderen Fahrzeugen. Die zweite Klassifizierung gibt an, inwieweit das Fahrzeug risikoreicher ist als sein Preis vermuten lässt. Den Fahrzeugen wird zunächst ein Risikofaktorsymbol zugeordnet, das mit ihrem Preis verknüpft ist. Wenn das Risiko höher (oder niedriger) ist, wird dieses Symbol angepasst, indem es auf der Skala nach oben (oder unten) verschoben wird. Versicherungsmathematiker bezeichnen diesen Vorgang als "Symbolisierung". Ein Wert von +3 bedeutet, dass das Auto riskant ist, während -3 bedeutet, dass es wahrscheinlich ziemlich sicher ist. Der dritte Faktor ist der relative durchschnittliche Schadenaufwand pro versichertem Fahrzeugjahr. Dieser Wert wird für alle Fahrzeuge einer bestimmten Größenklasse (zweitürige Kleinwagen, Kombis, Sportwagen usw.) normiert und stellt den durchschnittlichen Schaden pro Fahrzeug und Jahr dar.

Spalten:

- risikoniveau: -3, -2, -1, 0, 1, 2, 3
- normalisierter-verlustwert: 65 - 256
- marke: alfa-romero, audi, bmw, chevrolet, dodge, honda, isuzu, jaguar, mazda, mercedes-benz, mercury, mitsubishi, nissan, peugot, plymouth, porsche, renault, saab, subaru, toyota, volkswagen, volvo
- kraftstofftyp: diesel, gas
- absaugung: std, turbo
- türnummern: four, two
- körperform: hardtop, wagon, sedan, hatchback, convertible
- antriebsräder: 4wd, fwd, rwd
- motorstandort: front, rear
- radstand: 86.6 - 120.9
- länge: 141.1 - 208.1
- breite: 60.3 - 72.3
- höhe: 47.8 - 59.8
- leergewicht: 1488 - 4066
- motortyp: dohc, dohcvt, l, ohc, ohcvt, ohcvt, rotor
- anzahl-der-zylinder: eight, five, four, six, three, twelve, two
- motorgröße: 61 - 326
- kraftstoffsystem: 1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi
- bohrung: 2.54 - 3.94
- anschlag: 2.07 - 4.17
- verdichtungsverhältnis: 7 - 23
- pferdestärken: 48 - 288
- spitzendrehzahl: 4150 - 6600
- stadt-mpg: 13 - 49
- autobahn-mpg: 16 - 54
- preis: 5118 - 45400

In [2]:

```
headers = ["risikoniveau", "normalisierter-verlustwert", "marke", "kraftstofftyp", "absaugung", "türnummern", "körperform", "antriebsräder", "motorstandort", "radstand", "länge", "breite", "höhe", "leergewicht", "motortyp", "anzahl-der-zylinder", "motorgröße", "kraftstoffsystem", "bohrung", "anschlag", "ver
```

```
dichtungsverhältnis", "pferdestärken", "spitzendrehzahl", "stadt-mpg", "autobahn-mpg", "preis"]
```

```
raw_input_df = pd.read_csv('input/auto.csv', names = headers)
display(raw_input_df.head(10))
```

	risikoniveau	normalisierter-verlustwert	marke	kraftstofftyp	absaugung	türnummern	körperform	antriebsräder	motorstandort	radstandort
0	3	?	alfa-romero	gas	std	two	convertible	rwd		front
1	3	?	alfa-romero	gas	std	two	convertible	rwd		front
2	1	?	alfa-romero	gas	std	two	hatchback	rwd		front
3	2	164	audi	gas	std	four	sedan	fwd		front
4	2	164	audi	gas	std	four	sedan	4wd		front
5	2	?	audi	gas	std	two	sedan	fwd		front
6	1	158	audi	gas	std	four	sedan	fwd		front
7	1	?	audi	gas	std	four	wagon	fwd		front
8	1	158	audi	gas	turbo	four	sedan	fwd		front
9	0	?	audi	gas	turbo	two	hatchback	4wd		front

10 rows x 26 columns



In [3]:

```
csv_file_path = os.path.join('.', 'bronze')
csv_file_name = 'auto_bronze.csv'

if not os.path.exists(csv_file_path):
    os.mkdir(csv_file_path)

raw_input_df.to_csv(os.path.join(csv_file_path, csv_file_name), index=False)
```