**Metadata of Integrated Dataset**

The integrated dataset was generated by integrating 4 datasets of used cars from different car manufacturers. The integrated dataset contains 4 columns: brand, year, price (USD) and mileage (km).

The first dataset (“car\_price\_prediction.csv”) consists of used car data worldwide and was obtained from Kaggle (<https://www.kaggle.com/datasets/deepcontractor/car-price-prediction-challenge>) from the author: ‘Deep Contractor’.

The second dataset (“USA\_cars\_datasets.csv”) consists of used car data from the US and was obtained from Kaggle (<https://www.kaggle.com/datasets/doaaalsenani/usa-cers-dataset>) from the author: ‘Doaa Alsenani’.

The third dataset (“autoscout24-germany-dataset.csv”) renamed “germany\_cars.csv” during the data cleaning consists of used car data from Germany and was obtained from Kaggle (<https://www.kaggle.com/datasets/ander289386/cars-germany>) from the author: ‘Ander’.

The fourth and last dataset (“Belarus Used Cars Prices.csv”) consists of used car data from Belarus and was obtained from Kaggle (<https://www.kaggle.com/datasets/slavapasedko/belarus-used-cars-prices?resource=download>) from the author: ‘Slava Pasedeko’.

All datasets were of license CC0: Public Domain, meaning that the data may be used without prior permission, as it has been dedicated to the public and therefore, has no copyrights.