**DATA1002 Integrated Data**

**Group name:** DATA1002-CC-L1-G3

**Members:**

SID:510067773, UniKey: eyon5512

SID: 520562286, UniKey: tdra0337

 SID: 510594198, UniKey: skle4070

SID: 520604830, UniKey: asau0735

**Integrated Data**

The original 4 datasets where all retrieved from Kaggle but from different sources and authors. Three of the original datasets contained information from the price of used and new cars in three different countries (United States, Germany, and Belarus) the fourth dataset doesn’t have a clear country of origin but also contains different car brands, models, prices, year, and mileage.

After merging the datasets, the result consisted of a 119,952-row dataset with only three variables: Manufacturer, Mileage and Price. The Manufacturer contains only string values, Mileage contains integers, and Price contains floats.

The metadata of the resulting dataset and its licences are hence derived from the four original datasets.

Licence: CC0: Public Domain

Provenance: <https://www.autoscout24.com/>, AUCTIONEXPORT.com, machinehack.com