

Data Intake Report

Name: G2M insights for Cab Investment (XYZ)
Report date: 11.03.2024
Internship Batch: LISUM31
Version: 1.0
Data intake by: Jonas Lütticken
Data storage location: <https://github.com/DataGlacier/DataSets>

Tabular data details:

Total number of observations	359393
Total number of files	4
Total number of features	7
Base format of the file	.csv
Size of the data	20.663 KB

Total number of observations	49172
Total number of files	4
Total number of features	4
Base format of the file	.csv
Size of the data	1.027 KB

Total number of observations	21
Total number of files	4
Total number of features	3
Base format of the file	.csv
Size of the data	1 KB

Total number of observations	440099
Total number of files	4
Total number of features	3
Base format of the file	.csv
Size of the data	8.788 KB

Proposed Approach (this approach was done with every file):

- I checked the data for duplicates. I also checked that there were no missing values, every row had valid/readable data and that every column had the correct format.
- There are many possible ways of connecting the data from the files with each other, for example the City Data with the Cab_Data, which can be used with really every dataset.