

# 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:

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

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

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

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

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

- I cleaned and organised the data in columns, so that it would be easier to understand and analyze, using the Excel function “Text to Columns”. I created each column using the comma (“;”) as separator, since all the hypothetical columns were separated with a comma.
- To ensure that there were no duplicates in the files, I checked the data with the inbuild Excel function “Remove Duplicates”. I also checked that there were no missing values, every row had valid/readable data and that every column had the correct format.

- The Cab\_Data.csv file contains information about the two cab companies, giving information like: Transaction ID, Date of travel, Company, KM travelled, price charge and cost of trip.
- The Customer\_ID.csv file contains information about demographics of the clients: Customer ID, gender, age and income.
- The Transaction\_ID.csv file contains information about the payment method of the clients: Transaction ID, customer ID and payment method.
- The City.csv file contains information about the population and number of cab users per city.