

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 02/18/2024

Internship Batch: LISUM30

Version: 1.0

Data intake by: Yan-Ping Yu

Data intake reviewer:

Data storage location: <https://github.com/YanPing0227/yanping-my-repo>

## Tabular data details:

<b>Total number of observations</b>	49171
<b>Total number of files</b>	4
<b>Total number of features</b>	5
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	1.23 MB

<b>Total number of observations</b>	20
<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>	440098
<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>	8.58 MB

<b>Total number of observations</b>	359392
<b>Total number of files</b>	4
<b>Total number of features</b>	8
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	23.3 MB

1. In cases where duplicates contain partial or complementary information, the records of 3 tabular datasets are merged into 1 single, comprehensive dataset.
2. Also Check whether duplicated rows exist by using .duplicated() function