

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 10/14/2022

Internship Batch:

Version: 1.0

Data intake by: Glacier

Data intake reviewer: Yuchi Chen

Data storage location: <https://github.com/DataGlacier/DataSets>

## Tabular data details:

<b>Total number of observations</b>	359392
<b>Total number of files</b>	4
<b>Total number of features</b>	14
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	30 MB

## Proposed Approach:

- Data clean for int type, date type, and string type
- Merge files together based on primary keys and foreign keys
- Data quality check, including missing one
- Extract more features by feature engineering