

# Data Intake Report

Name: G2M Insight for cab investment firm

Report date: September 14, 2022

Internship Batch: LISUM13

Version: 1.0

Data intake by: Sok Leng Chan

Data intake reviewer: N/A

Data storage location: [DataGlacier/DataSets \(github.com\)](https://github.com/DataGlacier/DataSets)

## Tabular data details:

<b>Total number of observations</b>	359392
<b>Total number of files</b>	5
<b>Total number of features</b>	18
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	31 MB

**Note: Replicate same table with file name if you have more than one file.**

## Proposed Approach:

- Mention approach of dedup validation (identification)
  - Check the structure of the data and clean up the dirty data
  - Check for missing value and duplicated value
  - Check for the outliers
- Mention your assumptions (if you assume any other thing for data quality analysis)
  - There is different profit for different cab companies over the year
  - The age of customers might be different for both companies.
  - Different city have different profit margins for different cab companies.
  - There is different class to ride the different cab companies.
  - Different income class might have different profit for yellow cab and pink cab.