

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 13/08/2024

Internship Batch: LISUM36

Version:1.0

Data intake by: Stephanie Dawsonn-Andoh

Data intake reviewer:

Data storage location: [https://github.com/stephandoh/Sondoh\\_Glacier/tree/main/Week\\_2](https://github.com/stephandoh/Sondoh_Glacier/tree/main/Week_2)

## Tabular data details:

<b>Total number of observations</b>	49171
<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>	1.00 MB

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

<b>Proposed Approach</b>	<b>Assumptions</b>
Cleaned and standardized customer demographic data (e.g., gender, age, income).	Assumed that demographic data provided was accurate and complete.
Categorized customers into different age groups and income categories.	Assumed that categorization did not oversimplify the analysis.
Identified and flagged repeat customers in the dataset.	Assumed that customer IDs were consistently used across transactions.
Addressed missing values by imputing with median or mode.	Assumed that missing values were randomly distributed.