Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 15 May 2025
Internship Batch: LISUM45
Version:<1.0>
Data intake by: Priyanjali Patel
Data intake reviewer: Data Glacier
Data storage location: github

Tabular data details of the Master data:

Total number of observations	359392
Total number of files	5
Total number of features	19
Base format of the file	.csv
Size of the data	150.32 MB

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

Proposed Approach:
- Mention approach of dedup validation (identification)
  : To identify duplicate records in the dataset, I will check for exact duplicates across **all columns** in the data. This means that if two or more rows have identical values in every column, they will be considered duplicates and flagged for removal or further investigation.