

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 11/24/2023

Internship Batch: LISUM25

Version: 1.0

Data intake by: Amira Esmaeil

Data intake reviewer: N/A

Data storage location: https://drive.google.com/drive/folders/1_SkD7EqRsB_5N0y2vrgJKUvfiWGql7hq?usp=drive_link

Data Preparation:

Tabular data details: Total number of observations	359392
Total number of files	4
Total number of features	19
Base format of the file	.csv
Data name	Global_cab_data
Size of the data	46.1 MB

Proposed Approach:

- Global data formation

A global dataset was built by combining 3 different datasets which are Cab_Data.csv, Transaction_ID.csv and Customer_ID.csv and performing inner join with the column they have in common.

The Transaction_ID.csv and Cab_data.csv was inner join through the Transaction ID column present in both datasets was formed the Trans_cab.csv dataset. Then, this new dataset (Trans_cab.csv) was inner-joined with the Customer_ID.csv through a common Customer ID column to form our global_cab_data.csv dataset which details are listed in the table above.