

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 13.04.2023

Internship Batch:

Version: 1.0

Data intake by: Daria Lazorenko

Data intake reviewer:

Data storage location:

Tabular data details:

Cab_Data

Total number of observations	359392
Total number of files	
Total number of features	7
Base format of the file	.csv
Size of the data	20.1 MB

Transaction_ID

Total number of observations	440098
Total number of files	
Total number of features	3
Base format of the file	.csv
Size of the data	8.58 MB

Customer_ID

Total number of observations	49171
Total number of files	
Total number of features	4
Base format of the file	.csv
Size of the data	1.00 MB

City

Total number of observations	20
Total number of files	
Total number of features	3
Base format of the file	.csv
Size of the data	4.00 KB

Proposed Approach:

For dedup validation, I plan to compare records based on the unique identifiers, such as Transaction_ID and Customer_ID, and remove any duplicate records found. I assume that the data

has been collected and stored accurately and that there are no missing values or outliers. To ensure data quality, I will perform exploratory data analysis to check for any anomalies or inconsistencies in the data.