

Data Intake Report

Name: <G2M insight for Cab Investment firm>

Report date: <November 14, 2023>

Internship Batch:<LISUM27: 30 Oct 23 - 30 Jan 24>

Version:<1.0>

Data intake by:<Rodolphe Segbedji>

Data intake reviewer:<intern who reviewed the report>

Data storage location: <https://github.com/Roseco-crs/DataGlacier_Week2_G2Mstudy>

Tabular data details:

Cab_Data

Total number of observations	359392
Total number of files	1
Total number of features	7
Base format of the file	.csv
Size of the data	20.1MB

Customer_ID

Total number of observations	49171
Total number of files	1
Total number of features	4
Base format of the file	.csv
Size of the data	1MB

Transaction_ID

Total number of observations	440098
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	8.58MB

City

Total number of observations	20
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	1KB

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

Proposed Approach:

- Mention approach of dedup validation (identification)
- Mention your assumptions (if you assume any other thing for data quality analysis)

Note: Convert this doc in pdf and provide the link of pdf file in your dashboard.

Please do not forget to remove this section while converting the file into pdf.