

Data Intake Report

Name: EDA

Report date: 14.06.2022

Internship Batch: LISUM10: 30

Version: 1.0

Data intake by: Marii Torubara

Data intake reviewer: <intern who reviewed the report>

Data storage location: https://github.com/powidlowa/Data_Glacier

Tabular data details:

Cab Data.csv

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	<21.2MB>

Customer ID.csv

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	<1.05MB>

City.csv

Total number of observations	19
Total number of files	1
Total number of features	3
Base format of the file	<.csv>
Size of the data	<759MB>

Transaction ID.csv

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	<9MB>

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.