

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: Dec-12-2022

Internship Batch: LISUM16

Version: 1.0

Data intake by: Dong Han

Data intake reviewer: <intern who reviewed the report>

Data storage location:

[https://drive.google.com/drive/folders/1eco6ebjlnT5YIdRBCBWvEKd2v8J13nf4?usp=share\\_link](https://drive.google.com/drive/folders/1eco6ebjlnT5YIdRBCBWvEKd2v8J13nf4?usp=share_link)

## Tabular data details:

Cab Data.csv

<b>Total number of observations</b>	320532
<b>Total number of files</b>	1
<b>Total number of features</b>	7
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	21.2 MB

City.csv

<b>Total number of observations</b>	20
<b>Total number of files</b>	1
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	759 bytes

Customer ID.csv

<b>Total number of observations</b>	49171
<b>Total number of files</b>	1
<b>Total number of features</b>	4
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	1.1 MB

Transaction ID.csv

<b>Total number of observations</b>	440098
<b>Total number of files</b>	1
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	9 MB

**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)