

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 22 October 2021

Internship Batch: NLP02

Version:<1.0>

Data intake by: Chow Jun Wei

Data intake reviewer:<intern who reviewed the report>

Data storage location: <location URL eg: github, cloud>

## Tabular data details:

<b>Total number of observations</b>	359392
<b>Total number of files</b>	1 (Cab_Data.csv)
<b>Total number of features</b>	8
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	21 MB

<b>Total number of observations</b>	20
<b>Total number of files</b>	1 (City.csv)
<b>Total number of features</b>	4
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	759 Bytes

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

<b>Total number of observations</b>	440098
<b>Total number of files</b>	1 (Transaction_ID.csv)
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	8.6 MB

<b>Total number of observations</b>	359392
<b>Total number of files</b>	1 (cab_tran_cus.csv)
<b>Total number of features</b>	15
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	44 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)

**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.**