

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 11/14/2022

Internship Batch: LISUM15

Version:<1.0>

Data intake by: Xuyang Zhao

Data intake reviewer: Xuyang Zhao

Data storage location: <https://github.com/DataGlacier/DataSets.git>

## Tabular data details:

<b>Total number of observations</b>	359392
<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>	21.5MB
<b>File name</b>	Cab_Data

<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>	8.58MB
<b>File name</b>	Transaction_ID

<b>Total number of observations</b>	359392
<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>	1MB
<b>File name</b>	Customer_ID

<b>Total number of observations</b>	359392
<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>	759bytes
<b>File name</b>	City

## Proposed Approach:

- Use inner join to drop records with missing value when creating master data for EDA.
- The time period for analysis is from 31/01/2016 to 31/12/2018.