

# Data Intake Report

Name: G2M Cab Investment Report

Report date: 11<sup>th</sup> June, 2021 (11/06/2021)

Internship Batch: LISUM01

Version: 1.0

Data intake by: Udbhav Balaji

Data intake reviewer: <intern who reviewed the report>

Data storage location: GitHub URL: <https://github.com/udbhavbalaji/DataSets>

## Tabular data details:

<b>File Name</b>	Cab_Data.csv
<b>Total number of observations</b>	359392 rows
<b>Total number of files</b>	4
<b>Total number of features</b>	7 Columns
<b>Base format of the file</b>	Comma Separated Values (.csv)
<b>Size of the data</b>	21.2 MB

<b>File Name</b>	City.csv
<b>Total number of observations</b>	20 rows
<b>Total number of files</b>	4
<b>Total number of features</b>	3 Columns
<b>Base format of the file</b>	Comma Separated Values (.csv)
<b>Size of the data</b>	4 KB

<b>File Name</b>	Customer_ID.csv
<b>Total number of observations</b>	49171 rows
<b>Total number of files</b>	4
<b>Total number of features</b>	4 Columns
<b>Base format of the file</b>	Comma Separated Values (.csv)
<b>Size of the data</b>	1.1 MB

<b>File Name</b>	Transaction_ID.csv
<b>Total number of observations</b>	440098 rows
<b>Total number of files</b>	4
<b>Total number of features</b>	3 Columns
<b>Base format of the file</b>	Comma Separated Values (.csv)
<b>Size of the data</b>	9 MB

**Proposed Approach:**

- Initially, wanted to leave out date as the format didn't really help me too much in deciphering what the travel date actually was. I cleaned up the rest of the data and added the columns that I thought was necessary for the final analysis into a master data frame, then stored in Master\_Data.csv.
- **I assume that the data given is from 01/01/2016 till 31/12/2018. This is because the date of Travel field has 1095 unique values, which is the number of days between these 2 exact dates. Since our analysis is to be between 31/01/2016 and 31/12/2018, I will DROP the rows for the first 30 days in the dataset before beginning my analysis.**

**Note: This is Version 1.0 and will be updated before final submission.**