

# Data Intake Report

Name: <G2M insight for Cab Investment firm>

Report date: <13-07-2022>

Internship Batch:<Enter your batch code from Canvas course>

Version:<1.0>

Data intake by:<Taimoor Razi>

Data intake reviewer:<NA>

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>
<b>Total number of features</b>	<7 >
<b>Base format of the file</b>	<.csv>
<b>Size of the data</b>	<20.1 MB>

<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>	<1.0 MB>

<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>	<8.58 MB>

<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>	<4.0 KB>

## Proposed Approach:

- Clean the data by removing the commas from populations and user values and correct their data types to int.
- Merging datasets on common columns. Remove extra values which have missing columns.
- Check null values and duplicate values.
- Feature Engineering: Date of travel column to date time object. Creat new features.