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:**

|  |  |
| --- | --- |
| **Total number of observations** | <359392> |
| **Total number of files** | <1> |
| **Total number of features** | <7 > |
| **Base format of the file** | <.csv> |
| **Size of the data** | <20.1 MB> |

|  |  |
| --- | --- |
| **Total number of observations** | <440098> |
| **Total number of files** | <1> |
| **Total number of features** | <3 > |
| **Base format of the file** | <.csv> |
| **Size of the data** | <1.0 MB> |

|  |  |
| --- | --- |
| **Total number of observations** | <49171> |
| **Total number of files** | <1> |
| **Total number of features** | <4 > |
| **Base format of the file** | <.csv> |
| **Size of the data** | <8.58 MB> |

|  |  |
| --- | --- |
| **Total number of observations** | <20> |
| **Total number of files** | <1> |
| **Total number of features** | <3 > |
| **Base format of the file** | <.csv> |
| **Size of the data** | <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. Create new features.