Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 11/9/2022

Internship Batch:<LISUM15>

Version:<1.0>

Data intake by: Emily Yao

Data intake reviewer:<intern who reviewed the report>

Data storage location: <https://github.com/EMILYWYAO/Data-Glacier-Internship/tree/main/Week%202>

**Tabular data details:**

Cab\_Data

|  |  |
| --- | --- |
| **Total number of observations** | 359392 |
| **Total number of files** | 1st out of 4 total |
| **Total number of features** | 7 |
| **Base format of the file** | .csv |
| **Size of the data** | 20,663 KB |

CityData

|  |  |
| --- | --- |
| **Total number of observations** | 20 |
| **Total number of files** | 2nd out of 4 total |
| **Total number of features** | 3 |
| **Base format of the file** | .csv |
| **Size of the data** | 1 KB |

Customer\_ID

|  |  |
| --- | --- |
| **Total number of observations** | 49171 |
| **Total number of files** | 3rd out of 4 total |
| **Total number of features** | 4 |
| **Base format of the file** | .csv |
| **Size of the data** | 1,027 KB |

Transaction\_ID

|  |  |
| --- | --- |
| **Total number of observations** | 440098 |
| **Total number of files** | 4th out of 4 total |
| **Total number of features** | 3 |
| **Base format of the file** | .csv |
| **Size of the data** | 8,788KB |

**Note: Replicate same table with file name if you have more than one file.**

**Proposed Approach:**

* Mention approach of debut validation (identification)
  + I first identified five factors (gender, income, age, retention, and transaction mode) that I believed would have an impact on profit
    - I made hypotheses for them and explored them
  + For the data I checked for outliers, NA values, and duplicates and found there were only outliers. I merged all the datasets into one, MasterData, to better perform operations
* Mention your assumptions (if you assume any other thing for data quality analysis)
  + There were outliers in the dataset that were not considered
  + The mean was used instead of the median for profit analysis
  + Profit was calculated by only considering Price\_Charged and Cost\_of\_Trip

**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.**