Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 01/11/2024

Internship Batch: LISUM29

Version: 1.0

Data intake by: Shengyu (Felix) Huang

Data intake reviewer:

Data storage location:

**Cab\_Data.csv:**

|  |  |
| --- | --- |
| **Total number of observations** | 359,392 |
| **Total number of files** |  |
| **Total number of features** | 7 |
| **Base format of the file** | .csv |
| **Size of the data** | 20,663KB |

**City.csv:**

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

**Customer\_ID.csv:**

|  |  |
| --- | --- |
| **Total number of observations** | 49,171 |
| **Total number of files** |  |
| **Total number of features** | 4 |
| **Base format of the file** | .csv |
| **Size of the data** | 1,027KB |

**Transaction\_ID.csv:**

|  |  |
| --- | --- |
| **Total number of observations** | 440,098 |
| **Total number of files** |  |
| **Total number of features** | 3 |
| **Base format of the file** | .csv |
| **Size of the data** | 8,788KB |

**Proposed Approach:**

* Date should be converted in Cab\_Data from an excel format to Python datetime format.
* Transaction ID and Customer ID are primary that can use to merge table for analysis.