Data Intake Report

Name: <G2M Insight for Cab Investment Firm>

Report date: <October 14th, 2022>

Internship Batch:< LISUM14>

Version:<1.0>

Data intake by:<Nolan Pilozo-Hibbit>

Data intake reviewer:<Nolan Pilozo-Hibbit>

Data storage location: <location URL eg: github, cloud>

**Cab data details:**

|  |  |
| --- | --- |
| **Total number of observations** | 359,392 |
| **Total number of files** | 1 |
| **Total number of features** | 7 |
| **Base format of the file** | .csv |
| **Size of the data** | 20,663 KB |

**City data details:**

|  |  |
| --- | --- |
| **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** | 1 KB |

**Customer\_ID data details:**

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

**Transaction\_ID data details:**

|  |  |
| --- | --- |
| **Total number of observations** | 44,098 |
| **Total number of files** | 1 |
| **Total number of features** | 3 |
| **Base format of the file** | .csv |
| **Size of the data** | 8,788 KB |

**Proposed Approach:**

* Duplicate removal can be accomplished using basic Excel functions and built-in tools. The chosen method will be the **Data > Remove Duplicates** tool.
* Mention your assumptions (if you assume any other thing for data quality analysis)