Data Intake Report

Name: <G2M insight for Cab Investment firm>

Report date: <2022/06/14>

Internship Batch:<LISUM10: 30>

Version:<1.0>

Data intake by:<Henri Edwards>

Data intake reviewer:<Henri Edwards>

Data storage location: <<https://github.com/HenriEdwards/Data-Glacier>>

**Tabular data details: Cab Data**

|  |  |
| --- | --- |
| **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** | <22418KB> |

**Tabular data details: Customer\_ID**

|  |  |
| --- | --- |
| **Total number of observations** | <number of rows> |
| **Total number of files** | <1> |
| **Total number of features** | <4> |
| **Base format of the file** | <.csv> |
| **Size of the data** | <1027 KB> |

**Tabular data details: City**

|  |  |
| --- | --- |
| **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** | <1kb> |

**Tabular data details: Transaction\_ID**

|  |  |
| --- | --- |
| **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** | <8.7 ,MB> |

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

**Proposed Approach:**

* Mention approach of dedup validation (identification)
* Mention your assumptions (if you assume any other thing for data quality analysis)

**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.**