Data Intake Report

Name: G2M insights

Report date: 5/12/2022

Internship Batch: LISUM16

Version: 1.0

Data intake by: Chris Donaton

Data intake reviewer: N/A

Data storage location: github.com/rambles-tech/DG\_virtual\_internship/tree/main/week02/data

**Tabular data details: Transaction\_ID.csv**

|  |  |
| --- | --- |
| **Total number of observations** | 440 098 |
| **Total number of files** | 1 |
| **Total number of features** | 3 |
| **Base format of the file** | CSV |
| **Size of the data** | 8358 KB |

**Tabular data details: City.csv**

|  |  |
| --- | --- |
| **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 |

**Tabular data details: Cab\_Data.csv**

|  |  |
| --- | --- |
| **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** | 20,312 KB |

**Tabular data details: Customer\_ID.csv**

|  |  |
| --- | --- |
| **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** | 979 KB |

**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.**