Data Intake Report

Name: <EDA of G2M Dataset>

Report date: <21.05.2024>

Internship Batch:< 9395317>

Version:<1.0>

Data intake by:<Mustafa Mert Özyılmaz>

Data intake reviewer:<intern who reviewed the report>

Data storage location: < https://github.com/DataGlacier/DataSets>

**Tabular data details: Cab Data**

|  |  |
| --- | --- |
| **Total number of observations** | <359393 > |
| **Total number of files** | <1> |
| **Total number of features** | <7 Columns> |
| **Base format of the file** | <.csv> |
| **Size of the data** | <20.6 MB> |

**Tabular data details: City**

|  |  |
| --- | --- |
| **Total number of observations** | <21> |
| **Total number of files** | <1> |
| **Total number of features** | <3 Columns> |
| **Base format of the file** | <.csv> |
| **Size of the data** | <1 Kb> |

**Tabular data details: Customer ID**

|  |  |
| --- | --- |
| **Total number of observations** | <49172> |
| **Total number of files** | <1> |
| **Total number of features** | <4 Columns> |
| **Base format of the file** | <.csv> |
| **Size of the data** | <1 MB> |

**Tabular data details: Transaction ID**

|  |  |
| --- | --- |
| **Total number of observations** | <440099> |
| **Total number of files** | <1> |
| **Total number of features** | <3 Columns> |
| **Base format of the file** | <.csv> |
| **Size of the data** | <8.7 MB> |

**Proposed Approach:**

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