Data Intake Report

Name: Keith D. Smith

Report date:14 May 2024

Internship Batch: LISUM33

Version:<1.0>

Data intake by: Keith D. Smith

Data intake reviewer:

Data storage location:

**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** | 20,663 kb |
|  |  |

**City\_Data.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 |

**Customer\_Data.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** | 1,027 kb |

**Transaction\_Data.csv**

|  |  |
| --- | --- |
| **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** | 8,788 kb |

**Proposed Approach:**

* Mention approach of dedup validation (identification)

Data Deduplication will be rule based on customer ID or Transaction ID

* Mention your assumptions (if you assume any other thing for data quality analysis)

At this point, I assume all datasets are mergeable and fairly accurate.

**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.**