Data Intake Report

Name: Cab Investment data analysis

Report date: 2022-10-11

Internship Batch: LISUM14

Version: 1.0

Data intake by: Marija Babić

Data intake reviewer:

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

**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** | 20.1 MB |

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** | 759B |

Customer\_ID

|  |  |
| --- | --- |
| **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 MB |

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** | 8.58 MB |

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

**Proposed Approach:**

* Table primary keys:
  + Cab data – Transaction\_id
  + City data – City
  + Customer id data – customer id
  + Transaction id data – transaction id
* Cab data is the fact table and other tables are dimensions which were joined on the cab data table. After each join there was a step where unique values were checked which was the confirmation the the data have been joined properly and that no duplicates were detected.

**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.**