

Data Intake Report

Name: Insight for Cab Investment

Report date: 08/14/2022

Internship Batch: LISUM12

Version: 1.0

Data intake by: Jessica Gallo

Data intake reviewer:

Data storage location:

Tabular data details:

Cab.csv

| | |
|-------------------------------------|--------|
| Total number of observations | 359392 |
| Total number of files | |
| Total number of features | 7 |
| Base format of the file | .csv |
| Size of the data | 19.2MB |

City.csv

| | |
|-------------------------------------|-------------|
| Total number of observations | 20 |
| Total number of files | |
| Total number of features | 3 |
| Base format of the file | .csv |
| Size of the data | 608.0 bytes |

Customer_ID.csv

| | |
|-------------------------------------|--------|
| Total number of observations | 359392 |
| Total number of files | |
| Total number of features | 4 |
| Base format of the file | .csv |
| Size of the data | 1.5MB |

Transaction_ID.csv

| | |
|-------------------------------------|--------|
| Total number of observations | 440098 |
| Total number of files | |
| Total number of features | 3 |
| Base format of the file | .csv |
| Size of the data | 10.1MB |

Proposed Approach:

- Duplicate validation will be performed by checking for unique() values and by indexing the transaction and customer IDs
- Assumptions of this data include: data will be missing or incorrectly inputted and datatypes might need to be changed, such as the Date column to datetime format.