

Data Intake Report

Name: Cab Investments

Report date: 01/01/2023

Internship Batch: LISUM17

Version: 1.0

Data intake by: Nick Patterson

Data intake reviewer:

Data storage location:

Github location: <https://github.com/BagOniku/Cab-Investments/tree/main/Data>

Local Storage: User/Project-Folder/Data

Filename: City.csv

| | |
|------------------------------|-------|
| Total number of observations | 20 |
| Total number of data points | 60 |
| Total number of features | 3 |
| Base format of the file | .csv |
| Size of the data | 608 B |

Filename: Customer_ID.csv

| | |
|------------------------------|---------|
| Total number of observations | 49,171 |
| Total number of data points | 196,684 |
| Total number of features | 4 |
| Base format of the file | .csv |
| Size of the data | 979 B |

Filename: Cab_Data.csv

| | |
|------------------------------|-----------|
| Total number of observations | 359,392 |
| Total number of data points | 2,515,744 |
| Total number of features | 7 |
| Base format of the file | .csv |
| Size of the data | 19.8MB |

Filename: Transaction_ID.csv

| | |
|------------------------------|-----------|
| Total number of observations | 440,098 |
| Total number of data points | 1,320,294 |
| Total number of features | 3 |
| Base format of the file | .csv |
| Size of the data | 8.16 MB |

Proposed Approach:

- Transaction_ID.csv and Cab_Data.csv could be permanently merged into one table, I do not see a need to keep them separate. Combined data will be saved as Transaction_Data.csv
- City.csv and Customer_ID.csv will provide categorical boundaries for analysis and comparison within the Transaction data.
- Categorical boundaries will be created using buckets.