

Data Intake Report

Name: CAB Industry Exploratory Data Analysis

Report date: 6th September 2021

Internship Batch:LISUM03

Version:1.0

Data intake by:Abigael Nyakara Mogusu

Data intake reviewer:

Data storage location: <https://www.linkedin.com/in/abigail-nyakara001/>

1.0 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	20663 kb

2.0 Tabular data details:Customer_ID Data

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	1027 kb

3.0 Tabular data details:Transaction_ID Data

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	8788 kb

4.0 Tabular data details:City Data

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	759 Bytes

Proposed Approach:

- I used each table's unique identifier to merge it with another table. These would be through unique columns such as the customer Id or the transaction Id. All the four tables are joined to form one master dataset that is then used for the exploratory data analysis.