

Data Intake Report

Name: <Cab Investment Firm>

Report date: <2/21/2023>

Internship Batch:<LISUM18>

Version:<1.0>

Data intake by:<Nigar Mutallimova>

Data intake reviewer: -

Data storage location: <https://github.com/DataGlacier/DataSets/blob/main/Cab_Data.csv>

Cab_Data details:

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	<21,174 KB>

City data details:

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_ID data details:

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_ID data details:

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:

- I am checking for duplicates in 'Transaction ID' column after joining 4 datasets. Duplicates are not observed.
- Date column need to be changed to the needed format