

Data Intake Report

Name: <G2M insight for Cab Investment firm>

Report date: <2022/09/14>

Internship Batch:<LISUM13:30>

Version:<1.0>

Data intake by:<Juzheng Shi>

Data intake reviewer:< Juzheng Shi >

Data storage location: < <https://github.com/DataGlacier/DataSets> >

Tabular data details:

Cab_data.csv

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.csv

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.csv

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.csv

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:

- Mention approach of dedup validation (identification)
- Mention your assumptions (if you assume any other thing for data quality analysis)