

Data Intake Report

Name: <G2M insights for Cab investment>

Report date: <4/18/2022>

Internship Batch:<LISMU 08>

Version:<1.0>

Data intake by:<Uchenna Ilodigwe>

Data intake reviewer:None

Data storage location: <https://github.com/uchennailodigwe/VCWeek2>

Tabular data details:

| | Cab_Data.csv | Transaction_ID.csv | City.csv | Customer_ID |
|---------------------------------|--------------|--------------------|----------|-------------|
| Observations | 359392 | 440098 | 19 | 49171 |
| Total number of files | 1 | 1 | 1 | 1 |
| Total number of features | 7 | 3 | 3 | 4 |
| Base format of the file | .csv | .csv | .csv | .csv |
| Size of the data | 20663kb | 8788kb | 1kb | 1027kb |

Note: Replicate same table with file name if you have more than one file.

Proposed Approach:

- Mention approach of dedup validation (identification)
 - 1. Merge Files to get the necessary features for the analysis.
- Mention your assumptions (if you assume any other thing for data quality analysis)
 - The data is clean.