

# Data Intake Report

Name: XYZ G2M Cab Investment

Report date: 08<sup>th</sup> October 2021

Internship Batch:<

Version:1.0

Data intake by: Teddy Waweru

Data intake reviewer:

Data storage location: [https://github.com/teddywaweru/DataGlacier/tree/TeddyW\\_Cab-G2M-Investment/Week2\\_Cab\\_G2M\\_Investment](https://github.com/teddywaweru/DataGlacier/tree/TeddyW_Cab-G2M-Investment/Week2_Cab_G2M_Investment)

## Tabular data details:

Total number of observations	359392
Total number of files	4
Total number of features	15
Base format of the file	.csv
Size of the data	29.7 MB

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

## Proposed Approach:

- Mention approach of dedup validation (identification)  
*Checked on the data validity by utilizing Pandas Profiling Report, which confirmed on missing values.*  
*As for duplicated data, the data consisted of special index keys marked as UNIQUE, which ensured that there were no duplicates in the data*
- Mention your assumptions (if you assume any other thing for data quality analysis)