

Data Intake Report

Name: Cab Service Investment

Report date: 14 Jan 2023

Internship Batch: LISUM17

Version:1.0

Data intake by: Vamsi Krishna Oruganti

Data intake reviewer: Data Glacier

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

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	20.1 MB

Tabular data details: Customer_ID

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

Tabular data details: Transaction_ID

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

Tabular data details: City

Total number of observations	20 cities
Total number of files	1
Total number of features	3
Base format of the file	csv
Size of the data	4 KB

Proposed Approach: Data Deduplication