

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 31/01/2016 to 31/12/2018

Internship Batch: LISUM22

Version:1.0

Data intake by: Sukurat Salam

Data intake reviewer:

Data storage location: [https://github.com/Sukuratsalam/Data-Glasier-Intern-DatabScience-/tree/main/Week\\_2](https://github.com/Sukuratsalam/Data-Glasier-Intern-DatabScience-/tree/main/Week_2)

## 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	1.1 MB

### Transaction ID

Total number of observations	440098
Total number of files	1
Total number of features	3 columns
Base format of the file	.csv
Size of the data	9 MB

### City

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	4 KB

### 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	21.2 MB

## Proposed Approach:

- Merging the data based on some similar variables.
- Exploratory data analysis
- Hypothesis testing using t-test, correlation test with the assumption of independent observations.

