

Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: January 9, 2023

Internship Batch: LISUM17

Version: 1.0

Data intake by: Elissa Kuon

Data intake reviewer: Data Glacier

Data storage location: https://github.com/ekuon/Data_Glacier_Internship/tree/main/Week%202

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: 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

Tabular data details: Customer_Data

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

Proposed Approach:

- There are no missing and duplicated values within all 4 datasets
- Checked the skewness of each of the 4 datasets, and noted that the 'Population' variables are heavily skewed in the City Dataset
- Follow the assumptions of the hypothesis testing (ANOVA and T-tests)
 - Observations in each sample are iid, normally distributed, and equal variance