

# Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: 21-Jul-2021

Internship Batch: LISUM02

Version: 1.0

Data intake by: Aynur Cemre Aka

Data intake reviewer:

Data storage location:

[https://github.com/cemreaka/G2M\\_Insight for Cab Investment Firm Project/tree/main/Datasets](https://github.com/cemreaka/G2M_Insight_for_Cab_Investment_Firm_Project/tree/main/Datasets)

## 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	19.2+ 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	608.0+ bytes

## 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	1.5+ 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	10.1+ MB

**Proposed Approach:**

- None of data sets contain missing value.
- None of data sets contain duplicated value.
- Attribute names of the datasets are changed.
- New data sets which include new attribute names are saved as .pkl file.