

# Data Intake Report

Name: G2M Cap

Report date: 7-Mar-2021

Internship Batch: LISP01

Version:1.0

Data intake by: Mohammad Odeh

Data intake reviewer:<intern who reviewed the report>

Data storage location: <https://github.com/m-odeh/G2M-Data-Glacier-Internship->

## 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	21.5 MB

### City Data:

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	1 KB

### Customer Data:

Total number of observations	49172
Total number of files	1
Total number of features	4
Base format of the file	CSV
Size of the data	1 MB

### Transactions Data:

Total number of observations	440099
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:**

- Create a master data set that contains all the data.
- Perform EDA on the master data set and extract vital data such as profit and number of customers for each company.
- Create new features that can help in supporting the assumption and help later in modeling

**Assumptions:**

- All features are assumed constant and only the charged price and the cost is changing, this will help calculate the profit per each ride
- No outliers in the data sets and all the data provided is included in the study