

# Data Intake Report

Name: G2M insight for Cab Investment

Report date: June 8, 2022

Internship Batch: LISUM10

Version: 1.0

Data intake by: Max Elgart

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

## Tabular data details:

|                                     |          |
|-------------------------------------|----------|
| <b>File Name</b>                    | Cab_Data |
| <b>Total number of observations</b> | 359392   |
| <b>Total number of features</b>     | 7        |
| <b>Base format of the file</b>      | .csv     |
| <b>Size of the data</b>             | 21.2 MB  |

|                                     |           |
|-------------------------------------|-----------|
| <b>File Name</b>                    | City      |
| <b>Total number of observations</b> | 20        |
| <b>Total number of features</b>     | 3         |
| <b>Base format of the file</b>      | .csv      |
| <b>Size of the data</b>             | 759 bytes |

|                                     |             |
|-------------------------------------|-------------|
| <b>File Name</b>                    | Customer_ID |
| <b>Total number of observations</b> | 49171       |
| <b>Total number of features</b>     | 4           |
| <b>Base format of the file</b>      | .csv        |
| <b>Size of the data</b>             | 1.1 MB      |

|                                     |                |
|-------------------------------------|----------------|
| <b>File Name</b>                    | Transaction_ID |
| <b>Total number of observations</b> | 440098         |
| <b>Total number of features</b>     | 3              |
| <b>Base format of the file</b>      | .csv           |
| <b>Size of the data</b>             | 9 MB           |

## Proposed Approach:

- Ran a query to see if there are any duplicate values, there weren't.
- Just from looking at the describe of everything, it seems as though Yellow Cab does better as a company.