

# Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: October 13, 2022

Internship Batch: LISUM14

Version: 1.0

Data intake by: Elizabeth Banning

Data intake reviewer: N/A

Data storage location: N/A

## Tabular data details:

Cab\_Data:

<b>Total number of observations</b>	359,392
<b>Total number of files</b>	1
<b>Total number of features</b>	7
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	20.2 MB

City:

<b>Total number of observations</b>	20
<b>Total number of files</b>	1
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	759 B

Customer\_ID:

<b>Total number of observations</b>	49,171
<b>Total number of files</b>	1
<b>Total number of features</b>	4
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	1 MB

Transaction\_ID:

<b>Total number of observations</b>	440,098
<b>Total number of files</b>	1
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	8.58 MB

## Proposed Approach:

Duplicate values were investigated for all four files using Python's .duplicated method. No duplicates were present in any of the four tables.