

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: July 14<sup>th</sup>, 2022

Internship Batch LISUM11: 30

Version: 1.0

Data intake by: Yidan Gao

Data intake reviewer: Yidan Gao

Data storage location: [https://github.com/YidanGaoMeow/DG\\_week2](https://github.com/YidanGaoMeow/DG_week2)

## Tabular data details:

<b>Total number of observations</b>	440,098
<b>Total number of files</b>	4
<b>Total number of features</b>	14
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	30.3MB

## Proposed Approach:

- The first two steps are merging files and cleaning data. I detected and cleaned the missing values, duplicated values, and outliers.
- In the data visualization part, I conducted profitability comparison and user analysis.
- I would recommend yellow cab because it has better analysis results like user retention and total profit.