

Data Intake Report

Name: < G2M insight for Cab Investment firm Report >

Report date: <6/14/2022>

Internship Batch:<LISUM10:30 >

Version:<1.0>

Data intake by:<Jinwen Li >

Data intake reviewer:<intern who reviewed the report>

Data storage location: < https://github.com/jinwen0826/data_week2 >

Tabular data details:

Total number of observations	<848681 >
Total number of files	<4 >
Total number of features	<17 >
Base format of the file	<.csv>
Size of the data	<31.2 MB>

Note: Replicate same table with file name if you have more than one file.

Proposed Approach:

- Mention approach of dedup validation (identification)
Go through the rows and columns to check the data and then cleaned the data and convert string type data to float type, finally I merged data into one table.

- Mention your assumptions (if you assume any other thing for data quality analysis)

Hypothesis 1: Is there any seasonality in number of customers using the cab service?

Hypothesis 2: the number of transactions is related with city's population number?

Hypothesis 3: the number of users is related to city's population number?

Hypothesis 4: the price charged is related to city's average income?

Hypothesis 5: Average profit of trips is related to the city's average income?