

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 14th April 2023

Internship Batch: LISUM20

Version:<1.0>

Data intake by: Meghna Ijjapureddy

Data intake reviewer:

Data storage location:

Tabular data details:

Total number of observations	359392
Total number of files	
Total number of features	7
Base format of the file	.csv
Size of the data	21.2 MB

Total number of observations	20
Total number of files	
Total number of features	3
Base format of the file	.csv
Size of the data	759 B

Total number of observations	49171
Total number of files	
Total number of features	4
Base format of the file	.csv
Size of the data	1.05 MB

Total number of observations	440098
Total number of files	
Total number of features	3
Base format of the file	.csv
Size of the data	9 MB

Proposed Approach:

Dedup validation:

- Identified Null values if any, using isnull()
- Checked and removed duplicates using drop_duplicates()

- Identified if any uncleaned repeated data using `unique()`
- Identified the formats of the data using `dtype()` and changed the format or data types of those data which are in the wrong format.

Assumptions:

- Both the Cabs (Pink and Yellow) are introduced into the market at the same time.
- Price charged, Cost of Trip and Income are all indicated in the same currency units USD as this data belongs to US cities.