

G2M(Cab Industry) Data Intake Report

Name: <G2M(cab industry) >

Report date: <2 September, 2021>

Internship Batch:<LISMO2>

Version:<1.0>

Data intake by:<Francis Ngethe Ngari>

Data intake reviewer:<intern who reviewed the report>

Data storage location: <location URL eg: github, cloud>

Tabular data details: Cab_data_master file

Total number of observations	<359392>
Total number of files	<1>
Total number of features	<15>
Base format of the file	<.csv>
Size of the data	< 42.3MB >

Proposed Approach:

In the final master Data set, unique columns are :

- Month:is the month the trip happened.
- Year: the year in which specific cab trips were taken.
- Company:Names of cab companies.
- KM traveled:Distance covered in Km by the cabs
- Price charged: The total amount a customer is charged for distance covered.
- Cost of trip: The Expenses per trip.
- Margins: The profit made from a trip.

- Gender: Different sexes whom use the different cab services (Female and Male)
- Income: how much customers earn in Us dollars.

These columns outline the uniqueness of this dataset, while columns which are similar are Transaction_id, Customer_id and City are present in other datasets that I have merged. Dropped the city dataset. Derived columns are such as margins , month and year.