

# Data Intake Report

Name: <G2M insight for Cab Investment firm>

Report date: <14/03/2024>

Internship Batch:<LISUM31>

Version:<1.0>

Data intake by:<Antara Lole>

Data intake reviewer:<->

Data storage location: < >

## Tabular data details:

<b>Total number of observations</b>	<359392>
<b>Total number of files</b>	<4>
<b>Total number of features</b>	<12> <Additional columns added later>
<b>Base format of the file</b>	<.csv >
<b>Size of the data</b>	<39.3 MB >

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

## Proposed Approach:

- 1) Data validation checks done: Null/missing value checks, data type checks, checked if any values needed replacement, checked for outliers, etc.
  - a. No missing values present
  - b. 'Date of Travel' present in string format converted to date format
  - c. 'City' had a few entries with missing state information (e.g., ORANGE COUNTY). Added the state information
  - d. Added new columns such as Profit, Year of Travel, etc. for further analysis
  - e. Plotted boxplots for outlier checks
- 2) Data Visualisations:
  - a. Analysed company wise distribution for total distance travelled, total trips made, total profit, and customer preference for cab company
  - b. Checked the distribution of customers across geographies and which geographies were most preferred by customers for travelling
  - c. Performed time series analysis to check for the growth/decline in profit, distance travelled, etc.

## Next Steps:

- 1) For more robust, visualizations I plan to analyse the data using Tableau
- 2) Build a machine learning model for predicting profits for both the companies