

Data Intake Report

Name: G2M Insights for cab investment

Report date: 10 July 2024

Internship Batch: LISUM35

Version: 1.0

Data intake by: Gloria Jepkurui

Data intake reviewer:

Data storage location:

Tabular data details:

Total number of observations	359392
Total number of files	
Total number of features	17
Base format of the file	Csv
Size of the data	1 MB

Proposed Approach:

- Understand the dataset, identifying key columns necessary for analysis
- Identify duplicates in your data and handle them while maintaining data integrity
- Ensure all the columns data types are parsed correctly
- Check for missing values and handle them
- Perform statistical analysis to understand summary statistics and data distribution
- Use necessary visualization tools to gain insights on trends and data patterns
- Formulate and test hypothesis using statistical tool (e.g t-test, anova) or what you see fit

Assumptions

- Missing values are assumed to be random

- The column 'Transaction ID' is assumed to be a unique identifier for every transaction •
- The data covers specific time period (2016 - 2018)
- The customer behavior is consistent over different months in every year •
 - 'Price charged' and 'cost of Trip' shows actual prices and costs respectively, without fees and discounts