

Data Intake Report

Name: G2M Case Study Report

date: July 13, 2023

Internship Batch: LISUM23

Version: 1.0

Data intake by: Anivirudhan Ramesh

Data intake reviewer: N/A

Data storage location: Github

Tabular data details: Cab_Data.csv

Total number of observations	359,392 rows
Total number of files	1
Total number of features	7 columns
Base format of the file	.csv
Size of the data	20.1 MB

Tabular data details: City.csv

Total number of observations	20 rows
Total number of files	1
Total number of features	3 columns
Base format of the file	.csv
Size of the data	759 bytes

Tabular data details: Customer_ID.csv

Total number of observations	49,171 rows
Total number of files	1
Total number of features	4 columns
Base format of the file	.csv
Size of the data	1.1 MB

Tabular data details: Transaction_ID.csv

Total number of observations	440,098 rows
Total number of files	1
Total number of features	3

Base format of the file	.csv
Size of the data	9 MB

Proposed Approach:

Hypothesis 1: Barplot to show the average distance and price per transaction by company

Hypothesis 2 : Boxplot to show the distribution of cost of the trips by the payment modes

Hypothesis 3: Barplot to show the popularity of cabs per city

Hypothesis 4: Calculate the percentage of male and female customers of each company

Hypothesis 5: Correlation matrix and heatmap

Hypothesis 6: Analyzing the travel patterns by age

Hypothesis 7: Analyzing the travel patterns by city

Hypothesis 8: Popularity of payment modes by city

Hypothesis 9: Calculate the profitability of each company