Data Intake Report

Name**:** G2M insight for Cab Investment firm
Report date: 20 June 2021
Internship Batch: LISUM01
Version: 1.0
Data intake by: Payal Upadhyay
Data intake reviewer: Payal Upadhyay
Data storage location: https://github.com/payal-upadhyay/Week2

Tabular data details:

Cab_Data:

Total number of observations	359392
Total number of files	1
Total number of features	7
Base format of the file	.csv
Size of the data	19.2+ MB

City:

Total number of observations	20
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	608.0+ bytes

Customer_ID:

Total number of observations	49171
Total number of files	1
Total number of features	4
Base format of the file	.csv
Size of the data	1.5+ MB

Transaction_ID:

Total number of observations	440098
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	10.1+ MB

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

Proposed Approach:

- We are analyzing data sets using graphs to see patterns, trends and anomalies as well as testing hypothesis there we are conducting an Exploratory Data Analysis
- We made sure that there are no duplicate values in the table as well as no null values by using the is_null() function.
- No null values or duplicate values were returned

Assumptions:

- Profit = Price_Charged – Cost_of_Trip
- We encountered a few negative values for profit which were kept assuming passengers earned discounts on trips.

- In order to conduct hypothesis testing using t-test we assumed equal sample variances