Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 13th March 2023
Internship Batch: LISUM19
Version:
Data intake by: Hira Fahim
Data intake reviewer:
Data storage location: [DataGlacier/week2 at main · hirafahim/DataGlacier (github.com)](github.com)

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	20.663 KB

Tabular data details: 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	1 KB

Tabular data details: Transactional_ID

Total number of observations	440098
Total number of files	1
Total number of features	7
Base format of the file	.csv
Size of the data	8788 KB

Tabular data details: 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	1027 KB

Proposed Approach:
- There are no "Null values" in all the datasets.
- Assumptions:
    - Outliers are present in "Price_Charged" feature but due to unavailability of trip duration details, we are not treating this as outlier.
    - Profit of rides are calculated keeping other factors constant.
    - Only "Price_Charged" and "Cost_of_trip" features used to calculate profit.