Data Intake Report

Name: G2M investment
Report date: 14/10/2021
Internship Batch:<Enter your batch code from Canvas course>
Version:<1.0>
Data intake by:Betty Wairegi
Data intake reviewer:
Data storage location: https://github.com/bettytom/dataglacier/

Tabular data details:

Total number of observations	35932
Total number of files	4
Total number of features	16
Base format of the file	.csv
Size of the data	7.61MB

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

Proposed Approach:
- The files; customer and transcation were joined together and then the cab data and city data were also joined. Thereafter the joined customer data was appended into the new cab data.
- There were no missing values after joining the files and also there were no duplicates in the data
- However there were outliers detected on the Price charged, cost of trip, Km travelled and Users column but these did not affect the outcome of the analysis.