Data Intake Report

Name: G2M Insights
Report date: 3/11/2025
Internship Batch: LISUM43
Version:1.0
Data intake by: Katrina Kirby
Data intake reviewer: <intern who reviewed the report>
Data storage location: https://github.com/kirbykt/DataSets
Tabular data details:

Total number of observations	359392
Total number of files	4
Total number of features	14
Base format of the file	.csv
Size of the data	137.98 MB

Proposed Approach:
- Mention approach of dedup validation (identification)
  I validated duplicates by identifying duplicate records using 'pandas' method. Duplicates were removed (if any). I assumed that missing values existed and drop any that was in the data.
- Mention your assumptions (if you assume any other thing for data quality analysis)
  -Missing values are expected
  -Missing data in key fields such as price, data, transaction ID
  -I assume city population and cab usage numbers are accurate, unless outliers appear