Data Intake Report

Name: Cab Investment Insights G2M Strategy
Report date: 14/06/2025
Internship Batch: LISUM46- 30 May
Version:<1.0>
Data intake by: Ashwini B N (ashwinibn3@icloud.com)
Data intake reviewer:<intern who reviewed the report>
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	6
Base format of the file	.Csv file
Size of the data	15 MB

Tabular data details: City.csv

Total number of observations	20 rows
Total number of files	1
Total number of features	3
Base format of the file	.csv file
Size of the data	< 1MB

Tabular data details: Customer_ID.csv

Total number of observations	4912 rows
Total number of files	1
Total number of features	4
Base format of the file	.csv file
Size of the data	3 MB

Tabular data details: Transaction_ID.csv

Total number of observations	359,392 rows
Total number of files	1
Total number of features	3
Base format of the file	.csv file
Size of the data	2 MB

Proposed Approach:

- Merged all datasets to create a master dataset covering trips, customers, transactions, and city demographics.

- Performed data cleaning (duplicates, missing values, outlier detection).

- Engineered key features such as **Profit**, **Profit Margin**, **Price per KM**, and **Cost per KM**.

- Conducted visual exploratory analysis using bar plots, boxplots, line plots, and histograms.

- Analyzed patterns across **time**, **company**, **city**, **age**, **gender**, and **income** segments.

- Formulated and validated hypotheses based on business questions provided in the project brief.

- Identified performance differences to recommend the more viable cab company for investment.