Data Intake Report

Name: G2M Cab Invesment Firm
Report date: 14/12/2022
Internship Batch:LISUM16
Version:1.0
Data intake by:Efe Kara
Data intake reviewer:<intern who reviewed the report>
Data storage location: https://github.com/DataGlacier/DataSets/*

Tabular data details:

Total number of observations	Customer_ID.csv | City.csv | Cab_Data.csv | Transaction_ID.csv
Total number of files	4
Total number of features	4 | 3 | 7 | 3
Base format of the file	.csv
Size of the data	1,1MB | 759 bytes | 21,9MB | 9MB

Note: <u>Replicate same table with file name if you have more than one file</u>.

Proposed Approach:
- Mention approach of dedup validation (identification)
- **inner join on intersection columns without data loss**
- Mention your assumptions (if you assume any other thing for data quality analysis)
- Data quality generally not essential for EDA but validations and cleaning progress shown
- in related ipynb file.