Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 13.04.2023
Internship Batch: LISUM20
Version: 1.0
Data intake by: Viktorija Hajduk
Data intake reviewer:<intern who reviewed the report>
Data storage location: https://github.com/VHayduk/Data-Science-Internship/blob/main/Week2/taxi_EDA_V1.ipynb

Tabular data details:

Total number of observations	359392
Total number of files	1 – Cab_data
Total number of features	7
Base format of the file	.csv
Size of the data	21.1MB

Total number of observations	20
Total number of files	1 – City
Total number of features	3
Base format of the file	.csv
Size of the data	759 Bytes

Total number of observations	49171
Total number of files	1 – Customer_ID
Total number of features	4
Base format of the file	.csv
Size of the data	1.1 MB

Total number of observations	440098
Total number of files	1 – Transaction_ID
Total number of features	3
Base format of the file	.csv
Size of the data	9 MB

Proposed Approach:

- Dedup validation: running data sets through a Python function.
- Mention your assumptions: Some observations may be missing data in other features when merged with other tables.