Data Intake Report

Name: Cab Investment data analysis
Report date: 2022-10-11
Internship Batch: LISUM14
Version: 1.0
Data intake by: Marija Babić
Data intake reviewer:
Data storage location: https://github.com/marija0408/NLP-Data-Science-Internship/tree/main/Week%202
Tabular data details:

Cab_Data

Total number of observations	359392
Total number of files	1
Total number of features	7
Base format of the file	csv
Size of the data	20.1 MB

City

Total number of observations	20
Total number of files	1
Total number of features	3
Base format of the file	csv
Size of the data	759B

Customer_ID

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

Transaction_ID

Total number of observations	440098
Total number of files	1
Total number of features	3
Base format of the file	csv
Size of the data	8.58 MB

- Table primary keys:
    - Cab data – Transaction_id
    - City data – City
    - Customer id data – customer id
    - Transaction id data – transaction id
- Cab data is the fact table and other tables are dimensions which were joined on the cab data table. After each join there was a step where unique values were checked which was the confirmation the the data have been joined properly and that no duplicates were detected.