

Data Intake Report

Name: Taxi

Report date: Aug-27-2021

Internship Batch: LISUM03

Version: 1.0

Data intake by: Corey Hamren

Data intake reviewer: <intern who reviewed the report>

Data storage location: <https://github.com/TheHamhams/taxi>

Tabular data details:

Cab_Data.csv

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	21.2 MB

City.csv

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	759 bytes

Customer_ID.csv

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	1.1 MB

Transaction_ID.csv

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	9.0MB

Proposed Approach:

- No null values were found in any datasets.
- cab_merged dataset was created by merging Cab_Data, Customer_ID, and Transaction_ID.
- Profit was done by subtracting the Cost of Trip from the Price Charged.