

Data Intake Report

Name: XYZ Cab Investment Analysis

Report date: 14.10.2024

Internship Batch:

Version: 1.0

Data intake by: Alissa Alkhova

Data intake reviewer: <intern who reviewed the report>

Data storage location:

Cab_Data.csv

Total number of observations	359392
Total number of files	
Total number of features	7
Base format of the file	.csv
Size of the data	21.2 MB

Customer_ID.csv

Total number of observations	49171
Total number of files	
Total number of features	4
Base format of the file	.csv
Size of the data	1.05 MB

Transaction_ID.csv

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

City.csv

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

Identified and removed duplicate rows in each dataset using the `drop_duplicates()` function. The Date of Travel field was originally not in a datetime format. I have assumed the dates are based on an Excel serial date format and converted them accordingly to the standard datetime format.

I assume that the datasets provided are accurate, complete.