

# Data Intake Report

Name: Cab Company Data Analysis

Report date: 09/13/2023

Internship Batch: LISUM25

Version: 1.0

Data intake by: Ethan Feiza

Data intake reviewer: N/A

Data storage location: [here](#)

## Tabular data details:

<b>Total number of observations</b>	359392
<b>Total number of files</b>	1
<b>Total number of features</b>	18
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	37.2 MB

**Note: Replicate same table with file name if you have more than one file.**

## Proposed Approach:

- Mention approach of dedup validation (identification)

Exporting the four individual data sets into Microsoft Excel, I will run functions in each file to ensure no duplicated values in unique fields. Additionally, I will utilize the =ISNA() function to ensure there are no null values in any table records.

- Mention your assumptions (if you assume any other thing for data quality analysis)

I will inner join the four files on appropriate shared fields in Tableau. This will result in some lost transactional data (precisely 80706 observations) as not every record has information for the 18 descriptive fields.