Data Intake Report

Name: G2M insight for cab investment
Report date: 06.03.2021
Internship Batch: LISP01
Version: 1.0
Data intake by: Buse Gungor
Data intake reviewer:<intern who reviewed the report>
Data storage location: https://github.com/DataGlacier/DataSets.git

Tabular data details:

Dataset name	cab_data
Total number of observations	359392
Total number of files	<Number of files received>
Total number of features	7
Base format of the file	csv
Size of the data	19.2+ MB

Dataset name	city
Total number of observations	20
Total number of files	<Number of files received>
Total number of features	3
Base format of the file	csv
Size of the data	560.0+ bytes

Dataset name	customer_id
Total number of observations	49171
Total number of files	<Number of files received>
Total number of features	4
Base format of the file	csv
Size of the data	1.5+ MB

Dataset name	transaction_id
Total number of observations	440098
Total number of files	<Number of files received>
Total number of features	3
Base format of the file	csv
Size of the data	10.1+ MB

Proposed Approach:
- Mention approach of dedup validation (identification)
- Mention your assumptions (if you assume any other thing for data quality analysis)

My approach is in other pdf documentation.