Data Intake Report

Name: <Melis Tekin Akcin>
Report date: <20.06.21>
Internship Batch:<Enter your batch code from Canvas course>
Version:<LISUM01>
Data intake by:<Melis Tekin Akcin>
Data intake reviewer:<intern who reviewed the report>
Data storage location: <https://github.com/melis-ta/Week2_DataGlacier_MelisTekinAkcin>

Tabular data details:

Total number of observations	<359392>
Total number of files	Cab_Data
Total number of features	<7>
Base format of the file	<.csv >
Size of the data	<19.2MB >

Tabular data details:

Total number of observations	<49171>
Total number of files	Customer_ID
Total number of features	<4>
Base format of the file	<.csv >
Size of the data	<1.5MB>

Tabular data details:

Total number of observations	<440098>
Total number of files	Transaction_ID
Total number of features	<3>
Base format of the file	<.csv >
Size of the data	<10.1MB>

Tabular data details:

Total number of observations	<20>
Total number of files	City
Total number of features	<3>
Base format of the file	<.csv >
Size of the data	<608.0 Bytes>

Proposed Approach:

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

I ignore the outliers in population and users in City data set.