Data Intake Report

Name: <Cab Industry>
Report date: <21/07/2021>
Internship Batch:<LISUM02>
Version:<1.0>
Data intake by:<Alhamza Ibrahim>
Data intake reviewer:<Bilal Yildiz>
Data storage location: <Github>

Tabular data details:

Total number of observations	<359392>
Total number of files	<1 , Cab_Data >
Total number of features	<7>
Base format of the file	<.csv >
Size of the data	<21.5 mb>

Total number of observations	<43171>
Total number of files	<1 , Customer_ID>
Total number of features	<4>
Base format of the file	<.csv >
Size of the data	<1 mb>

Total number of observations	<20>
Total number of files	<1 , City>
Total number of features	<3>
Base format of the file	<.csv >
Size of the data	<561 bytes>

Total number of observations	<440098>
Total number of files	<1 , Transaction_ID>
Total number of features	<3>
Base format of the file	<.csv >
Size of the data	<8.58 mb>

Proposed Approach:

Customer_ID column is used in both filels (Transaction_ID and Customer ID) and that helps with merging these 2 fils together , the Trasniction_id colum can be found in the Cab_Data file as well , finally, we can merge the cities info from the City,csv file since we have the colum City in both City.csv and Cab_Data.csv.
And now we have our final data set that contains the above mentioned 4 files in one single file.