

Data Intake Report

Name: <Customer Segmentation>

Report date: <7/25/2021>

Internship Batch:<LISUM01>

Version:<1.0>

Data intake by:<Anye Igwacho>

Data intake reviewer:<Data Glacier>

Data storage location:

<<https://drive.google.com/drive/folders/1bfCpJIKmp6IHxiLPWvOS2nU1dc24pViB>>

Tabular data details:

Total number of observations	<100000>
Total number of files	<1>
Total number of features	<47>
Base format of the file	<.csv>
Size of the data	<154MB>

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

Proposed Approach:

- Mention approach of dedup validation (identification)
Dedup in MATLAB,python, rows with NAN and empty cells were deleted
Normalized features, format strings

