

Data Intake Report

Name: <Customer Segmentation>

Report date: <7/25/2021>

Internship Batch:<LISUM01>

Version:<1.0>

Data intake by:<Neil Erwin Mauricio Villadoma>

Data intake reviewer:<Data Glacier>

Data storage location: <

https://drive.google.com/file/d/1rnBMZ7uC9I1_dZ38Qal77EcPQJO65Kje/view?usp=sharing>

Tabular data details:

| | |
|-------------------------------------|-----------|
| Total number of observations | <1000000> |
| 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

