

# Data Intake Report

Name: Customer Segmentation

Report date: 10/04/21

Internship Batch: LISUM03

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](https://drive.google.com/file/d/1rnBMZ7uC9I1_dZ38Qal77EcPQJO65Kje/view)

## Tabular data details:

cust\_seg.csv

<b>Total number of observations</b>	1000000
<b>Total number of files</b>	1
<b>Total number of features</b>	47
<b>Base format of the file</b>	csv
<b>Size of the data</b>	154 MB

## Proposed Approach:

- Mention approach of dedup validation (identification)
- Rows with NAN and empty cells were deleted
- Normalized features, format strings