Data Intake Report

Name: Customer Segmentation

Report date: 9/16/2022

Internship Batch: LISUM12

Version:<1.0>

Data intake by: N/A

Data intake reviewer: Brennan Clinch, Kutay Selçuk, Rohit Sunku, Zhan Shi

Data storage location: https://github.com/zhanshi1997/dgintern/tree/main

**Tabular data details:**

|  |  |
| --- | --- |
| **Total number of observations** | 1000000 |
| **Total number of files** | 1 |
| **Total number of features** | 48 |
| **Base format of the file** | .csv |
| **Size of the data** | 168.6MB |

**Proposed Approach:**

* Go through the ‘ncodpers’ column to check if there’re any repeated records of customers.
* When doing the data cleaning, It’s not necessary to remove all the null values (N/A) based on the features we take into consideration.
* Different combination of features might be used as the criteria of categorization, but all of them should include same kind of features.
* The values under the last 24 columns (starting from ind\_ahor\_fin\_ult) should be either 0 or 1. Null values will not be counted when the feature is considered for categorization.