Data Intake Report

Name: Cross Selling Recommendation

Report date: 26.04.2021

Internship Batch: LISP01

Version: 1.0

Data intake by: Buse Gungor

Data intake reviewer:<intern who reviewed the report>

Data storage location: https://drive.google.com/file/d/16-nzZR91-ijrfjUcI2PniTpOgrvFAykA/view

**Tabular data details:**

|  |  |
| --- | --- |
| **Dataset name** | train |
| **Total number of observations** | 13647309 |
| **Total number of files** | <Number of files received> |
| **Total number of features** | 48 |
| **Base format of the file** | csv |
| **Size of the data** | 4.9+ GB |

|  |  |
| --- | --- |
| **Dataset name** | test |
| **Total number of observations** | 929615 |
| **Total number of files** | <Number of files received> |
| **Total number of features** | 24 |
| **Base format of the file** | csv |
| **Size of the data** | 170.2+ MB |

**Proposed Approach:**

* Mention approach of dedup validation (identification)
* Mention your assumptions (if you assume any other thing for data quality analysis)

**My approach is in other pdf documentation.**