

Data Intake Report

Name: Cross Selling Recommendation

Report date: 15/8/2024

Internship Batch: LISUM35

Version:1.0

Data intake by: Munirah Alfahaid

Cross Selling data details:

Total number of observations	929615
Total number of files	1
Total number of features	24
Base format of the file	csv
Size of the data	110.3 MB

Proposed Approach

1. Data Understanding

To begin, we will load the dataset and conduct an initial exploration to familiarize ourselves with the data. Key steps include reviewing the first few rows of the dataset, assessing data types to ensure alignment with the expected format, and generating summary statistics for both numerical and categorical columns.

2. Data Cleansing and Transformation

Data cleansing will focus on handling missing values, either through imputation or removal, depending on the extent and significance of the missing data. We will also address any outliers in critical columns such as `age` and `renta` to ensure they do not skew the analysis. Necessary data type conversions will be performed, particularly for date columns like `fecha_alta`. Additionally, categorical variables will be encoded appropriately to prepare the dataset for subsequent analysis.

3. Exploratory Data Analysis (EDA)

Exploratory Data Analysis will be conducted to uncover patterns and relationships within the data.