

Team members' details

Group Name: Cross-Selling Recommendations

Name: Sri Saikamal Priyank Samireddi

Email: kp.samireddi@gmail.com

Country: United Kingdom

College: Bayes Business School

Specialisation: Data Analyst

Problem description

XYZ Credit Union has good performance in selling banking products to new customers, but struggles with cross-selling additional products to existing customers. They want to increase cross-selling to boost revenue.

Business understanding

- The goal is to identify potential products to cross-sell to existing customers.
- This will increase revenue and customer retention.
- Need to analyse customer demographics, income, current product holdings, transaction history, etc.
- Identify insights like which customer segments have higher cross-selling potential.
- Suggest marketing strategies tailored to customer profiles.

Project lifecycle along with the deadline

Data Understanding: 3 days

Receive data files, inspect data quality, document sources, metrics, etc.

Data Preparation: 5 days

Cleaning, and transforming data to an analysis-ready state.

Exploratory Analysis: 6 days

Analyse data to identify trends, correlations, customer segments, etc.

Model Development: 5 days

Build models to predict the likelihood of cross-buying different products.

Results Presentation: 1 days

Create an executive presentation with key insights and recommendations.

Dashboard Development: 5 days

Develop an interactive dashboard for ongoing tracking of KPIs.

Final Delivery: 3 days

Wrap up analysis, recommendations, files, and documents.

Data Intake report

Name: G2M Strategy

Report date: 19-08-2023

Internship Batch: [LISUM23: 30](#)

Version: 1.0

Data intake by: Sri Saikamal Priyank Samireddi

Data intake reviewer:

Data storage location: <https://drive.google.com/file/d/16-nzZR91ijrfjUcI2PniTpOgrvFAyKA/view>

Tabular data details:

Total number of observations	13647309
Total number of files	1
Total number of features	48
Base format of the file	.csv
Size of the data	2.5 GB

- 52 Features(2 columns removed and 6 columns engineered)
- Timeframe of the data: 2015-01-28 to 2015-02-28
- Total data points: 13647309 rows

Assumptions:

- Missing or invalid values for numerical columns like age, renta, etc. can be replaced with the mean/median value. Removing them entirely may bias the analysis.

- For categorical columns like sexo, ind_empleado, etc. invalid values can be replaced with the mode or most common value.
- Columns like fecha_alta, ult_fec_cli_1t, fecha_dato which capture date information are unlikely to have missing values. If they do, those rows can be removed.
- The columns starting with 'ind' seem to be indicator variables denoting possession of a banking product. Missing values can be assumed as 'No' or 0 values.
- Segmento indicates customer segments; missing values can be replaced with the most common 'Individuals' segment.
- Columns like pais_residencia, cod_prov, nomprov may have some abbreviated values that need expansion to full names.
- There are some columns like canal_entrada, ind_actividad_cliente that seem important for analysis and any missing values should be replaced carefully.
- For any remaining missing values, imputation techniques like KNN can be used to fill them in.
- Outliers in numerical columns need investigation before removal to avoid data loss. Capping at the 99th percentile could be an option.
- Text columns may need standardisation of spellings, casing, abbreviations, etc.

GitHub Repo link

<https://github.com/kamalpriyank/Cross-Selling-Recommendations>