Group Name: D_for_Data
Name: Dylan Kua
Email: dk3332@tc.columbia.edu
Country: USA
Specialization: Data Analyst

Problem Description:

XYZ Credit Union in Latin America excels in selling banking products but faces challenges in cross-selling—most customers only hold a single product. This situation shows their opportunity to increase revenue from existing clients. As a data analyst, the task is to analyze customer data and identify actionable insights that can help XYZ Credit Union boost cross-selling without relying on machine learning.

Business Understanding:

The main objective is to improve cross-selling by encouraging existing customers to purchase additional banking products. This involves:

- Understanding customer behavior and preferences.
- Identifying factors that influence product ownership.
- Suggesting targeted strategies to promote relevant products to different customer segments.

Key business goals:

1. Increase the average number of products per customer.
2. Enhance customer satisfaction and engagement.
3. Boost profitability through more comprehensive service offerings to current customers.

With this data, we can investigate factors such as:

- Customer segmentation based on income, seniority, or activity level.
- Customer behavior (e.g., which segments buy which products).
- Opportunities for bundling products based on customer profiles.

Project Lifecycle:

1. Data Collection & Exploration (1 Week)
   - Collect and load the data provided.
   - Explore the dataset to understand its structure, missing values, and inconsistencies.
   - Identify key variables relevant to cross-selling (e.g., customer demographics, existing product ownership, and segmentation).

2. Data Cleaning & Preprocessing (1 Week)
   - Handle missing values and outliers (e.g., customer age, income).
   - Convert categorical variables (e.g., customer segments, employee index) into usable formats.
   - Ensure data consistency across relevant variables (e.g., income ranges, customer seniority).
   - Deliverable: Cleaned and preprocessed dataset, ready for analysis.

3. Exploratory Data Analysis (EDA) (1 Week+)

- Create new variables or features that might help in understanding cross-selling opportunities (e.g., customer tenure in months, activity index).
- Segment customers based on factors such as income level, seniority, and product ownership.
- Perform an in-depth analysis of the data to find patterns in customer behavior.
- Identify which products are commonly bought together, customer profiles for each product, and potential cross-selling opportunities.
- Analyze correlations between customer demographics and product ownership.
- Deliverable: visualizations of EDA findings

4. Customer Segmentation & Strategy Development (1 Week+)
- Segment customers into distinct groups based on the insights gathered during EDA (e.g., by income, age, seniority, or product usage).
- Identify high-potential customer segments that could benefit from cross-selling efforts.
- Develop tailored strategies for each segment, recommending specific products or bundles based on their profiles.
- Deliverable: Segmentation analysis and cross-selling strategies report.

5. Recommendations & Presentation (5 Days)
- Present actionable recommendations for XYZ Credit Union on how to increase cross-selling (e.g., offering personalized product bundles, targeting high-potential segments).
- Provide insights into which products should be prioritized for cross-selling based on customer data.
- Deliverable: Final presentation with recommendations and an implementation plan.

Total Project Duration: 6 weeks

Data Intake Report

Name: Data Analyst: Cross selling recommendation - How to increase cross selling of Banking Products
Report date: 2024.8.28
Internship Batch: LISUM35
Version:1.0
Data intake by: Kua Hong Rui
Data intake reviewer:
Data storage location:

Tabular data details:

Test.csv

Total number of observations	929, 615
Total number of files	
Total number of features	24
Base format of the file	.csv
Size of the data	107.7 MB

Train.csv

Total number of observations	13, 647, 309
Total number of files	
Total number of features	49
Base format of the file	.csv
Size of the data	2.13 GB

Proposed Approach:

- Check missing values for every column in each data set.
- Used drop_duplicates function on python to remove duplicate rows.
- Due to the large size of data, Train data will be sufficient for analysis.
- Adjust consistent data types for columns.