Group Name: Data Detective

Name: Syed Ahmad Sohail

Email: syedahmad.sohail@ucalgary.ca

Country: Canada

College/Company: Data Glacier

Specialization: Data Analyst

Problem description:

XYZ Credit Union is a banking institution located in Latin America, and it's doing well in selling different types of banking products such as credit cards, deposit accounts, retirement accounts, and safe deposit boxes to its customers. However, despite their success in selling these products, the bank is facing a challenge of not being able to sell more than one product to their existing customers. This means that the bank is not performing well in cross-selling, which is an important strategy to increase revenue and customer loyalty. Cross-selling is the practice of encouraging customers to purchase additional products or services beyond their initial purchase. In this case, the bank is missing out on the opportunity to sell more of their products to their existing customers, which could lead to increased revenue and stronger customer relationships. As a result, the bank has approached ABC analytics to help them solve this problem. The goal is to identify ways to increase cross-selling and provide actionable insights that the bank can implement to improve its cross-selling strategy. By doing so, the bank can improve its overall performance and maintain its competitive edge in the market.

Business understanding:

XYZ Credit Union is a successful banking institution in Latin America that offers a variety of financial products to its customers. While they have been successful in selling their products, they have noticed that they are not able to sell more than one product to their existing customers. This means that customers are not buying additional products from the bank, which can hurt the bank's profitability and reputation. To tackle this challenge, XYZ Credit Union has reached out to Data Detective for assistance. The goal is to identify ways to increase cross-selling, which is the practice of encouraging customers to purchase additional products beyond their initial purchase. The bank hopes to improve its cross-selling strategy to not only increase revenue but also strengthen customer relationships. By doing so, they hope to stay competitive in the market and continue to provide quality financial services to their customers.

Project lifecycle along with deadline:

Week 1 (Due: March 26th 2023):

During the first week of the project, the focus will be on understanding the data that we have for analysis. We need to identify the type of data we are working with, such as structured or

unstructured, and determine the problems in the data. These problems could include missing values (NA values), outliers, skewed data, etc. We will also discuss the approaches that we will use to overcome these problems, such as imputing missing values using mean/median/mode, model-based approaches, or Weight of Evidence (WOE). Each team member will work on a different approach to handling missing values and outliers.

Week 2 (Due: April 2nd 2023):

During the second week of the project, we will work on data cleansing and transformation. We will share our code on GitHub and each member will work on different techniques to clean the data, such as imputing missing values or handling outliers. For example, if one team member is using mean to impute values, another member may experiment with segmented or model-based approaches. We will also use different NLP (Natural Language Processing) techniques to clean the data, such as regex (regular expressions) and Python.

Week 3 (Due: April 9th 2023):

During the third week of the project, we will perform exploratory data analysis (EDA) on the cleaned data. This will involve creating various visualizations and summary statistics to gain a better understanding of the data and identify any patterns or trends. We will also perform feature engineering and try different feature selection techniques. Based on the findings from the EDA, we will provide our final recommendations to the XYZ Credit Union on how they can improve their cross-selling strategy.

Week 4 (Due: April 16th 2023):

In this week, the team will prepare an EDA presentation for business users. The presentation should include important insights gained from the exploratory data analysis conducted in the previous week. The team should focus on highlighting trends, patterns, and relationships in the data that can help the business understand its customers better. The final slide of the presentation should be dedicated to technical users and should contain recommended models for this data set.

Week 5 (Due: April 23rd 2023):

In this week, the team will select a base model and then explore one model of each family. For example, if it is a classification problem, then the team should select one model for linear models, one model for ensemble, one model for boosting, and other models if time permits, like stacking. The team should ensure that the selected models fit the business requirements. For example, if the business does not want a black box model, then the team should only select models that can be used to explain the prediction. The team should also evaluate the performance of each model based on relevant metrics and select the best model for further refinement.

Week 6 (Due: April 30th 2023):

As it is a group assignment, schedule a call with your team to discuss the solutions presented by each member. Evaluate and select the solution that best meets the project requirements and objectives. Prepare a PowerPoint presentation to showcase the selected solution, including its methodology, data analysis process, and insights gained from the project. The presentation should also highlight the recommended models and their respective performances. Make sure to include a detailed description of the solution, key findings, and actionable recommendations. It should be presented in a clear and concise manner, taking into consideration the audience (business stakeholders, technical team members, etc.). The presentation should be delivered within the agreed-upon timeline and should be accompanied by a written report summarizing the key points covered in the presentation. The final report should include a detailed explanation of the data analysis process, challenges faced, solutions implemented, and the overall impact of the project.