Project Proposal

Class:	AY2020 S2	Team:	1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / **10** /11/12

Project Title

Remedy recommendation engine for retaining high churn-risk bank customers

Business Problem or Opportunity Statement

Identify customers who are highly likely to churn and recommend the top personalised remedy action(s) that can make these high-risk customers stay as clients.

Justification

Churn represents big opportunity cost:
1. Lost revenue
2. Increased CAC to re-acquire the same number of customers

With the financial services sector being highly saturated, existing players face several challenges in differentiating their service offerings. The ratio of customers switching over to other financial service providers has increased significantly due to a wide range of alternative financial products and competing providers. As a result, customer churn rates today have gone high.

In financial services, churn is of particular concern to companies with non-binding contracts, like credit card companies, insurance agencies, credit unions, and banks. For these organizations, attrition rates as high as 25-30% are not uncommon, and even companies with some type of annual contract may experience attrition rates around 5-7%. With a tool that can identify the clients that are highly likely to churn, financial institutions can take actions to prevent losing these clients before it's too late.

Proposed Approach and POC

We have a 3-step process in our proposed approach:
1. First, we make use of our dataset to identify and rate customers based on their expected churn rate.
2. Next, we establish key parameters that affect their churn rate, and identify socio-economic patterns that might be of help in tailoring our strategies.
3. Lastly, we adapt our strategies and approaches to reduce their churn rate e.g. improving the UX for millennials *a la* OCBC FRANK, improving customer outreach for seniors etc.

Our solution allows the user to set a customised tolerance for churn-risk floor and ceiling threshold and identify customers whose risk of churning within that user-defined threshold (i.e. if the user only wants to target low-hanging fruits, he will want to set a lower threshold while on the other hand if the user wants to target urgent flight-risk cases, he will prefer to identify customers that have high churn-risk.) In addition to identifying such customers, the engine is able to recommend remedy actions, ranked by magnitude of impact, that are personalised to each at-risk customer.

Contextual Questions:
1. What does research literature say about why customers churn in real life?
2. What pattern of behaviour do customers show when they are about to churn?

Definitional Questions:
3. Which columns of the data are useful for the purpose of the project?
4. What does it mean when a data row indicates that the customer is retained yet his bank balance is 0?
5. How is "isActive" defined?
6. Are most customers in the data still customers, or have most of them churned? If most of them had churned, it might not be a good idea to give recommendations to remedy churn but instead identify trends among the customers that churn.
7. Is the data under the geography column defined by the person's citizenship or by the geographical location where the person registered with the bank?
8. What factors affect the credit score and how is it calculated? Is credit score a linear scale? Will the credit score have a different meaning in different countries? What is a "good" credit score? Are the credit scores dependent on the other columns e.g. income?

Insight Questions:
9. What is the nature of the bank products? Do all customers start off a default number of products?
10. Are all customers worth the effort to carry out remedy actions? Can we assign a value to each customer based on their credit score to determine the threshold of a worthy customer? What would be the determining factors of this value and what profile would a high value customer have? What is the breakeven point?
11. Is the estimated salary based on the other indicators or it is independently acquired?
12. Does increasing tenure tend to lead to higher risk of churn or the other way around?
13. How do we know what the best remedy is? Can we identify previously high churn risk customers who were successfully retained?
14. Might the gender column introduce bias into the data?
15. How can we replicate conditions for retained customers for customers that are likely to churn?
16. Is there a possibility for an early warning system, i.e. the ability to accurately identify specific customers who are about to churn?
17. What metrics can we best use to predict customer attrition?
18. What are the biggest pain points in our current process that lead towards customer churn?

Predicting Churn for Bank Customers
https://www.kaggle.com/adammaus/predicting-churn-for-bank-customers

Important References

1. https://www.qualtrics.com/blog/customer-churn-banking/

Talks about the factors that lead to customers churning from the bank, mainly in the form of poor customer service and banks also not keeping themselves close to the ground on what customers find lacking in banks' services (e.g. due to not having frequent enough customer experience surveys).

Also mentions the factors that make customers attracted to a bank: low fees, breadth of products (mortgage loans, credit card schemes), return rates (e.g. for ETF products), high availability of branch locations, having high quality customer service. This may suggest that it would be useful to try to predict what stage of life the bank customer is at, so as to determine what needs he/she has that the bank can step in to fulfill.

Further, it also means that the bank may have a customer loyalty weakness that is relevant across all of its customers. It may help to improve in one area of its metrics, almost regardless of who the customer is.

Looking another way, it may also mean that the bank is good at handling a particular strata of customers, given its range of products, for example.

2.   https://www.bain.com/insights/what-loyalty-leading-banks-do-differently-slideshow/

Says that banks that have higher customer loyalty tend to be those that allow customers' issues to be settled online as often as possible, because that eases customers' convenience

3.   https://core.ac.uk/download/pdf/6504665.pdf

Paper that talks about loyalty programmes. Different customers will perceive different products differently. So different customers may need a different number of products, perhaps like higher income customers need a higher average number of products. It also mentioned that a retained customer is not necessarily a loyal one. There are in fact a few classes of retained customers, including: trapped, truly loyal, accessible and risky. Can the data set allow us to infer which class does each customer belong to?

4.   https://www.quora.com/What-are-the-different-products-offered-by-banks

Gives an idea of the typical types of products offered by a retail bank to individual customers.

5.   https://www.experian.com/blogs/ask-experian/credit-education/score-basics/500-credit-score/

Explains the factors that lowers your credit score aka FICO credit score and how to bring it up. Also gives us a sense of what is a "good", "average", and "bad" FICO score.

Related Work:

1. [Why Do Customers Stop Doing Business With a Bank?](): Medium article detailing a step-by-step customer churn analysis on a similar dataset
2. [Customer Churn Analysis](): TDS article detailing a step-by-step customer churn EDA and churn prediction using decision trees
3. [Behavioral Attributes and Financial Churn Prediction](): Research paper on customer churn analysis based on customer characteristics and behavioral patterns
4. [The Complete Guide to Checking Account Churn Prediction in BFSI Domain](): AnalyticsVidhya article on the background of customer churn in banking/financial institutions and brief overview of machine learning techniques to analyze churn trends
5. [Developing a Prediction Model for Customer Churn from Electronic Banking Services using Data Mining](): Research paper on using SVM techniques to predict customer churn via customer interaction with e-banking services
6. [Customer Churn Prediction Using Machine Learning: Main Approaches and Models](): KDNuggets article on possible machine learning approaches to analyzing customer churn

Background Knowledge:

1. [Addressing Customer Churn in Banking](): Detailed article by Tiger Analytics on the background, causes, and possible solutions to reducing customer attrition

[How Banks Can Close the Back Door on Attrition](): BCG article on the background, causes, and possible solutions to reducing customer attrition

---

Project Schedule

The project will be separated into 3 main stages:

1. Data Preprocessing, Cleaning and Exploratory Analysis
   1. To ensure that only entries with complete and accurate data is used
   2. To identify any interesting distributions, correlations or patterns
   3. To develop primary hypotheses through data exploration and visualisation

2. Development of Engine
   a. To develop a predictive model of the likelihood of customer churn based on key metrics
   b. To suggest remedial actions which will result in a lower likelihood of churn

3. Validation and Verification of Results
   a. To validate using test data that the predictive model is able to identify customers with a high risk of churn
   b. To verify that the predictive model is able to provide effective remedial actions to lower the overall bank churn rate
   c. To produce a concise report of key findings

The timeline can be found in **Appendix A**.

Appendix A: Timeline and Key Milestones

Project Proposal
Feb 28, 2021

Development of Engine
Mar 19, 2021

Verification and Validation Complete
Apr 2, 2021

Project Submission
Apr 4, 2021

	Week 7	8	9	10	11	12

Data Preprocessing and Cleaning
Mar 1 - Mar 3 — Data Cleaning and Visualisation
Mar 3 - Mar 5 — Exploratory Data Analysis

Development of Engine
Development of Predictive Models — Mar 8 - Mar 19
Development of Remedial Actions — Mar 15 - Mar 19

Verification and Validation
Verification of Predictive Model Performance — Mar 22 - Mar 26
Validation of Real-World Usability — Mar 29 - Apr 2
Development of Final Report — Mar 22 - Apr 2