Case Assignment: Predicting and Preventing Customer Churn at QWE

MMA On-Line Datathon 2018

Background and Data:
The main objective of this exercise is to understand which factors are affecting customer churn at QWE and to be able to identify, with reasonable accuracy, those customers who are most likely to leave. Please see the case for further description of the background and business objectives.

The data is provided in a separate Excel file. Column C contains information on whether the customer has churned or not. The data description is provided in the case.

Your overall task is to analyze customer churn at QWE, come up with insights as to what drives up churn risk for some customers, and suggest analytical support for proactive customer retention programs.

Some particular questions you may want to address:

1. Mr. Well (the case protagonist) believes that two important predictors of customer churn are Customer Age (i.e., tenure with QWE) and CHI score. Does your data analysis support these beliefs?
2. Is there a natural customer segmentation with respect to churn risk that QWE should be thinking about? If so, what is it? What churn factors are particularly important in different segments? Or does the same set of factors impact all customers, and thus segmentation is not particularly useful?
3. A proactive customer retention program should be focused on a small subset of particularly risky customers. Can you identify such a subset? How accurate is your churn prediction mechanism for this subset? (Hint: think and discuss what "accuracy" means in this context)
4. For the riskiest group identified in the previous part, what characteristics (other than churn risk) separate them from the rest of QWE customers? What should customer retention communications aimed at this group focus on? If different communications should be used for different subgroups, make sure to estimate the importance of each subgroup to QWE.
5. It is estimated that an average ARPU (Average Revenue per User) for QWE is about $100 per month. An outbound call-based program is proposed to improve customer retention. Each phone call (assuming right-party-connect is made) will cost QWE around $10. Can your predictive model(s) be used to support such a program? If so, which customers should be contacted and what financial returns can be expected (make sure you state your assumptions).

Your submission should be in the form of a presentation to QWE management addressing some or all of the questions above. The presentation should be limited to 10 pages and non-technical in nature; it can be followed by an Appendix containing more technical descriptions of your work. While you can also submit your code in a separate file, the presentation should be self-contained (i.e., it should not be necessary for the reader to open your code).

Hints / Suggestions: your first step should be to try to get some intuition of the relationship between churn and customer age. A very useful step may be to bin the age variable into some reasonable number of categories (e.g. 10) and then analyze how the churn risk changes with

the category. Another useful step is to examine your potential predictors. While there are no missing values, there are a lot of 0's and it is likely (though not clear from the case) that many 0's actually represent missing data.

Note that Mr. Well is suggesting that the relationship between Churn and Age is not linear. If he is right, it may be useful to construct different models for different age groups (though you should base this decision on the relationship you observe in the data, not on someone's hunch).