Marketing Analytics
David M. Muir

Directions: Read the description for the brief and then thoroughly yet concisely address the questions in a report. The report for the brief should be typewritten in standard formatting and be no longer than ve (5) pages, excluding an appendix. Include all tables, gures, calculations, and R script in the appendix. Label any tables and gures appropriately. Note: Be sure to follow the format as adapted from the Syllabus:
(i) Executive Summary; (ii) Purpose Statement; (iii) Basic Description of Statistical Methodology; (iv) Analysis of Data Findings; (v) Conclusions and Managerial Recommendations; and (vi) an Appendix.

Due Date: Wed, March 7, 2018, 12:00am

PROJECT

Retaining Customers at Pilgrim Bank

1.1 Introduction

It was February 8, 2017 and Alexis Green was nishing her rst month as an analyst in Pilgrim Bank's online banking group (https://www.pilgrimbank.com/).

Lexi had a meeting with Erica Dorstamp, head of IT services, in the building next door. \Erica," she announced while stepping into her o ce, \I need to analyze customer retention over the weekend and was hoping you could help me access some data."

Erica was used to such requests. \No problem. We've just pulled and cleaned data on a random sample of about 32,000 customers for a separate project. I can give you that dataset right away, depending on what other variables you want included. Are you alright starting with the random sample of 32,000? What sort of information do you need besides customer retention data?"

\If you would be so kind as to give me their pro tability for year ends 2015 and 2016 as well as whether they use the online channel, including online bill pay, that would be great."

\Alright, I'll get this set up for you immediately," Erica agreed. \I'll also put some key demographic information in there for you."

By the day's end, Lexi had her hands on the data. Erica successfully had captured the data of customers from the 2015 sample in a second time period as of December 2016. Scanning the data le, Lexi noticed many customers in her 2015 sample, in excess of 10%, had left the bank in the past year. Lexi thought she should determine whether online banking had any impact on retention. It seemed as if everyone who worked for the bank assumed customers who used online banking were more likely to remain with the bank, but, according to her boss, Ravi Raman, the assumption had never been tested.

Lexi's data le included all the 2015 and 2016 data, including demographic information about customers. For customers who had left the bank the 2016 data was blank. Lexi looked forward to making sense of the data and portraying the retention story in a clear, e cient manner. In particular, Lexi wanted to know (i) whether the use of the online channel in uenced customer retention; and (ii) if it was possible to predict the likelihood of retention for individual customers, including new customers who join the bank. Lexi felt con dent that with this understanding she could make a well-reasoned recommendation to the management committee regarding its customer Internet strategy. Lexi knew her Lerner training was about to pay o .

Lexi knew the rst step would be to clean the data set by creating an outcome variable and by properly factorizing the categorical variables so their natural ordering was preserved. You will employ the raw data le (.csv) for all your analyses, but will need the Excel le (.xlsx) to understand how the raw data are coded, i.e., you will need to know how the variables in the data are de ned.

1.2 Questions

1. How well does customer channel use in 2015 predict whether customers stay with the bank through 2016? Does knowing the demographics of a customer in 2015 help predict customer retention in 2016? (Hint: Use the full data set to analyze the predictors that best suit the model.)

2. Employ a training-and-test regimen on the ve classi cation methods to determine the most suitable model with which to predict customer retention. What classi cation method works best and why?

3. Can you make predictions about individual customers in terms of future retention? If so, what predic-tion about customer retention does your preferred classi cation method make for a 22 year old living in Northeast Texas who earns $65,000 and who currently is interested in joining the bank?

4. What should Lexi Green and the management of Pilgrim Bank do regarding customer retention in relation to the prediction model you built? (Hint: What variables are under management's control?) Do you recommend any marketing programs that may encourage customer loyalty or satisfaction?