***Background:***

This case examines the fictitious company, High Note, which is an online music streaming service where the majority of users are free subscribers.  The “freemium” model allows for users to build a profile and use the service for free.  They hear advertisements at regular intervals.  Paid subscribers receive more features and hear no ads.  You should think about how freemium can be monetized and what is new with this freemium model.  In the case, the High Note employees turn to customer data to try and figure out how to increase paid subscriptions.

***Instructions:***

The questions for this exercise are based upon the assigned reading for this class, Freemium.  Please complete this reading before answering these questions.  Links for the dataset and R script are provided here:

[freemium\_Analysis\_Part2.R](https://canvas.cmu.edu/courses/9526/files/3618460/download?wrap=1)

[High Note data csv.csvPreview the document](https://canvas.cmu.edu/courses/9526/files/3508802/download?wrap=1)

[High Note data dictionary.xlsxPreview the document](https://canvas.cmu.edu/courses/9526/files/3508800/download?wrap=1)

This script guides you in conducting an exploratory data analysis and a classification and regression tree to build a predictive model that helps you understand what users are most likely to upgrade from free to paid subscriptions.

This exercise is to be completed in your group as listed on Canvas.

Good answers may require assumptions of facts that may not be presented in the case.  You are welcome to make these assumptions, but please state these assumptions and briefly justify why that are reasonable.  Also, you may use whatever resources you can locate to provide further information about this industry or the web in general.  Please reference your sources.

Your response must be typed, double spaced, with one-inch margins, and a 10 to 12 point font size, and must be a PDF file.  This writeup must not exceed 3 pages in length.  You may attach exhibits, tables, and/or graphs to support your arguments.  These supporting materials must be referenced in the text and do not count toward the 3 page limit.

***Required:***

Lisa has received your previous memo, and has some concerns.  Specifically she has responded as follows:

Thank you for your previous memo.  I’m excited about using a data-based approach to convert more free users to paid premium subscribers.  I’ve decided that this should be the focus of our project.  The investors are putting a lot of pressure on management to improve the number of paid subscribers quickly.

I was told that there are new Machine Learning techniques like Decision Trees that we could try instead of Logistic Regression.  Would you give me a comparison between your logistic regression model and a new decision tree?  Please just show me your best results.  Would you give me a recommendation of which one to use, and give me some evidence so when I make a presentation to top management I can convince them that our model is a good one.

I did some checking about our dataset.  Our dataset has three periods, CURRENT period spanning 3(?) months, the PRE period spanning three months with corresponding variables suffixed by 1 (e.g. delta1\_friend count is the number of new friends made in the three month period that PREcedes the CURRENT period), and the POST three month period that follows the CURRENT period. The POST period is suffixed by 2 (e.g. delta2\_songsListened is the number of songs listened to in the POST.  Instead of just looking at the current period should we use the change from the historical period?  I think using the changes might be better than looking at the current levels.  For example, isn’t is more important to understand that a customer is using our platform more rather than just looking at total usage?  Please do an analysis and give me your opinion.

Finally, top management has signed off on a new promotional offer.  They are willing to let us give a free subscription for three months to any customer that we want to target.  To make your analysis easier please assume that revenues are $3 per month for paid subscribers and we receive $0.125 per month from advertisers for our free subscribers.  Let’s assume that our promotional offer would give three months for free and if we get a customer to subscribe that they will stay with us for one year (or twelve months).  Also to keep the analysis simple let’s assume that there are no costs of making the offers.  But I am worried about the opportunity costs associated with giving offers to customers that would have subscribed even without the promotion, as well as losing out on the advertising revenue for customers that we give the promotional offer to but don’t sign up in the future.  Which users should we target and what would the results be?  If we assume that our sample of current users is representative of new customers, let's test the results on the data that you have -- do you think this is the correct way to do it?

Unfortunately, we need the results by Monday so please get to work on this right away.

Write a memo to respond to Lisa.  My learning goal for this exercise is not for you to provide me tables and output from R, but to use your analysis and model to interpret the data and tell a story that persuades the reader about motivating users to go “from free to fee”.

Your memo will be evaluated on the following four dimensions (10 points for each dimension for a total of 40 points):

Content: How well did your memo answer one of the questions posed?

Analysis: Did you support your recommendation?

Concept: Did you demonstrate good knowledge of data science principles and data mining concepts?

Overall: Is your recommendation persuasive?