

Midterm

[Submit Assignment](#)

Due Nov 30 by 12:59pm **Points** 40 **Submitting** a file upload

Midterm Exam

Three-Week Take Home Exam (Due: November 30, 12:59 pm, just before class)

Introduction

The “freemium” business model — widely used by online services such as LinkedIn, Match.com, Dropbox, and music-listening sites — divides user populations into groups that use the service for **free** and groups that pay a **fee** for additional features. Key points related to the freemium model:

- Free accounts are monetized using online advertising
- Premium subscribers (those paying a fee) are typically 24 times more profitable than free users
- However, premium subscribers are rare

High Note is an anonymized real music streaming company --- similar to Last.fm, Spotify or Pandora --- that uses a freemium business model. You are provided with High Note customer data, including:

- Demographic characteristics such as age, gender and country
- Social network characteristics such as number of friends a user has on the network
- Engagement level data on activities performed when using the service, which include the number of songs the user has listened to, playlists created, “shouts” sent to friends, etc.

The High Note case description can be downloaded from the link below:

[High Note case for instructors 2-0.pdf](#) 

Your Objective

Given the higher profitability of premium subscribers, it is generally in the interest of company to motivate users to go from “free to fee”; that is, convert free accounts to premium subscribers. Your task in regards to this case is to analyze the data for potential insight to inform a “free-to-fee” strategy.

Literature shows that *peer influence* and *user engagement* can affect users’ decisions to pay for a premium subscription. Using the High Note data, you can predict the decision to buy using both types of variables. Your results will quantify the effect of social engagement on revenue, as well as how valuable a premium subscriber can be in a freemium social community.

The data set for this analytics exercise can be accessed at this link:

[HighNote Data Midterm.csv](#) 

The data dictionary is as follows:

Midterm Questions

1. Summary statistics: Generate descriptive statistics for the key variables in the data set, similar to the table on the last page of the case. (Note that your table will look different because the data set you are analyzing is different from the one used to generate the table in the case.) Analyze the differences in the mean values of the variables, comparing the adopter and non-adopter subsamples. What tentative conclusions can you draw from these comparisons?
2. Data Visualization: Generate a set of charts (e.g., scatter plots, box plots, etc) to help visualize how adopters and non-adopters (of the premium subscription service) differ from each other in terms of (i) demographics, (ii) peer influence, and (iii) user engagement. What can you conclude from your charts?
3. Propensity Score Matching (PSM): You will use PSM to test whether having subscriber friends affects the likelihood of becoming an adopter (i.e., fee customer). For this purpose, the "treatment" group will be users that have one or more subscriber friends (`subscriber_friend_cnt >= 1`), while the "control" group will include users with zero subscriber friends. Use PSM to first create matched treatment and control samples, then test whether there is a significant average treatment effect. Provide an interpretation of your results.
4. Regression Analyses: Now, we will use a logistic regression approach to test which variables (including subscriber friends) are significant for explaining the likelihood of becoming an adopter. Use your judgment and visualization results to decide which variables to include in the regression. Estimate the odds ratios for the key variables. What can you conclude from your results?
5. Takeaways: Discuss some key takeaways from your analysis. Specifically, how do your results inform a "free-to-fee" strategy for High Note?

When you are done, please upload your work (including any R or Python programs, charts, and description of results) using the Canvas dropbox.