January 12, 2017

CLASSPASS

MacKenzie's Project Design

Project Background

- What's the project about? What problem are you solving?
  - ClassPass is always trying to acquire new customers with various promotions across many channels. We currently base our subscriber projections on only two factors promotion and channel. I am solving the problem we have of not fully understanding conversion rates and what makes a promotion user convert to a full-fare subscriber (the conversion rate). I want to explore what other factors play an important role in this conversion.
- Where does this seem to reside as a machine learning problem? Are you predicting some continuous number, or predicting a binary value?
  - This project falls into a classification or clustering machine learning problem. I am predicting the continuous number of the conversion rate of a population by examining the binary value of convert/not convert for promotion users.
- What kind of impact do you think it could have?
  - The impact potential for this project is high which is why I chose it! The more we understand this population the more we can encourage certain behaviors during a promotion in the hopes that is will increase the likelihood to convert. More conversions will result in more subscriber growth which will be needed in order to reach our aggressive high growth plan for the year.
- What do you think will have the most impact in predicting the value you are interested in solving for?
  - I think that promo users taking highly rated classes will have the most impact on predicting conversion.

Problem Statement and Hypothesis

- Problem Statement
  - Determine the association between ClassPass user behavior during a promotion period with the likelihood to convert to full-fare subscribers.

- Hypothesis
  - Promo users that are more active on the ClassPass platform (using the app, taking classes, etc.) are more likely to convert than users that do not use or rarely use the platform.

Data Set Available

Data Set	Fields Available
Users Data	User ID, First Name, Last Name, Billing Zip, MSA/city
Lead Data	Lead Created Date, Lead Initial URL, Lead Marketing Channel, Lead UTM campaign, Lead UTM medium, Lead UTM content, Lead UTM source
Acquisition Data	User Acquisition Date, Acquisition Initial URL, Acquisition Marketing Channel, Acquisition UTM campaign, Acquisition UTM medium, Acquisition UTM content, Acquisition UTM source, Promo alias, Lead to Acquisition time
Reservation Data	Class Start Datetime, Class Name, Class Genre, Class Rating, Instructor, Instructor Rating, Reservation Created Datetime, Reservation Cost, Venue Tier, Venue Zipcode, Venue Amenities (Class Supplies, Food/Beverages, Lockers, Mats, Parking, Shop, Showers, Towels)
Plan & Pricing Data	Promotion Price, Promotion Plan Type, Payment Method
Engagement Data	Event Tracking for Mobile app, App Opens, Home Card Views, Impressions, Engagement with Recommendations (created from Machine Learning!), Completion of Onboarding Flow

Domain Knowledge

- What experience do you already have around this area?
  - I currently work on the Business Intelligence Team at ClassPass where I have supported ad-hoc Marketing requests over the past year. I am very familiar with our strategy for acquiring new subscribers and how to analyze the data.
- Does it relate or help inform the project in any way?
  - Yes this helps inform the project as I understand some of the high level trends and am already familiar with the strengths/weaknesses of the data.
- What other research efforts exist?
  - In August, I built a Growth Planning excel model that is large (almost 100 MB) and hard for people to understand. The file has frequently broken as multiple people are using it and the accuracy has ranged each month from 75% to 95%. Prior to this the model was in Google sheets and had much lower accuracy ~70%.
  - The head of Growth has looked at class usage before but did not notice large differences in conversion rates.
  - Several analysts have looked at pieces of the conversion process, but no holistic analysis has been done.

Project Concerns

- What are the assumptions and caveats to the problem?
  - Assumptions are that the historical data can be used predict future performance even though our business model has changed over the years from an Unlimited product at a lower price point to a tiered product at a higher price.
- What data do you not have access to but wish you had?
  - I wish I had access to other Gym/Boutique fitness studio data on conversion funnels. I also wish I could figure out if someone was taking classes with friends/ people that they know.
- What is already implied about the observations in your data set?
  - Implied is that all promo users would want to convert to full-fare if they properly experienced ClassPass
- What are the risks to the project? What's the cost of your model being wrong?
  - The risks to the project are that subscriber conversion is too complex of variable to predict
  - Cost of the model being wrong is lower the conversion rate of members by encouraging behavior that actually takes away from the user experience and likelihood to convert.
- Is any of the data incorrect? Could it be incorrect?
  - Yes always possible that data is incomplete or inaccurate, but there are no huge red flags I know about
  - We have a known problem of trial fraud (where someone signs up for multiple trials with multiple email addresses)

Outcomes

- What do you expect the output to look like?
  - I expect my model to predict the conversion rate of various groups of promo users with similar channel, promotion, engagement, and trial activity.
- What does your target audience expect the output to look like?
  - My target audience does not have expectations on the output and simply wants what is most valuable to the business.
- What gain do you expect from your most important feature on its own?
  - I want to gain deeper knowledge on the best way to split classes taken and ratings to create appropriate groupings.
- How complicated does your model have to be?
  - I think the simpler the model the better as many people on the Marketing do not have an analytical background.
- How successful does your project have to be in order to be considered a “success"?
  - For the project to be considered a “success” by me, I will have to apply the lessons and models I have learned in class to my company’s data. My dream would be to uncover an insight that can help improve our conversion rates!
- What will you do if the project is a bust (this happens! but it shouldn't here)?
  - I will learn from my mistakes and figure out what the next steps need to be in order to make

Outstanding Questions

- What questions do you have about your project? What are you not sure you quite yet understand? (The more honest you are about this, the easier your instructors can help)
  - I am still confused by clustering and how to test different features in the cluster
    - How could I test different class cut-offs for the clusters?
  - Is there a best order to test the various models I want to run?
  - Is it wrong to exclude trial fraud people (at least the ones I can identify) from my dataset?
  - What are the best ways to translate my model into predictions I can use for forecasting?

Thanks!