**Project Problem and Hypothesis**

* What’s the project about? What problem are you solving?
  + The project is about predicting whether or not a patient will show up for their dental appointment. The problem I am trying to solve is decrease the number of missed unconfirmed cancelled appointments by identifying the patients that are at risk of not showing up for their appointment. I would like to send an appointment reminder confirmation to these patients to decrease the number of no shows in order for practices to better manage their schedules and get the most productivity out of their business.
* Here does this seem to reside as a machine learning problem? Are you predicting some continuous number, or predicting a binary value?
  + Yes this does seem to be a machine learning problem as predicting whether or not a patient will show up for their appointment entails predicting patient behavior and as practices grow in their patient numbers and receive new patients, that’s new behavior that will need to be incorporated and learned over time. Although the problem is binary, whether or not a patient will show up for their appointment, yes or no, I’ll be using continuous values to determine the confidence in which a patient will show up.
* What kind of impact do you think it could have?
  + I think this will have an impact for the dental practice by generating more revenue as there will be less empty slots and increase productivity for the dentists in a given practice. However, I also think this will enhance the value of our Lighthouse Dental product especially if the feature that targeted a high risk patient is accurate. Although it will be more difficult for a practice to see the effects since they wouldn’t know if a patient was going to cancel to begin with, we can measure the high-level impact or success that the product is having by measuring various KPIs prior to the feature being turned on and after for a given practice. An example KPI could be the percent of patients with an appointment.
* What do you think will have the most impact in predicting the value you are interested in solving for?
  + I think what will have the most impact in predicting show rate will be whether or not an appointment has been cancelled or an appointment was not completed. A cancelled appointment however, is not a measure of failure. If we see an increase in cancelled appointments, it may show that we are targeting the right patients and they are cancelling so the practice can fill the slot with someone else. A failure could be indicated as a patient that has confirmed they will show up for an appointment but they do not actually complete the appointment. Or a patient that has neither cancelled/confirmed/completed an appointment but has scheduled.

**Datasets**

* Description of data set available, at the field level (see table)
* If from an API, include a sample return (this is usually included in API documentation!) (if doing this in markdown, use the javacription code tag)

**Domain knowledge**

* What experience do you already have around this area?
  + I haven’t worked extensively with this data but have done reporting for our Lighthouse product. However, the analysis that was done was more on a QA point of view. I have yet to do a deep analysis on patient and appointments. I’m also not a dentist that has had issues with patient scheduling.
* Does it relate or help inform the project in any way?
  + The analysis I’ve done for the LH product will help to an extent. I’m aware of some of the data issues that currently exist. It won’t help from a domain knowledge perspective, but from a data cleaning perspective, I can better determine whether or not the data is usable and get a better sampling of the patients/appointments that will be used to train/test the model.
* What other research efforts exist?
  + Use a quick Google search to see what approaches others have made, or talk with your colleagues if it is work related about previous attempts at similar problems.
    - The analysis I’ve done for the LH product will help to an extent. I’m aware of some of the data issues that currently exist; however, I will need to confer with the Product Managers that have the most domain knowledge at my company to get a better sense of what expectations they see as they’ve spoken with the practices and get feedback on how to improve our product to better assist them in their day-to-day operations.
  + This could even just be something like “the marketing team put together a forecast in excel that doesn’t’ do well.”
    - Currently, I do not believe there have been any product developments or plans that will do this. But I believe it would be a valuable feature if accurate to enhance the value of the product. Lighthouse Dental already has high customer retention and investment from the company so this will only benefit if implemented properly and predicts accurately.
  + Include a benchmark, how other models have performed, even if you are unsure what the metric means.
    - Show rate – measure of confidence in which a patient will show up to their appointment.

**Project Concerns**

* 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’m unsure about a true measure of success in regards to higher show rate or higher confirmation predictions. I think being able to target a high risk patient and have them confirm they are cancelling is still a success but may develop another layer of complication. Should I stick with one or the other?
* What are the assumptions and caveats to the problem?
  + 1. What data do you not have access to but wish you had?
       - Some data that would be helpful to have but don’t have is whether or not a practice has a cancellation policy. I think this would be a strong cofounding factor that would affect an appointment cancellation but it’s not data we track.
    2. What is already implied about the observations in your data set? For example if your primary data set is twitter data, it may not be representative of the whole sample (say, predicting who would win an election)
       - Many of our dental customers primarily use Dentrix as their PMS system. This could skew our data if applied to other practices that uses there PMS systems.
* What are the risks to the project?
  + What’s the cost of your model being wrong? (What’s the benefit of your model being right?)
    1. The cost of this model being wrong is very low. Say we identified a low-risk patient, meaning the patient would have shown up to the appointment anyways, receiving additional messaging to confirm their appointment would probably not have adverse effects. However, depending on how product wants to use this data, such as bad messaging could adversely affect if the patient feels they are contacted too often.
  + Is any of the data incorrect? Could it be incorrect?
    1. The main measures I’m using to determine the success of the model are completed, confirmed, and cancelled appointments. These metrics are manually inputted and therefore I have to assume that the inputs are accurate.

**Outcomes**

* What do you expect the output to look like?
* What does your target audience expect the output to look like?
* What gain do you expect from your most important feature on its own?
* How complicated does your model have to be?
* How successful does your project have to be in order to be considered a “success”?
* What will you do if the project is a bust (this happens! but it shouldn’t here)