

PatientPop Data Team Take Home Challenge

Background:

At PatientPop, we provide tools to help healthcare providers (e.g., doctors, dentists, chiropractors, therapists) manage their online reputations. One key data team initiative involves assessing the performance of these tools in order to help optimize our products and quantify the value they bring to providers.

For example, one of our tools is an automated feedback survey sent to patients after their appointments. Practices use a “short” version of the survey during their first 6 months with PatientPop and then must switch to one of two “long” versions after that. The goals of these surveys are to generate 1) more online reviews for a practice and 2) more positive reviews (higher ratings). The Data team was asked to analyze the effectiveness of each of the “long” versions and recommend which “long” version would be best suited to each practice.

In this challenge, we would like you to answer several questions related to this request using sample data similar to what actually would be used in this initiative. The challenge is intended to assess your ability to analyze data using analytics, statistics and/or machine learning methods and to tell informative data stories (including visualizations!) with the insights generated by your analysis.

Please answer as much as you can, spending no more than 4 hours on this project. It is fine if you do not have time to answer all of the questions. We will evaluate whatever you complete.

Data:

[see attached file]

- **review_id:** unique identifier for each patient-submitted review of a provider at a given practice
- **review_date:** date the review was submitted
- **rating:** number of stars given to the provider in the review; stars range from 1-5, with 5 being high, no stars given = 0
- **practice_id:** unique identifier for each healthcare practice
- **n_providers:** number of providers working at a practice
- **n_providers_surveys:** At some practices, not all providers choose to send surveys to their patients. Thus, this field is less than or equal to n_providers.
- **version:** which version of the “long” survey a practice chooses (v1 or v2)
- **short_start_date:** date a practice joins PatientPop and begins sending the “short” survey to its patients
- **long_start_date:** date on which a practice begins sending the “long” survey to its patients. This will be 6 mos after joining PatientPop.
- **tech_type:** Code for the electronic health records system used by each practice. The type of system used allows different levels of integration with PatientPop’s tools.
- **specialty:** primary specialty of a practice (e.g., internal medicine, dermatology, pediatrics, general dentist)
- **zip:** practice headquarters location
- **city:** practice headquarters location
- **state:** practice headquarters location

Questions:

1. Do the “long” surveys result in significantly more reviews than the “short” surveys? Which “long” survey has a greater impact (v1 or v2)?
2. What kinds of practices do best with “long” survey v1? Would you recommend “long” survey v2 to any specific kinds of practices?
3. Do ratings change significantly with either “long” survey? How would you predict which practices will have success, defined as at least a 0.3 star increase in their “long” survey average monthly rating vs their “short” survey average monthly rating?
4. Would you recommend replacing the “short” survey with either (or both) of the “long” surveys for any or all practices? Provide reasons to support your answer. If you do not have enough data to make a recommendation, what additional data would you need?