ANALYSIS OF NO-SHOW APPOINTMENTS IN BRAZIL

A. Project Overview

A1. Research Question

Is there a correlation between patients receiving a reminder text for their appointment and if they show up to their appointment?

A2. Context and Background

Patients not showing up for their appointments can be a wasted time for the practice and for the patient. Being able to decrease the amount of no shows would help medical practices run more efficiently and ensure that patients are able to arrive to the appointment they scheduled. This situation would help doctors and medical practices across Brazil to discover if text reminders are really worth sending. Sending text reminders can be an additional cost to the practice, and if there is no difference in the rate of no-show appointments, then text reminders can be removed. However, if there is a significant difference in text reminders decreasing the amount of no shows, more medical practices in Brazil can implement these types of reminders.

A3. Summary of Three Published Works

1. "Evaluating the Impact of Patient No-Shows on Service Quality"
   This article discusses the issues that patient no-shows have on the healthcare industry and utilizes various methods to address these issues. Marbough discusses multiple factors that contribute to no-shows, such as patient behavior, financial situations, and scheduling policies. The article goes on to describe many negative impacts that no-shows have on the healthcare industry, including inefficient use of resources, loss of revenue, delayed treatments, longer wait times, and a decrease in patient satisfaction. One study conducted at a vascular laboratory "found that a no-show rate of 12% can cost the laboratory a gross loss on $89,107 annually" (Marbough). Overall, Marbough describes several reasons for no-shows as well as the negative impact of no-shows and how to start addressing this issue. While there are many possible causes for patient no-shows, it is clear that more solutions need to be identified in order to help decrease the number of no-shows.
2. "The Effect of No-Show Appointments on Patients and Hospitals"
   In this article, Tsernov first explains what a no-show is and how prevalent they are, citing a survey that "concluded that as many as 42% of patients skip their appointments" (Tsernov). They go on to explain several reasons why people don't show up to their appointments, including financial costs, long wait times, transportation issues, and poor medical literacy. On the other hand, hospitals lose valuable time and resources that could have been focused elsewhere when waiting

for a patient to not show up, with surgeons losing up to several hundred dollars per missed visit. Tsernov argues that hospitals need to employ the use of analytics to help identify and lessen the number of no-shows, otherwise hospitals will continue to lose more money. This article shows that the number of no-shows is increasing, and therefore it is even more important to find and address causes of no-shows. Discovering if text reminders are helpful in combatting no-shows would be important information for health professionals to know.

3. "Missed Appointments, Missed Opportunities: Tackling the Patient No-Show Problem"
   This article highlights the issue of patient no-shows in the healthcare industry and discusses the negative impacts no-shows have on the industry and the individual. Jain explains that a patient who misses one appointment is "70% more likely not to return within 18 months" (Jain). The article discusses multiple methods to combat patient no-shows, such as improving transportation, implementing reminders, as well as using data driven solutions. Jain goes on to stress the importance of building relationships between healthcare providers and their patients to improve patient satisfaction and outcomes. By combining data driven solutions and personalized healthcare, providers can reduce the number of patient no-shows. While this article is focused on data driven solutions and personalized care, looking into the correlation and possible causation of text reminders can help to improve no-shows from one side of this two-pronged solution.

A4. Summary of Data Analytics Solution

The solution this project is presenting is to identify whether there is a correlation between receiving a text reminder for an appointment and patient no-shows. This will help healthcare providers to see if implementing a reminder system in their organizations is beneficial or simply a waste of time. The output of this solution will be two charts, one to show the correlation between the two factors and one bell curve that shows the difference between average no-show rates under the null hypothesis. Along with these two visual outputs, there will also be a report of the findings using logistic regression to model the situation. This project will use data from patient appointments in Brazil but could be further used to explore data from other countries as well to see if there is a benefit globally as well.

A5. Benefit to Organization and Decision-Making Process

The two primary benefits to healthcare organizations would be a decrease in lost time and money. By decreasing the number of no-shows with a text reminder system, providers will be more efficient and able to see more patients each day. Using a text reminder system can also be automated so employees don't have to spend time texting or calling each patient to remind them of their upcoming appointments.

B. Data Analytics Project Plan

B1. Goals, Objectives, and Deliverables for Project

The goal of the project is to see if there is a correlation and a possible causation between receiving a text reminder and patient no-shows. The objectives are as follows:

- Determine if there is a correlation between text reminders and no-shows.
    - The deliverable for this objective is the findings of what type of correlation, if any, exists between the two. This will consist of a bar chart that compares the average rate of no-shows of when a patient receives a text reminder versus receiving no text reminder.
- Determine if text reminders have a causal relationship with no-shows.
    - The deliverable for this objective is to find the p-values from several different methods to see if text reminders cause a decrease in patient no-shows. The several different methods of calculating the p-value are as follows. One, an A/B test that provides a sampling distribution of 10000 simulations under the null hypothesis (receiving or not receiving a text reminder has no effect on no-shows) of that will be visualized as a bell curve with a red line to represent the actual observed difference between the average no-show rates. Second, a Z-test will be used to calculate the z-score and p-value based on the data. Third, logistic regression will be used and a model will be fit using the Noshow data and a summary of the model will be output.

B2. Scope of Project

The scope of this project will include analysis and visualizations from a Python application. This project will use patient appointment data as input. The output of this application will be visualizations and numbers that explore the correlation and causation between text reminders and no-shows. The scope of this project will not include other factors that may have a correlation or causation with patient no-shows. The scope will also be limited to data provided from healthcare appointments in Brazil, not data from other countries or other types of organizations.

B3. Project Planning Methodology

This project will use the Agile project methodology. This methodology consists of four major principles with those being: adaptive, goal-oriented, integrated, and learnable. This project will be adaptive and change as needed throughout the analysis process. This can include many things that may need to be flexible throughout the project, such as what libraries and techniques used in analysis, or changing how the findings are visualized. This project is incredibly goal-oriented, with the two clear goals set of exploring the correlation and causation between text reminders and no-shows. These goals will be met within their specific timeframes as well. While this project does not apply as much to the integrated principle since there is only one person working on it, the results and applications from this project can be applied on different projects, such as healthcare appointment data from other countries, or data from non-healthcare related appointments. The final principle, learnable, is very important to this project. This project is focused on learning about the relationship between reminders and no-shows, and how to further learn and improve this analysis throughout. This project is the first step to many of diving in and exploring the different relationships between various factors and patient no-shows.

B4. Timeline for Milestones

Milestone	**Projected Start Date**	**Projected End Date**	**Duration**
Establish requirements for analytics process	6/22/2023	6/22/2023	1 day
Acquire and clean data	6/22/2023	6/22/2023	1 day
Code analysis and visualizations	6/23/2023	6/24/2023	2 days
Test application	6/24/2023	6/24/2023	1 day
Code any revisions/final touches	6/25/2023	6/25/2023	1 day

B5. Resources and Associated Costs

1. Appointment data: no cost
2. Python IDE (Jupyter Notebook): no cost
3. Work hours: no cost

Since I am using public data, a free IDE, and using my own labor, there are no associated costs with the resources used for this project.

B6. Measurable Criteria for Success

The measurable criteria for success I will use are the associated visualizations produced to show any correlation, as well as the respective p-values produced from my multiple analyses of any causation. The project will need to determine if there is any correlation, and what type, between the two factors. If it can not produce an answer to that, then this is not a success. The project will also need to produce p-values to show the statistical significance of text reminders to no-shows through various techniques. If p-values cannot be calculated, this is not a success.

C. Design of Data Analytics Solution

C1. Hypothesis

Sending a reminder text about the appointment will have a negative correlation with no-shows. Patients receiving a text reminder will also cause a decrease in no-shows.

C2. Analytical Method

I will use descriptive data analysis techniques to find if there is a correlation between text reminders and no-show appointments. For example, I will employ bootstrapping and logistic regression to see if there is any statistical significance between receiving a text reminder or not receiving one. I will use Python and custom code to help in creating, cleaning, and analyzing the dataframe. Descriptive analytics is used to help determine trends in data, which is exactly what this project aims to do with determining the correlation between text reminders and patient no-shows.

C3. Tools and Environments

This project will use Jupyter Notebook, a Python integrated development environment. Python is very commonly used to gather, clean, and analyze data since it has a variety of libraries perfectly suited for data analysis. Jupyter Notebook will be used specifically since the project can use well defined chunks of code with output directly underneath it, which helps to keep all outputs and visualizations organized.

C4. Methods and Metrics

The metric that will be used to determine statistical significance is the p-values of several analysis techniques. The p-value will be used to see if there is any significant difference between the project's hypothesis and the null hypothesis, which is that sending a text reminder will have no effect on patient no-shows. If the p-value is less than 0.05, that means that there is a significant effect between text reminders and no-shows. This metric is most appropriate to determine if there is a causal relationship between receiving a reminder and patient no-shows.

C5. Practical Significance

The practical significance of this solution is that text reminders have a direct cause on a patient showing up to their appointment. If sending a reminder to every patient before their appointment will decrease the number of no-shows, this can help healthcare organizations to combat the increase in patient no-shows. This will help providers know whether implementing a text reminder system can result in saved time and money for their organizations.

C6. Tools and Graphical Representations

The graphical representations this project will use are a bar chart to compare the average number of no-shows when a patient receives a text reminder and when they do not receive one. A bar chart is used since there are the two options: receive a text or don't, and the average number between these two options will help to see if there is any significant difference between averages. A bell curve will also be used to help visualize the possible causation between text reminders and no-shows. The bell curve will help to show the distribution of no-shows under the null hypothesis, which we can then compare to the observed difference from the data. Another visualization will be a table that is produced from the logistic regression, in which we can see the p-value from the model as well as other values. All three visualizations will be used to clearly demonstrate the correlation and causation of text reminders and patient-shows.

D. Description of Dataset

D1. Sources of Data

The source of this data is Kaggle.com. This dataset is owned by Joni Hoppen and Aquarela Analytics, but is publicly available on Kaggle.com. Under the CC BY-NC-SA 4.0 license, I am allowed to use this data as long as it is not for commercial use and give appropriate credit to the owners. I am not using any other datasets.

D2. Appropriateness of Data

This dataset is appropriate because it provides data on appointment details from healthcare organizations in Brazil. The two columns that are imperative to this project are already there, being whether a patient received a text reminder and if the patient did not show up to their appointment. While there are other columns, most of these are irrelevant to the project and are dropped from the analysis.

D3. Data Collection Methods

This data was collected by downloading the csv file of the dataset off of Kaggle.com and then reading it into a dataframe in Python.

D4. Observations on Quality and Completeness of Data

While I did not create this dataset, it was fairly clean. There were a few rows that I removed due to outliers in the age column, but the data was complete in the two columns that were imperative to the project. I did change the data from 'yes' and 'no' in the Noshow column to 1 and 0 to help with the analysis.

D5. Data Governance, Privacy, Security, Ethical, Legal, and Regulatory Compliances

Since the data was public on Kaggle.com, this project is freely able to use it under the license listed in D1 as long as it is not for commercial use. This data does have the name and age of each patient, as well as some chronic health issues, but these columns are not used or included in this project. There are no major privacy, security, ethical, or legal compliances that this project needs to follow since the data is publicly available and is using columns that do not include any personal information.

E. Sources

Project Data: https://www.kaggle.com/datasets/joniarroba/noshowappointments

Marbouh, Dounia. "Evaluating the Impact of Patient No-Shows on Service Quality - PMC." *PubMed Central (PMC)*, 4 June 2020, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7280239/.

Tšernov, Kirill. "The Effect of No-Show Appointments on Patients and Hospitals | Qminder." *Qminder*, Qminder, 6 June 2017, https://www.qminder.com/blog/queue-management/no-shows-affect-hospitals/.

Jain, Sachin H. "Missed Appointments, Missed Opportunities: Tackling The Patient No-Show Problem." *Forbes*, Forbes, 6 Oct. 2019, https://www.forbes.com/sites/sachinjain/2019/10/06/missed-appointments-missed-opportunities-tackling-the-patient-no-show-problem/?sh=163e0afd573b.