

# Data Management and Data Analytics Capstone Topic Approval Form Capstone Topic Approval Form

The purpose of this document is to help you clearly explain your capstone topic, project scope, and timeline. Identify each of the following areas so you will have a complete and realistic overview of your project. Your course instructor cannot approve your project topic without this information.

**Student Name:** [REDACTED]

**Student ID:** [REDACTED]

**Capstone Project Name:** Analysis of No Show Appointments in Brazil

**Project Topic:** This project will examine data on medical appointments made in Brazil of 2016 in relation to the whether a patient showed up or not.

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

**Hypothesis:** Sending a reminder text about the appointment will have a negative correlation with no shows.

**Context:** 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.

**Data:** I will use data collected on scheduled appointments in Brazil of May 2016. This dataset already exists and can be downloaded off of Kaggle.

The dataset I will be using has information collected on scheduled medical appointments and the personal details of the patient. Some information included is the patient's age, gender, the location of the appointment, and of course if the patient showed up to the appointment.

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 restricted, private, or proprietary data.



**Data Gathering:** Describe the data-gathering methodology you will use to collect data. I will download the dataset csv file from Kaggle and then clean it as needed.

**Data Analytics Tools and Techniques:** 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.

**Justification of Tools/Techniques:** Using bootstrapping is an appropriate way to first create random samples using the data and then analyzing to see any differences in the rate of no shows when receiving a text reminder or not receiving one. Logistic regression is a great way to help predict and classify between two categorical outcomes, which in this case is showing up or not showing up to an appointment. Both of these techniques will help to answer [REDACTED] question in different ways.

**Application Type, if applicable (select one):**

- ☐ Mobile
- ☐ Web
- ☒ Stand-alone

**Programming/Development Language(s), if applicable:** Python

**Operating System(s)/Platform(s), if applicable:** N/A

**Database Management System, if applicable:** N/A

**Project Outcomes:** The project will determine if there is a correlation between the text reminders and a patient showing up to their appointment. It will include the code used in the data analysis, as well as a report to show the findings.

**Projected Project End Date:** 6/30/2023

**Sources:** No Show Appointment Data:  
<https://www.kaggle.com/datasets/joniarroba/noshowappointments>

---

#### Human Subjects or Proprietary Information

Does your project involve the potential use of human subjects? (Y/N): N

Does your project involve the potential use of proprietary company information? (Y/N): N

---

#### STUDENT SIGNATURE

[REDACTED]

**By signing and submitting this form, you acknowledge** that any cost associated with the development and execution of your data analytics solution will be your (the student) responsibility.



---

**TO BE FILLED BY A COURSE INSTRUCTOR**

**The capstone topic is approved by a course instructor.**

**COURSE INSTRUCTOR SIGNATURE:**

A handwritten signature in blue ink, appearing to read "Jim Ashe".

Jim Ashe, Ph.D. Mathematics

**COURSE INSTRUCTOR APPROVAL DATE:**

Friday, June 16, 2023

**Project Compliance with IRB (Y/N): Y**

