DS 5100 Project Proposal: Analysis on COVID-19 Patients with Pre-existing Conditions

Group 5: Alex Bass, Connie Cui, Peumali Surani Withanage, Seth Galluzzi

Computing ID: ujb3bu, qqv3uu, upp2dh, vzw6yk

Introduction

Our project focuses on exploring how pre-existing medical conditions may impact COVID patients in hospitals, specifically on death rates.

The Data

Our dataset was obtained from an online dataset released by the Mexican government found at https://www.gob.mx/salud/documentos/datos-abiertos-152127. This dataset was then translated and posted on Kaggle, https://www.kaggle.com/madan44/covid19-patient-precondition-dataset-cleaned, and we will be working with this dataset specifically for the project. The dataset primarily consists of information about the hospitalized patients in Mexico such as their age and sex, as well as whether or not they have specific underlying conditions (i.e. diabete, asthma, hypertension, etc.), and whether or not they died during their hospitalization. We chose this dataset specifically due to its abundance in information regarding patients and their pre-existing medical conditions before they were hospitalized from COVID-19. Our preprocessing procedure will include checking the dataset for any null values and deciding whether or not to remove them for future analysis.

Experimental Design

Generally, we plan to explore how preexisting conditions impact COVID patients in hospitals. The data set we found is already relatively cleaned, but will need more wrangling and subsetting to produce analyses desired. As the main portion of our analysis, we plan to show a total of 3 visualizations of COVID-19 death rates, first, showing overall death rates given by various pre-existing conditions such as diabetes, obesity, and pneumonia in bar charts and two similar charts where the user can specify seeing death rates of preexisting conditions broken out by age or gender. Visualizing the overall death rates allows for important analytical comparison of which conditions are more deadly. This analysis will also suggest how these preexisting

conditions affect an individual given gender or age. To avoid overwhelming the user with data points and charts, the program will prompt the user after the first chart (the one showing death rates overall) and ask if the user would like to see the death rates broken out by gender. If the user responds "Yes," the next chart will be generated. The same process will occur for the gender chart. The user can respond "No" to stop seeing charts. We hope this format can be useful and informative to the user in understanding the effect of COVID-19 on preexisting conditions overall and across various ages and gender.

Project Management

        While our team members all intend to be involved with all processes and steps of our project in one way or another, we have set specific focuses and aspects of the project for each of our team members to work on. Alex will focus on the data visualizations of the project, Peumali will focus on data cleaning and preparation, Seth will focus on testing and debugging, and Connie will focus on the user input aspect of our project. As a group, we plan to meet weekly to discuss our progress, observations, and understanding of the project data. Finally, after completing our individual tasks, we will work together to prepare a presentation of our findings.

Results

        Results will be discussed mainly in three ways: a visual representation of the main results, statistical summaries, and written explanations. These three methods have standalone advantages and complement each other to explain the results comprehensively. Our study's main conjecture is that pre-existing conditions make the hospitalized COVID patient outcome worse. The adverse outcomes among the COVID patients with pre-existing conditions have been well documented throughout the pandemic, and we expect to test this hypothesis using a comprehensive dataset with various characteristics of hospitalized COVID patients. Moreover, not every pre-existing condition impacts the outcomes of the COVID patients equally. Some pre-existing conditions have a more severe impact than others. Therefore, we expect to rank and order the pre-existing conditions in terms of severity.

Testing

One key aspect of our project is testing and debugging our code. While multiple eyes will preview each line of code produced, we will also implement unit testing methods to check for possible errors with the user input portion of our project. This is especially important in cases such as when the user does not necessarily input an option that we have offered them or does something that may lead to some other type of exception or error. For example, if we ask the user a yes or no question and they input some other response, we need to have code that will prepare for these sorts of situations and unit tests to ensure that both our code and potential errors are all handled properly. We will also include multiple sanitization procedures of user input to reduce errors.

Outcome

During this project, we hope to gain a better understanding of how preexisting conditions impact COVID patients in hospitals. We also hope to create clear visualizations that inform others how preexisting conditions can impact COVID patients in hospitals. Finally, we hope to strengthen our understanding of python as a tool that helps us gather and present data efficiently and effectively. To accomplish this we will compare the death rates of COVID patients with preexisting conditions, to the death rates of COVID patients without preexisting conditions. We will also compare the death rates of COVID patients with different preexisting conditions (i.e. diabetes, asthma) to each other. To make this data even more useful for others, we will create opportunities for users to interact with visualizations and explore the impact different conditions have on COVID patients in hospitals. Furthermore, users will be able to explore how variables such as age, race, and gender impact the death rate of patients with preexisting conditions. By completing this project, we will learn more about how preexisting conditions impact COVID patients, we will inform others of the risks patients with preexisting conditions experience, and we will learn more about Python as a tool to help us present our new found knowledge.