

Project Proposal

Aiden Manuel

November 2023

1 Introduction

This is my project proposal for the final project in CS2704 Data Analytics in Python. In this document, I will briefly outline my starting hypothesis, how I plan to test that hypothesis, the databases I will be accessing, and the results I will be expecting. This serves as the rough outline for my project, and is being done before I have put anything to code, to give myself an idea of what exactly I need to accomplish. Let us now examine what exactly I plan to do, and how I plan to do it.

2 Topic

The topic of my data analysis will be the people's reactions and beliefs regarding the COVID-19 vaccinations. It is well-known that during the pandemic there was a large anti-vaccine movement across the world, but predominantly I heard about it in the United States, thanks to their widespread media coverage. People were rallying behind the idea that with the COVID-19 vaccinations came a myriad of different side effects.

From my own experience, I heard people claiming everything from getting a fever and headache after their vaccine, to actually getting more sick than COVID-19 would have made them after receiving the vaccine. My understanding is that because the vaccines were produced by multiple companies with a very short trial window, in an attempt to push them out as fast as possible, people became very suspicious about the quality and nature of these vaccinations. Even doctors and healthcare professionals seemed to be wary to some extent, from what I recall.

3 Hypothesis

3.1 What is my hypothesis?

Anecdotal evidence is fine and all, but I want to put some concrete statistics down, and find out more about the public consensus on vaccines, which I can do fairly simply through skimming publicly available databases. However, on

top of that, I want to see what sort of factors influence someone's likelihood to be wary of vaccinations. The specific hypothesis which I would like to test is:

The level and quality of education someone receives is related to their beliefs around the COVID-19 vaccinations in the United States.

3.2 How do I aim to prove that?

To prove this, I am going to compare the beliefs regarding vaccines of specific populations throughout the United States to the data on their public schooling. Metrics I will be interested in are high school graduation rates, average high school GPA, student engagement, and other metrics which speak to the level and quality of the education a student receives.

I will then calculate a "quality of education" based on some weighted factors as an aggregated statistic. After I have determined what my metric for quality is, I will look at regions with varying levels of vaccine suspicion, and compare that to their educational quality metric. Ideally, I will be able to calculate some linear regression for the relationship, or perhaps some more complex relationship will present itself which I can build a model from.

4 Databases

The databases I will be using were obtained from the U.S. Government's Open Databases. I found one database which comes from the Monthly Outcome Survey, specifically from January 2021 to April 2023. It specifically assessed the "beliefs, intentions, and behaviors relevant to COVID-19 vaccination at a point in time", that quote being taken directly from the metadata.

The other database I found is a measure of Public School Characteristics in the NCES Common Core of Data for elementary and secondary schools in the United States. This includes data about the student demographics, grade spans, and other administrative information. This will hopefully fulfill exactly what I need for my analysis.

5 Expected Results

I expect that in populations with a lower quality of education, the suspicions and wariness surrounding the COVID-19 vaccine will be heightened. That is to say, there will be an positive correlation between quality of education and trust of the COVID-19 vaccine.

My reasoning for this is that I believe people who are provided with a poor education often find ways to educate themselves, and nowadays that occurs via the internet. In such a scenario, it is very easy for someone to stumble across a forum or group of people with a specific belief which isn't based in science, and then get trapped in the echo chamber of poorly informed opinions.

6 Concluding Statements

I do hope to have some interesting results come out of this, and if I have time, I may try to measure the COVID-19 vaccine trust factor against some other sort of statistic, say minimum wage, or percent of blue collar workers. I think there are a lot of interesting avenues to pursue here, but I want to start (hopefully) simple with the education perspective.

My github repository can be found here. Thank you for considering my proposal.