**Jude Nwosu**

**Final Project v2**

# Project Design Write-up and Approval Template

Follow this as a guide to completing the project design write-up. The questions for each section are merely there to suggest what the baseline should cover; be sure to use detail as it will make the project much easier to approach as the class moves on.

**### Project Problem and Hypothesis**

**\* What's the project about? What problem are you solving?**

The project I have chosen to undertake is on Healthcare Expenditure in the United States of America. The project involves predicting/forecasting healthcare cost increases/decreases in the US health systems in the coming year based on available/historical data.

The purpose of my project is to try and provide as much clarity to the average citizen on the cost of Healthcare Expenditure relative to national debt, and what the financial outlook may be in the healthcare industry. It is assumed that such information will also assist individuals plan ahead.

**\* Where does this seem to reside as a machine learning problem? Are you predicting some continuous number, or predicting a binary value?**

This project resides in the Time Series Analysis section of Machine Learning. By its nature, the project predictions shall focus on some continuous number.

**\* What kind of impact do you think it could have?**

As alluded to in the “What’s the project about” question above, this project aims to provide clarity to the average citizen on the cost of Healthcare Expenditure, and what the financial outlook may be in the healthcare industry.

**\* What do you think will have the most impact in predicting the value you are interested in solving for?**

The level of healthcare utilization by the citizens.

**### Datasets**

\* Description of data set available, at the field level (see table)

**### Domain knowledge**

*\* What experience do you already have around this area?*

I have approximately 5 years of experience in the healthcare industry which I strongly believe is an invaluable asset for the nature of this project.

*\* Does it relate or help inform the project in any way?*

It does help inform the project in a solid way because my knowledge in the healthcare industry due to the years spent in it provides me with the wherewithal to understand the nature of the problem/project undertaken.

*\* What other research efforts exist?*

I did a quick google search on this project topic and came across several sites related to same research topic. One particular page from the CDC seemed quite interesting (https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/proj2016.pdf ) as it approached the topic from a wholly economic standpoint; measuring it against the nation’s GDP. The report provides summary that stretches over a ten (10) year period (2016 to 2025).

*\* Include a benchmark, how other models have performed, even if you are unsure what the metric means.*

The key projections from the site indicates that the US Healthcare Expenditure will grow at a rate of 1.2 percentage faster than Gross Domestic Product (GDP) annually from 2016 to 2025.

**### Project Concerns**

*\* What questions do you have about your project? What are you not sure you quite yet understand? (The more honest you are about this, the easier your instructors can help).*

So far, I feel confident about this successfully completing this project, although I suspect I might need additional table(s), which I have to join with the Healthcare Expenditure data table to be able to harness the full potential of this project.

*\* What are the assumptions and caveats to the problem?*

*\* What data do you not have access to but wish you had?*

For some reason, I’m thinking I will need some economy-related data that may assist in comparing my results with how the nation’s economy is doing. This may not be the case, but I want to ensure that I am prepared for any eventualities.

*\* What is already implied about the observations in your data set? For example, if your primary data set is twitter data, it may not be representative of the whole sample (say, predicting who would win an election)*

*\* What are the risks to the project?*

*\* What's the cost of your model being wrong? (What's the benefit of your model being right?)*

As a start, this is a project that I plan to use in showcasing my newfound skill/expertise in the field of Data Science; hence my decision to pick a project that cuts across two (2) of the most vital sectors of the economy, Finance and Healthcare. This will inevitably assist in my either breaking into the Healthcare or Finance job market, or to search for consulting projects in the machine-learning field.

*\* Is any of the data incorrect? Could it be incorrect?*

Since the data is coming from the United States government ([www.data.gov)](http://www.data.gov)), I would trust that the information is as accurate as possible.

**### Outcomes**

*\* What do you expect the output to look like?*

The fact that I intend on using Time Series Analysis for this project, I expect the output to mirror results, or trend closely to what other reputable organizations have already done.

*\* What does your target audience expect the output to look like?*

I would suspect that the target audience mostly want to see how the impending changes to the Affordable Care Act would impact my findings/forecast. They will obviously want to see a clear summary of what the future offers regarding Healthcare Expenditure and how it might impact them, and also to include visualizations that would assist them put the results/report in perspective.

*\* What gain do you expect from your most important feature on its own?*

Don’t understand this question. Clarify with Mr. Naumann

*\* How complicated does your model have to be?*

At this stage, I cannot really give a precise answer to this question. It will all depend on how/what the initial results would look like. At that stage, I will decide on how to escalate the model to be used in the project.

*\* How successful does your project have to be in order to be considered a "success"?*

Success is obviously subjective but I would want my project to mirror what other reputable organizations, particularly the CDC have in done already. This will involve incorporating visualizations where possible, forecasting Healthcare Expenditure for the next five (5) years, especially with the impending changes to the Affordable Care Act (ACA).

*\* What will you do if the project is a bust (this happens! but it shouldn't here)?*

Eat a burrito, sleep, and then try working on another project. The key issue is acquiring the skillset necessary to build a career in Data Science.