Capstone 2 Project Proposal:
“Mental Health and Social Media: Suicide rates in correlation with Social Media”
By: Vincent Tolentino

The world today is filled with technological innovations that help society connect, exchange, and interact with each other on a level like nothing before. But how does this affect the mental states of the population? Does increased screen time affect the mental health of society? Does a deeper level of connection through technology have a negative or positive impact on the people that are connected? I plan to dive deep into these questions and more through this Capstone.

The specific problem I want to explore is the rise in social media and its effect on on suicide rates in the population. I want to explore the correlations between certain social media sites and applications and the prevalence of suicide rates in a given population. There are numerous accounts of the use of social media affecting a population's mental health. I want to identify several social media companies and their user base and find any correlations between mental health, specifically suicide rates. Specifically, I want to explore *TikTok*, *Facebook*, *Instagram*, *Twitter*, and *LinkedIn* to find any links between the two and the mental health status of a given population, mainly the United States, which can be seen as a First World country whose population utilizes all forms of social media. I want to be able to identify if any of those companies have a positive or negative effect on the US population’s mental health through suicide, and with those results, construct a model to predict future trends that may occur. Scraping data from the internet regarding social media user bases, screentime, and other potential factors, in addition to governmental suicide and mental health statistics may aid in finding these results and predicting future trends.

The use of social media has been on the rise for decades, as more and more people have access to the technology to utilize them. One can speculate that its use is necessary in everyday life to stay connected with family and work, and is a vital part of a person’s leisure time. Starting in the early 2000s, social media has been able to connect people from around the world with a click of a button, revolutionizing the way we communicate and interact. This has not only had an impact on social relations, but the world economy, and has brought the entire world closer together. However, the prevalent use of social media in our society may, or has already started to become, a double edged sword. With data only spanning two decades, the long term effects of social media are now being brought to light, and many studies conducted so far have

alluded to a negative overall sentiment with regards to the continued and increasing use of social media, especially when directed towards children and younger adults.

In order for a project like this to succeed, we must first take a deep dive into the analytics of *TikTok, Facebook, Instagram, LinkedIn,* and *Twitter* to pull as much pertinent information we can. Once that information is set, we must also obtain population information of the US and its suicide rates. It may also be necessary to obtain additional mental health statistics as they come by, and in order to gauge a baseline rate prior to the rise of social media, we would also collect data going back several decades. Once this data is compiled, analysis and predictive models will be utilized in order to pinpoint any significant ticks in suicide rate compared to the social media data. To succeed, we must be able to correctly identify the negative impact social media has, and predict future trends based on this data.

The scope of our solution space will be constricted between the social media sites above, as well as the population of the United States. Utilizing worldwide data may prove to be too difficult and may fog the landscape of what we are trying to accomplish. Other countries, for example, may not have the prevalence of social media due to technological or cultural restraints. However, we may still hit some roadblocks in our analysis, as information may be restricted to students like myself. The data is also only restricted to the inception of the company, which only goes back a few years, which may hinder our predictive models, since the length of time is limited. These constraints may cause some unpredictability within our model to some degree. Having not dived too deep into data acquisition for the social media companies I intend to use, it is difficult to say there will be the necessary data available to me, so these social media companies may change over the course of the project.

The stakeholders of this would be governmental institutions and private institutions that deal with the mental health of the population. The goal of institutions like this is to analyze the mental health of the nation and make decisions based on that analysis. Based on our models and projections, we may be able to incite policies and mental health recommendations when it comes to social media usage, depending on the outcomes of our models. Projects like this can influence the decision making of these institutions and, hopefully, implement the necessary changes to create a more positive outlook on mental health in the United States. Additional stakeholders within this project will likely be the Springboard faculty as well, including my mentor, Ricardo Alanis-Tamez and my advisor Cindy Martinez.

The data that will be used for this project is a long list. *Twitter* user bases, as well as its financial data may need to be scraped from various internet sources. I plan to utilize multiple dataset search engines including Google Data Search, Quandl, the US Government websites, Kaggle, Wikipedia, and others. This also goes for the user bases of *TikTok, Facebook*, *Instagram* and *LinkedIn*. Data about the population may be easier to come by, as the US Government supplies this data over the web, however, specific mental health data and suicide rates may need to be obtained from multiple sources. The data will be contained within a repository that I have set up in GitHub. In addition to this proposal, it will also contain the code used to pull, clean, and analyze the data as well as the predictive models I will utilize for my results. A slide deck and project report will also be included in the repository once the data is compiled and analyzed.