The Relationship between Global Crises and the Suicide Rate

University of Pittsburgh School of Computing and Information, 135 N Bellefield Ave, Pittsburgh, PA 15213

Tom Greene, Shailey Gulrajani, Adnaan Hasan

INFSCI 0510 Data Analysis,
Dmitriy Babichenko, University of Pittsburgh Fall 2021

thg25@pitt.edu
spg35@pitt.edu
ash124@pitt.edu

Abstract:

This paper describes a research project comparing socio-economic information with suicide rates by year and country. Our research question is twofold. We ask how the suicide rate has changed throughout the years and if we can predict who and where the most at-risk individuals are at. We also ask how global financial and banking crises have affected the suicide rate. The first part of our research question is interesting to us because we want to see how the past suicide statistics have impacted the present. The second part of the research question is interesting to us because we want to see if financial and banking crises have greatly increased the suicide rate throughout the world. We consider our project useful because it can potentially help us figure out where this issue is the most frequent and therefore where the most suicide prevention needs to occur. The people who would benefit the most from our project are those who work in suicide prevention and other mental health professionals.

For this project, we use Python's Pandas to manipulate, understand, and represent the data since it is a powerful tool that performs data analysis for the user. One feature we use to visualize and represent our data is a line graph. Other features that we use include "iloc"/"loc", "summary", and missing data functions. "iloc" and "loc" are used to isolate certain columns, "summary" is used to get certain values from data (such as the mean), and the missing data functions are used so that we can clean up messy data (such as if an outlier should not be there).

Keywords – Suicide, Fincancal, Crisis

I. Introduction

Mental health is an extremely important topic in this day and age. In recent memory, it has become a very popular topic of discussion. As a whole, awareness for topics like anxiety, OCD, depression, and the point of this paper, suicide, has increased exponentially. The CDC estemates that 1.3 million adults attempt suicide every year [10].

Despite the increase in awareness, there is still little that is understood about suicide and suicidal tendencies. Because it is such an abstract and grim concept, most people do not want to thoroughly educate themselves on a topic like this. This is one of the reasons why topics, like causation and prevention, get lost in translation about mental health awareness.

Causation is a big topic of confusion when it comes to discussing suicide. Being a mental illness, it is not necessarily triggered by a single, constant variable. This is where it is important to note that correlation is not equal to causation. Despite not being able to predict the individual, there are ways to see causation with groups. When it comes to global crises, it is somewhat a given that national levels of happiness are low. For example, the US financial crisis of the 1920s was called the "Great Depression". Therefore, there is a way to see how much more the rate of suicide increases based on what kind of global crisis is happen.

To figure out whether there is a correlation between global crisis and the rate of suicide, we decided to look at different types of crises and international suicide rates in the past 50 years. Both of the datasets we used, called "Suicide Rates Overview 1985 to 2016"[1] and "Global Crises Data by Country"[2], had data from numerous different countries around the world. Since, as previously mentioned, we aimed to find out where in the world individuals are mostly likely to commit suicide, finding datasets that had countries as a variable was extremely crucial.

In addition, we were glad to find a dataset that had a long time span, since a myriad of crises and other events have occurred over that time. We were also glad to find a separate dataset that showed the different types of crises.

We hope that the readers of this paper can better understand how different crises have impacted the suicide rate across the world through our methodology, results, and discussion sections.

II. Methodology

A. Preparations

We analyzed this data using the “Pandas” software library written for the programming language *Python*. We started with two different datasets; the “Global Crisis” dataset [2] and the “Global Suicide Rate” [1] dataset. We spent a lot of time cleaning and transforming these datasets to coincide with one another.

The key to get these datasets to work with each other were two columns that they both contain; “Year” and “Country”. These would be used as the foundation to link our sets together. However, nothing ever comes easy with code because we had to help more with compatibility than simply using the “merge” function of Pandas. The dataset containing global suicide had many rows with the same “Year” value because it listed different groups data on suicide such as age group and gender, which we did not want. To combat this we combined everything so that every row would be a different year. This reduced the dataset by a multiple of 12 (we divided appropriate columns by this number to find average). For the “Global Crisis” we had to drop every row with a year that contained a value before 1985. This was because the “Suicide Rate” dataset does not contain years before 1985. The final major thing done for preparing the merge was dropping a lot of unneeded columns from the dataset (for ease later). After this we finally merged the data.

B. Classification

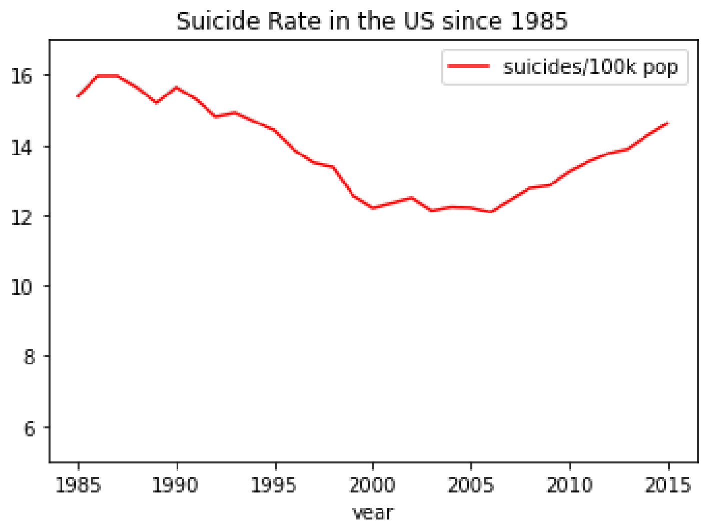
To help with the prediction models we made a third dataset based on our merged set from earlier. This one is grouped/ ordered by year instead of by country. We also made a dummy variable column to classify if a year had an unusual amount of average suicide or not based on the average amount per year. This allowed us to create a classification model to predict an answer to this question.

C. Missing Data/ Outliers

We ran into a good amount of missing variables throughout our data, especially with the important features pertaining to the different types of crises. Luckily, the only parts where there were missing values, happened to be dummy variables which are fairly easy to fill up. We assumed that every row that didn't have a value recorded meant that it was a zero. The reasoning behind this is because it is safer to assume that a crisis did not happen rather than assume one did in fact happen. Also, a person is more likely to record something that is uncommon (a crisis) rather than something that is normal (stability) which is why missing values would probably mean nothing happened that year. Finally, there were two missing values in the “Annual percentages of average consumer prices” column, in Argentina 2014 and Argentina 2015. Filling these would not be as easy as dummy variables so we decided to do some research on what happened these years and found that Argentina went through a dramatic crisis during these years. This caused the average consumer price to be impossible to record due to its flexible nature. This caused us to drop Argentina from the dataset as a whole.

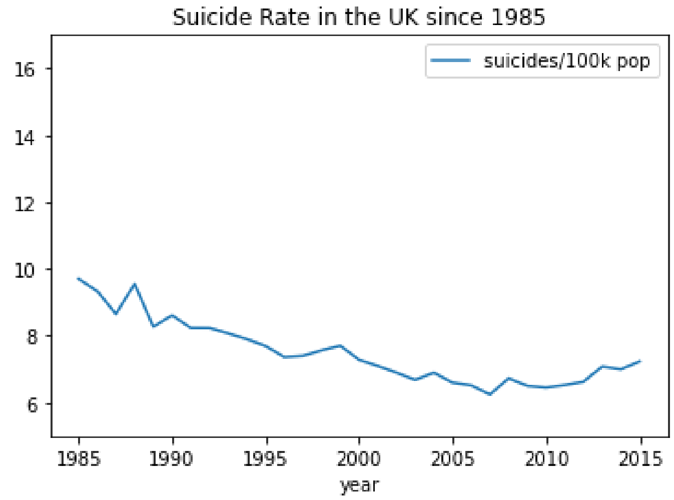
D. Exploring the data

Now that our data is fully prepared, cleaned, and polished it is time to explore the set a little bit. We were interested to see different countries rate of suicide and how they compared to others to give context to the predictive models.

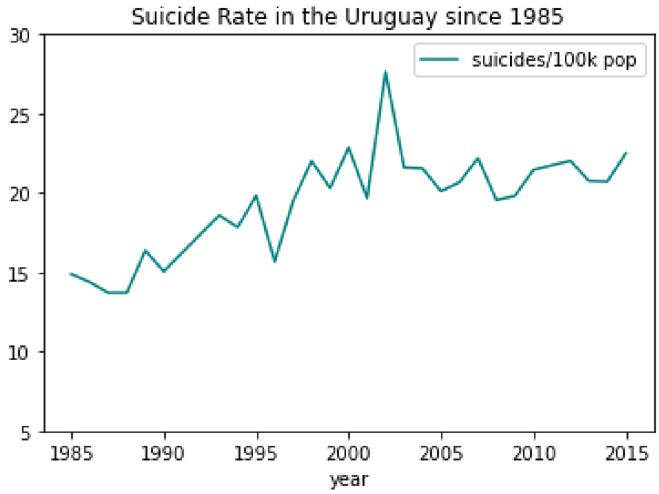


Here is the rate of suicide in the United States since 1985. Interestingly, it had been at a constant decrease since the start of our dataset but after approximately 2008, it is going up. An inference we could make about this is a result of the Financial Crash the US (and many other countries) experienced in 2007 - 2008. As a side note, this inference is completely by eye and we have not researched far into this.

Here are a few more:

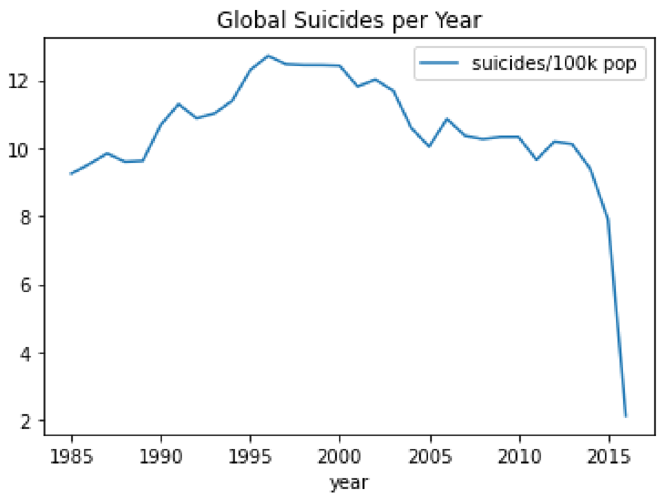


As you can see there has been less dramatic changes in the UK than the US.

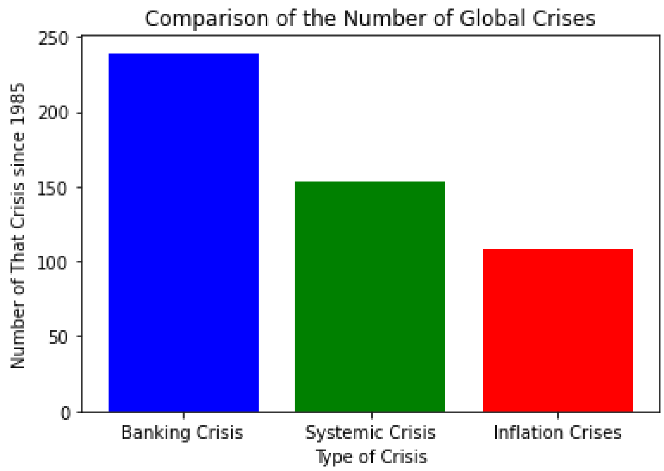


Sadly, it is evident from this graph that the suicide rate in Uruguay is significantly higher than other countries. This may be found to be a result of their financial situation

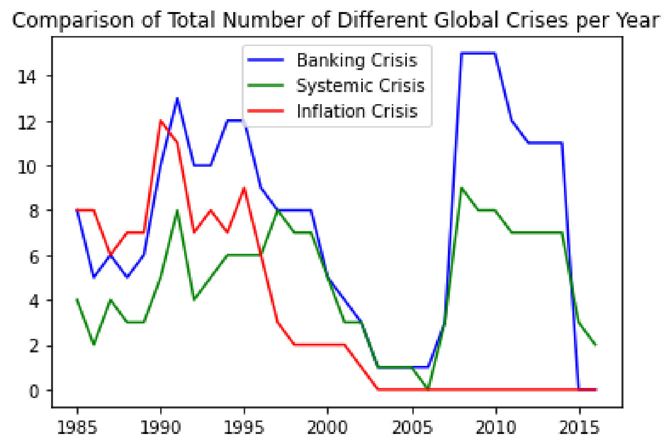
There was a large peak in the middle of the 1990s which could possibly be a result of the “Troubles”. This was a conflict between Northern Ireland and Great Britain. Many acts of terrorism were committed in this time.



Here we have a graph over the average global suicide rate per year. You can see it peaked in the 1990s, which is not what we personally expected. There is a dip in 2015 because the data is unfinished.



Here is a graph of the total number of crises that have happened per country globally. Banking Crises are the most common while Inflation Crises are least common.



These are the different types of crises that happen per year. One interesting thing is that there was a huge spike in 2007-2011, which is most likely because of the US Financial Crash.

III. Results

A. SVM

*accuracy score - **55%***

The ideal machine learning algorithm that we chose to use was SVM. The reason for using the SVM classifier is that it is the best suited for extreme case binary classification[3]. Also the classes in our data set were separable so this helped in our process. We completed an SVM model, chose the X variables and single Y variable and trained it. Our goal in this process was to see with what accuracy we would be able to predict the likelihood of suicide rate increase based on our financial crisis. Our accuracy score for SVM was 55%. We recognize that this is not the best and there are many reasons why this is the case that will be discussed in our discussion portion.

B. Naive Bayes

*accuracy score - **59%***

The second machine learning algorithm that we chose was Naive Bayes. The reason for choosing this classification method is Naive Bayes is fast and easy and also works well with multiclass predictions [4]. It was understood by our group that SVM was the best way to go about the process, but even though Naive Bayes wasn't best, we decided that the commonality of its use would be an interesting aspect to observe in our second classification. The accuracy score of our Naive Bayes classifier was 59%. This also was not a very desirable feat, but as we understand we may have bit more than we can chew in this project, an in-depth discussion of our faults will take place in the discussion section.

C. Variables of Focus with Abbreviations

Statistics:

- **country** – column that specifies location
- **year** – Year statistic is from (1985 - 2016)
- **Banking Crisis** – Whether or not banking crisis occurred that year
- **Systematic Crisis** – Whether or not a systematic crisis occurred that year
- **Currency Crises** – Measures Currency Crisis
- **Inflation Crisis -** Whether or not an inflation crisis occured
- **suicides/100k pop -** Amount of suicides per 100k population
- **HDI for year [5] -** Human Development Index. This is a measure of three dimensions of human development. These three dimensions are
    1. a long and healthy life
    2. access to education
    3. a decent standard of living
- **GDP per Capita [6]** - Gross Domestic Product. This is a monetary value of all goods and services produced within a country over a specific timeframe. "per capita" means it is being compared to a metric of population.
- **Exch_USD -** This is the exchange rate per US dollar ($)
- **Domestic_Debt_In_Default -** a measure of whether the country was in debt domestically

IV. Discussion

A. *Limitations of knowledge*

The results of this project are not necessarily what we predicted, the reason for this is that we did not have a complete grasp of the regression and classification system. When we went into this project we had a very useful idea that we wanted to bring to light. The fact that there may be a way to predict suicide rate per year when a financial crisis was in play. The reality of this situation is that we as a group were in over our heads and we realized it when it was too late. We understood there were potential possible ways to make this happen, but that would bring us to our next issue.

B. *Limitation of dataset*

For this project we would need datasets that were nonbinary to make a prediction of what the suicide rate would be. We were intending to solve a regression problem. We were setting up for this problem without fully understanding the dataset at hand. Our dataset, especially the financial crisis dataset, was set up to answer a question of classification. A yes or no question that we were not prepared to answer. We wanted to answer a question that we did not have the means to answer. We understood what needed to be done and put the dataset to the job. At the end of the process we had to change to a yes or no question. That question was one that changed the entirety of our project and the scope of this goal. The reason we could not fix this in a timely manner brings us to our next issue.

C. *Team communication*

When working on a project that involves code, you must have a process of seamlessly communicating and collaborating. The work that went into this project lacked this seamless communication and collaboration in every sense. With almost all members of this group opting for an online option rather than returning to class due to COVID-19, there was no seamless way to collaborate. We had many zoom meetings piecing together what this project would look like in theory. This however was one of the issues that we hadn't foreseen. Theory is one thing, but code is bound to have problems. None of us are familiar with any process where we could work on the code together, so there was a halt on this process. When we eventually decided we had to meet in person, we ran into many issues. One of the issues was the data didn't line up with our intended goal. However the group had to continue in this process, as the data had already been cleaned and meeting as a group and getting this project completed was very difficult. Also meeting in person was difficult to organize. Our schedules didn't line up, and we elected to meet much less than we should have to recognize these issues as soon as they appeared. This was our fatal flaw which led to all the others.

V. Conclusion

The thought process in which we went into this project was a powerful one. According to statistics, there is "roughly one death by suicide every 40 seconds. [9]. Through this project we attempted to help predict an issue that clearly is in dire need of resolution. The fact of the matter is unfortunately we were unable to fully attain this goal, for various reasons. Although our prediction model is successful to a degree, there is clearly room for improvement. This insight is one that should be carried onto future projects as this is a subject that needs to be researched more thoroughly. This problem is not going away. The overall suicide rate in the U.S. has increased by 35% since 1999 [11]. One valuable aspect of this project is that it raised awareness for this very important subject.

Acknowledgment

We would like to thank Dmitriy Babichenko, the professor of this section of INFSCI 0510. We would also like to thank Harvard School of Business [2], The United Nations Development Program [7], and the World Health Organization (WHO) [8] for gathering the data that was implemented in this study.

References

[1] Rusty. "Suicide Rates Overview 1985 to 2016." *Kaggle*, 1 Dec.2018,https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016.

[2] "Data." *Global Crises Data by Country - Behavioral Finance & Financial Stability - Harvard Business School*, https://www.hbs.edu/behavioral-finance-and-financial-stability/data/Pages/global.aspx.

[3] Yadav, Ajay. "Support Vector Machines(SVM)." *Medium*, Towards Data Science, 22 Oct. 2018, https://towardsdatascience.com/support-vector-machines-svm-c9ef22815589.

[4] Chauhan, Gaurav. "All about Naive Bayes." *Medium*, Towards Data Science, 8 Oct. 2018, https://towardsdatascience.com/all-about-naive-bayes-8e13cef044cf.

[5] Roser, Max. "Human Development Index (HDI)." *Our World in Data*, 25 July 2014, https://ourworldindata.org/human-development-index.

[6] Fernando, Jason. "Gross Domestic Product (GDP)." *Investopedia*, Investopedia, 7 Dec. 2021, https://www.investopedia.com/terms/g/gdp.asp.

[7] "Human Development Reports." | *Human Development Reports*, http://hdr.undp.org/en/indicators/137506.

[8] "Suicide Prevention." *World Health Organization*, World Health Organization, https://www.who.int/health-topics/suicide#tab=tab_1.

[9] "Suicide Statistics and Facts." *SAVE*, https://save.org/about-suicide/suicide-facts/.

[10] Fontana, Pat. "How Raising Awareness Helps Suicide Prevention Efforts." *South Miami Recovery*, 24 Feb. 2021, https://www.southmiamirecovery.com/awareness-suicide-prevention/.

[11] "Suicide Prevention Awareness Month." *NAMI - National Alliance on Mental Illness*, https://www.nami.org/get-involved/awareness-events/suicide-prevention-awareness-month.

[12]