Predicting Suicide Rates based on Availability of Mental Health Facilities

# **Jennifer Ruiz Winter 2020 GitHub Portfolio:** <https://github.com/jfruiz15/DataSciencePortfolio>

# Which Domain?

This project comes from the healthcare domain. Information in this domain is used to directly impact healthcare research and treatment for patients. Please see reference list below for annotated information.

# Which Data?

I will be analyzing the World Mental Health and Suicide data set. The data is available to the public through Kaggle through the following link: <https://www.kaggle.com/twinkle0705/mental-health-and-suicide-rates?select=Age-standardized+suicide+rates.csv>.

# Research Questions? Benefits? Why analyze these data?

For the purpose of this project, I will be exploring factors that predict higher suicide rates within a country. I will use prediction and classification analyses to better understand the data. My research questions are:

1. What impact do widespread access to mental health facilities have on suicide rates?
2. What countries have the highest rates of suicide?
3. Do lower rates of day treatment mental health facilities predict higher suicide rates?
4. What factors are the greatest indication of suicide rates within a country? What factors have the least impact in predicting suicide rates?

I chose to approach the data with these research questions as they have not been explored in this manner previously. Understanding how the availability and access to mental health facilities impacts suicide rates can lead to initiatives to promote treatment and lower suicide rates. I believe analyzing these factors could provide new insight which would be beneficial to the overall understanding and prevention of these events.

# What Method?

For this project my plan is to use a multivariate regression models to better understand the factors that most influence suicide rates. The dataset is coded numerically allowing for easy implementation of regression models that will address the research questions above. This method should provide valuable insight and areas of focus in mental health services availability as it is related to suicide rates.

# Potential Issues?

This dataset is complex with a large amount of data spread across multiple csv files. Merging the data into a single dataframe will take some time. Also, with the size of the dataset, training the regression model may take considerable time. I will need to prepare for this to ensure delivery on time.

# Concluding Remarks

Suicide rates continue to climb around the world. Access to mental health treatment facilities and staff can have an impact on suicide rates within a country. State or country focused research on factors influencing mental health have been conducted but very few worldwide analyses related to facility availability have been conducted. In this project, the author seeks to examine factors such as availability of in-patient and outpatient services, number of mental health staff providers, and the overall availability of mental health services to better understand their relationship with suicide rates. The author will use multivariate regression analysis to achieve this goal.

References:

Amr, T. (2016). Predict the Future with Regression Analysis. Retrieved on January 16, 2021, from https://gr33ndata.medium.com/learn-regressions-analysis-23b789bf2c36

This article discusses both linear and multivariate regression analyses using sklearn.

Obi Tayo, B. (2019). Bad and Good Regression Analysis. Retrieved on December 16, 2020 from https://medium.com/towards-artificial-intelligence/bad-and-good-regression-analysis-700ca9b506ff

This article discusses the nuances of regression analysis. It also provides a more in-depth explanation of various methods for regression. A sample Python implementation is also provided.

Kansal, H. (2020). Multivariate Linear Regression. Retrieved on December 5, 2020 from <https://medium.com/dev-genius/multivariate-linear-regression-dfd18a26431d>

This article discusses the benefits of using a multivariate linear regression model vs. a multiple regression model with health-related data. The author highlights the complexity of this data domain and the unique challenges to predicting in this category.

Menon, P. (2017). Data Science Simplified Part 5: Multivariate Regression Models. Retrieved on December 5, 2020 from <https://towardsdatascience.com/data-science-simplified-part-5-multivariate-regression-models-7684b0489015>

This article outlines how to interpret multivariate regression models as well as the challenges to working with such models. The author highlights common interpretation mistakes and methods to check model accuracy.

Penumudy, T. (2021). A Comprehensive Guide to Logistic Regression. Retrieved on January 16, 2021 from https://medium.com/analytics-vidhya/a-comprehensive-guide-to-logistic-regression-e0cf04fe738c

This article discusses the benefits of using a logistic regression vs. a linear regression for large datasets. The nuances of each regression model are outlines as are use cases.

Grover, P. (2017). Intuitive Interpretation of Random orest. Retrieved on December 5, 2020 from <https://medium.com/usf-msds/intuitive-interpretation-of-random-forest-2238687cae45>

This article outlines 4 major methods of interpretation for random forest models. The author explains each method as well as its benefit to creating an overall interpretation for models.

Sung-Schenck, J. (2020). Suicide Prediction-Part 1. Retrieved on January 16, 2021 from https://medium.com/@juhee.sungschenck/suicide-prediction-part-1-29798c643b29

This article outlines prediction methods for working with suicide rates data. It discusses potential areas of weakness in analysis and provides guidance about data transformation.

Sung-Schenck, J. (2020). Suicide Prediction-Part 2. Retrieved on January 16, 2021 from <https://medium.com/@juhee.sungschenck/suicide-prediction-part-2-c23050b761a4>

This article outlines features selection methods and target variable creation. Also discusses other potential supervised and unsupervised learning methods that work with the data.

Umair, A. (2020). Data Science: Global Suicide Rates Analysis & Prediction Model. Retrieved on January 16, 2021 from <https://medium.com/@ali.umair/data-science-global-suicide-rates-analysis-prediction-model-ali-umair-65c202347932>

This article discusses possible research questions and provides a sample analysis for working with suicide data.

Asif, C. (2020). Mental Health & the Covid-19 Pandemic. Retrieved on January 16, 2021 from <https://gasif8622.medium.com/mental-health-the-covid-19-pandemic-79e4e46dcd7b>

This article discusses the impact pandemic events on mental health. It also explores how natural disasters, war, and catastrophic stress impact individual mental health.