

**Table of Contents`**

[**Abstract** 1](#_Toc90847930)

[**Research Question** 2](#_Toc90847931)

[**Literature Review** 2](#_Toc90847932)

[**DATA PROCESSING** 4](#_Toc90847933)

[**Data Prep & Cleaning** 4](#_Toc90847934)

[**Data Description** 4](#_Toc90847935)

[**DATA EXPLORATION** 5](#_Toc90847936)

[**Data Clustering** 5](#_Toc90847937)

[**Clusters on World Map** 7](#_Toc90847938)

[**Visualizations to understand variables** 8](#_Toc90847939)

[**MULTIPLE LINEAR REGRESSION** 10](#_Toc90847940)

[**Model 1** 10](#_Toc90847941)

[**Model 2** 11](#_Toc90847942)

[**PREDICTIVE MODELLING** 13](#_Toc90847943)

[**Logistic Regression** 13](#_Toc90847944)

[**Decision Tree:** 13](#_Toc90847945)

[**K-Nearest Neighbor:** 15](#_Toc90847946)

[**RECOMMENDATIONS** 15](#_Toc90847947)

[**LIMITATIONS** 16](#_Toc90847948)

[**APPENDIX** 17](#_Toc90847949)

[**REFERENCES** 22](#_Toc90847950)

# **Abstract**

*This research is aimed at understanding how the suicide mortality rate differs across under-developed, developing and developed countries. Moreover, the analysis also highlights the impacts of culture, religion, individualism, and collectivism on the suicide rates. With the aid of data visualization, unsupervised and supervised learning mechanisms, the analysis of the data delves into the multiple external and internal factors affecting the suicide mortality rates across the globe. The countries are clustered using K-means clustering. Additionally, it uses techniques like multiple linear regression to assess significant variables and predictive modelling like KNN to predict suicide mortality rate during COVID. Thus, giving us important insights into policy around mental health and suicide prevention.*

# **Research Question**

With increased research on mental health and fundamental breakthroughs in the field of psychology, the issue of suicide prevention has become an important and pressing policy issue as it is **the second most common cause of death.** We aim to understand how the suicide mortality rate differs in **developing, developed and underdeveloped** countries and test whether **internal** factors affect it more or the **external** factors. There is a significant gap in research in addressing the issue of rising suicides in **individualistic and collectivistic societies**; especially research that addresses this issue across multiple development level nations rather than just a single group within the same economic growth level with similar development and technology. It is high time we begin to understand why suicide happens across the world in order to prevent it, because it is futile to address a problem if it’s causes aren’t understood first multidimensionally. Therefore, this paper attempts to answer this research question:

*Which variables (external and internal) significantly affect the suicide mortality rate across nations in differing social and economic growth groups (pre and during the COVID-19 pandemic) and how can this knowledge help us in determining better suicide prevention policy for a mentally healthier and prosperous world?*

# **Literature Review**

The suicide mortality rate is higher in the developed world with late-stage capitalist secular systems. Two leading examples of this are Japan and USA. In a study, it was observed that from 1985-2006, there has been a consistent increase in suicide rates in, but it was also observed that there is a distinct significant increase in suicides of men between the ages of 50 to 64 years (Odagiri, Uchida, and Nakano, 2011). It has also been observed that the age trajectories of mortality rate are heavily weighted by death due to degeneration of physiology and development, the age trajectory in suicide mortality rates stems from social conditions (including group integration, interpersonal conflict, life fulfillment levels, and career progression) rather than biological resistance. This implies that suicide mortality rates will differ largely with temporal, cultural, and national differences in social conditions (Pamper and Williamson, 2001) and that research done on economically advanced nations cannot be entirely applicable in other regions of the world. The patterns of gender and unemployment leading to suicide in Japan, for example may or may not occur in countries like Pakistan. Cultural differences - **like religions, and collectivistic vs individualistic sociability** - thus become defining characters in suicide risk.

There is also the additional fact that countries with greater **income inequality and exclusion of youth from productive roles**, will have higher suicide rates in the youth. This derives from a sense of disadvantage caused by **unemployment** and lack of opportunities (pamper and Williamson, 2001). More equal societies with greater opportunity equalities are likely to fare better in terms of suicide mortality rates than the ones which are not. **Culture** also plays a vital role apart from economic growth, with collectivist cultures having higher suicide rates (Oskin, M., 2018). Then there is the obvious impact of differing economic growth and GDP in determining suicide mortality rates. **Higher GDP** countries generally tend to also be more politically stable, which is an important aspect in youth suicides which often gets sidelined. More financially well-off countries promise stability and a minimum quality of life that less financially well-off countries cannot. GDP is the direct determinant of net cumulative output (and thus wealth) of a nation, i.e., the higher the GDP the higher a country’s financial/economic standing (Johnson, A., 2018). This is why GDP distinction – among other reasons – becomes central to the discussion of suicide mortality rates. Moreover, social well-being of the citizenry is highly affected by the economic growth rate or GDP growth of a country.

Other factors to consider are the **happiness levels and HDI** (Human Development Index) countries. The happiness index is an ordinal measure of a populous’ self-assessment of its satisfaction. It is allotted to a country by surveying its citizens and asking them to rank their self-satisfaction on a scale of 1-10; the values are then aggregated to come up with an overall average level of happiness for a country. The subjectivity of this measure is what sets it apart from other standards of well-being. The happiness index and HDI are particularly relevant to this study for two reasons:

*1) GDP isn’t fully representative of self-satisfaction because it only assumes utility is derived from monetary expenditure.*

*2) They have the potential to explain differences among countries within the same GDP range and yet largely different suicide mortality rates.*

Scandinavian countries like Finland tend to do well on the happiness index despite having lower **GDP growth rates** than countries like USA or Canada (owing to their social welfare economic structures). It is also interesting to note that these countries have largely public provision of both healthcare and education; what is more noteworthy in our research is that mental healthcare is largely government territory rather than private therapy which is less sought after (Halliwell et al., 2018). And suicide literacy helps improve suicide rates (Cruwys et al., 2018). Literacy on the other hand, increases the risk of suicide across 33 European countries (Marušic, Khan and Farmer, 2000). While alcohol has been reported to increase the risk of male suicide through firearms in USA, another trend observed in past two decades or so is prescription drug abuse. It has become endemic driven primarily by an increased prescription and consequent abuse of opioid medications. This has led to a reversal of the previously high but rapidly declining relationship between drug and alcohol abuse and suicide (Borgschulte, Corredor-Waldron and Marshall, 2018). Addiction of drugs is more likely to induce suicidal thoughts.

Lastly, **sociability** is a major impact factor in the recovery of schizophrenic patients (who are generally high risk for suicide). Their disease not only affects them but their relatives and caregivers largely. How a society diagnoses and treats schizophrenics is somehow more important than the actual symptoms of the disease itself and the outcomes of the disorder. The label itself comes with so much stigma, which differs in different nations based on their mental health literacy (Marušic, Khan and Farmer, 2000). This is one factor that shows that in countries which are culturally and religiously rich there are suicides based on societal standards (Eskin, M., 2018).

# **DATA PROCESSING**

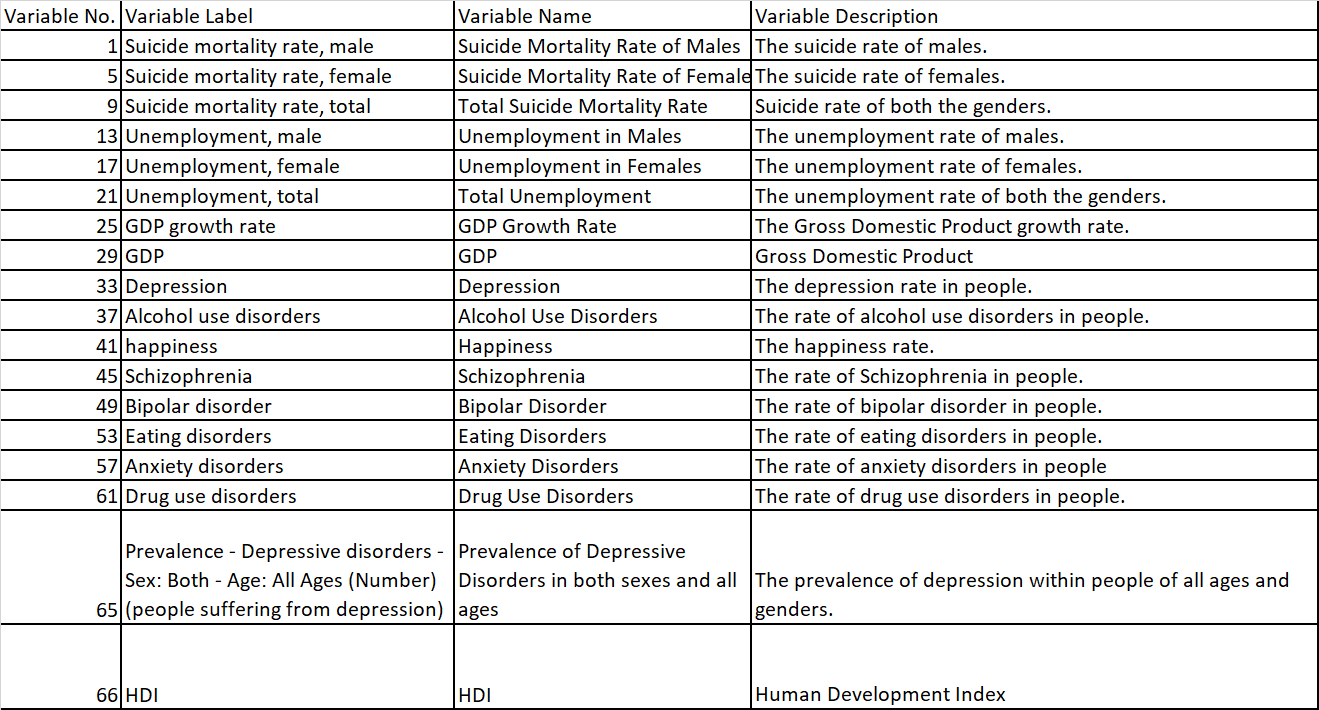
## **Data Prep & Cleaning**

After doing secondary research, we collected raw data from multiple sites like WHO, Nationsonline, World Bank, UNICEF etc for variables affecting suicide mortality rate. As the data was collected from different sites, there was mismatch among countries, so we merged the data using loops to make it coherent. The NA values were removed to clean the data set and in the end we were left with complete data of 129 countries. Lastly, we scaled our data because the values had varying ranges. Data post COVID-19  was gathered for prediction analysis only. Data prep pre-exploratory & predictive analyses can be seen in R Script files.

## **Data Description**

The data for each variable was found between the period of 2016 to 2019 as it was the most recent data available and to better understand the impact of time on the sucide rate we decided to present our dataset as **panel data**. Some of these variables we identified were external (like GDP growth rate, HDI etc.) while others are internal (Depression, Anxiety, Alcohol use etc.).

The variables are mentioned below:



# **DATA EXPLORATION**

Since a part of our research question was mainly based on finding out how the suicide mortality rate differs across developing, developed and underdeveloped countries, we used the unsupervised method of data clustering and visualizations to understand the impact of suicide rates on different countries based on their degree of development. The degree of development was measured using the HDI rank and the data for it was collected from the World Bank.

## **Data Clustering**

*We applied both k-means and hierarchical clustering techniques to our data set and the analysis for both are given below*

*Text

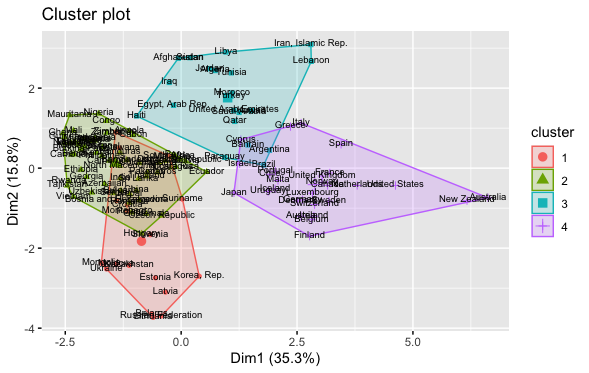
Description automatically generated with medium confidence****K means clustering:*** We wanted to group countries in a way that the countries falling in one cluster are somewhat similar in their HDI ranking. This helped us understand their development ratio as well which was the primary basis of our analysis. After examining the clusters for k=2 to k=10, the value of k=4 was the best to fit our analysis. This is because k=4 gave our 4 heterogeneous clusters and the clusters themselves were homogeneous in terms of their HDI rank. Moreover, on examining further we found out that the clusters were made in a way that countries falling in cluster 2 promote individualism, countries falling in cluster 1 are culturally rich whereas those falling in cluster 3 are seen to be collectivist in nature. Thus, we chose k=4 on a subjective basis because it will give us extremely valuable insights when we move further in our analysis.

A picture containing diagram

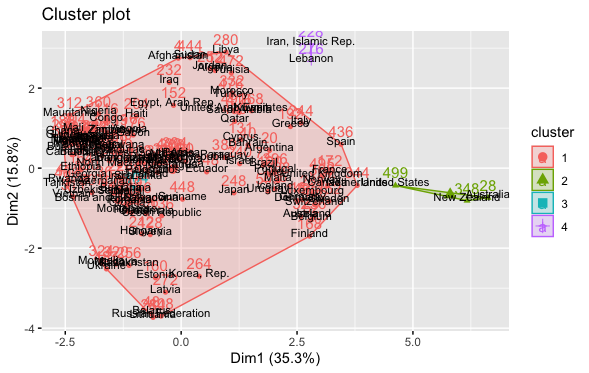
Description automatically generated***Hierarchical Clustering:*** We did not go with hierarchical clustering as a technique because firstly it did not produce distinct clusters according to our criteria which was the development ratio of countries. On observing the cluster formed with K ranging from 2-10, we found out that there was no homogeneity among clusters which could fit our analysis as we proceed. For instance, the clusters, when k=4, showed that through k means clustering there was less overlap between countries in terms of HDI ranking and through hierarchical clustering, there was a lot of overlap between countries.

The heterogeneity of both clustering techniques can be seen below (**Figure 1 - K Means & Figure 2 - Hierarchical**):

***Figure 1 (K-Means Clustering)***



***Figure 2 (Hierarchal Clustering)***



### **Clusters on World Map**

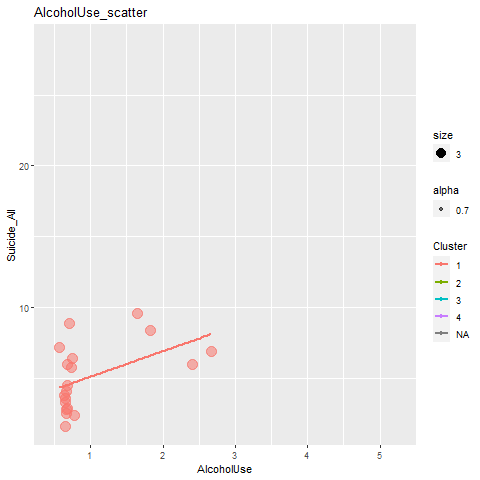
If we take the k=4 clustering technique, the diagram below shows the overall cluster distribution. **Cluster 1** includes countries such as Saudi Arabia, Northern African countries, some Latin American, and other middle eastern countries. These are what we have described as ***culturally rich*** and ***religiously conservative*** countries so there is homogony within the cluster. **Cluster 2** includes developed countries such as Japan, Australia, New Zealand, Canada, Ireland, Norway, countries in the United Kingdom and the USA. These are what we have described as economically developed. Furthermore, these countries promote ***individualism***, so the cluster is relatively homogeneous. **Cluster 3** includes Asian countries such as Pakistan, India, Sri Lanka, China, some American countries like Colombia and some African countries such as Nigerar and Kenya. Most of the countries in this cluster are developing countries and some are developed countries. These are what we have described as family-centric and conservative countries so there is homogony within the cluster. **Cluster 4** includes countries such as Zimbabwe, Mongolia and Kazakhstan and some European countries such as Ukraine. These are what we have described as economically developing countries. Furthermore, these countries either have a history of civil war or political rivalry with other countries. Many countries included in this cluster were either colonized by Russia or Britain so there is homogeneity within the cluster.

*Figure 3 – cluster distribution*

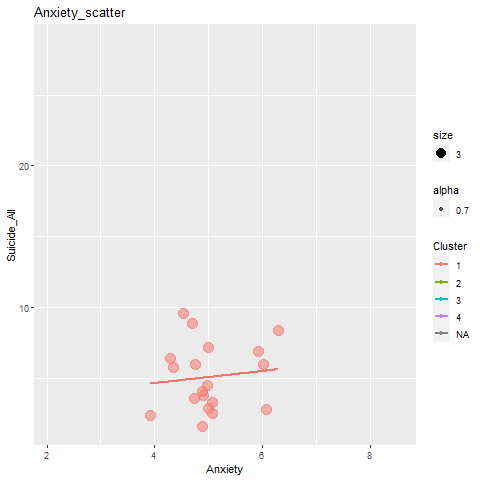
Map

Description automatically generated

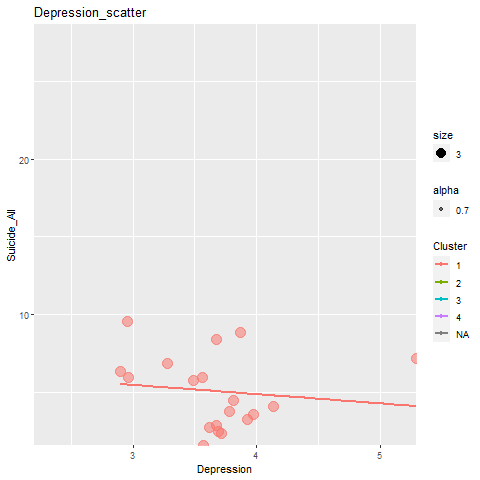
## **Visualizations to understand variables**

***Visualization # 01:***In cluster 1 and 2, alcohol use has a direct positive relationship with suicide rates (gradient almost identical). This relationship is less apparent in cluster 3 and not apparent in cluster 4 at all.

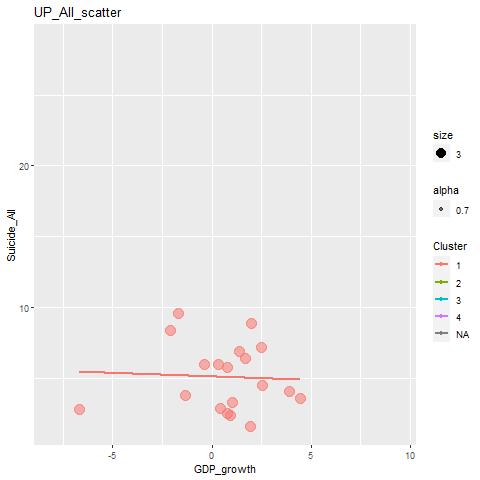
**Interpretation & Significance:** This is interesting because it re-enforces the literature review’s claim that alcohol use has a direct relationship with suicide rates, especially noticed in USA which lies in cluster 2. Countries in Cluster 4 have more regularized legal use of alcohol. Another interesting thing about the alcohol consumption in cluster 4 countries is that it is rather a social consumption of alcohol rather than an individual addiction. Countries in cluster 2 are more individualistic and conspicuous alcohol consumption in these countries is indicative of addictions which and as our literature review suggests are directly related to the suicide rate.

***Visualization # 02:*** Anxiety is highly impactful on suicide mortality rate in cluster 4. The impact is lesser in cluster 1 and even lesser in 3. The range of anxiety is most varied in cluster 2 and it doesn’t impact he suicide mortality rate a lot in this cluster.

**Interpretation & Significance:** Countries in cluster 4 are generally collectivistic or communal countries with a lot of sociopolitical upheaval and distance or conflict in terms of the religious landscape. This leads to a stronger relationship between anxiety and suicide rates. Cluster 2 countries are puritan countries whether they be pre-dominantly Christian or Muslim, and anxiety or mental distress thus does not translate to suicide because of the notion of suicide as a sin and surrounded by extreme taboo. Thus, we can see that our results are endorsing the ideas of theory that religion and sociability largely impact the suicide rate.

***Visualization # 03***: On the left, you can see the impact of depression on suicide rates in the four clusters. Unsurprisingly, the strongest relationship between the determinant and the suicide mortality rate is in cluster 4. The cluster 1 shows a negative relationship between depression and suicide rate.

**Interpretation & Significance:** Depression has the highest impact on the suicide mortality rate in cluster 4; countries that have been colonized and suffered from civil war violence, but economically well off (with a few outliers). On the other hand, depression has an inverse impact on suicide rates in cluster 1 countries which are predominantly Muslim majority countries. Here we see the effect of strong religious taboos when it comes to suicide, meaning that depressed people are more likely to live on given the stakes for committing suicide are so high in their faith. Both these findings are in line with our literature review.

***Visualization # 04***: In the animation on the right, you can observe the impact of GDP growth rate on suicide mortality rate across our four clusters. It is evident that GDP growth rate is most strongly inversely related in cluster 2. in cluster 1 and 3, the decrease in suicide rate with increasing GDP rate is not that obvious. But the shocking discovery comes from cluster 4 in which increasing GDP growth leads to increasing suicide rates.

**Interpretation & Significance:** With increasing GDP growth rate, there is a sharp decline in suicide rates in developed countries like Japan, USA, and the UK. Thus, in individualistic economies people are more likely to do well mentally if they are employed and can value themselves highly in terms of money and success. Thus, in periods of economic recession, there will be more suicides in these countries. Cluster 4 has Scandinavian countries, central Asian states, and some south African nations; the value systems in these countries are more reliant on external metrics like happiness index or self-satisfaction so it makes sense that higher GDP growth doesn’t prevent suicides. In fact, according to our literature, higher GDP growth in these countries means higher literacy and that also leads to higher suicide morality rate.

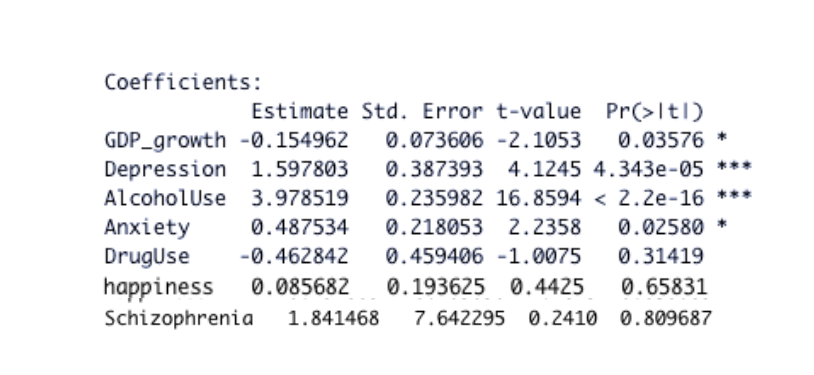
# **MULTIPLE LINEAR REGRESSION**

The use of Multiple Linear Regression was carried out to evaluate the impact on suicide mortality rate by using multiple **internal and external exploratory variables**. MLR was employed to find out the significant variables which impacted the suicide mortality rate the most. We **used fixed effect technique** in MLR to analyze the impact of varying variables over different time intervals. This helped us answer our research question “*Which external and internal variables impact the suicide mortality rate the most?*”. We carried out MLR twice to better understand the data.

## **Model 1**

Chart

Description automatically generatedFor the first model, we inserted the data for all countries and all variables to see which ones are more significant and contribute towards the variation of suicide rate, but the results were a little out of place. This is because we did not cater to **multicollinearity**, and it can lead towards broad confidence intervals which ultimately gives less reliable probabilities in terms of independent variables’ effect. Therefore, to cater to this we created a correlation plot and removed those independent variables which had **high correlations among them**.

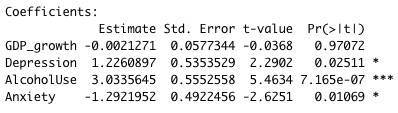
Bipolar and Eating disorders were removed as they were highly correlated. GDP was also removed and instead only GDP growth was taken as they both essentially lead towards the same interpretation. Now, we ran regression again and got the results as shown:

The results show that we have **4 significant variables** among which one is an external factor, and the others are internal factors. *Changes in GDP growth, Anxiety, Depression and Alcohol Use* correlate to shifts in suicide mortality rate. Among the significant variables, the variable which causes the most variation in suicide mortality rate is “Alcohol Use”. Our model tells us that **1 unit increase in Alcohol Use causes 4-unit increase in suicide mortality rate.** Additionally, **1 unit increase in depression causes 1.59-unit increase in suicide mortality rate.** Our first model has solved our first research question which is that what internal and external factors affect the suicide mortality rate, we dive deeper into the analysis and do regression on the clusters that we obtained through k-means. We are doing this to analyze which variables (among those which already came out to be significant) are significant in which clusters and what’s their overall impact.

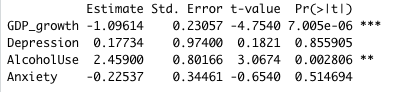
## **Model 2**

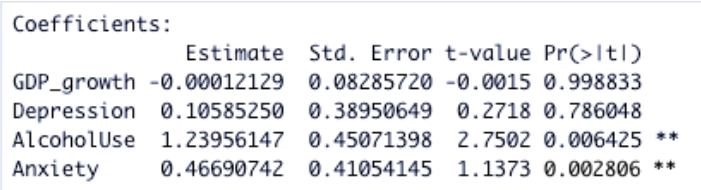
For model 2 of regression, we analyzed the significant variables across different clusters which were made according to the degree of development of countries. This was done to answer our second research question which was that *whether variables affecting suicide rate are same across all countries or they differ*. *Our analysis also answers why do the variables differ across various clusters*. We ran regression on all 4 clusters and the results for all of them are below

***Cluster 1:***

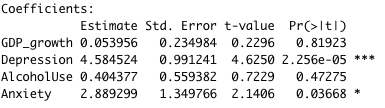
****The cluster 1 contains Middle Eastern developing countries which are **culturally and religiously very rich in nature**. The figure from regression output shows the significant variables. We can see that in these countries the significant variables are all **internal in nature (Anxiety, depression, alcohol use).** This is because in countries belonging to cluster 1, there are **high societal standards** which cause internal frustration among individuals leading them towards suicidal thoughts***. 1 unit increase in alcohol use causes a 3-unit increase in suicide mortality rate***. This is because in countries like Saudi Arabia, Afghanistan, Iraq, Qatar, Libya, etc., where religion of Islam is followed, the use of Alcohol is prohibited. Hence, if individuals fall a prey to this habit, they are publicly shamed which leads towards suicidal tendencies; thus, increasing suicide mortality rate.

***Cluster 2:***

The regression results of cluster 2 countries are shown in the figure. The cluster 2 contains **developed countries** and on observing these countries we can see that they are more **modern and individualistic** in nature according to their IND score. Every individual in this country has a high expectation bar set for him/herself and are very engrained towards their **financial growth.** This is because people belonging to societies in which individualism is promoted are more sensitive to personal achievements, especially monetary ones. Not only this, in countries like the US, Canada, Austria, New Zealand, etc. individuals are on their own from a very young age unlike countries from cluster 1. The concept of “success” is ingrained in them so much that failure brings **insecurity** among them and according to our literature review insecurity is one of the main drivers of suicide among individuals especially the youth. Furthermore, the use of Alcohol is the highest in these countries and the individuals often get so addicted that they don't know where to draw the line. Once you are under this state of mind you often end up taking your life. This analysis is also supported by our literature review.

***Cluster 3:***

The countries of this cluster belong to the **lower economic bracket** and are mainly considered to be **developing** in nature. The regression results are shown in the figure. Some of the countries like Ethiopia, Georgia, Chad, etc. in this cluster are so poor that people don't even get anything to eat. The countries struggle economically and don't have enough resources to secure a better future for themselves. Under such conditions, the feeling of getting anxious gets so worse that people often reach the extent of hopelessness where they comfortably take their lives. This explains the significance of variable “Anxiety” in the regression model of cluster 3.

***Cluster 4:*** The countries belonging to this cluster are mainly **collectivist** according to their IND score and are **developing** in nature. In collectivist societies such as that of Mongolia, Kazakhstan, Zimbabwe, Ukraine, etc. people are relatively more sensitive to societal experiences. Thus, when they don't achieve that validation, they have been striving for, they get depressed and anxious. Depression, in turn, is considered to be one of the most significant contributors towards suicide mortality rate. Depression is considered to be a source of distortion that induced the sense of negativity leading towards increased suicide rate. Thus, we can analyze from the results that in individualistic culture depression is the main source of suicide and this was more prominent in countries which are neither developed nor under-developed but are in their transition phase.

# **PREDICTIVE MODELLING**

There are several supervised machine learning techniques which are available for the analysis of data. Two Techniques fit our data set: ***one was KNN and the other one was decision trees***. KNN was able to give us results that fit our research question and that’s why we proceeded with the analysis using this technique. Moreover, KNN is a very comprehensive method which gives us results that are easily interpreted.

## **Logistic Regression**

For the data prep part of prediction analysis, the suicide mortality rate was categorized into the following categories based on ranges

1. *Extremely High*
2. *High*
3. *Low*
4. *Extremely Low*

This shows that our dependent variable is categorical, but it is not binary in nature and logistic regression is applied when we must make predictions about classification in binary terms or predict probabilities of an event. Thus, this technique was discarded as it did not qualify to fit our analysis or our dataset.

## **Decision Tree:**

The decision tree approach classifies data into different hierarchies by splitting it at different intervals and giving a clear visualization of data. This approach recognizes several splits in the data and makes the analysis easier, however it also has its drawback which we had to face. We built 2 models of decision trees. In the first model we incorporated all variables and in the second model we only added those variables which came out to be significant in MLR. The accuracy of both models was **25% and 43%** respectively which is a very low number. **In both the models, we can see that for the sake of fitting the data, even noisy data, the tree kept producing new nodes which made it extremely complex to interpret**. Thus, our decision tree technique was discarded due to overfitting which led towards its incapability to make generalized predictions on untrained data.

**MODEL 1**

Diagram

Description automatically generated

***Model 1***

***Model 2***

**MODEL 2**

Diagram

Description automatically generated

## **K-Nearest Neighbor:**

****KNN technique is used wherever simplicity and accuracy is required to a great extent. We used this technique to make predictions for the suicide mortality rate in the year 2020 as data for it was unavailable. The 2020 data for variables which came out to be significant was collected online to make accurate predictions. Before doing this, a train and a test set was made with a **65-35%** split. The predictions came out to be ***78%*** accurate.

****Since this accuracy is really high, we decided to apply this model to the 2020 data set for significant variables so that the supervised learning model can provide us with suicide mortality for the year 2020. To our surprise, there were no countries which had an “**extremely high**” suicide rate for the year 2020. This can be linked to the on-set of the pandemic. People were so distracted in keeping themselves safe from the virus that they did not think much about the problems of life. Moreover, obviously there was depression and anxiety due to quarantine but the effects of it were balanced with the feeling that “***everyone is in it together***”. This is one thing that kept people going in the pandemic. Furthermore, the suicide rates mainly declined for developed countries US, Canada, Australia, etc. because in these countries the suicide rates depended mainly upon external factors like GDP growth. As compared to other countries, **the developed countries** had enough resources to deal with this crisis; therefore, it wouldn't be wrong to say that their *crisis management helped them get through this tough time by giving people hope that they can depend on the government*. On the other hand, **underdeveloped countries** faced an increase in the suicide rates because people were left on their own as the government itself faced downfall. Thus, increasing depression and **anxiety leading towards increased suicide mortality rate**. The predictions for 2020 suicide rate are given on the right in the first prop table:

*The model that we made through KNN gave us good result in terms of accuracy which shows that the variables that we selected in the data exploration phase and while running multiple linear regression allowed us in building a model that could predict the suicide mortality for the year of 2020.*

# **RECOMMENDATIONS**

Our analysis pointed out how different variables can affect the suicide mortality rate differently across the globe. Not only this, the ratio of individualism vs collectivism was also incorporated to better interpret the dataset. The factors which were the most significant came out to be Anxiety, Depression, GDP growth and Alcohol Use. By reviewing our findings, these are some of the policy recommendations we came up with:

1. Out of these factors **Alcohol use** was seen to be prominent in both developed and developing nations. Thus, there is a pressing need to curb its use and make it illegal above a certain extent. Moreover, the *use of Alcohol also leads towards depression and anxiety* which again increases the suicide mortality rate. Hence, we can say that by controlling its use we can majorly bring down the suicide rate. This can be done by ***setting up rehabilitation centers*** which can professionally cure people who have fallen prey.
2. To control the psychological factors, proper channels need to be set-up. In developing countries, people are so focused on their growth that anything that doesn’t satisfy them leads towards negative thoughts. Thus, for such people ***career planning and counselling centers*** should be set up by the government. Cheaper well-provided psychological counselling and suicide helplines are also helpful alternatives.
3. **Foreign aid through funds** should be given to the under-developed countries so people can at-least not be anxious and depressed about the necessities of life like food, shelter, etc.
4. In the Middle Eastern and other traditionally conservative countries, **a cultural shift through education** can help people give each other the space to function according to their own wants and needs. Community centers established in the developed world would help lower depression & anxiety rates.
5. Higher suicide literacy and **awareness** through government policy should be encouraged.
6. **Youth rehabilitation centers** should be established for young people struggling with addiction due to lack of career opportunities. Jobs should be provided for high-risk youths.

# **LIMITATIONS**

Suicide mortality rate depends on several ***biological, social, psychological, and environmental factors*** and it is humanly impossible to cater to all these factors. Not only this, but these factors are also different for different people, and anything might be a triggering point for someone committing suicide. Thus, definite causality cannot be assumed between the number of deaths caused by suicide and other selected variables through mere “correlation”. Additionally, there might exist other aspects which are either undiscovered or differ from person to person pronouncing the causal relation with suicide and are beyond the scope of this project. Moreover, for our predictive analysis we only catered the variables which were significant in our MLR although there could have been multiple other undiscovered factors which were triggered during the pandemic. As pandemics are very rare, little research has been done on them which limits us in evaluating impacts of variables and thus distorts our predictions.

# **APPENDIX**

**Hierarchical Clustering dendrogram (k=4)**

A picture containing text, antenna

Description automatically generated

**Suicide Mortality Rate relation with variables and line of best Fit**

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

**CLUSTERS ON MAP AND SIGNIFICANT VARIABLES**

CLUSTER 1

***Chart

Description automatically generated***

CLUSTER 2

***A picture containing chart

Description automatically generatedA picture containing chart

Description automatically generated***

CLUSTER 3

***A picture containing text, different

Description automatically generatedChart, scatter chart

Description automatically generated***

CLUSTER 4

***Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated***

**Cross-table for KNN**

***Table

Description automatically generated***

**Accuracy for decision tree models**

******

******

# **REFERENCES**

Borgschulte, M., Corredor-Waldron, A. and Marshall, G. (2018). A Path Out: Prescription Drug Abuse, Treatment, and Suicide. SSRN Electronic Journal.

Cruwys, T., An, S., Chang, M.X.-L. and Lee, H. (2018). Suicide literacy predicts the provision of more appropriate support to people experiencing psychological distress. Psychiatry Research, 264, pp.96–103.

Eskin, M. (2018). SUICIDE AND CULTURE. BAU Journal -Health and Wellbeing BAU Journal -Health and Wellbeing, [online] 1. Available at: https://digitalcommons.bau.edu.lb/cgi/viewcontent.cgi?article=1014&context=hwbjournal [Accessed 17 Dec. 2021].

Helliwell, J., Layard, R., & Sachs, J. (2018). World Happiness Report 2018, New York: Sustainable Development Solutions Network.

Johnson, A. (2018). Gross domestic product. [online] Dgff.unctad.org. Available at: http://dgff.unctad.org/chapter2/2.1.html [Accessed 13 Dec. 2021].

Marušic, A., Khan, M. and Farmer, A. (2000). P02.345 Literacy rate, gross domestic product, and suicide rate in 33 European countries. European Psychiatry, 15(S2), pp.415s415s.

Marušic, A., Khan, M. and Farmer, A. (2000). W 13. Negative attitudes and discrimination against severe mentally ill patients. European Psychiatry, 15 S2, pp. 219s.

Nationsonline.org. (2018). Countries of the First World. [online] Available at: https://www.nationsonline.org/oneworld/first\_world.htm [Accessed 13 Dec. 2021].

Odagiri, Y., Uchida, H. and Nakano, M. (2011). Gender Differences in Age, Period, and Birth-Cohort Effects on the Suicide Mortality Rate in Japan, 1985-2006. Asia Pacific Journal of Public Health, [online] 23(4), pp.581–587. Available at: http://www.jstor.org/stable/26723741 [Accessed 17 Dec. 2021].

Pampel, F.C. and Williamson, J.B. (2001). Age Patterns of Suicide and Homicide Mortality Rates in High-Income Nations. Social Forces, [online] 80(1), pp.251–282. Available at: http://www.jstor.org/stable/2675538 [Accessed 17 Dec. 2021].