Alcohol Consumption and Suicide Rates Worldwide

Anisha Khandelwal

Masters of Business Administration, A.P.J. Abdul Kalam Technical University

Introduction

During the recent times, Alcohol consumption has massively increased in the society with equational rise in suicidal tendency and mood swings related to it. Two very important reasons for Alcohol; consumption is-

- Stress due to work load
- Due to social influence

In-order to cope up with the stressful life and deal with more challenging work environment, people are adopting more unhealthy lifestyle. This in turn is creating a change in mental well -being and in turn, people are more prone to suicidal tendencies. It is high time to create awareness about the alcohol consumption menace and it's impact on society in today's scenario to prevent any mental (suicidal rates) and physical imbalance in coming years.

Now, this study context aims at finding some more reasons and factors which are responsible for increased Alcohol consumption and suicidal rates in the society like –

- Population
- Employment rate
- Literacy rate
- Happiness quotient on score board
- Income of the person.

The concept study aims to create awareness about alcohol consumption in present day and reduce the upcoming health issues mostly suicidal rates, mood swings and other drastic mental health issues related to it in future. Arresting the causative factors in present day will prevent the secondary complications in future. This is the main concern put up through this study.

Methods that can be used to reduce Alcohol consumption and related Suicidal rates-

- Withdrawal method – Taking a back seat in consuming less alcohol and withdraw from buying and having Alcohol at home.
- Education- Educating and spreading a word in areas like rural , Organisations, Educational Institutions for teenagers,
- Regulating and Restricting the use and availability of Alcohol in the market.
- Implementing screening and Intervention programmes in health services.
- Providing accessible and affordable treatment for alcohol use dis-orders.
- Policy making.

Historical Background

Way back in 1849, a Swedish physician coined the term "Alcoholism". And in 1647, A Greek monk was the first to document that chronic alcohol misuse was resulting into medical dis-orders like seizures, paralysis and psychiatric issues. In 2001, 2.8 million number of cases was calculated related to Drinking Alcohol. In 2001, 594 people died due to suicide by Alcohol overdose. And it was seen that men counted more in consuming alcohol rather than women population.

So, approximately 3-15% of alcoholics commit suicide. Over 25% teenagers or school students get suicidal thoughts over Alcohol use.

Measures used to treat Alcohol over use in historical times-

- Transtheoritical model of motivation for behavioural change. This was successful in treating alcohol addicts when they were in the extreme stage.
- Prohibition of alcohol through ban of import, export, production, transport and sale of alcohol.
- Making of policies in 1920 and 1930's led to decrease in suicidal rates due to alcohol use.

So, the need of the hour is to implement new norms and regulations to save the lives of many . As, it is disabling people. The problem has hit hard to the society and should be arrested before it cripples down the society.

Literature Review

Techniques which were mostly used to treat Alcohol Use Dis-order are many. Still the importance of these techniques are benefitted by the people now. More precisely, techniques like-

*Behavioural Therapy- It teaches to cope with skills, includes how to manage stress and how to change the thoughts that cause to drink.

* Injected Medication- Anti-Alcohol medications were induced to the person as a general first line medication.

* Alcohol withdrawal with drug substitution with Benzodiazepams were used.

* The concept of 12 step recovery phase was implemented when the person reached the stage of "Hitting Rock Bottom".

To evaluate and reach to a comparison of various scores and the variability for the study, different analytical techniques are used like the "Random Forest Regressor" to find out the impact of various factors like Urban and Rural population, Employment rate, Emotion (Happiness) score, literacy rate.

 To identify the various dependant and independent variables, summarization and data type of each variable was clarified.

The various data types of the variables are discussed as below-

• Country- The Unique Identifier

• Income per person- The gross domestic income per capita and the cost of living was taken into account.

• Alcohol consumption- Alcohol consumption per adult, average consumption per litres (age 15+) was taken.

• Suicide rate – Age specific, suicides per 100000 per population.

• Employee rate - Total number of people employed per year, with age specified.

• Urban rate- The urban population calculation (Using World Bank Population Control).

• Literacy rate:-Amount of people literate, age specified (age 15+).

 •Happiness score- Factors contributing to Happiness score board, freedom, life expectancy, GDP level, etc.

In the study, many null values found could hamper the on -going process of the model, so to eliminate these null variable, engineering method is employed. The boxplot method was used after skewness checking of variables. Of all the variables, three out of all had outliers and were removed by two methods of "Quantile based flooring" and "Cupping technique"

After the data analysis of all variables, the highest correlation was found between alcohol consumption and suicidal rates. To understand better about alcohol use in geographical area, Scatter plot was taken indicating direct impact on one another. The Random Forest Regressor statistics which usually fits all the classifying decision tree in sub variables in dataset and uses average values for results, also showed scores less than 1 , for which it makes regression a good fit for this problem.

Problem Statement

In the recent years, alcohol abuse has become a menace to the society, and so has suicides. Alcohol dependence is an important risk factor for suicidal behavior. Mood disorder is a more powerful risk factor for suicide among problem drinkers as age increases. Two common reasons for alcohol consumption are: (a) people drink alcohol to cope with stress, and (b) people drink alcohol because of social influences. In today's competitive world, people around the world often follow an unhealthy lifestyle by drinking alcohol in huge quantities to cope up with their stressful lives. Excess levels of stress and following such unhealthy lifestyle also contribute to the rising suicide rates. Moreover, peer pressures and social influences can also lead to increase in alcohol consumption as well as suicide rates. It is important to be aware of the toll alcohol consumption can take on our bodies, as well as on our mental health.

The paper focuses at finding out how alcohol consumption and suicide rates in all countries of the world are affected by various factors like income, employment, urban population and literacy rates. To

understand the impact of these factors, random forest regressor will be used, along with finding the correlation of these variables.

Objectives

The paper aims at finding out how alcohol consumption and suicide rates are impacted by various other factors like urban population, income per person, employment, literacy rates and happiness scores in various countries, as well as how much impact they have on each other. This study is done to bring awareness about the prevention of alcohol consumption and suicides.

Methodology

In this paper, Random Forest regressor will be used to find the impact of all the factors like employment, urban population, income per person, literacy rates and happiness score on alcohol consumption and suicide rates.

Steps involved in the analysis:

1. Obtaining dataset

2. Importing libraries and dataset

3. Data cleaning and Imputation

4. Data analysis

5. Data visualization

6. Data modelling

1. Obtaining dataset:

The data was collected from Kaggle (https://www.kaggle.com/sansuthi/alcohol-consumption), an open-source platform where free datasets are available. This dataset was combined with literacy rates of adult population, which was taken from World Bank, and happiness score dataset from Kaggle (https://www.kaggle.com/ajaypalsinghlo/world-happiness-report-2021).

2. Importing libraries and dataset:

The following libraries are imported for the analysis, visualization and modelling:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import plotly.express as px

from sklearn.model_selection import train_test_split

from sklearn.impute import SimpleImputer

from sklearn.ensemble import RandomForestRegressor

from sklearn.metrics import mean_absolute_error

import shap

Importing dataset:

alccons = pd.read_csv(" alcohol_final_dataset.csv")

The dataset contains 210 rows and 8 columns. There seems to be enough variables for the analysis. In order to identify the dependent and independent variables from the dataset, it is important to summarize the dataset and find out the datatypes of each variable.

alccons.info() #for checking column information

the following are the columns in the dataset:

```
Int64Index: 210 entries, 0 to 209
Data columns (total 8 columns):
 #  Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0  country         210 non-null    object
 1  alcconsumption  210 non-null    float64
 2  incomeperperson 210 non-null    float64
 3  suicideper100th 210 non-null    float64
 4  employrate      210 non-null    float64
 5  urbanrate       210 non-null    float64
 6  literacyrate    210 non-null    float64
 7  Happiness score 210 non-null    float64
dtypes: float64(7), object(1)
```

From this table, 2 dependent variables can be derived, i.e., alcohol consumption and suicide per 100th, according to the objectives of the study, and the remaining variables are independent. A brief description of the dataset variables is as below:

- country: Unique Identifier

- incomeperperson: Gross Domestic Product per capita in constant 2000 US$. The inflation but not the differences in the cost of living between countries has been taken into account.

- alcconsumption: Alcohol consumption per adult (age 15+), litres Recorded and estimated average alcohol consumption, adult (15+) per capita consumption in litres pure alcohol

- suicideper100th: Suicide, age adjusted, per 100 000 Mortality due to self-inflicted injury, per 100 000 standard population, age adjusted

- employrate: Total employees age 15+ (% of population) Percentage of total population, age above 15, that has been employed during the given year.

- urbanrate: Urban population (% of total) Urban population refers to people living in urban areas as defined by national statistical offices (calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects)

- literacyrate: literacy rate, adult total (% of people ages 15 and above)

- Happiness score: The average ladder score obtained from six factors: levels of GDP, life expectancy, generosity, social support, freedom, and corruption.

3. Data cleaning and imputation:

Null values can negatively impact the performance of the model, so it is necessary to remove the null values before moving ahead with the analysis. The dataset contains null values in moderate amounts, among all the variables, literacy rates contain the highest number of nulls.

alccons.isnull().sum() #to check no of null values in each variable

```
country             0
alcconsumption     23
incomeperperson    20
suicideper100th    20
employrate         35
urbanrate           7
literacyrate      173
Happiness score    66
dtype: int64
```

In order to remove these null values, feature engineering is required. So, imputation will be followed in this process, i.e., the null values will be replaced with the mean in each variable. The null values were checked after the imputation, to see if all the null values were replaced or not.

alccons.isnull().sum() #to check if any null values exist

country   0

alcconsumption   0

incomeperperson   0

suicideper100th   0

employrate   0

urbanrate   0

literacyrate   0

Happiness score   0

The next step was feature engineering, which is another part of the data cleaning process. After imputing the null values, outliers need to be identified and removed. For this process, skewness was checked for each variable in the dataset.

print(alccons.skew())

alcconsumption   0.655499

incomeperperson   3.414677

suicideper100th   1.496543

employrate   0.130740

urbanrate   -0.019165

literacyrate   -1.142354

Happiness score   -0.141419

In this output, it can be observed that 3 variables, i.e., incomeperperson, suicideper100th and literacyrate are highly skewed distributions. This means that they have more outliers than the other variables in the dataset. This was observed with the help of a boxplot for both the variables before implementing the outlier removal process, as shown below.
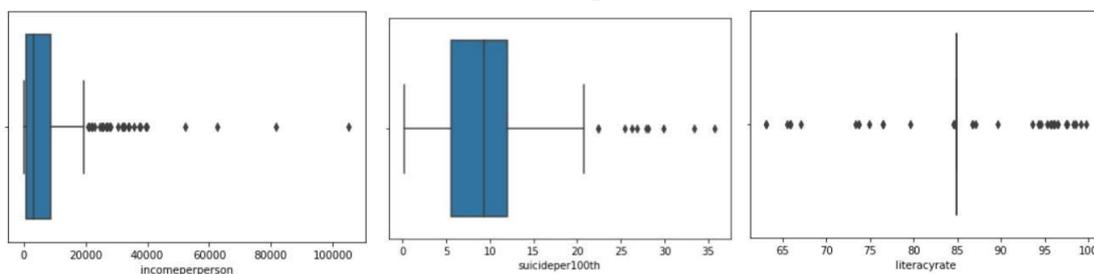
If the skewness value is less than −1 or greater than +1, then it is considered to be a highly skewed distribution. Since these 3 variables are greater than +1 and less than -1, they are highly skewed, and the outliers of both these variables are checked through box plot. From the above plots, it is seen that three of these variables are having outliers. To remove these outliers, quantile-based flooring and capping technique was implemented. The flooring (e.g., the 10th percentile) for the lower values and capping (e.g., the 90th percentile) for the higher values were done. The calculated 10th and the 90th

percentile values were used to remove the outliers. Then the skewness for these variables were calculated after the outlier removal. The boxplots were plotted as below, post outlier removal.
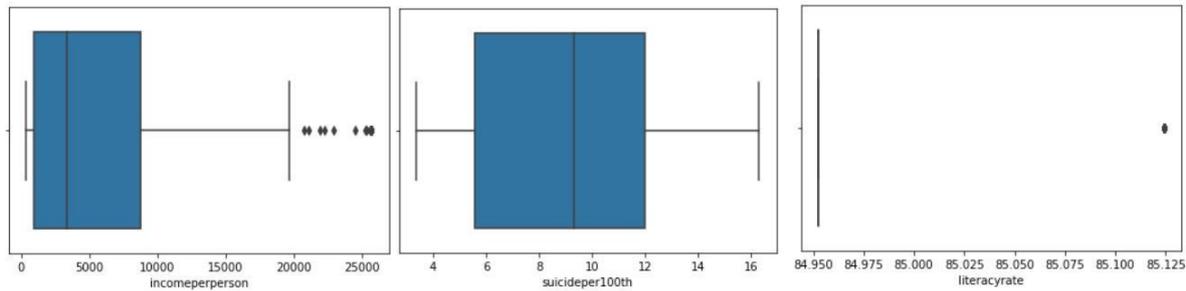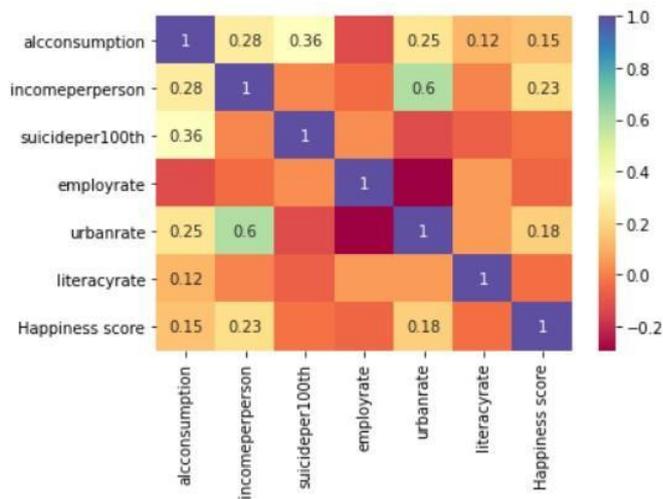

	Alconsumption	incomeperperson	suicideper100th	employrate	urbanrate	literacyrate	Happiness score
count	210.000000	210.000000	210.000000	210.000000	210.000000	210.000000	210.000000
mean	6.689412	7239.853636	9.074012	58.867429	56.769360	84.969552	5.538618
std	4.622165	8329.331004	4.091378	9.537075	23.442215	0.051692	0.898460
min	0.030000	358.541558	3.351656	32.000000	10.400000	84.952362	2.523000
25%	3.185000	949.473779	5.574041	53.725000	36.965000	84.952362	5.077500
50%	6.689412	3389.537977	9.323386	58.867429	56.769360	84.952362	5.538618
75%	9.575000	8740.966076	11.974608	63.775000	73.605000	84.952362	5.981750
max	23.010000	25673.001782	16.302258	83.199997	100.000000	85.124257	7.842000

A significant decrease in the points on the box plot can be observed, which implies to a reduction in outliers in these variables. Now, the mean, standard deviation, min, max, and the number of values in the variables were checked.

4. Data Analysis:

For the analysis, the initial step was to check the correlation between variables, with the help of a heatmap. So, a heatmap was plotted for the dataset, with all the variables. There were variables with positive, negative and zero correlation. The correlation heatmap ranged from -0.2 to 1.0, and distinct color palette was used for clarity. Since only positive correlation needs to be identified from the heatmap, another heatmap was plotted with a condition to make only the correlation values of those variables visible on the map, which were above 0.1. The correlation heatmap obtained is as shown below.

It is observed that the highest correlation is between suicide rate and alcohol consumption. i.e., 0.36. There is also a positive correlation between alcohol consumption and income per person, urban population, literacy rate and happiness score, whose values are all above 0.1. Since the dataset contains multiple variables, it can be seen that many variables share positive correlation with alcohol consumption. To understand this better, multivariate analysis was done, and the plots were marked by a regression line. From this regression line to find the relations between the variables of the dataset, a lot of findings can be drawn, but since the study is focused at the variables' impact on alcohol consumption and suicide rates, this will be better understood in the data modelling phase.

5. Data Visualization:

As the paper focuses at studying the impact of the variables on alcohol consumption and suicide rates, it is important to understand and interpret visualizations of the countries with the highest and lowest alcohol consumption and suicide rates, and how these 2 variables, in turn, impact each other.
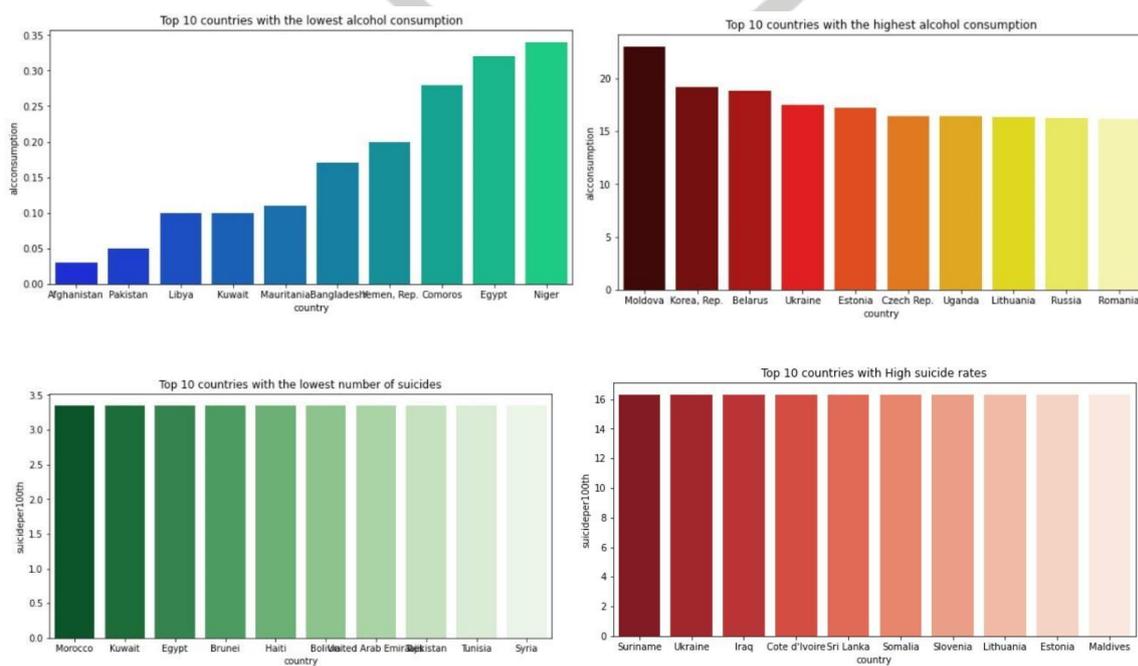
From the above bar charts, the highest alcohol consumption is in Moldova, and Suriname is having the highest suicide rate. On the contrary, Afghanistan has the lowest alcohol consumption rate, and Morocco is having the lowest suicide rates. The relationship between alcohol consumption and suicide rate can be studied by plotting them on a scatter plot.

Suicide rate vs Alcohol consumption
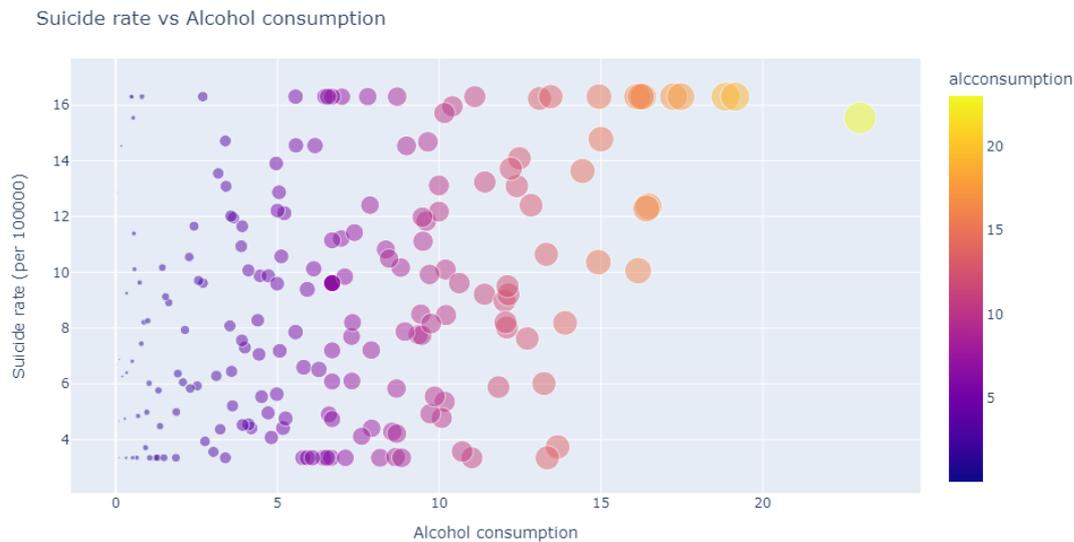
From the scatter plot, it can be said that both these variables are having an impact on each other. But this will be better understood in the data modelling phase.

6. Data Modelling:

Splitting data into train and test data:

In this step the entire dataset was split into train and test data. The alcohol consumption variable was stored in the y value and all other features were stored in the x values for train and test. Enough observations were considered in the train and test datasets for the data modelling process. Then the random forest regressor was used for data modelling. A random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. For this, the number of trees were taken as 100, random state instance was set to 22 and the minimum number of samples in newly created leaves was set to 10 and the minimum number of samples required to split an internal node is set to 0.7, and then the model was fitted to the train and test data. Then the regression scores were checked, which were 0.371435637083401 and 0.4089479567584784 for train and test respectively. The best possible score is 1.0 and it can be negative (because the model can be arbitrarily worse), but in this case, it is not negative but less than 1, so it is relatively a good fit.

Role of Mental Health in Preventing Suicide

Mental Health is the foundation for keeping a balance between the physical and psychological aspect. If the person is mentally stable and is alert enough to understand the difference between right and wrong then he/she can have control over alcohol and can prevent themselves from suicides.

Some therapeutic manoeuvres that can be adapted to eradicate this problem in future are-

*Following a Holistic Approach – Maintaining a good lifestyle, mindfulness eating, mindful watching, mindful thinking, physical activities, counselling, timely examinations and assessments can help.

*Techniques like relaxation, yogic methods, deep breathing exercises, meditation , music therapy

* Awareness and Educating people.

Results

Mean absolute error was predicted for both train and test data. Mean Absolute Error is a model evaluation metric used with regression models. The mean absolute error of a model with respect to a test set is the mean of the absolute values of the individual prediction errors on over all instances in the test set. Each prediction error is the difference between the true value and the predicted value for the instance. It is usually used when the performance is measured on continuous variable data. It gives a linear value, which averages the weighted individual differences equally. The lower the value, better is the model's performance. The mean absolute error values obtained for the train and test data are 2.7184469670914795 and 3.1227602641403425 respectively. This shows that the model's MAEs are fairly small given that our data's range, which suggests that the model is great at prediction. The same methods were followed in case of suicideper100th variable. To explain the output of the model, the SHAP library was used. SHAP (SHapley Additive exPlanations) is a game theoretic approach to explain the output of any machine learning model. It connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions. With the help of this library, the impacts of other variables over alcohol consumption and suicide rates can be understood through summary plots.
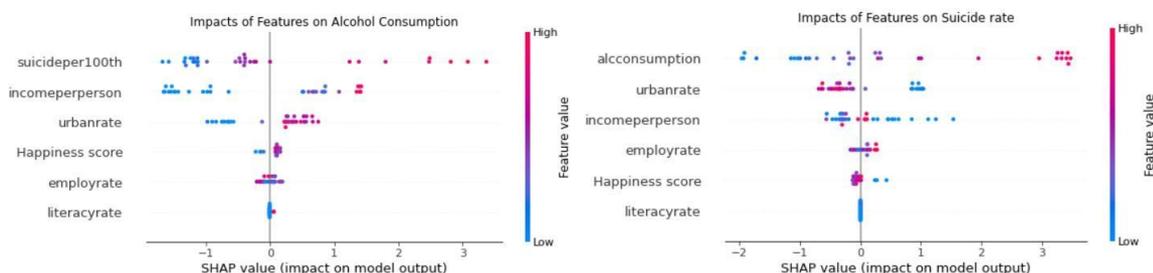
From both the plots above, it can be observed that suicide rates are lower in case of low alcohol consumption, and higher with high alcohol consumption, which means that they are directly proportional to each other. The next factor is income per person, and it can be seen that alcohol consumption is low in low-income countries and high in high income countries, whereas suicide rates are highest among the middle-income countries, moderately high in low-income countries, and lowest in high income countries. A high urban population leads to an increase in alcohol consumption, but a lower urban

population leads to high suicide rates as shown in the plots. In case of employment rate, it can be seen that alcohol consumption is moderately high and very high in countries with high as well as low employment rates respectively, whereas it is low in countries with moderate employment rates, and suicide rates are high with higher employment rates and moderate with low employment rates. It is also observed that alcohol consumption is more in countries with higher literacy rates and much lower in countries with lower literacy rates, whereas suicide rates are low in both cases of literacy rates. The last feature is happiness score, in which it can be observed that higher happiness scores lead to more alcohol consumption as compared to low happiness scores, whereas suicide rates are high in case of lower happiness scores.

Conclusion

In this paper, the impact of alcohol consumption on suicide rates, as well as the impact of many other social and economic factors on suicide rates and alcohol consumption were studied, to understand which other factors impact both these aspects the most, and how its effects can be reduced. From the results, it is clear that the people who consume alcohol the most, are more likely to commit suicide, because intoxications can cause hallucinations and suicidal tendencies among the alcohol consumers. Since people with low income cannot afford alcohol, its consumption is low as compared to those who are earning more, and can afford expensive alcoholic drinks in large quantities. Suicide rates are lower among people with higher earnings as compared to people with lower earnings, as people in low-income countries have a lot of liabilities and debts which is usually not possible for them to clear.

An interesting pattern was noticed in case of income per person, which is a positive correlation between urban population and income per person, which shows that the people with increase in their income will relocate to urban areas thereby increasing the urban population, and since the urban areas are more influenced by the western culture of drinking alcohol, so the alcohol consumption is more in the cities with high population. But suicide rates are drastically high among lower urban population, as people in cities with less population may tend to feel more isolated and depressed. Alcohol abuse can be observed among the unemployed population, whereas, suicides are drastically high among the employed and working class due to possible severe workloads and issues which has a negative impact on their mental health. Literacy rate does not have much impact on the alcohol consumption and suicide rates, so these factors are having a low impact, which implies that a person may consume alcohol or commit suicide irrespective of his/her education level. But surprisingly, it is seen that a higher happiness score has led to higher alcohol consumption, as people in most countries include alcohol consumption during certain festivals or celebrations. Suicide rates are low with high happiness scores, which means that people who are satisfied, have a positive attitude towards life and a good mental health, whereas a low happiness score means that those people are more prone to committing suicides due to mental health issues.

It can be concluded that neither alcohol consumption nor suicide rates are solutions to our problems. More awareness is required on the negative impacts' alcohol has on our bodies as well as on our mental health, so that its consumption can be minimized to a great extent. One must follow a healthy lifestyle and have a positive attitude towards life in order maintain a healthy mental state and prevent suicides.

References

[1] Agnus M.Kim "Alcohol Consumption & suicide Rate : A cross sectional analysis in 83 countries(2021),[online] Academia

[2] Emine M. Inelmen1 , Marialuisa Gazerro2 , Erol Inelmen3 , Giuseppe Sergi1 , Enzo Manzato1, "Alcohol consumption and suicide: a country-level study" (2010), [online] http://inelmen.boun.edu.tr/emi/publ10/view.pdf

[3]L.Sher, "Alcohol Consumption and suicide" (2005), [online] https://academic.oup.com/qjmed/article/99/1/57/1523792

[4] Maurizio Pompili, Gianluca Serafini, Marco Innamorati, Giovanni Dominici, Stefano Ferracuti, Giorgio D. Kotzalidis, Giulia Serra, Paolo Girardi, Luigi Janiri, Roberto Tatarelli, Leo Sher, David Lester (2010), " Suicidal behaviour and Alcohol Abuse", [online] https://www.mdpi.com/1660-4601/7/4/1392

[5] Sharon C. Wilsnack, Richard W. Wilsnack, Arlinda F. Kristjanson, Nancy D. Vogeltanz-Holm, Michael Windle, " Alcohol use and suicidal behavior in Women : Longitudinal Patterns in a US National Sample" (2006),
[online] https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1530-0277.2004.tb03602.x

[6] Yury E. Razvodovsky, "Alcohol consumption and Suicide in Belarus, 1980–2005" (2011), [online] http://www.suicidology-online.com/pdf/SOL-2011-2-1-7.pdf