Suicides and Homicides in Different Countries Analysis IEEE PAPER

Soffyan Ali
x13114531
Software Development, 4th Year
BSc. Honours in Computing
National College of Ireland
Dublin, Ireland
x13114531@student.ncirl.ie

***Abstract* - The Data has been analyzed on mental health of suicides and homicides in different countries. The visual graphs are represented by combining the web scrapped Homicides data and Suicide data along with Mental health survey. Data interrogations in between these three dataset are showing the age groups of crimes and them seek help before the incident and or got treatment.**

Keywords – Suicides, Homicides, Mental Health and Crime

I. INTRODUCTION

Around the world every minute there has been estimates that approximately around one million die each year from suicides. It's horrible to see someone killing themselves and not many people seeks help or share their deep inside pain. And some out there are killed by Homicides where in the global study on Homicide 2013 estimated that 6.2 per 100,000 population. [1]

However, suicide and homicides arises from many social cultural factors that can occur through individual crises (e.g. unemployment, loss of loved one, or they just need to share with someone). In the analysis of these incidents there has been survey combined to find out how many out there males and females seek help and get treatment. The Homicide data has been crawled from web and then combined with Suicides and Mental health survey.

II. LITERATURE REVIEW

In the research, several statistics survey found on suicides and homicides towards mental illness. One of the related topics has been analyzed in a technical workplace to find out the employees mental illness and examine the frequency of health performance in work environment. [1] In the survey, one of the main attribute is Employee that has been analyzed by treatment from the company. The study shows that 40% of employees are not aware of the mental health care options in the workplace and 25% are not sure. And half of the Employees underwent the treatment, the rest half didn't. Also the most interesting analytic graph is that the employer hasn't provided resources for employee to seek help of mental health issues which should be first priority for the company. In the suicide and homicide data, the survey was combined to analyze that how many victims has reached out to seek help and got their treatment regarding the mental illness, along with the city population and the age group that the crimes occurred.

III. DATA

Overviews of the datasets used for analysis are detailed below. Beginning with the Homicides web crawled data through Wikipedia of the 50 cities that have the highest murder rates in the world of all the cities. [1] In the Homicide data there are 6 variables in total and 2 variables were removed Rank, Homicidesper. Homicide column and City Population titles were change.

The Suicide dataset had 27820 rows and 12 variables which were downloaded from Kaggle. [1] The unnecessary variables were removed which were 8 in total. They were age, suicide100pop, country/year, HDI for year, gdp for year, gdp per year and generation column. The Country and Country population titles were changed.

In survey dataset which also was download from Kaggle. [1] The dataset had 27 variables and 1259 rows. Certain column were removed Timestamp, Gender, state, no employees, tech company, care options, wellness program anonymity , leave, co-workers, supervisor, mental health interview, mental vs. physical, obs consequence, comments, remote work, work interfere, and family history. Only 9 variables were used for analysis.

After these three dataset which were then combined by country and it came out over a million (1454748) and 15 variables in total.

Homicides attribute and description:-

Attributes	Description
City	50 cities CHAR
Country	(Primary key) Which has 10 countries CHAR
Homicides	as numeric
City Population	as numeric

Suicides attribute and description:-

Attributes	Description
Country	(Primary key) 1 to 100 countries CHAR
sex	Male/Female Factor
suicides_no	as Integer
Country Population	as Integer

Survey attribute and description:-

Attributes	Description
Age	1 – 100 as numeric
Country	(Primary key) 1 - many CHAR
self-employed	Yes/No factor
treatment	Yes/No factor
Employed	4 levels Never, Often, Rarely, Sometimes factor
Social Benefits	3 levels Don't Know, No, Yes factor
seek_help	3 levels Don't Know, No, Yes factor
mental_health_consequence	3 levels Maybe, No, Yes factor
phys_health_consequence	3 levels Maybe, No, Yes factor

IV. METHODOLOGY

The KDD (Knowledge Discovery in Database) is used to extract knowledge from Big Data which has been applied in this project. KDD process is the analysis step in Data mining where data is cleaned, integrated, selected, transformed, data mining, pattern evaluated and knowledge presented.

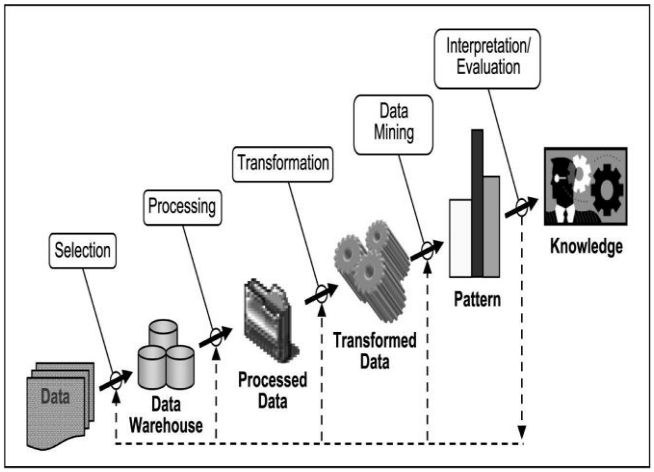


Fig1: - KDD process steps [1]

Below are the steps that were taken in the discovery process:

- **Data Cleaning –** In this step, the data was cleaned by removing missing values, special characters and unnecessary columns.
- **Data Integration –** In this step, suicide, homicides, mental health survey was combined by Country as Primary key.
- **Data Selection –** In this step, relevant data to produce the analysis were retrieved from the database. The suicide no, homicides, city population, sex, age, seek help were retrieved.
- **Data Transformation –** In this step, the data algorithm was chosen to match the goals of process to a data mining such as classification.
- **Data Mining –** In this step, run the chosen algorithms on data to extract the pattern.
- **Pattern Evaluation –** In this step, data patterns were made such as visualization graphs. Also the TopTen map reducer has been used to see most crimes in city.
- **Knowledge Presentation –** Finally In this step, the discovered knowledge us represented such as this report.

V. IMPLEMENTATION AND ARTITECTURE

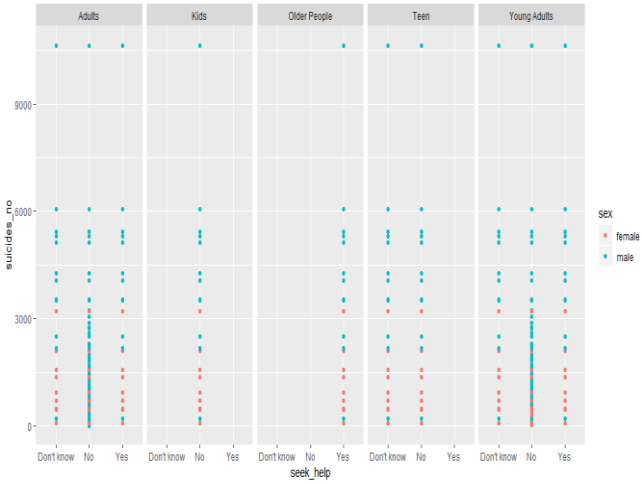
- Two dataset namely Suicide and Survey on Mental Health are used.
- One dataset was web scrapped from Wikipedia of Homicides.
- Merging three datasets was not easy as it had to be somehow related to each other.
- The dataset was later combined by Country.
- Initially data had many missing values and special characters which were removed.
- The column Titles were changed to make it more sense able and data was restructured.
- The data was over a million when merged, and then the rows were reduced to 100,000.

- Furthermore various data mining techniques were applied to get a real insight of data.
- The suicide and homicide were mutated together called crime column.
- The Age was in numbers which then divided into Age groups like (Kids, Teen, Young Adults, Adults, and Older People).
- Libraries used for visualization are ggplot, ggpubr, magrittr, dplyr.

VI. RESULTS

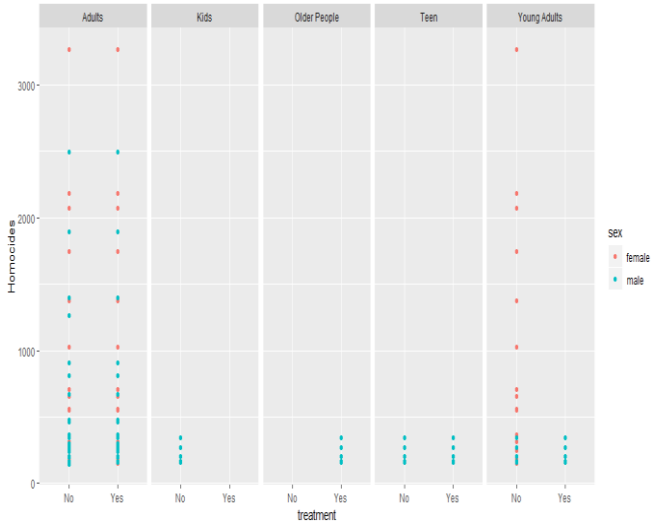
Number of graphs are discussed later which are combined by three datasets. Suicide and Homicides are merged and shared Country as a reference key. Based on the mentioned keys which are discussed below:-

A. Number of Suicides by Age Group who Seeked Help



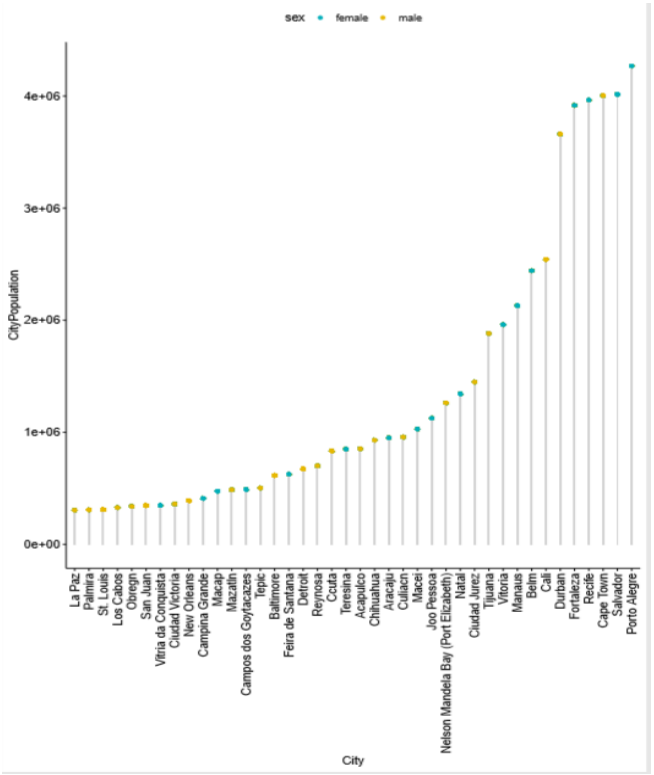
As the graph shows above that Young Adults and Adults has high percentage of Male who hasn't seek any help regarding the suicide. The purpose of using this visual graph is to find out the reason behind the suicide. The survey shows that more people need to be aware of their mental health and reach out to seek help which can stop them from suicide.

B. Number of Homicide by Age Group that had treatment



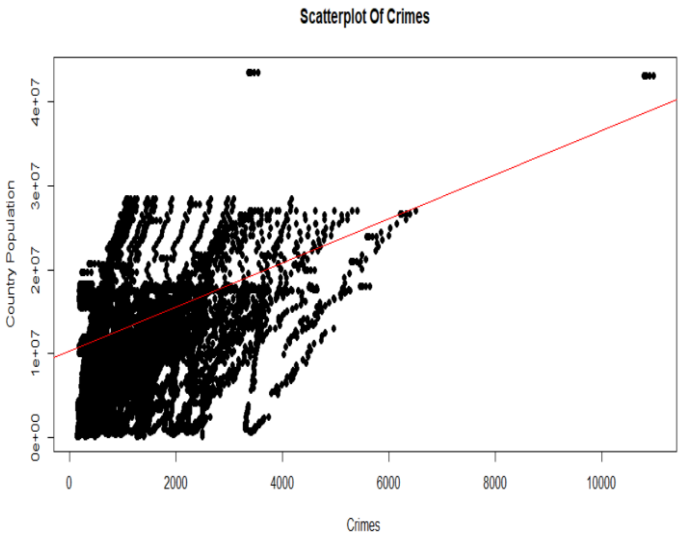
The analysis of this graph visualizes that the Young Adults females had no treatment and only males had gone for treatment for mental illness as a homicide victims. Also the graph shows that there are only males in age group of kids, teen and older people.

C. Lollipop Gender population in City



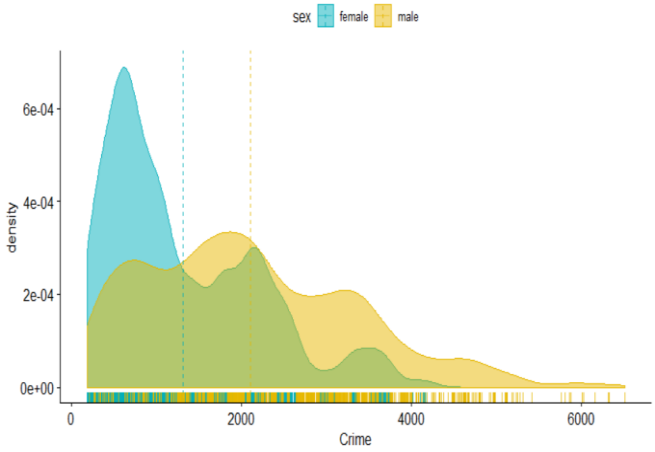
In the above Lolipop graph shows in which City has most Males and Females. Porto Alegre has the most females and Cape Town has the most Males. This Lolipop graph can be very useful to show numerical variable with categorical varaible and the visuals look great.

D. Scatter Regression Plot of Crimes [1]



In the Scattered plot above, it shows how many Crimes happening in the city along with a regression line. Between 0 to 4000 there is bunch of crime happening.

E. *Sexes of Crime* [1]



Crimes are happening many in World. This analysis graph visualizes the mean of Males and Females Crime. Graph shows that Males most crime are just over 2000 and Female 1700.

F. *Map Reduce Results*

- During the implementation of MapReduce, algorithm was implemented rather than using inbuilt functions of python. The MapReduce function was given example through NCI college examples.
- As a huge amount of data available in Fulldata.csv file, the Crime variable has been extracted to Run TopTenMapper.
- Out of TopTen the outcome was sorted and the highest crime in the Countries is 6508.
- Then the TopTen crime outcome was Reduce it to 5 attributes of crime.
- MapReduce has been hard to understand and further learning has to be applied in later research.

VII. Conclusion and Future Work

The Project has been a great journey of learning new experience, while working on dataset I used web crawling for the first time. Working on this project I've gain insight about Data mining and how to deal with Big Data. Learned how to break the problem into pieces and apply my skill set of analysis on the problem. It took quite a while to web scrap a dataset through JSON calling an API. But I couldn't find the relationship in between the dataset so I ended up not scrapping through JSON. In future, I would like to work on JSON and work on more realistic Big Data like extracting and learning new unique graphs and methods. For my future project, I would like to work on my final year project which is fetching the data from GoogleFit wearable device. And analyze the data through JSON which will be a fantastic experience.

VIII. References

[1] "List of cities by murder rate," Wikipedia, the free encyclopedia, 10 December 2018. [Online]. Available: https://en.wikipedia.org/wiki/List_of_cities_by_murder_rate. [Accessed 8 12 2018].

[2] "List of countries by intentional homicide rate," 10 December 2018. [Online]. Available: https://en.wikipedia.org/wiki/List_of_countries_by_intentional_homicide_rate. [Accessed 8 12 2018].

[3] "List of countries by intentional homicide rate," Wikipedia, the free encyclopedia, 21 December 2018. [Online]. Available: https://en.wikipedia.org/wiki/List_of_countries_by_intentional_homicide_rate. [Accessed 12 December 2018].

[4] DSangeetha, "Analysis on Survey of Mental Health," 21 July 2018. [Online]. Available: https://www.kaggle.com/devisangeetha/analysis-on-survey-of-mental-health/notebook. [Accessed 25 December 2018].

[5] Szamil, "WHO Suicide Statistic," 14 September 2018. [Online]. Available: https://www.kaggle.com/szamil/who-suicide-statistics. [Accessed 12 December 2018].

[6] J. W. &. Sons, "Methodology," SML Solutions, 23 May 2011. [Online]. Available: http://smlsolutions.blogspot.com/p/methodology.html. [Accessed 20 December 2018].