

The background features a dark blue and black abstract design. A white line graph with yellow circular markers is visible on the left side. In the center, there is a large, stylized 'L' shape composed of two overlapping bars, one light blue and one grey. To the right of this shape, the title 'SUICIDE RATE OVERVIEW' is written in a bold, white, sans-serif font. Below the title, the text 'Submitted By: Anuja Yawar' is written in a smaller, white, sans-serif font. The overall aesthetic is modern and data-oriented.

SUICIDE RATE OVERVIEW

Submitted By: Anuja Yawar

Introduction

Suicide is a critical issue in modern society. Early detection and prevention of suicide attempts should be addressed to save people's life. Suicide rates vary across different groups of people, geographic regions, over time, etc. Identifying differences in suicide rates across groups often suggests hypotheses for why those differences occur. For example, the western and southeastern states have higher suicide rates, on average. If after determining a trend or pattern in suicidal ideation, for example, school officials may decide to implement a peer counseling program. The school should then track the number of students reporting suicidal ideation on the Youth Risk Behavior Survey or talking to school counselors in order to evaluate whether the peer counseling program is having the desired effect.

Data set

Source-Kaggle

Size - 2588kb

Link-

<https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016>

Suicide Rates Overview 1985 to 2016: This dataset is compiled dataset pulled from four other datasets linked by time and place. The data set by built to find the signals correlated to increased suicide rate among different cohorts globally, across the socio-economic spectrum. The dataset contains approximately 27,000 rows and 12 features like country, year, sex, age group, count of suicides etc.



Idea

- Perform EDA using PySpark.
- To visualize some of the insights gained from the EDA
- To store the data in AWS or MongoDB: my main motive to store the data in one of these platform is to get hands on either of the platform.
- Implement Machine Learning Model (At this moment I m not sure which machine learning model I'll be implementing. Project updates will be presented on the next due date)



EDA & Visualization



AWS / MongoDB



Machine Learning

By looking at the data set, following insights can be produced using Pyspark (some examples):

1. Mean suicides number by gender.
2. Number of suicides in a given year
3. Suicide number by generation and sex
4. Highest number of suicides in a given Federation
5. Train the model using a Machine Learning Algo.

These are just a rough idea of what I have thought so far. I will be incorporating more ideas to evaluate this data set to my best.

I am also looking forward to use AWS or Mango Db platform to explore data storage techniques. Like, how do we use services provided by these platform to storage and access the data etc..