Machine Learning Engineer Nanodegree

Capstone Proposal

Ahmed Nasser
NOV 9st, 2019

Proposal

Domain Background

According to The World Health Organization (WHO) close to 800 000 people die by suicide every year. Furthermore, for each suicide, there are more than 20 suicide attempts. Suicides and suicide attempts have a ripple effect that impacts on families, friends, colleagues, communities and societies. Suicides are preventable. Much can be done to prevent suicide at individual, community and national levels.

Problem Statement

Suicide Prevention., I want to inspect the reasons behind Suicides and how it evolved through the years and similarities between countries according to those reasons to give some indicator on the right direction to prevent suicides.

Datasets and Inputs

This compiled dataset pulled from four other datasets linked by time and place, and was built to find signals correlated to increased suicide rates among different cohorts globally, across the socio-economic spectrum.

References

United Nations Development Program. (2018). Human development index (HDI). Retrieved from http://hdr.undp.org/en/indicators/137506
World Bank. (2018). World development indicators: GDP (current US$) by country:1985 to 2016. Retrieved from http://databank.worldbank.org/data/source/world-development-indicators#
[Szamil]. (2017). Suicide in the Twenty-First Century [dataset]. Retrieved from https://www.kaggle.com/szamil/suicide-in-the-twenty-first-century/notebook
World Health Organization. (2018). Suicide prevention. Retrieved from http://www.who.int/mental_health/suicide-prevention/en/

Solution Statement

The main objective of the project is to build a model to Cluster the countries to show how the similarities evolved since 1985. Given data about the countries, the model can cluster the countries according to their similarities using K means

Benchmark Model

I will drop the population data at my solution and make a model with it later to test how my clustering goes

Evaluation Metrics

I'll use accuracy score as the Metrics to evaluate the model

Project Design

The project will consist of the following steps:

1. Data processing , Exploration and correlation
2. Cluster analysis using K-means .