

Project Proposal – Final Project

Linear Regression
EGY-World



Business Understanding

Motivation

As a dual citizen of the United States and Egypt, I am strongly connected and passionate for the well being and resources for these nations.

Problem

The issue at hand is we do not know the life expectancy or other measures of health in Egypt. We are unsure of how social, and health related factors has had an affect on the population. This study will help gain a baseline understanding of Egypt's past and current state so we may project how various factors have played a role in people's decline.

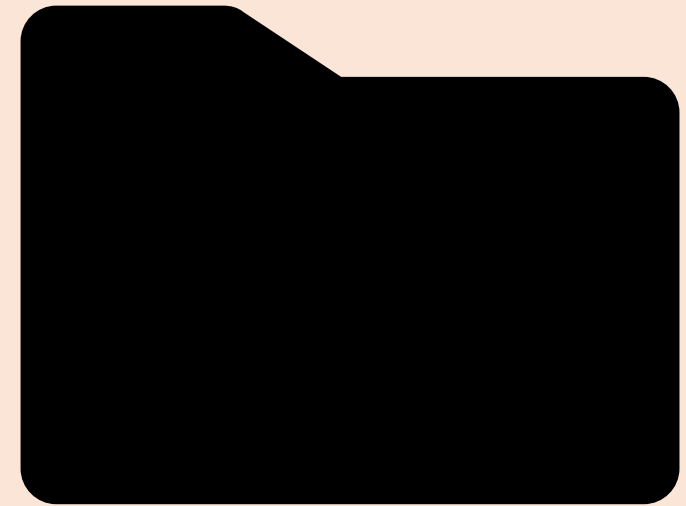
Industry/Realm

Social Development

Health

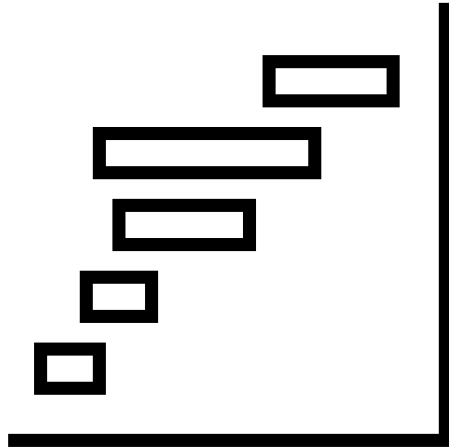
Data Understanding

- The data is sourced from The Global Health Observatory (GHO) data repository under World Health Organization.
- This data is from the time period of 2000 to 2015.
- In a nutshell, this data consists of immunization factors, mortality factors, economic factors, social factors and other health related factors.
- Due to time constraints, this dataset is already prepopulated and found online.
- Features include life expectancy, infant deaths, BMI index, population, HIV/AIDS, schooling, and much more.



Data Preparation

- There are null values in this data, so finding them and understanding how best to tackle will be reviewed.



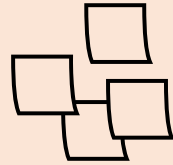
Modeling

- Linear regressions or multiple linear regressions are best suited for this project.
- This is a regression problem due to the nature of the data by years and country.
- The target variable will be life expectancy.



Evaluation

- I will be looking at statistical inferences such as MAE, MSE and RMSE.



Tools/Methodologies

- Linear regression will be the main modeling algorithm concept for this project.
- If time permits logistic regressions will also be utilized.