

Hackverse 2022 – Life Expectancy Prediction

&

Statistical Analysis on factors influencing Life Expectancy

Requirement:

Life Expectancy Prediction: Based on features predict life expectancy of various countries

Problem Statement Context & Scope:

Although there have been lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates. It was found that effect of immunization and human development index was not taken into account in the past. Also, some of the past research was done considering multiple linear regression based on data set of one year for all the countries. Hence, this gives motivation to resolve both the factors stated previously by formulating a regression model based on mixed effects model and multiple linear regression while considering data from a period of 2000 to 2015 for all the countries. Important immunization like Hepatitis B, Polio and Diphtheria will also be considered. In a nutshell, this study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well. Since the observations this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

Content

The Global Health Observatory (GHO) data repository under World Health Organization (WHO) keeps track of the health status as well as many other related factors for all countries. The data-sets are made available to public for the purpose of health data analysis. The data-set related to life expectancy, health factors for many countries has been collected from the same WHO data repository website and its corresponding economic data was collected from United Nation website. Among all categories of health-related factors only those critical factors were chosen which are more representative. It has been observed that in the past 15 years, there has been a

huge development in health sector resulting in improvement of human mortality rates especially in the developing nations in comparison to the past 30 years.

Inspiration

A simple dataset which most of you are familiar with is given in the google classroom

The data-set aims to answer the following key questions:

1. Do all the features available in the dataset really affect the Life expectancy? What are the predicting variables actually affecting the life expectancy?
2. Should a country having a lower life expectancy value(<65) increase its healthcare expenditure in order to improve its average lifespan?
3. How does Infant and Adult mortality rates affect life expectancy?
4. What is the impact of schooling on the lifespan of humans?
5. Does Life Expectancy have positive or negative relationship with drinking alcohol?
6. Do densely populated countries tend to have lower life expectancy?
7. What is the impact of Immunization coverage on life Expectancy?

- **Build life expectancy prediction model**
- **Build a simple UI using either flask or Django or any other**

Dataset: Attached separately in the google classroom. Participants should use only the provided dataset.

Deliverables:

- 1) Solution document for problem statement (Idea/Approach, Design, Execute and Test)
- 2) Details of EDA.
- 3) Solution code in Jupyter notebook (Follow coding standards ie: Title of task, add comments etc..)
- 4) Build at least 2 models (1st model with initial accuracy, 2nd model with improved accuracy)

Good to have

- 1) Add code description in jupyter notebook
- 2) Try models with different algorithms
- 3) Compare models
- 4) Select final model for prediction

- 5) Simple UI to demo your results

Duration: 24 Hours.