

## **Capstone Two**

### **Factors Driving Life Expectancy : Do immunization matters?**

#### **1 Context**

In view of the current pandemic, vaccination does not seem to find the consent of some people in the world. A look at the impact of vaccination on life expectancy could be important to highlight. But also, to identify the other factors which contribute to the improvement of life expectancy in the world. Moreover, It has been observed that in the past 15 years, a huge development in the health sector resulted in improvement of human mortality rates especially in the developing nations in comparison to the past 30 years. 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 dataset related to life expectancy, health factors for 193 countries has been collected as well as its corresponding economic data from 2000-2015. Among all categories of health-related factors, only those critical factors were chosen which are more representative. The final dataset consists of 22 Columns and 2938 rows which means 20 predicting variables. All predicting variables were then divided into several broad categories: Immunization related factors, Mortality factors, Economical factors, and Social factors.

From this project, we are interested in developing a model which predicts life Expectancy based on factors identified. We could break down with the follow questions:

- What is the trend of life expectancy from 2000 to 2015?
- Are immunisation features related to mortality features (as infant-deaths, under five year old deaths, and adult deaths) ?
- What are the factors that impact the trend of life expectancy?
- Do immunisation factors impact the increase of life expectancy as compared to other factors?
- What are the factors that affect the life expectancy model ?
- What was the level of life expectancy over one decade (2005 -2015)? And what were the top 30 countries with a great increase of life expectancy over this decade?
- What is the 95% confidence interval of life expectancy ?
- Is the average mean of life expectancy from developed countries greater than that of developing Countries?

#### **2 Criteria for success**

The model of life expectancy will take into account all the features and data between 2000 and 2015. We will assess how the life expectancy model behaves with regard to immunisation factors as compared to others.

### 3 Scope of solution space

The life expectancy model will focus on immunization factors. It would be also interesting to get a particular look on deaths with regards to immunisation and their implication on the model.

### 4 Constraints within solution space

The dataset has important missing values to handle (table 1). This could impact the model depending on the imputation technique chosen.

Table 1: Overview of missing data of some features of the dataset

Features	Percentage of missing value
Measles	33.458135
infant_deaths	28.863172
under_five_deaths	26.718856
Population	22.191967
percentage_expenditure	20.796460
Hepatitis_B	18.822328
GDP	15.248468
Income_composition_of_resources	10.108918

### 5 Stakeholders to provide key insight

This project is conducted around exchange and discussion with experts from Springboard and mentors.

### 6 Key data sources

<https://www.kaggle.com/kumaraarshi/life-expectancy-who>