World Health Organisation

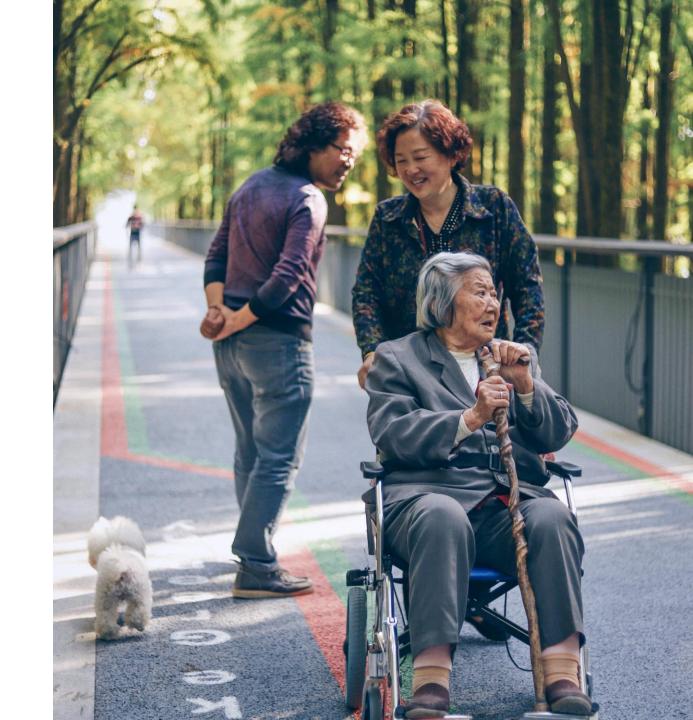## Life Expectancy Measurements

Project Scope & Motivation

Life expectancy is a statistical measurement used to estimate an individual's lifespan.

On an individual level, life expectancy is crucial in determining one's plans, support and care. On a larger group level, it holds paramount socio-economical implications.

On a country level, it is used to derive insights, analytics and further studies to better understand the needs & risk factors of populations.

World Health Organization

Your goal

Your Data Analytics team has been tasked with providing an estimate prediction of life expectancies across countries on the globe.

The data has been provided by WHO, and it contains records between 2000 and 2015 across 183 countries.

Collected data has been omitted in some cases, and as such some of the existing measurements for life expectancy are missing. You may ignore those cases

Your goal

One of the main focuses of this project is data integrity. Several countries have previously expressed concerns with sharing some of their sensitive data, such as medical records, as they brought about unwanted financial implications when correlated with their quality of life measurements and hindered social developments.

"As such, your team has been tasked with <u>*constructing two models*</u>*: One that uses the* <u>*least information*</u> *necessary to make a prediction, as well as a* <u>*more elaborate one*</u> *that can be used if states decide to share more sensitive data."*

Part of your task will be using your judgement as ethical data practitioners to determine which features may/should be used.

Output

You've been asked to produce a function that takes in relevant population statistics (features) and makes a prediction on the average life expectancy.

Your function should also prompt *"Do you consent to using advanced population data, which may include protected information, for better accuracy? (Y/N)"* to decide which model is to be used.

For transparency, we've also employed the services of another contractor team to compare results and cross-validate.

Learning Objectives

————

- **Build two continuous predictive models (linear regression) - One best performing, one minimalistic**

- **Deploy both into an interactive function**

- **Pay attention to ethical considerations**

- **Deliverables:**
  - **File(s) containing your code/work**
  - **Single file containing your function**

We've managed to get an early insight into our competitors. The other team has successfully generated a semi-robust model with an RMSE of 2

We should consider that as our baseline and beat it if we want to secure further collaboration with WHO.