



WORLD LIFE EXPECTANCY

JIAHUI BONG, CELINE TAN, RHEA KENNETH
Tutorial Group : DSF3



TABLE OF CONTENTS

01.

Introduction

02.

Data Cleaning

03.

EDA + ML (Global)

04.

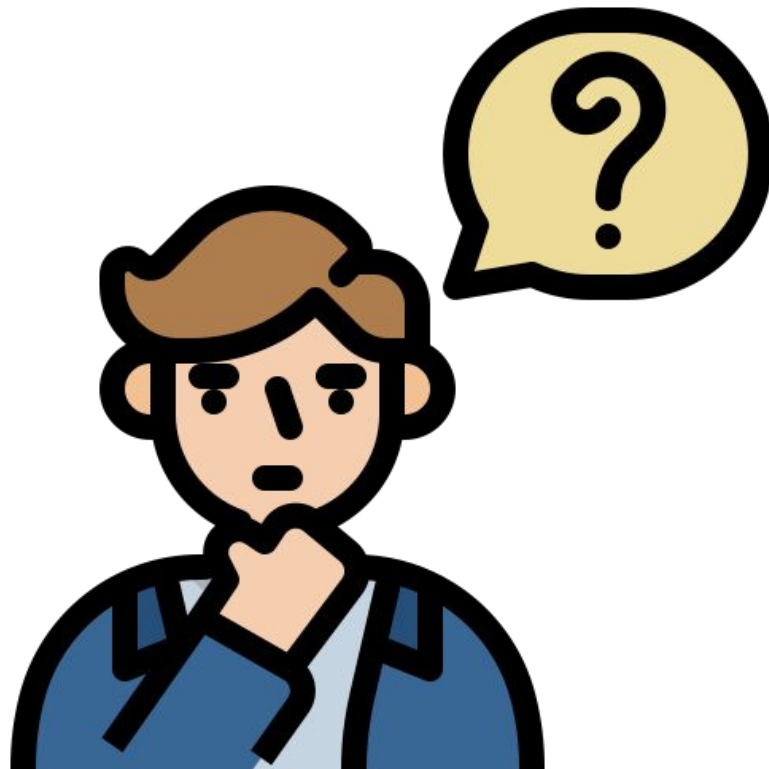
EDA + ML (Regional)

05.

Conclusion

01

Introduction





Motivation

- Stereotype Thinking : Developed countries have higher life expectancy while less developed countries have lower life expectancy
- Eg. Africans have lower life expectancy due to malnutrition or other diseases
- What factors contribute to such inequality?



Dataset Used



MMATTSON · UPDATED 2 YEARS AGO

▲ 18

New Notebook

Download (245 kB)




WHO national life expectancy

Subset of factors influencing average life expectancy

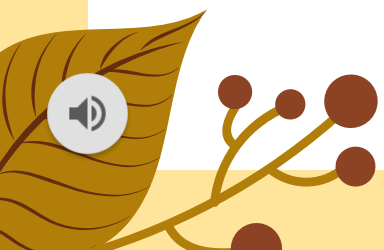


- Dataset from kaggle
- Made up of GHO (Global Health Observatory) and UNESCO (United Nations Educational Scientific and Culture Organization) Information





Does life expectancy inequality really exists and if yes, what are the indicators of such scenarios and are there possible solutions to narrow the inequality?

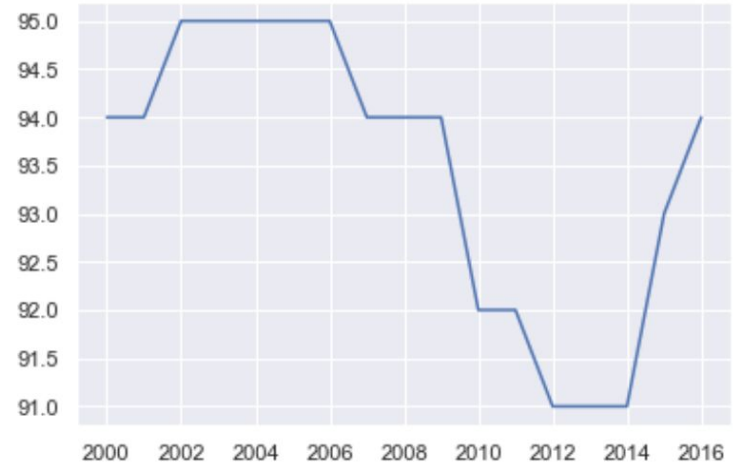
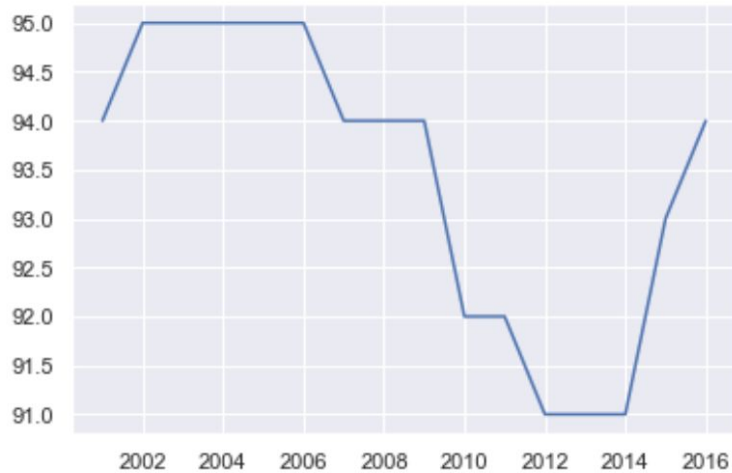


02

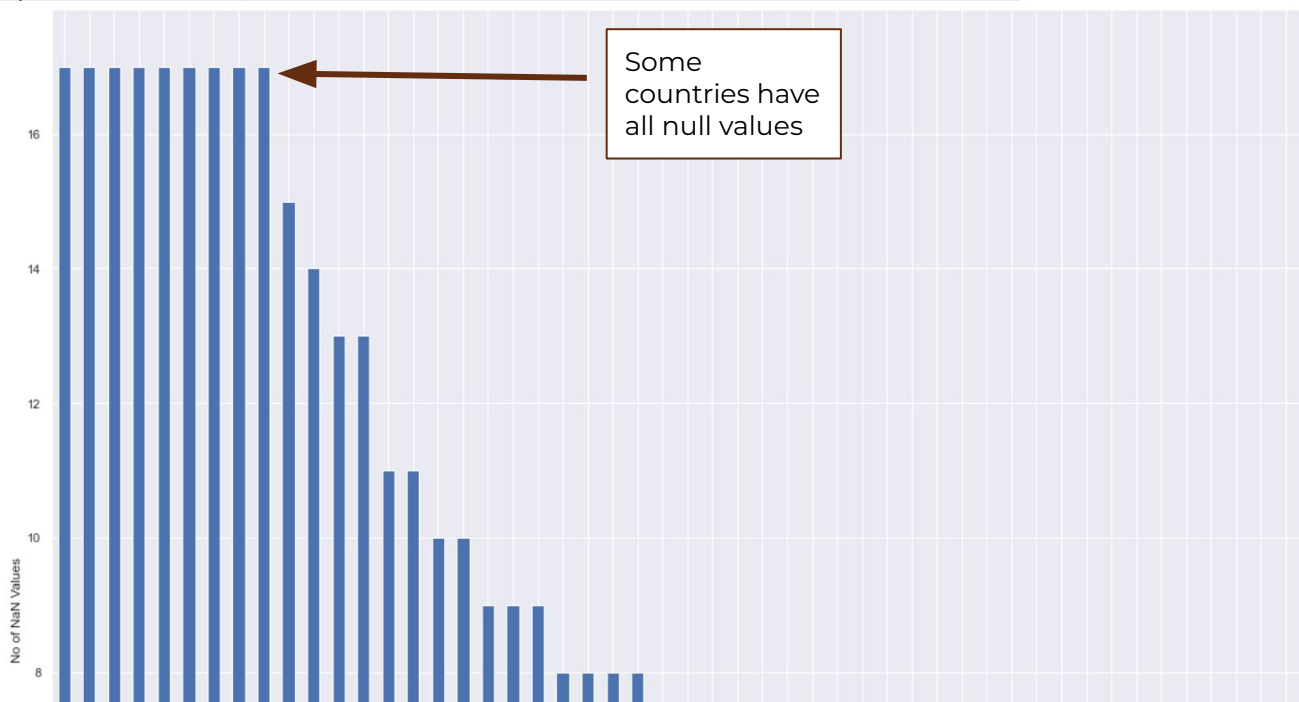
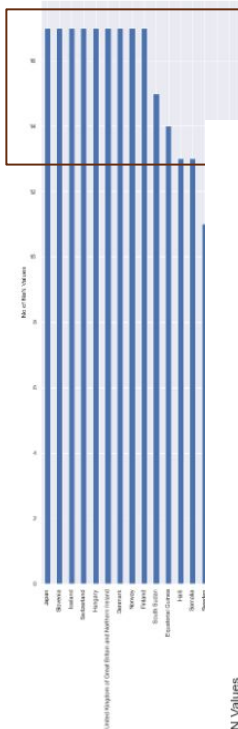
Data Cleaning



Linear Interpolation Method



Bar Plot of Number of Null Values vs Countries





Fill_na method

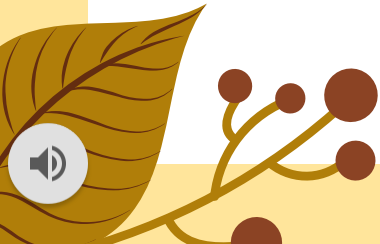
- Datas are in time series
- Use linear interpolation method to fill in the null values
- If there are more than 10 years of values are missing then will leave that country's data unfilled (which will be removed later after EDA)





03

Global Life Expectancy

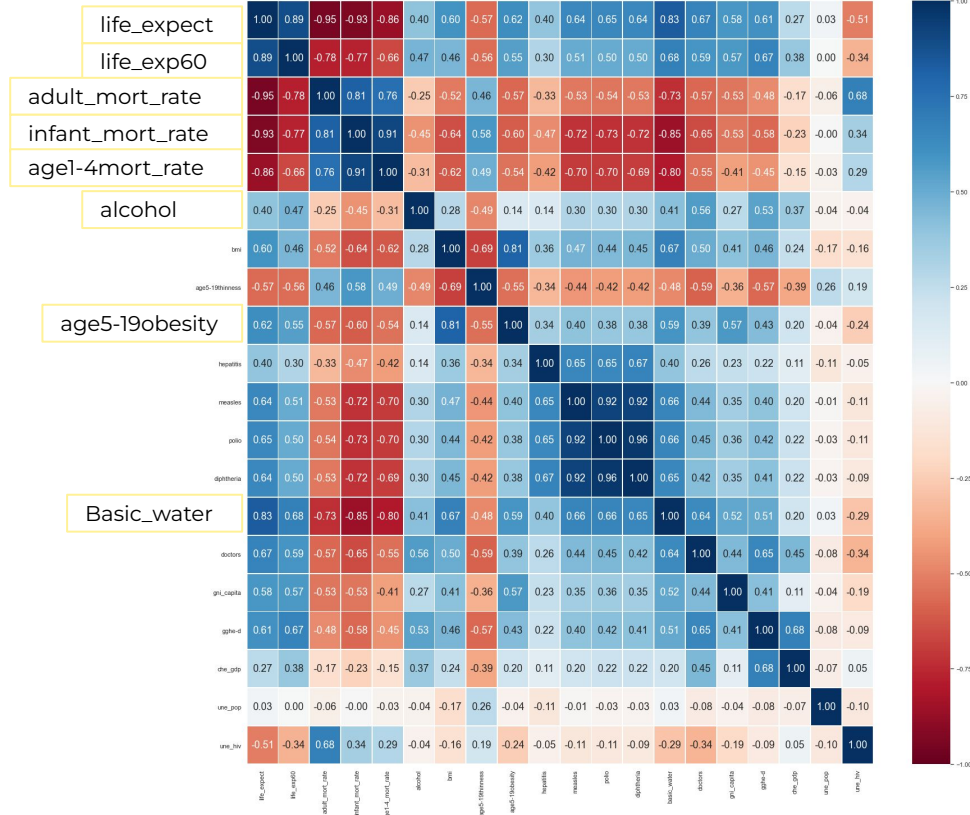


Exploratory data analysis

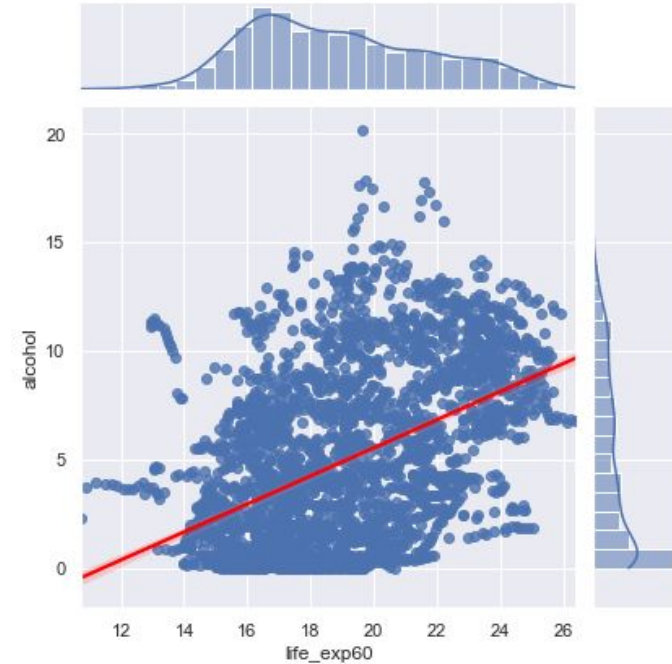
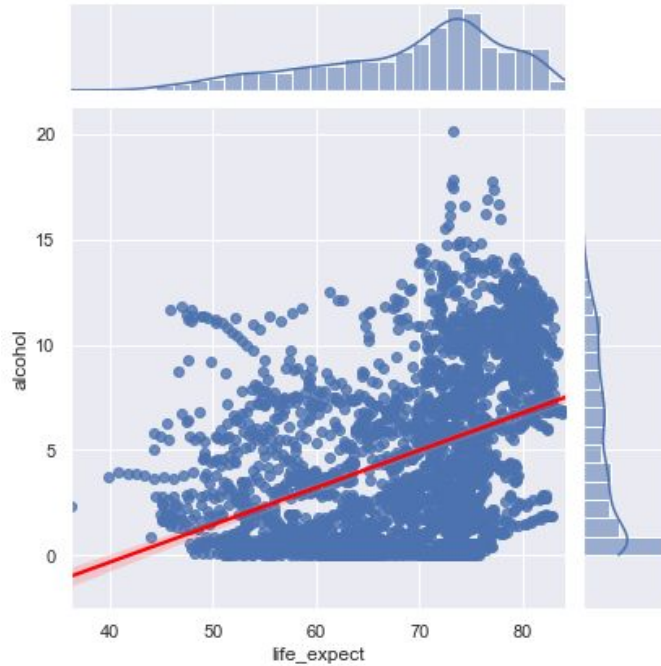
- 1) Heatmap
- 2) Insights
- 3) Estimated top predictors of life expectancy



Heatmap



Alcohol



Median of alcohol
consumption

4.02 liters per capita



Insights



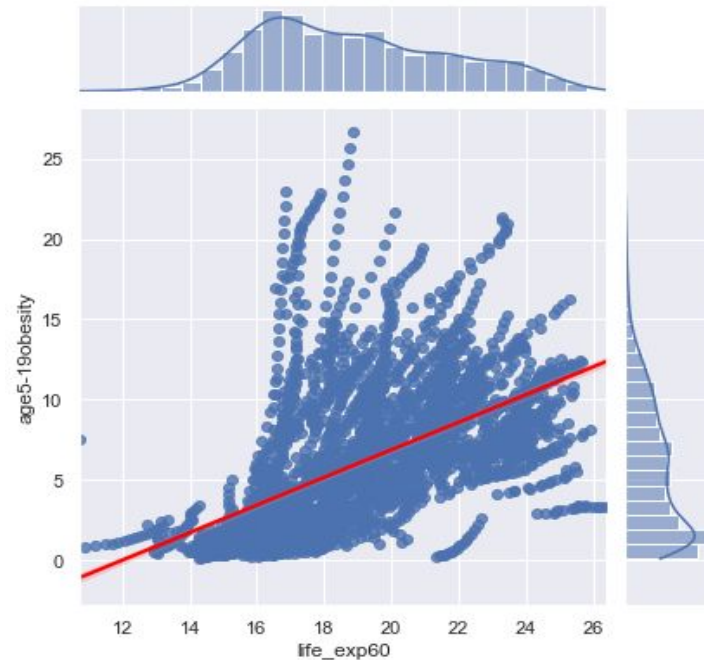
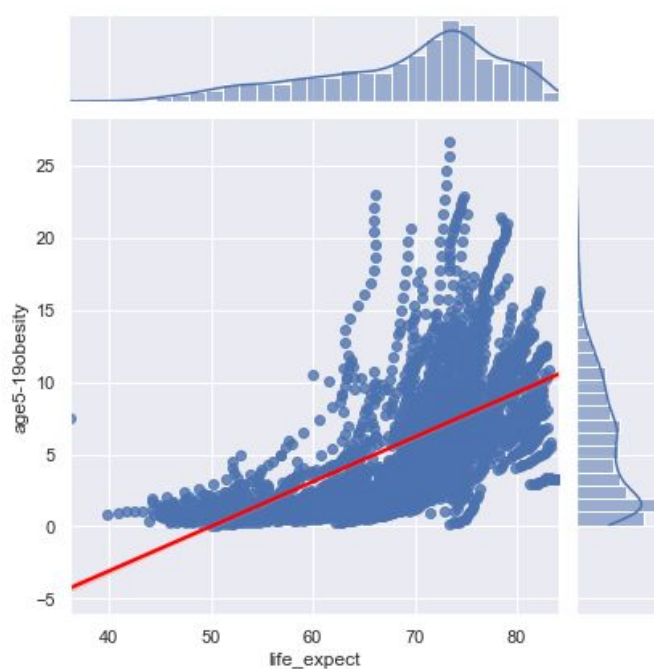
Expectation

Alcohol is associated with health issues and has a **negative relation** with life expectancy

Reality

Alcohol in **moderation** can have beneficial health effects leading to longevity

Obesity rates from ages 5-19



Median of obesity rate: 5.20 %

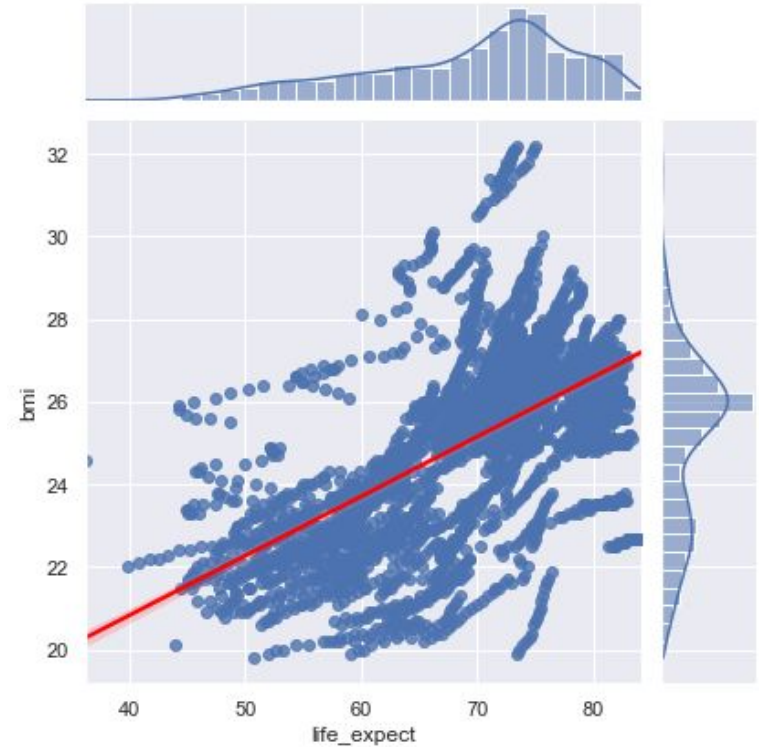
Insights

Expectation

Obesity is considered a health issue and is expected to have a **negative relation** with life expectancy

Reality

- 1) The **BMI** of those **>30** makes up only about **5.97%** of population **on average**.
- 2) BMI also indicates **muscle mass** and **not only body fat**



Machine Learning

- 1) Multivariate Linear Regression
- 2) Random Forest Regressor
- 3) Gradient Boosting Regressor



RESPONSE : Life Expectancy At Birth

Train Data

Score	Linear Regression	Random Forest	Gradient Boosting
R-squared value (R^2)	0.986	0.999	0.999
Mean Squared Error (MSE)	1.140	0.056	0.028

Test Data

Score	Linear Regression	Random Forest	Gradient Boosting
R-squared value (R^2)	0.986	0.995	0.996
Mean Squared Error (MSE)	1.160	0.361	0.316



RESPONSE : Life Expectancy At Birth

Train Data

Score	Linear Regression	Random Forest	Gradient Boosting
R-squared value (R^2)	0.986	0.999	0.999
Mean Squared Error (MSE)	1.140	0.056	0.028

Test Data

Score	Linear Regression	Random Forest	Gradient Boosting
R-squared value (R^2)	0.986	0.995	0.996
Mean Squared Error (MSE)	1.160	0.361	0.316



Comparing MSE for train and test data

Response = life_expect



RESPONSE : Life Expectancy At Age 60

Train Data

Score	Linear Regression	Random Forest	Gradient Boosting
R-squared value (R^2)	0.824	0.990	0.999
Mean Squared Error (MSE)	1.482	0.082	0.001

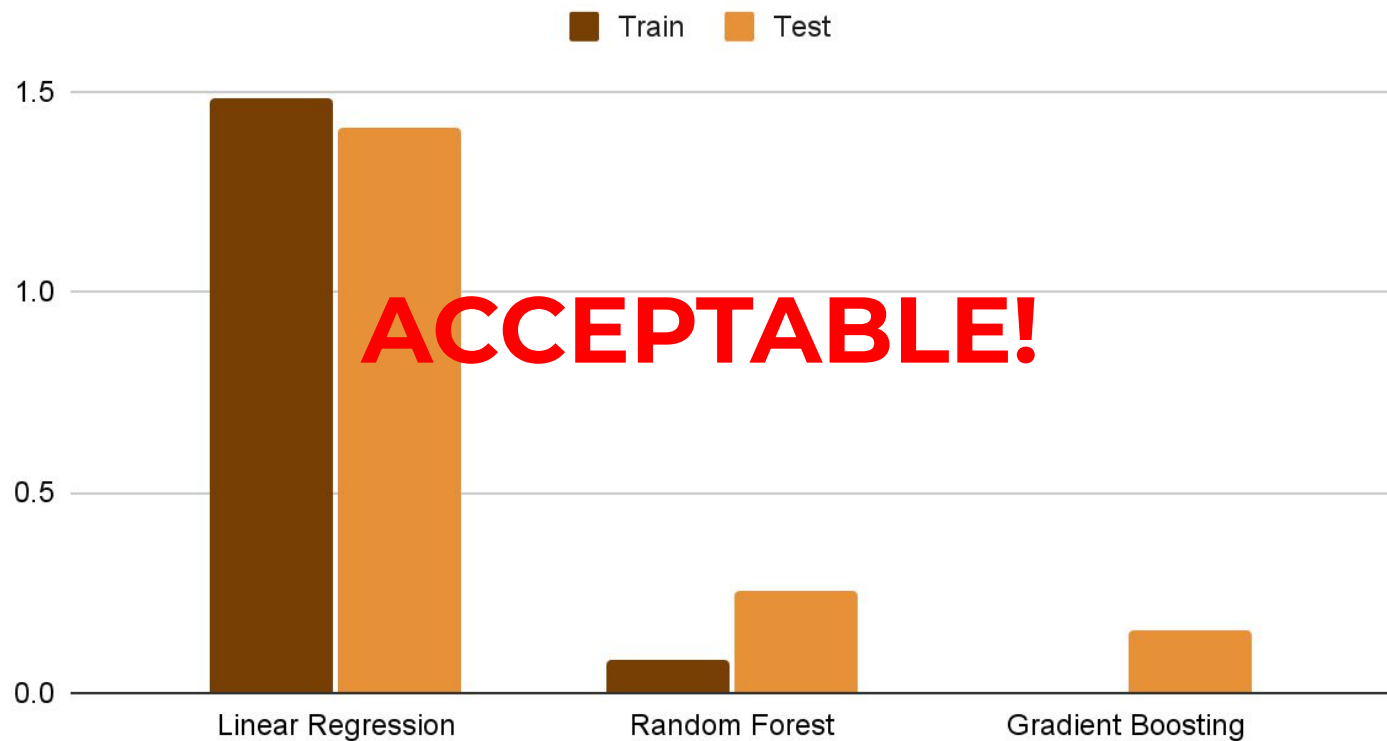
Test Data

Score	Linear Regression	Random Forest	Gradient Boosting
R-squared value (R^2)	0.848	0.972	0.983
Mean Squared Error (MSE)	1.412	0.254	0.159



Comparing MSE for train and test data

Response = life_expect60



Top Predictors

Top 3 predictors for each
response variable inferred from
GBR



Gradient Boosting Regressor

Life Expectancy At Birth

Age 1-4 Mortality Rate



Adult Mortality Rate



Infant Mortality Rate



Life Expectancy At Age 60

Adult Mortality Rate



Age 5-19 Thinness



Region (Americas)



04 Regional Life Expectancy

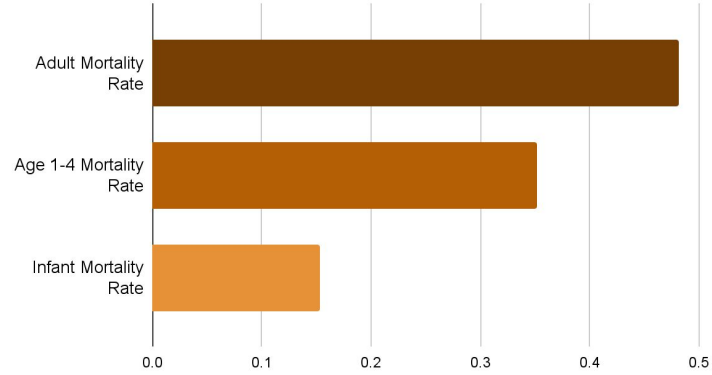


Exploratory data analysis

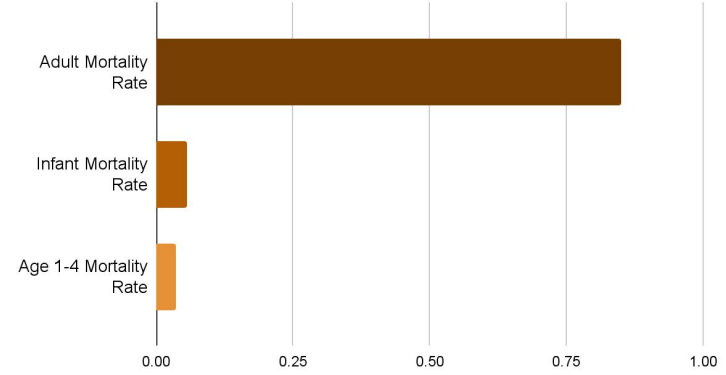
Top predictors for each region's life expectancy at birth inferred from GBR



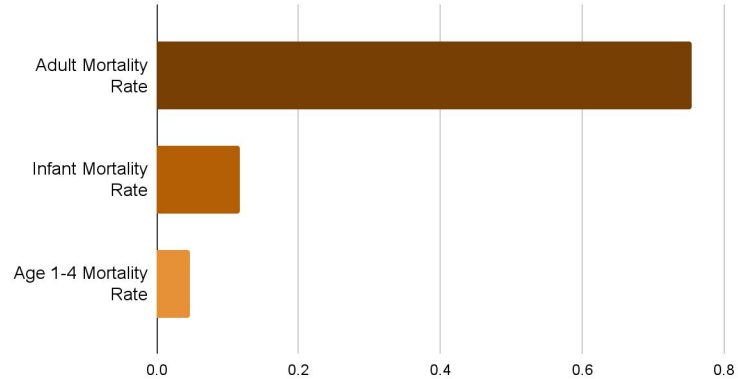
Top Predictors for Africa



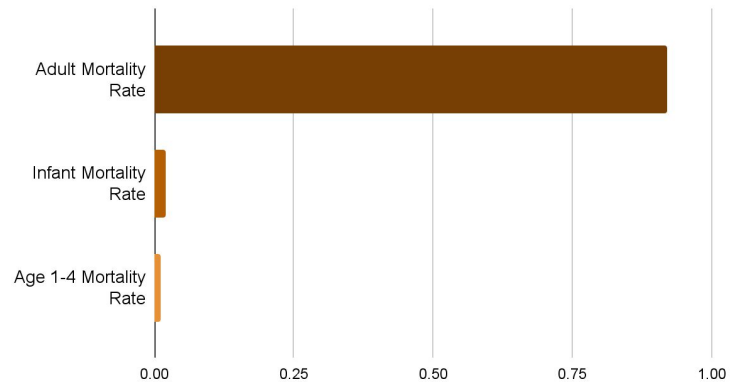
Top Predictors for Europe



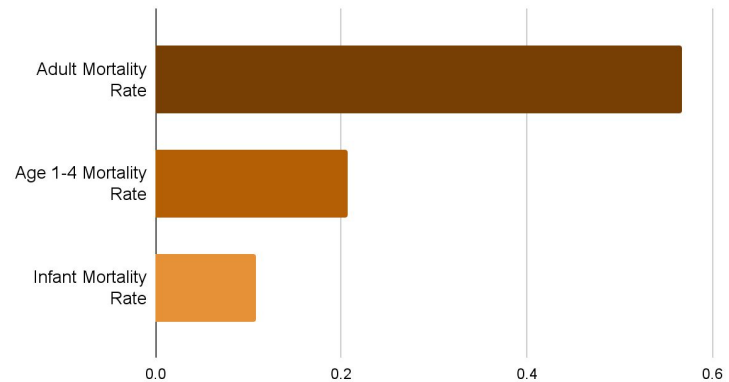
Top Predictors for Western Pacific



Top Predictors for Americas

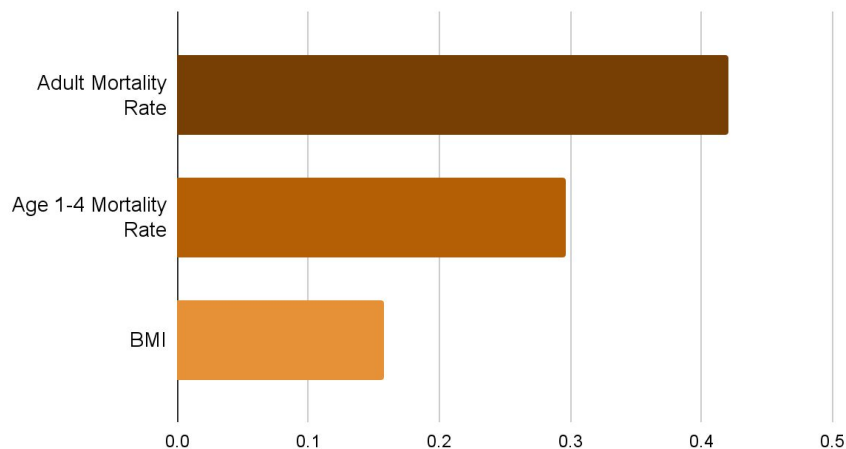


Top Predictors for Eastern Mediterranean



EASTERN MEDITERRANEAN

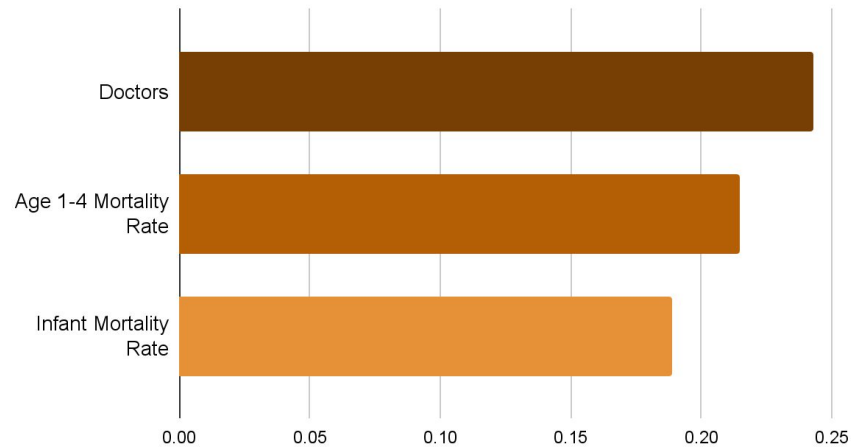
Top Predictors for Eastern Mediterranean



Score	Train Data	Test Data
R-squared value (R ²)	0.999	0.991
Mean Squared Error (MSE)	0.019	0.305

SOUTH-EAST ASIA

Top Predictors for SEA



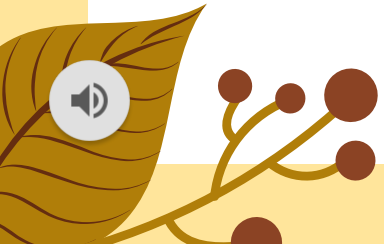
Score	Train Data	Test Data
R-squared value (R^2)	0.999	0.930
Mean Squared Error (MSE)	0.004	0.643



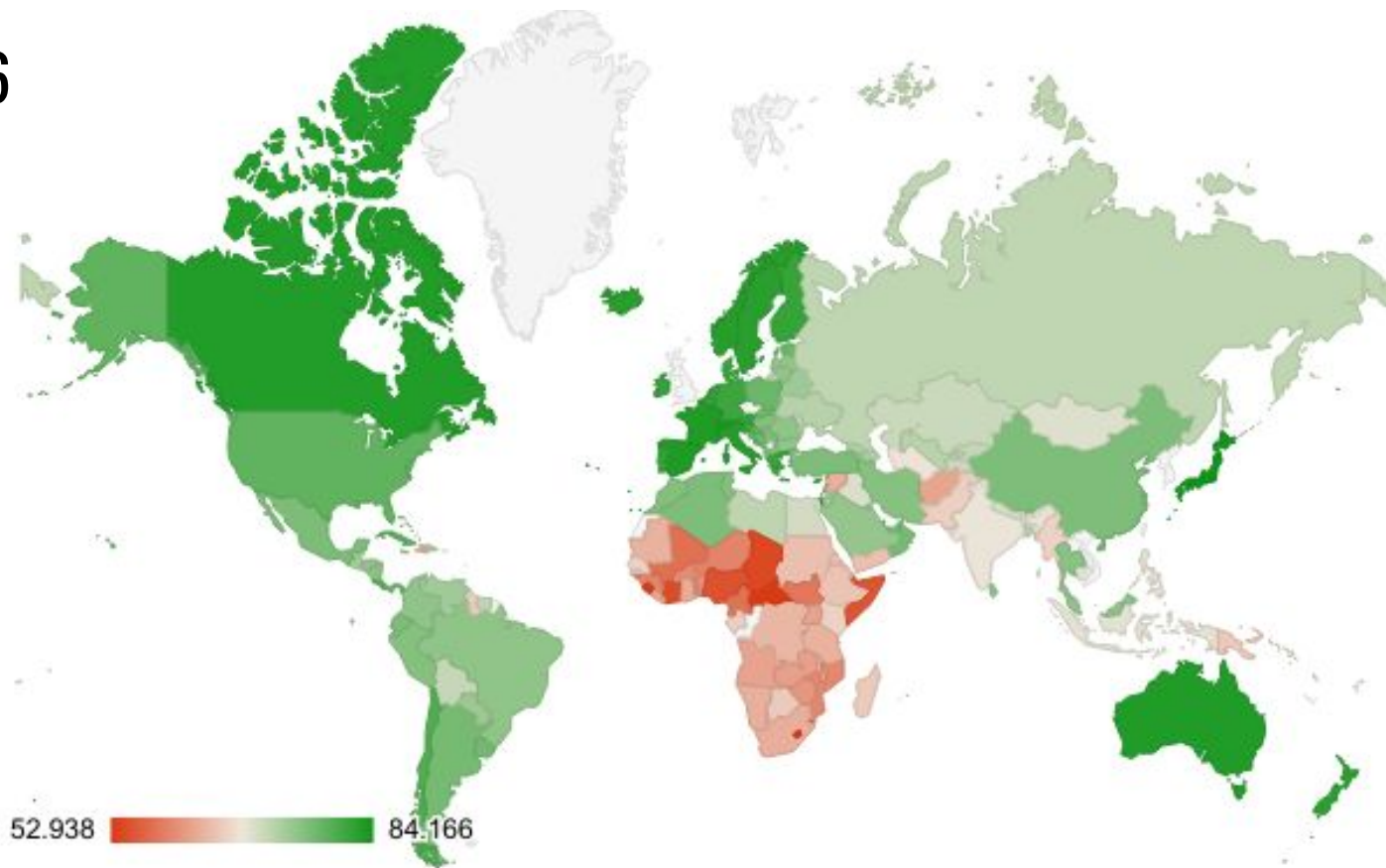


05

Conclusion



2016



52.938 84.166

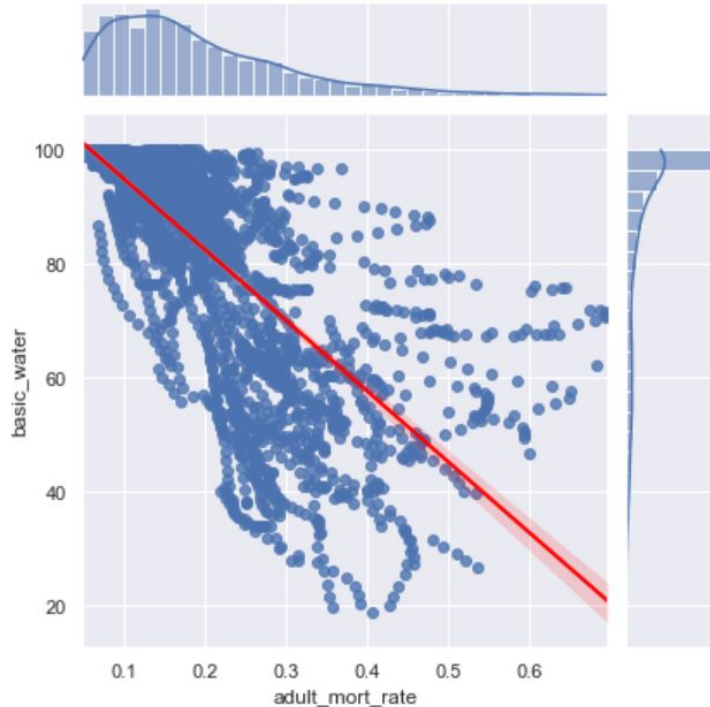


Indicators affecting life expectancy

- Mortality rates are common top predictors across regions
- High negative correlation between life expectancy vs mortality rate of different group age
- Decrease mortality rate but **HOW?**



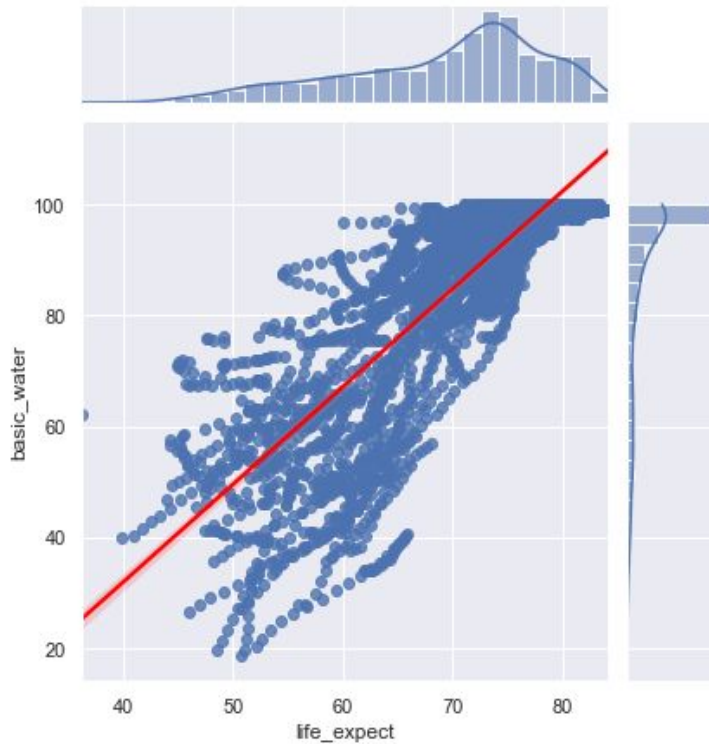
Adult Mortality Rate vs Basic Water



- High negative correlation
- Rationale : Water acts as solvent to transport nutrients to body cells



Basic water

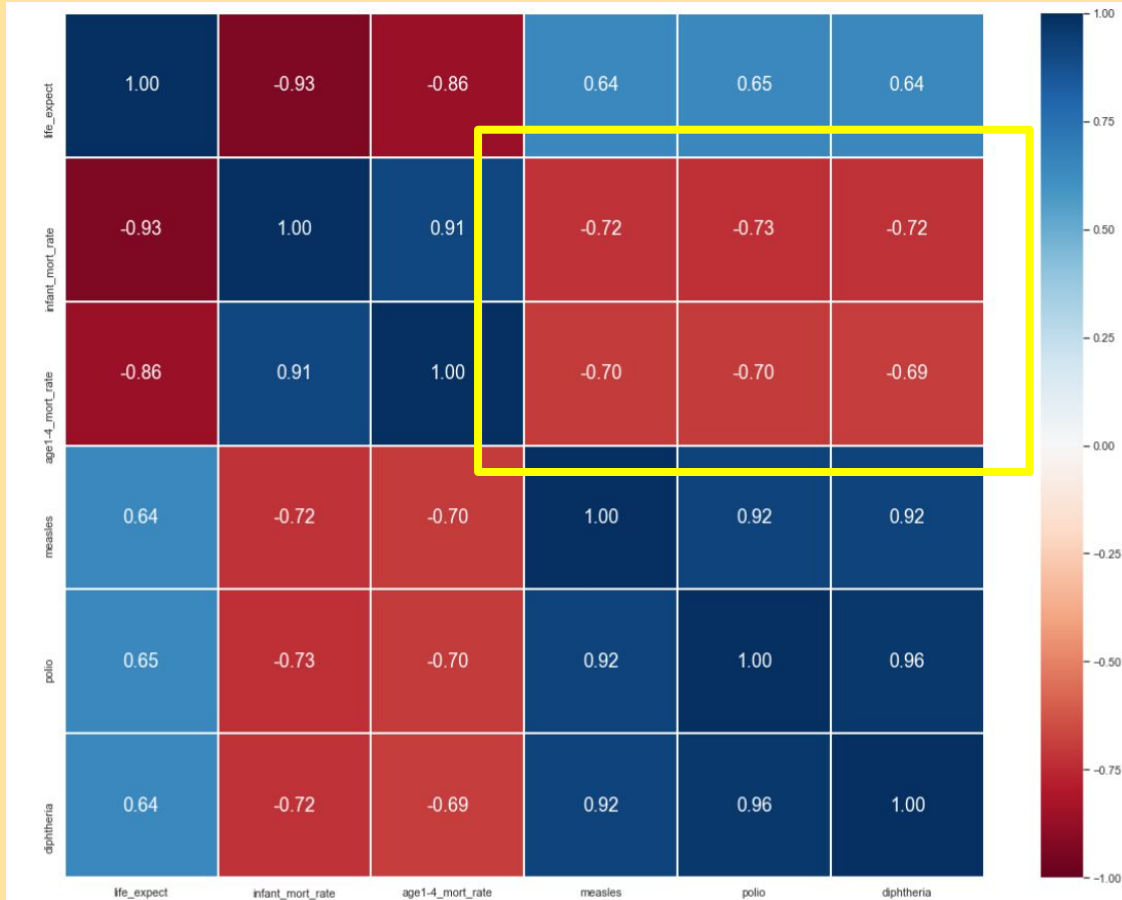


Correlation with
life_exp:

0.83



Correlation between child mortality rate and immunization



→ High negative correlation between child mortality rate vs immunization

→ Children are at high risk for Measles, Polio and Diphtheria



Possible Implementation

Goal

To improve life expectancy in less developed country and narrow the inequality gap



Water Infrastructure

- Government should allocate more budget to water infrastructure



Vaccination

- Raise awareness on the importance of immunisation in children

Limitation of Proposed Solution



- Sudden drop in life expectancy in **Haiti** in **2010**
- Caused by a **7.0 magnitude earthquake**
- **NO** possible implementation when low life expectancy is caused by **unpredictable circumstances**
- Need to take **geographic factors** into consideration

THANK YOU!

