



IMPACT OF SOCIOECONOMIC FACTORS ON LIFE EXPECTANCY

PRESENTED BY
Makenna Barton, Nivedita Arvind,
Saman Ateeq

PROBLEM STATEMENT

Based on an initial analyses of Life Expectancy data provided by the WHO, there is an observable difference between the mean life expectancies of individuals between Developed and Developing countries (**78 years for the former and 67 for the latter**), and an observable difference in immunization coverage. We hope to predict life expectancy and analyze whether there is a statistically significant relationship between Immunization coverage and life expectancy based on correlation analysis and will conduct a difference in means two sample Z test.

HYPOTHESIS

H0: Immunization Coverage has no impact on Life Expectancy

H1: There is a statistical significance between immunization coverage and Life Expectancy.

CI: 95%

Dataset info after
handling missing values

data.info()

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 2938 entries, 0 to 2937  
Data columns (total 22 columns):
```

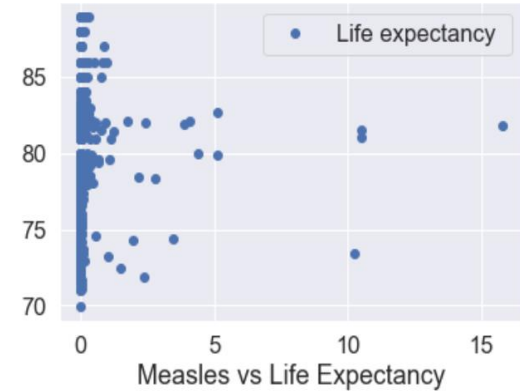
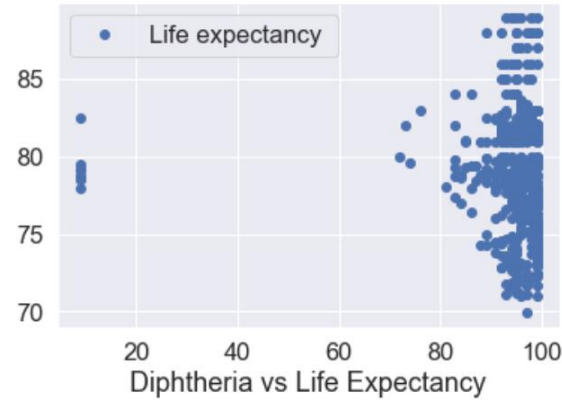
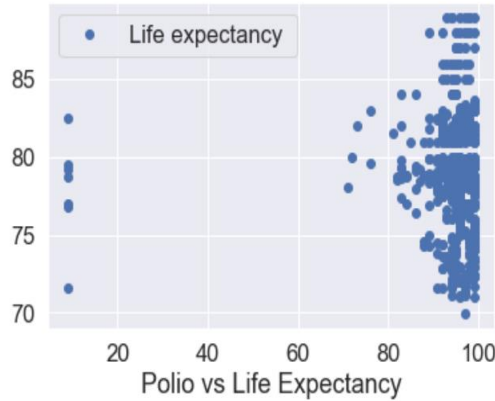
#	Column	Non-Null Count	Dtype
0	Country	2938 non-null	int32
1	Year	2938 non-null	int64
2	Status	2938 non-null	int32
3	Life expectancy	2938 non-null	float64
4	Adult Mortality	2938 non-null	float64
5	infant deaths	2938 non-null	int64
6	Alcohol	2938 non-null	float64
7	percentage expenditure	2938 non-null	float64
8	Hepatitis B	2938 non-null	float64
9	Measles	2938 non-null	int64
10	BMI	2938 non-null	float64
11	under-five deaths	2938 non-null	int64
12	Polio	2938 non-null	float64
13	Total expenditure	2938 non-null	float64
14	Diphtheria	2938 non-null	float64
15	HIV/AIDS	2938 non-null	float64
16	GDP	2938 non-null	float64
17	Population	2938 non-null	float64
18	thinness 1-19 years	2938 non-null	float64
19	thinness 5-9 years	2938 non-null	float64
20	Income composition of resources	2938 non-null	float64
21	Schooling	2938 non-null	float64

Dataset Details

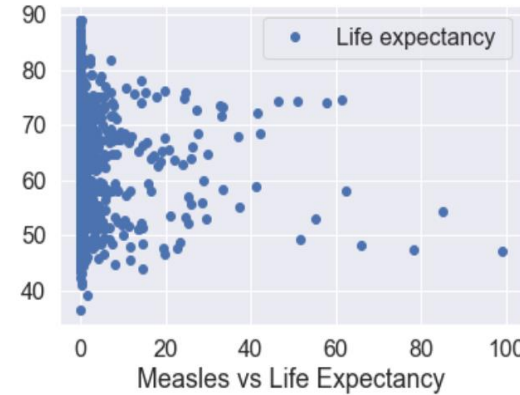
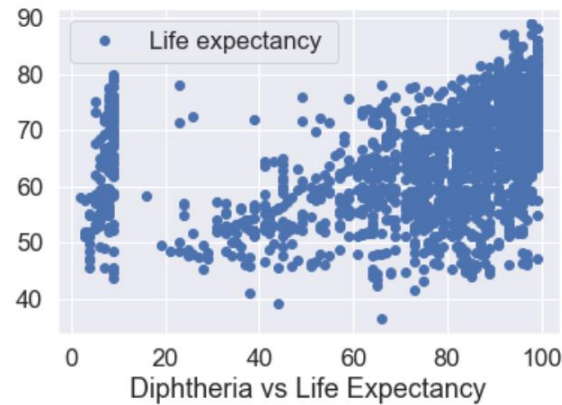
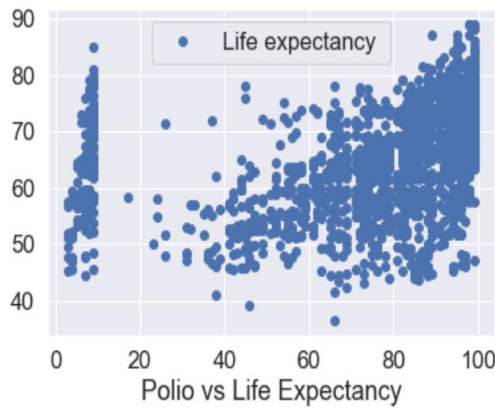
Link	WHO (https://www.kaggle.com/kumarajarshi/life-expectancy-who)
Total Instances	2938
Total Attributes	22
Target Variable	Life Expectancy
Variables of Interest	Status (Developed or Developing) Hepatitis B (Immunization % coverage among 1-year-olds) Polio (Immunization % coverage among 1-year-olds) Diphtheria (Immunization % coverage among 1-year-olds) Measles (Number of reported cases per 1000 population)

Data Exploration (Scatter Plot)

Developed Countries

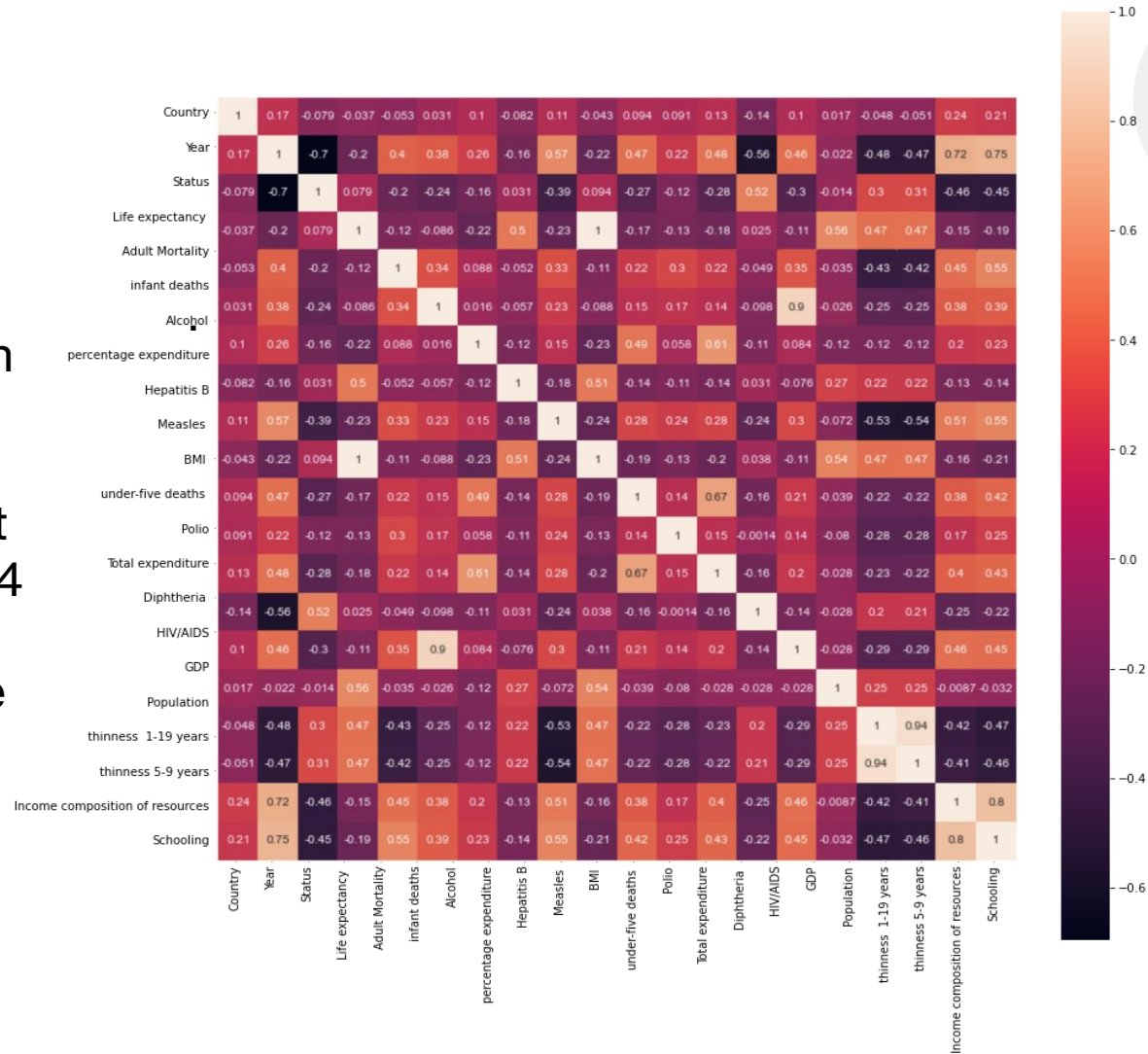


Developing Countries



Incorporated Data Science Concepts and Techniques

- Model Selection: Decision Tree Regression*, Logistic Regression**, or Linear Regression Model
- Correlation Analysis: Heat Map
- Principal Component Analysis: at 0.75 (7 components) and 0.95 (14 components) thresholds
- p-value Evaluation and Inference



* R2: 0.74

Mean Absolute Error: 4.6803659975488925
 Mean Squared Error: 48.09977438137236
 Root Mean Squared Error: 6.93540008805349

**Precision & Recall Score: 89%

Thank you!

Questions?

