Data Analytics Bootcamp
Project 1

Explanatory data analysis of life-expectancy based on health and development indicators per region

Group 1:
Gwendoline Grenu
Peter Lehel
Temitope Ajose-Adeogun
Farheen Oomatia

Table of Content

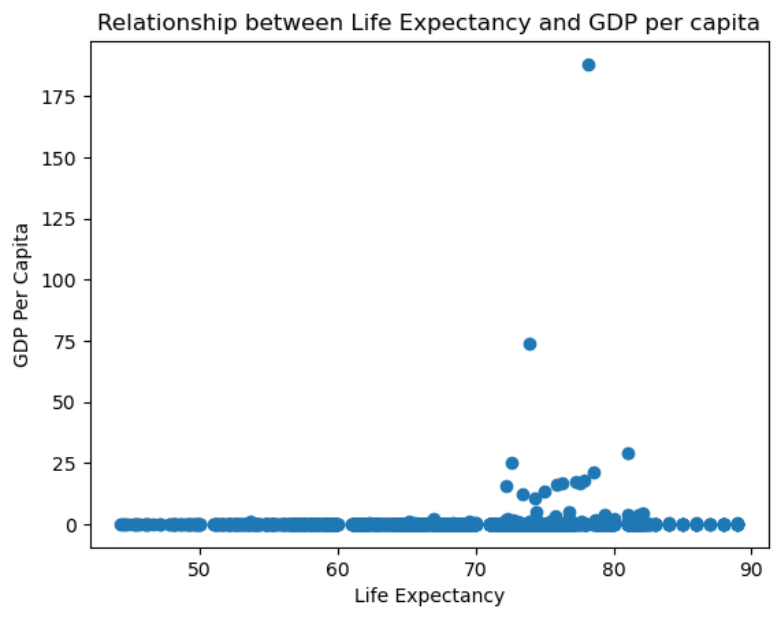
1. Aim of the project - Peter
2. Relationship between GDP and Life Expectancy across the entire dataset – Temi
3. Data on diseases analysed for Asia & Africa– Gwen
4. Life expectancy in the American regions – Farheen
5. Limitations and Conclusion

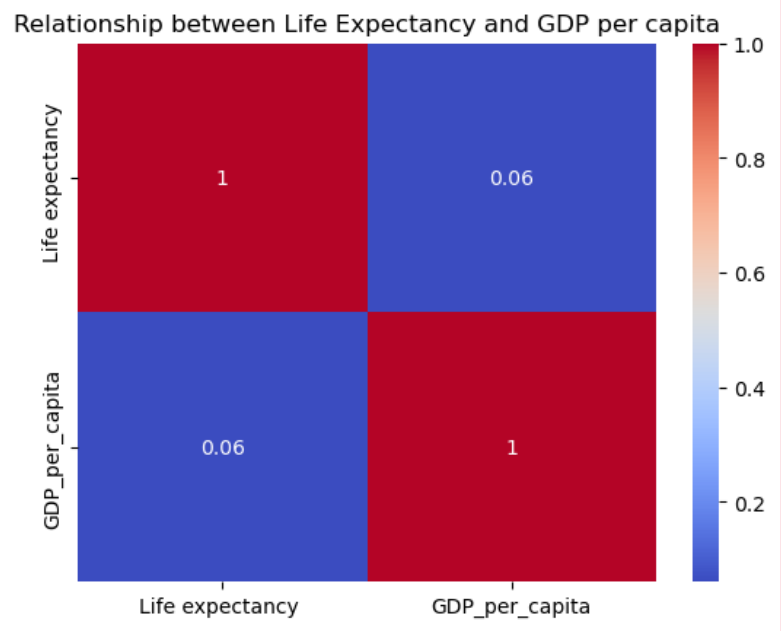
Aim of the project

- Dataset overview – which indicators are we looking at
- Overview of life expectancy change over a decade
- Correlation of GDP, life expectancy and health factors for selected regions
- Life expectancy comparisons within each continent

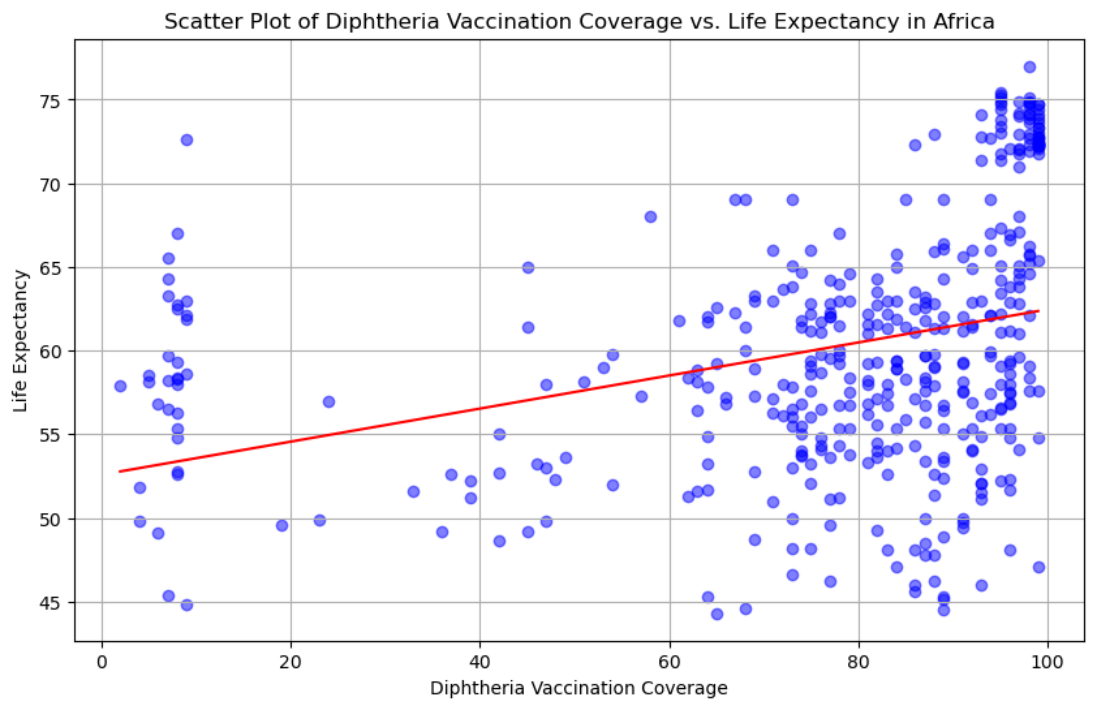
Relationship between Life Expectancy & GDP per capita across the entire dataset

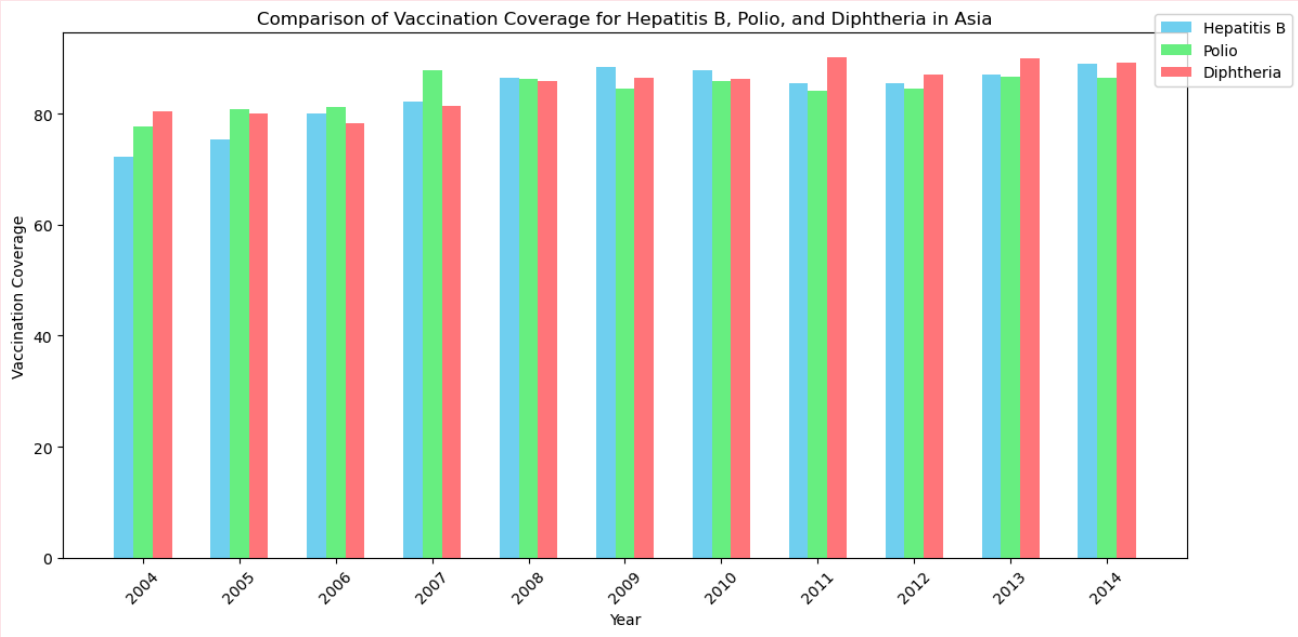
- Life expectancy is mostly between 70-80 years.
- There is a weak positive correlation between life expectancy & GDP per capita in the dataset. This means that although both life expectancy and GDP per capita tend to go up in response to one another, the relationship is not very strong.
- This suggests that as income per household rises in these countries, there is a rise, albeit slight, in the life expectancy of the country's inhabitants.
- GDP per capita increases the life expectancy through increasing economic growth and development in a country and thus leads to the prolongation of longevity.

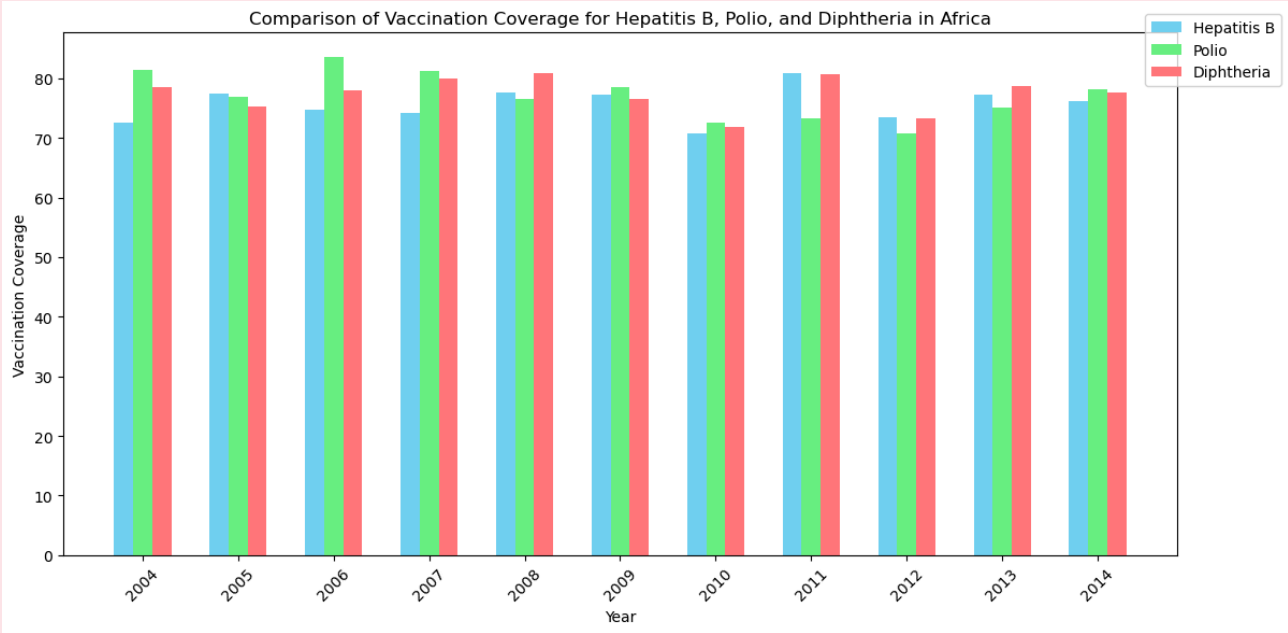




Comparison of vaccination coverage for Hepatitis B, Polio and Diphtheria for Asia and Africa



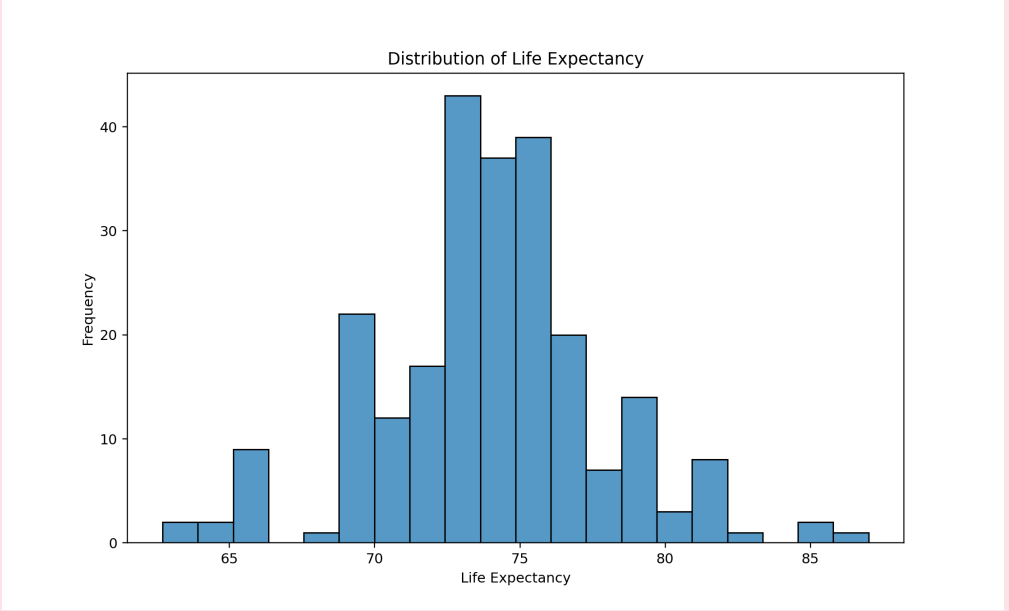




Life expectancy overview across the Caribbean, North, Central and South America

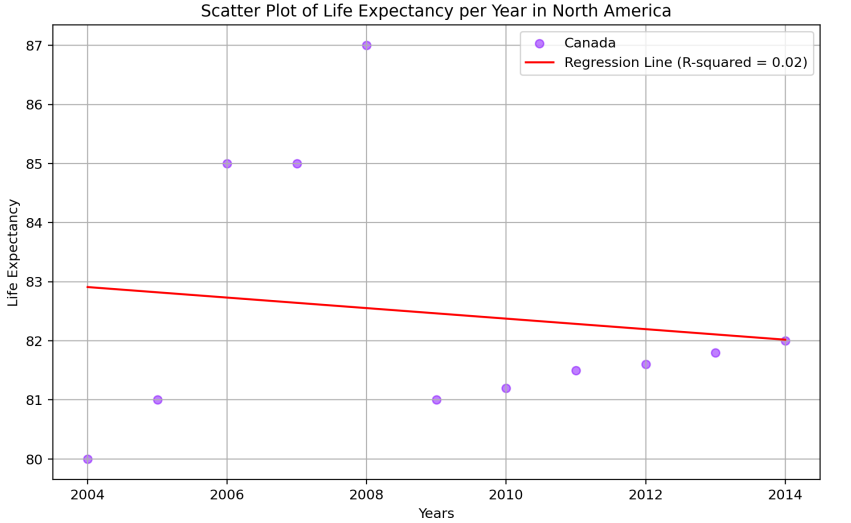
1. Life expectancy is around mostly at around 71-76 years

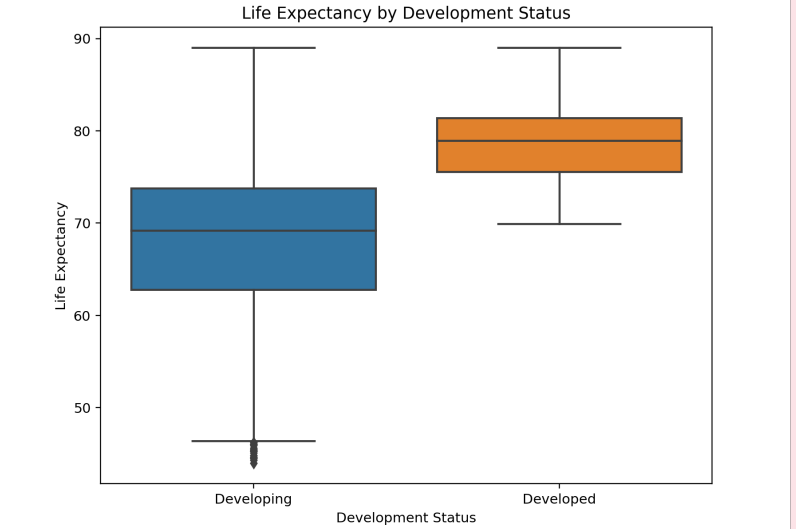
2. As a general trend, across all countries in the Caribbean, Central and South America, the life expectancy rose between 2004 and 2014. Only North America's regressions line shows a decline.
- only Canada, no other country within the region
- unexplained spikes in life expectancy skewed the graph



3. Life expectancy overview across developing and developed countries -> developing countries show a wider distribution of LE compared with developed countries

4. More outliers probably due to infant death





Limitations and Conclusion

- Missing data from some countries
- Missing USA from North America
- Some indicators were not analysed due to loss of resource
- Dataset accuracy - eg. Country status not as expected
- Reliability of source
- Political and environmental factors missing which might explain unexpected spikes or drops in data

- Resources:

https://www.kaggle.com/datasets/arunjangir245/life-expectancy-data