Emissions and Life Expectancy

Looking into the relationship
between emissions per capita and
life expectancy

This project set about looking into the relationship between countries **emissions per capita and life expectancy**

Project Goal

To look at relationships between emissions and life expectancy. I was interested to see if there is a correlation between the two. Increased emissions over the years can suggest a country is developing and therefore such things as education and medical care would likely improve.

Key Objectives

To explore:

- a country's average life expectancy as emission levels increase.
- Correlation between emissions and life expectancy.
- Time after the initial increase in a country's emissions and a rise in life expectancy.
- Any decline in the rate of increasing life expectancy or a decrease once emission levels get to a certain amount per capita?
- How emissions and life expectancy compare across the world?

Datasets

https://www.kaggle.com/datasets/ulrikthygepedersen/life-expectancy

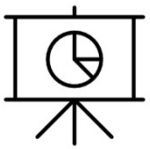
https://www.kaggle.com/datasets/thedevastator/global-fossil-co2-emissions-by-country-2002-2022

https://www.statista.com/statistics/1040159/life-expectancy-united-kingdom-all-time/#statisticContainer

Tools

Skills & Procedures

-Exploratory analysis through visualizations (scatterplots, correlation heatmaps, pair plots and categorical plots)
- Geospatial analysis using a shapefile
- Regression analysis
- Cluster analysis
- Time-series analysis
- Analysis narrative and final results



Preparation

- Sourcing data

- Conduct exploratory visual analysis using relevant Python libraries

- Define questions to explore based on understanding of what the data contains

Analysis

- Source a shapefile containing location data that corresponds to the location data
- Data cleaning and wrangling in Python
- Creating a data profile of summary statistics
- Conduct a geospatial analysis by creating a choropleth map using Python libraries
- Create visualizations using Python
- Prepare data for a regression analysis
- Split data into a training set and a test set
- Run a linear regression on the data and analyse the model performance statistics
- Prepare data for a cluster analysis
- Use the elbow technique to determine the optimal number of clusters
- Run the k-means algorithm
- Source time-series data relevant to project data via an API
- Conduct a Dickey-Fuller test and plot autocorrelations to test for stationarity.
- Perform differencing to stationarize non-stationary data
- Create dashboard of findings on Tableau

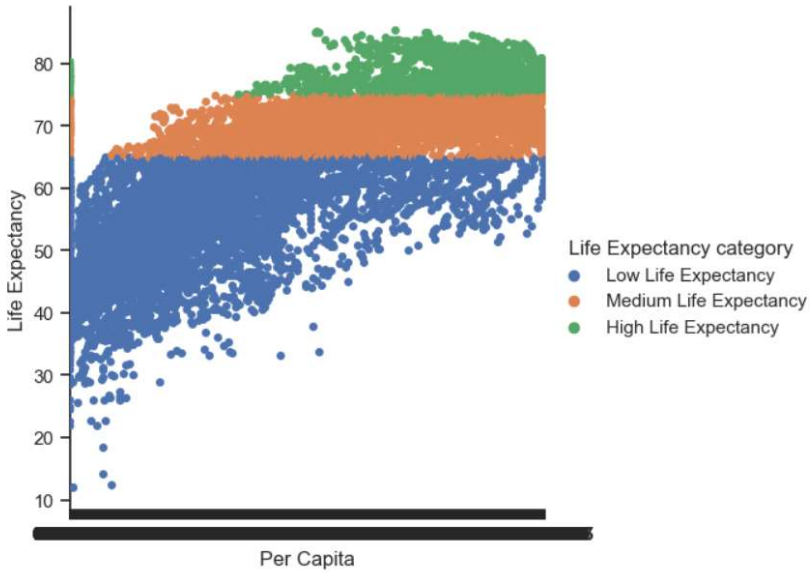
Results & Recommendations

- To further this analysis it would be suggested to look into data on a more individual basis, as I did with the UK, from developed, emerging markets, and developing countries.

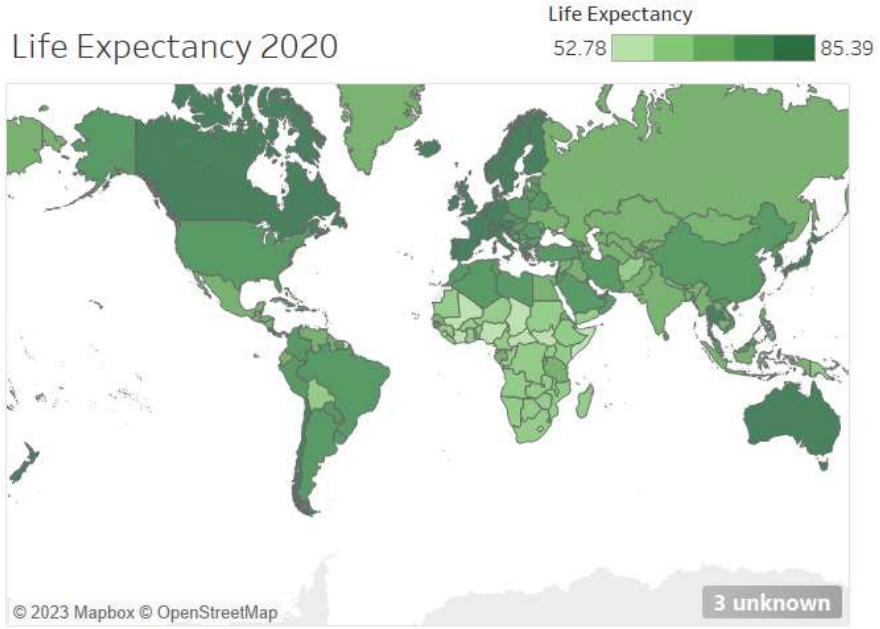
- Once having analysed a handful of countries from each different category, it will be interesting to see what patterns emerge regarding life expectancy and emissions per capita. For developed countries there is data out there for life expectancy going back 200 years such as that used for the UK in this analysis.

- As the world moves towards more green energy, it might be of benefit to not only look at emissions but also productivity or GDP as extra markers.

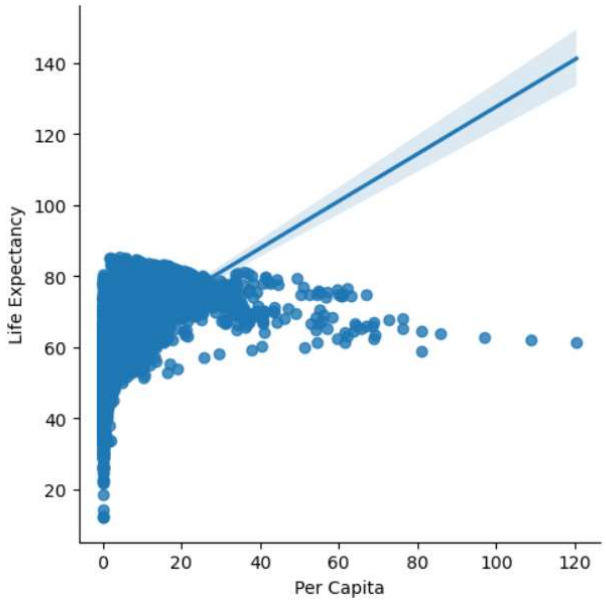
VISUALISATIONS



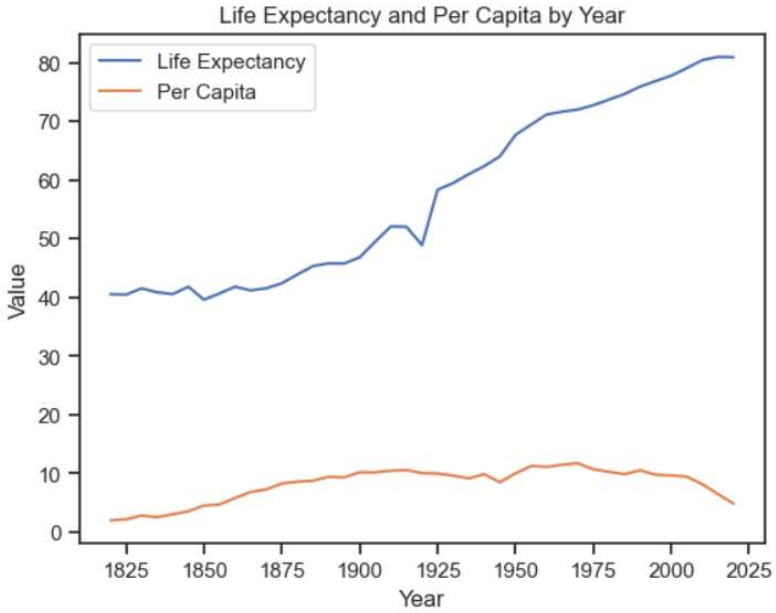
In this categorical plot, the most dense collection of High Life Expectancy appears to coincide with the highest emissions, and the largest collection of Low Life Expectancy with low emissions per capita. Interestingly, it appears that the highest life expectancies are in the mid-range of emissions per capita, but only marginally.



VISUALISATIONS



This scatterplot shows a positive association between the variables, but not a strong relationship, and the data is non-linear



This is an interesting representation of life expectancy and emissions per capita over time in the UK. They both increase simultaneously, emissions level off in the 1900's whilst life expectancy continues to rise apart from a dip following ww1, possibly due to improved infrastructure and living conditions, healthcare etc, by 2020 life expectancy is at is highest level ever, whilst emissions per capita have reduced to their lowest levels since the 1800's. The y-axis is representing age in years and metric tons of emissions.

Additional Sources

Storyboard
https://public.tableau.com/views/EmissionsandLifeExpectancy/Story1?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

Analysis and Visualisations
https://github.com/danfradat/EmissionsLifeExpectancy

Python code
https://github.com/danfradat/EmissionsLifeExpectancy