

CONTENT:

- Project Title:Exploratory Data Analysis and Predictive Modeling of Life Expectancy
- Name: Solomon Okechukwu Egbe
- Course Name:IBM Data Science
- Date:Friday August 29, 2025

This project analyzes global life expectancy trends using the Gapminder dataset, applies SQL for summary statistics, visualizes patterns with plots and maps, and predicts life expectancy from GDP using regression

Executive Summary

Short summary of the project objectives, methods, and key findings

This project explores the relationship between GDP and life expectancy for countries worldwide. Using data visualization, SQL analysis, interactive mapping, and predictive modeling, we identified patterns across continents, highlighted countries with high and low life expectancy, and built a regression model to predict life expectancy based on GDP.

Introduction

Content

Dataset: Gapminder

Key variables: country, continent, year, lifeExp, pop, gdpPercap

Objective: Analyze life expectancy trends and predict life expectancy from GDP.

We explore patterns in life expectancy and GDP across continents and years.

The Gapminder dataset contains country-level data from 1952 to 2007. The goal is to analyze trends in life expectancy and understand how GDP influences it, while also using SQL and visualization techniques for insights

Data Collection & Wrangling

Content:

Steps:

- Load Gapminder dataset using Plotly Express (px.data.gapminder())
- Cleaned and filtered data for analysis
- Created subsets for continents, years, and maximum/minimum values

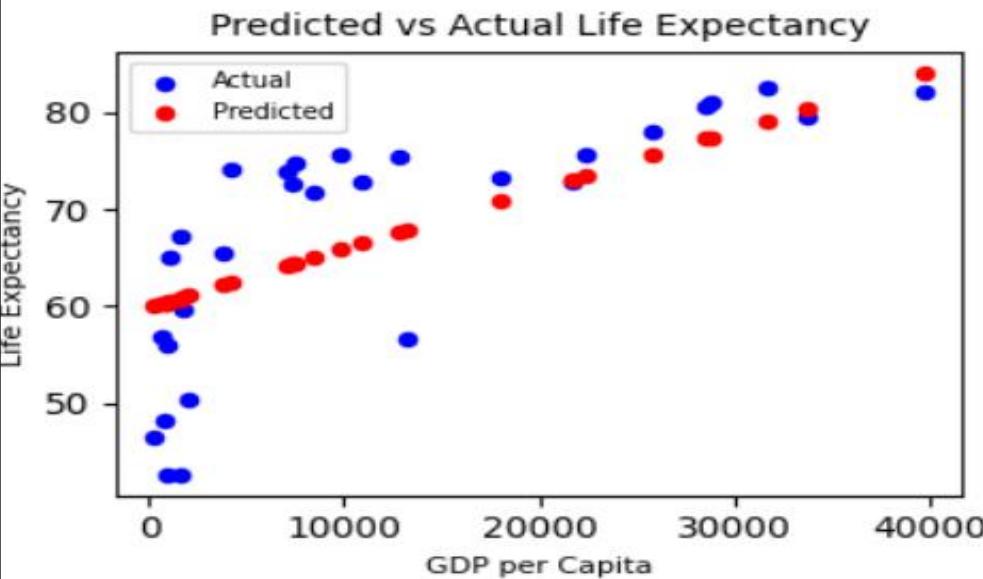
The dataset was loaded and cleaned in Python using Pandas. Key columns were selected for analysis, and subsets were created for SQL queries, EDA visualizations, and predictive modeling.

EDA Interactive Analysis Methodology

We performed Exploratory Data Analysis (EDA) on the Gapminder dataset to understand patterns in life expectancy, GDP per capita, and population across continents and countries. Interactive analysis was implemented using Plotly and Folium to allow dynamic visualization of trends and geographical patterns. We explored relationships between GDP and life expectancy, identified continent-level differences, and highlighted countries with extreme values.

Predictive Analysis Methodology

We applied a linear regression model to predict life expectancy based on GDP per capita. The model was trained on the Gapminder dataset, allowing us to estimate life expectancy for countries given their economic indicators. The predictive approach helps identify patterns and potential outliers where actual life expectancy differs from model predictions.

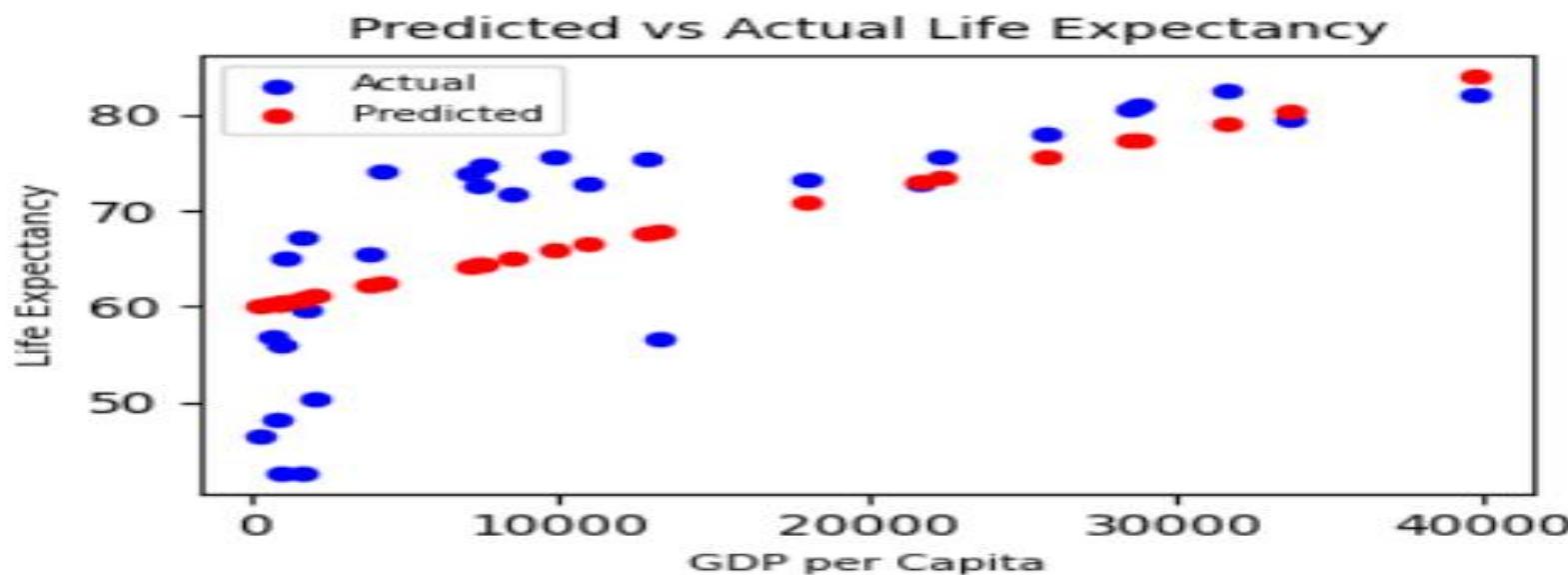


```
avg_lifeExp_per_continent = pd.read_sql_query(query, conn)  
avg_lifeExp_per_continent
```

continent	avg_lifeExp	Icon
0 Oceania	74.33	
1 Europe	71.90	
2 Americas	64.66	
3 Asia	60.06	
4 Africa	48.87	

EDA Visualization-Scatter Plot

The scatter plot shows the relationship between GDP and life expectancy for each continent. Larger population countries are represented with bigger dots, showing trends and outliers across regions.



EDA WITH SQL QUERIES RESULT

SQL queries helped identify countries with the highest and lowest average life expectancy per year and continent. This allowed us to observe regional differences and extreme values in the dataset.

	year	max_lifeExp		continent	avg_gdp
0	1952	72.67	0	Africa	2193.75
1	1957	73.47	1	Americas	7136.11
2	1962	73.68	2	Asia	7902.15
3	1967	74.16	3	Europe	14469.48
4	1972	74.72	4	Oceania	18621.61

Interactive Map (Folium)

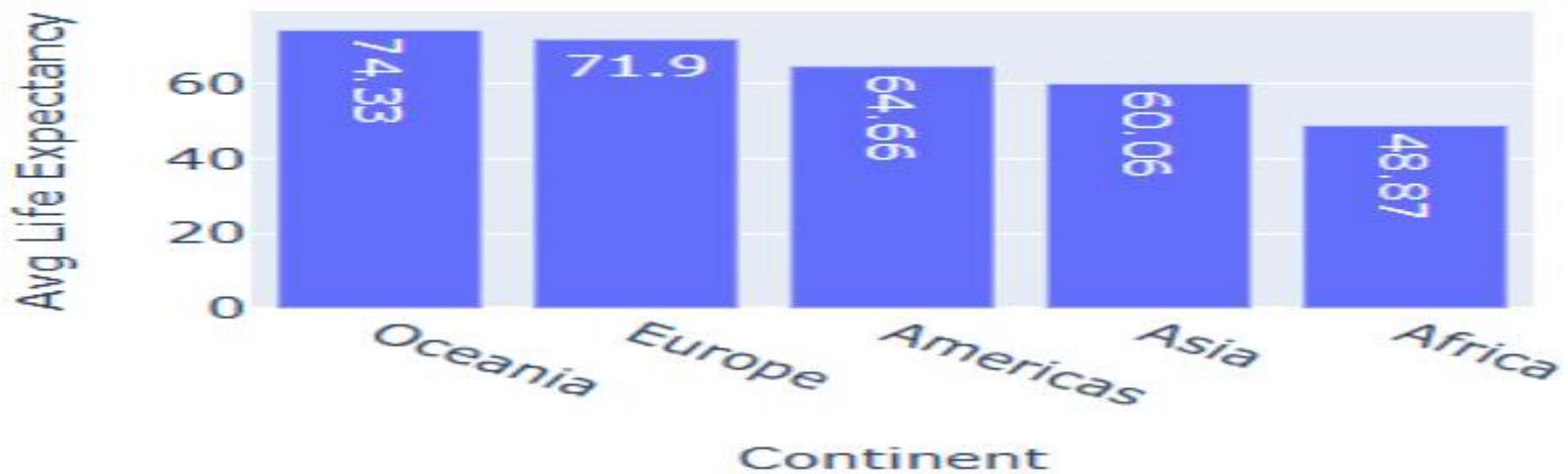
The interactive map shows countries with marker size representing population. Hovering over markers displays country names and life expectancy, providing a visual overview of population distribution



PLOTLY DASH RELATED SLIDES

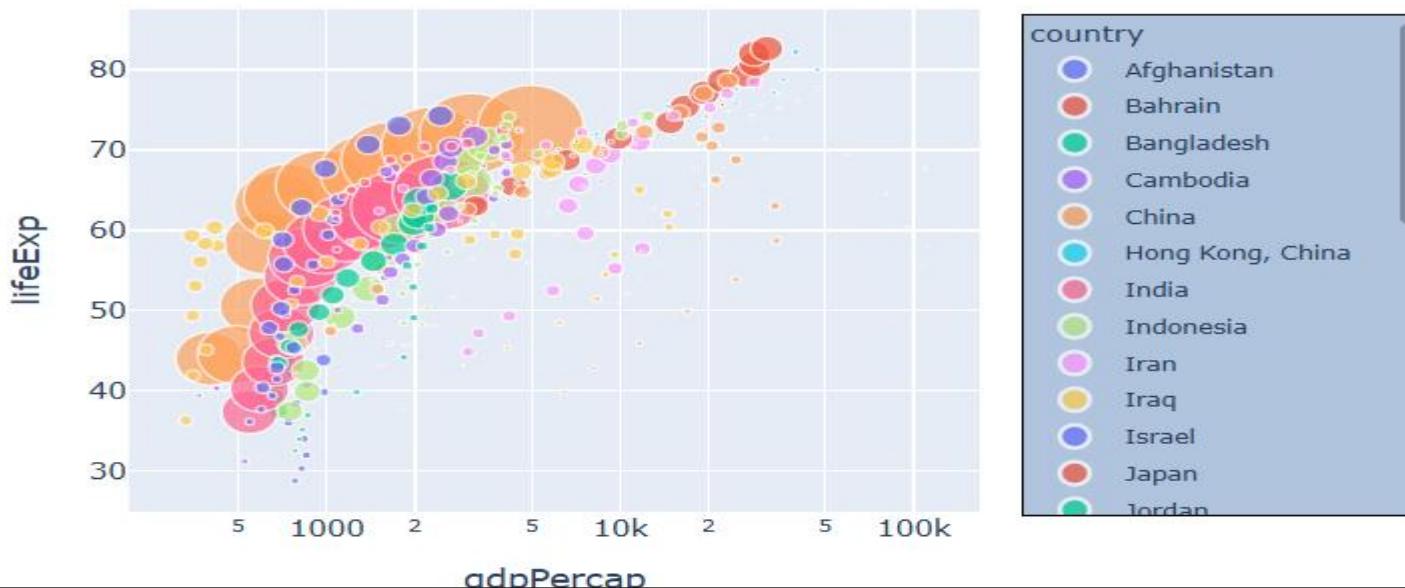
We built an interactive dashboard using Plotly Dash to allow users to explore the Gapminder dataset dynamically. Features include a dropdown to select continents, interactive scatter plots, and real-time updates of visualizations. This dashboard provides a user-friendly interface to observe trends and compare countries across different metrics and their life expectancy.

Average Life Expectancy per Continent



Predictive Analysis: Life Expectancy from GDP

A linear regression model was built to predict life expectancy based on GDP per capita. The plot shows a positive correlation: countries with higher GDP generally have higher life expectancy



Conclusion

In conclusion, GDP is a strong indicator of life expectancy across countries. EDA and predictive modeling reveal global health trends, while interactive visualization enhances understanding of population distribution and disparities

Creativity & Innovative Insights

Additional insights include highlighting countries with exceptional improvements in life expectancy, creating multi-column tables for better readability, and designing compact charts for effective slide presentation.