**Team Name : Hack Pack**

**Life Expectancy Analysis Report**

**Introduction**

This report presents a comprehensive analysis of global life expectancy statistics using data from the UN, World Bank Group, World Health Organization, and the CIA World Factbook. The analysis highlights life expectancy trends by age, sex, and over time, revealing disparities and improvements among countries. Additionally, the concept of Healthy Life Expectancy (HALE) is discussed.

**1. Data Sources Overview**

* **Table 1:** UN estimates of life expectancy at different ages (2023)
* **Table 2:** Changes in life expectancy from 2019 to 2023
* **Table 3:** World Bank Group historical trends (2014–2022)
* **Table 4:** World Health Organization (2019) life expectancy and HALE
* **Table 5:** CIA World Factbook (2022) by sex and sex gap
* **Table 6:** Life expectancy estimates in the OCED for 2022

### **2. Project Overview**

This project focuses on analyzing life expectancy data for various countries, with emphasis on differences between males and females across time and age groups. The goal is to understand:

* Gender-based longevity gaps
* Time-based changes in life expectancy
* Trends across different geographic regions

Data will be scraped and compiled from **Wikipedia**, cleaned using Python and regular expressions, analyzed statistically, visualized using data plots, and stored in a **MongoDB** database. The project concludes with an optional Streamlit dashboard to display interactive insights.

### **3. Project Objectives**

We aim to answer questions like:

* Which countries have the **highest** and **lowest** life expectancy?
* Where is the **male-female life expectancy gap** largest or smallest?
* How has life expectancy **changed from 2019 to 2023** in various countries?
* Are countries improving or declining in health over time?

**4. Methodology**

**🟠 Data Extraction**

* Scrape tables using **BeautifulSoup** and **Pandas** .
* Export and store raw tables in Excel format.

**🟠 Data Cleaning and Regular Expressions**

* Handle nulls, invalid or missing values.
* Normalize country names and formats.
* Use **Regex** to:
  + Extract numeric values from mixed formats.
  + Parse columns with multiple values into structured forms.

**🟠 Data Analysis**

* Compute:
  + Mean, median, and standard deviation of life expectancy.
  + Gender gap by country and year.
  + Year-over-year growth or decline.
* Compare:
  + Continents (Asia, Europe, etc.)
  + WHO vs World Bank vs CIA reports

**🟠 Data Visualization**

* **Bar charts** for male vs. female life expectancy.
* **Line graphs** for year-over-year trends.
* **Heatmaps** to show global life expectancy distribution.
* **Histograms/Boxplots** for gap analysis.

**🟠 Data Storage**

* Final processed data stored in **MongoDB** for flexible querying and scalability.

**Bonus: Streamlit Web App**

* A simple dashboard for exploring:
  + Top/bottom countries
  + Male-female gap over time
  + Country-specific graphs