# README – Life Expectancy Dataset (Custom)

This dataset was custom-created to support research and model development for predicting life expectancy based on multiple lifestyle and health-related features. Due to the lack of any publicly available dataset that includes all of the required attributes, this dataset was manually generated and refined using SQL queries in a MySQL environment.

## 📌 Purpose

To build a predictive model (using XGBoost) for estimating life expectancy based on an individual's demographic details and lifestyle behaviors.

## 📈 Why a Custom Dataset?

• No existing public dataset covered the full set of desired features, including:  
 - Age, Gender  
 - BMI, Sleep Hours, Exercise Hours  
 - Work Stress Level, Medical History  
 - Smoking, Alcohol Consumption  
 - Social Life Level  
  
• These features were selected after reviewing academic literature and health studies linking these factors to lifespan and health outcomes.  
• This dataset reflects a scientifically plausible simulation of how these features correlate with life expectancy.

## 🔍 Methodology

• The dataset was created using SQL procedures to insert and update 4992 records.  
• Country and gender-specific base life expectancy values were derived from trusted demographic data sources.  
• Adjustments to life expectancy were made based on:  
 - Peer-reviewed medical studies on smoking, alcohol, exercise, diet, sleep, etc.  
 - Global epidemiological research for distribution patterns.  
• Data validation included sanity checks, consistency analysis, and reference matching to real-world statistics.

## 🧠 Model Usage

This dataset was used to train an XGBoost regression model capable of predicting an individual's life expectancy based on the aforementioned features. The model helps demonstrate the value of interpretable machine learning in public health and personalized wellness applications.

## 📚 References

The dataset was created referencing:  
- Worldometers, WHO, The Lancet, Nature, PubMed, Harvard Health, and other peer-reviewed sources.  
See the attached 'LifeExpectancy\_Research\_References.docx' for a complete list.