

Predicting Life Expectancy: A Machine Learning Approach

By: Maher Salman & Adan Butto
Advisors: Dr. Naomi Unkelos-Shpigel & Dr. Samah Idrees Ghazawi



Background & Problem

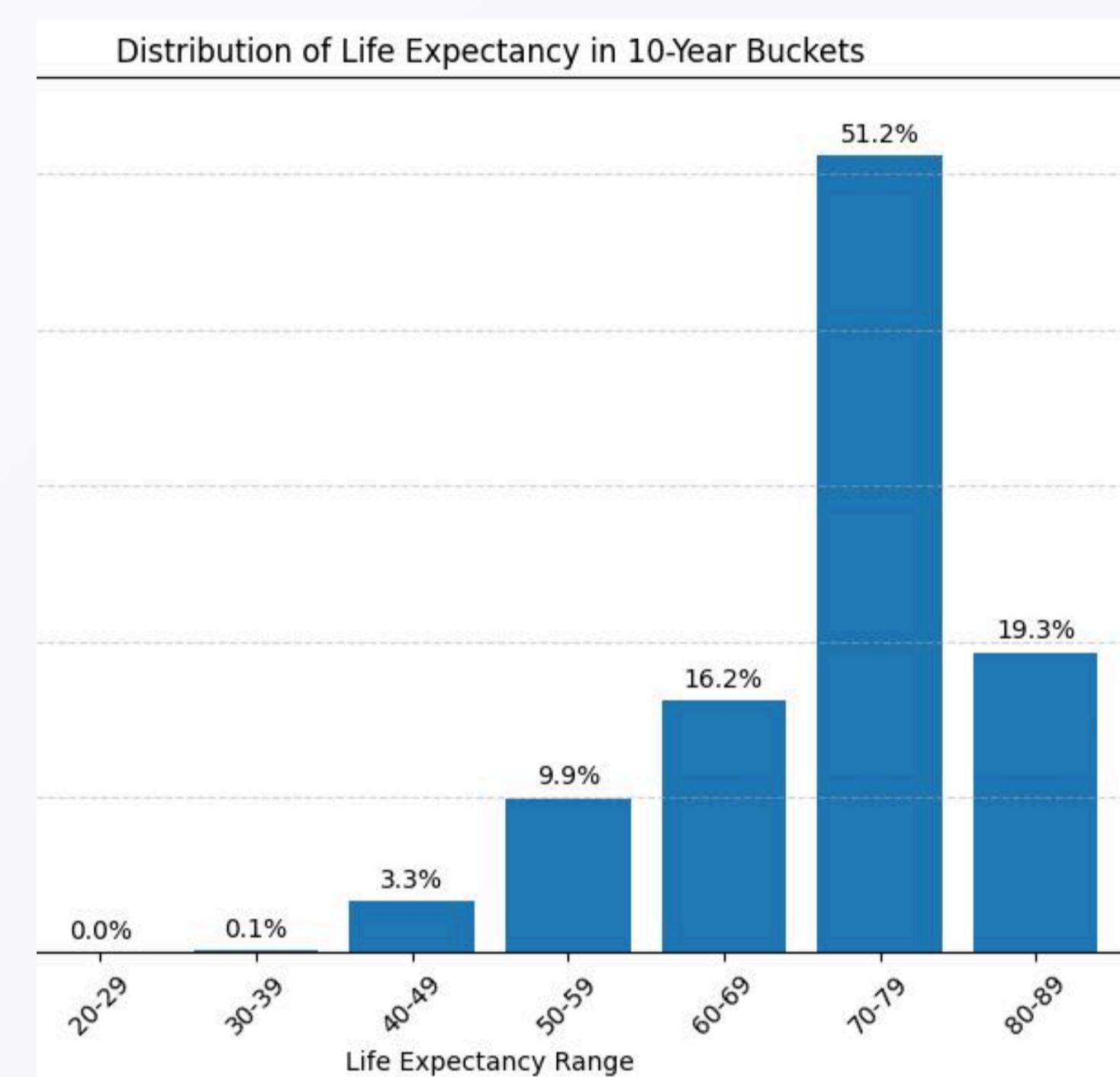
Life expectancy (LE) is a critical measure of public health, influenced by personal, lifestyle, and medical factors.

- ❖ No public individual-level LE dataset exists
- ❖ Complex non-linear relationships between indicators
- ❖ Need for accurate ML-based predictions
- ❖ Existing models rely on country-level averages

Challenge: Build comprehensive dataset & accurate ML life expectancy prediction

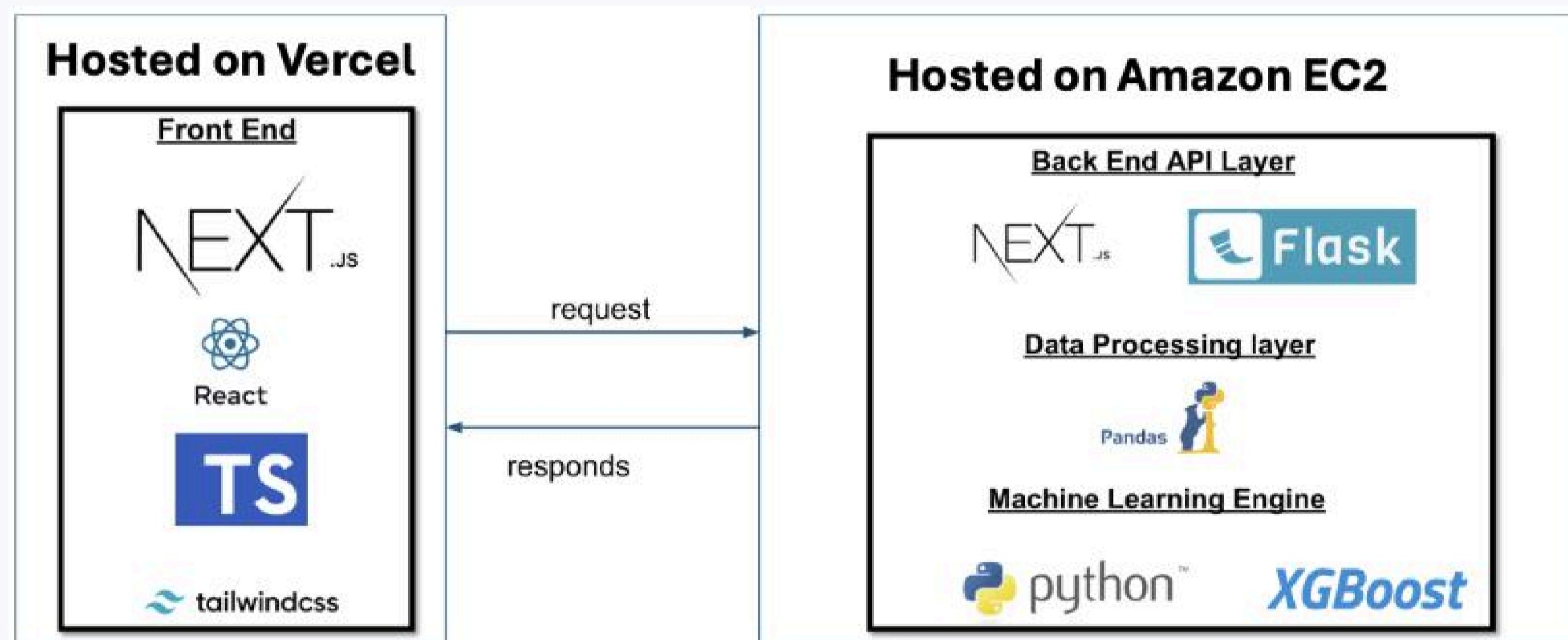
Dataset Generation

We created a realistic synthetic dataset using CDC and WHO data to estimate individual life expectancy. A scoring system based on Personal Info, Lifestyle, Medical History, and Preventive Care simulated real-world lifespans.



Key insights: smoking and depression reduced life expectancy, while physical activity increased it. Most predicted lifespans ranged from 60–89 years, reflecting actual trends.

Solution Architecture

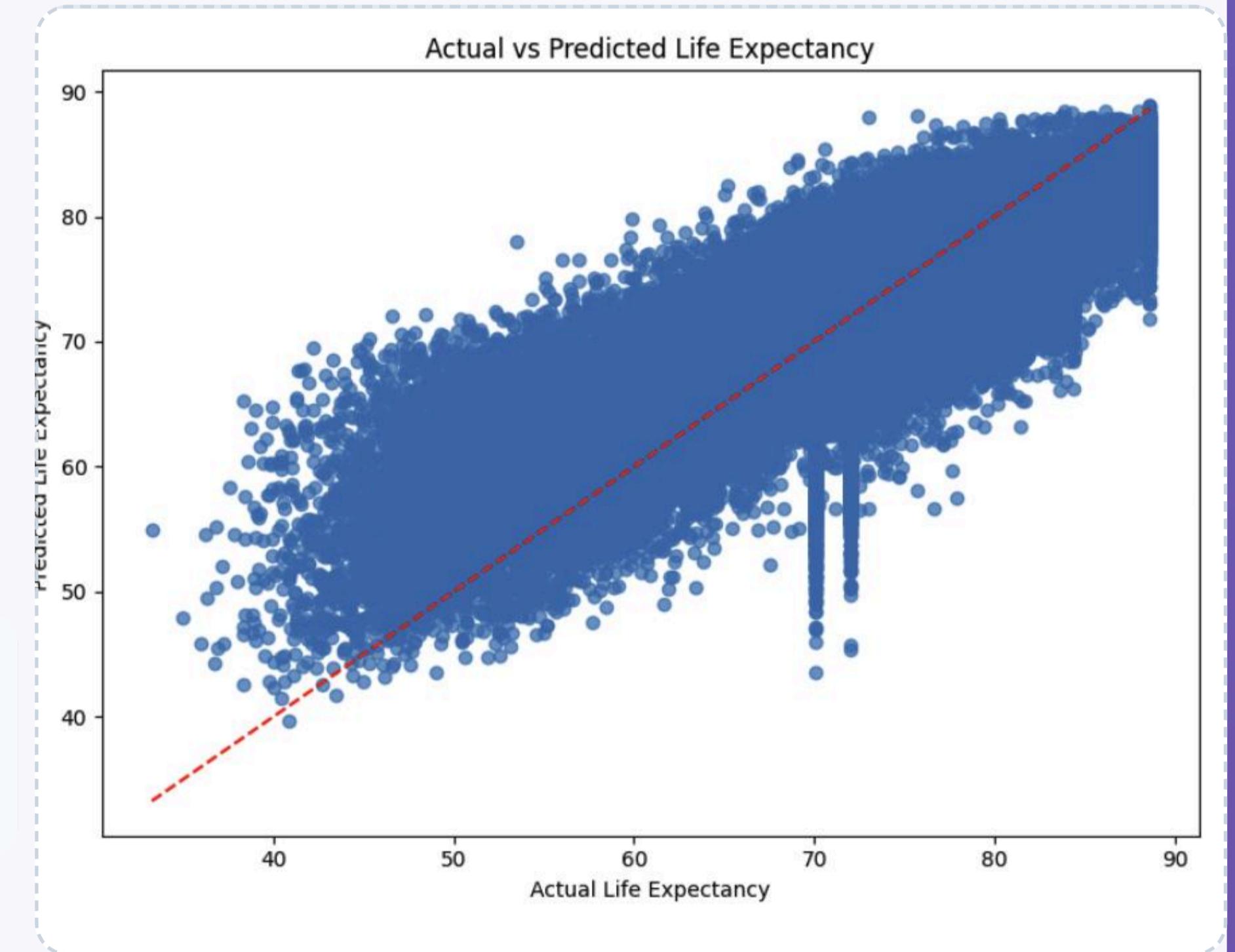


Machine Learning Model

To estimate individual life expectancy, we used XGBoost, a gradient boosting algorithm well-suited for tabular health data.

- Final model achieved:

RMSE: 5.73 MAE: 4.4 R²: 0.65



User Interface & Key Features

Multilingual Support

English, Hebrew, and Arabic interfaces

5-Step Form

Personal Info → Lifestyle → Medical History → Preventive Care → Review

Personalized Health Tips

Evidence-based recommendations based on user inputs

Interactive Simulator

Visualize impact of lifestyle changes on life expectancy

78

SUS Score
Good to Excellent

32

Participants

- ✓ Easy to use - Highest rated
- ✓ Quick to learn - Minimal training
- ✓ Well integrated - Consistent UI

Improvements Made

- Added multilingual support (Hebrew & Arabic)
- Split medical sections for better visibility
- Added personalized health tips section