Final Lecture Report: Lecture 1

Life Expectancy Analysis Report

Submission Date: May 9, 2025
Submitted by: **Kaan Yazıcıoğlu (97364)**

1. About the Data

Our dataset has:

- 2938 total records

- 2350 records for training

- 588 records for testing

- 4 main factors: GDP, Total expenditure, Alcohol, and HIV/AIDS

Some data was missing:

- GDP: 448 empty values

- Total expenditure: 226 empty values

- Alcohol: 194 empty values

- HIV/AIDS: no missing values

- Life expectancy: 10 empty values

2. What I Found in the Data

- The lowest life expectancy is 36.3 years

- The highest is 89.0 years

- Most countries have life expectancy between 63.1 and 75.7 years

The countries with the best life expectancy are:

1. Belgium (89.0 years)

2. Finland (89.0 years)

3. France (89.0 years)

3. How Well Each Factor Predicts Life Expectancy

When I tested each factor alone:

1. HIV/AIDS:

- Best single predictor (30.9% accurate)

- Strong negative relationship with life expectancy

2. GDP:

- Second best single predictor (18.5% accurate)

- Positive relationship with life expectancy

3. Alcohol:

- Third best predictor (15.4% accurate)

- Moderate positive relationship

4. Total Expenditure:

- Weakest predictor (5.1% accurate)

- Very weak relationship

When I used all four factors together:

- Training data: 52.0% accurate

- Test data: 50.5% accurate

- Much better than any single factor!

4. How Much the Model Gets Wrong

- Usually off by about 5.16 years

- Sometimes predicts slightly higher than real life expectancy

- Predictions can vary by about 6.70 years

5. How Things Are Connected

I found that:

- HIV/AIDS has the strongest link to life expectancy (negative relationship)

- GDP is the second most important factor (positive relationship)

- Alcohol is the third most important

- Total expenditure has the weakest connection

6. What I Learned

1. Using all four factors works much better than using just one

2. HIV/AIDS is the most important factor for predicting life expectancy

3. The model is now twice as accurate as before

4. Money (GDP) is still important, but health factors (HIV/AIDS) matter more

7. What I Did

I:

- Loaded and cleaned the data

- Fixed the missing information

- Made simple models for each factor

- Made a bigger model using all four factors

- Checked how well everything worked

- Made this report

8. Problems and Ideas for Next Time

1. The model is better but still not perfect

2. Maybe we could:

  - Look at more health-related factors

  - Try different ways to handle missing data

  - Study how life expectancy changes over time

  - Use more complicated math to make better predictions

9. Why I Chose These Factors

1. HIV/AIDS: Shows the strongest relationship with life expectancy

2. GDP: Important economic indicator

3. Alcohol: Health-related factor

4. Total expenditure: Healthcare spending indicator

10. Comparison of Models

- Simple models: 5.1% to 30.9% accurate

- Multiple regression model: 50.5% accurate

- Big improvement when using all factors together!