Data Processing

Project description

May 16, 2024

Full Name	Student ID	University
William Bærenholdt	15348857	UvA

Amsterdam

Project description

This project deals with the life expectancy of women and men in Denmark, both in terms of the difference between them and the trends over recent years. The initial scientific question "Is there a gender in Denmark that lives longer than the other?" is investigated using visual graphs from the Bokeh Python package to determine if this is the case. It is noted that if there is a gender that lives longer, it is clearly women. Additionally, another scientific question arises: "Are Danes, in general, living longer and longer?" based on data. It is added to the investigation whether there is statistical evidence that women live longer than men, which is concluded to be the case.

It is noted that a significant part of the project has also been to find appropriate data and literature to calculate various aspects. The dataset used includes mortality and exposure data from the Danish statistical server called "Statistikbanken" from the years 1837 to 2023.

In the project, great emphasis has been placed on well-organized and correctly implemented code. Reference is made to the courses SP1, SP1, and DP, where efforts have been made to adapt the coding style learned throughout the courses. Additionally, considerable time has been spent creating visually appealing graphs to make the project more engaging.

The project largely guides the reader through its content. Further elaboration on the project is therefore not provided here, as it would merely repeat the project itself. Much effort has been put into ensuring that the project stands as an independent work and can be read without this description or the accompanying notebook.

A requirement in the project formulation was that something new should be learned during the project. To develop the project, I have learned to use two different Python packages: Bokeh and SciPy. In addition to this new information, I have also naturally utilized a significant portion of the aforementioned courses and therefore use both the Python packages Numpy and Panda data frames.