DAT565/DIT407 Assignment 1

Vaibhav Talari
talari@chalmers.se

Stefán Ólafur Ingimarsson
stefanla@chalmers.se

2024-09-05

This report is submitted by group 25 for assignment one in *Introduction to Data Science & AI*.

Problem 1 - Dependency ratio

We are working with the dataset **swedish_population_by_year_and_sex_1860-2022.csv**. This dataset provides population data by age and gender from 1860 to 2022, offering a view of demographic trends spanning more than half a century and a half.

In this analysis, we categorize the population into two main groups: the productive population and the dependent population. The productive population, or labor force, includes individuals of working age, generally defined as those between 15 and 64 years old.

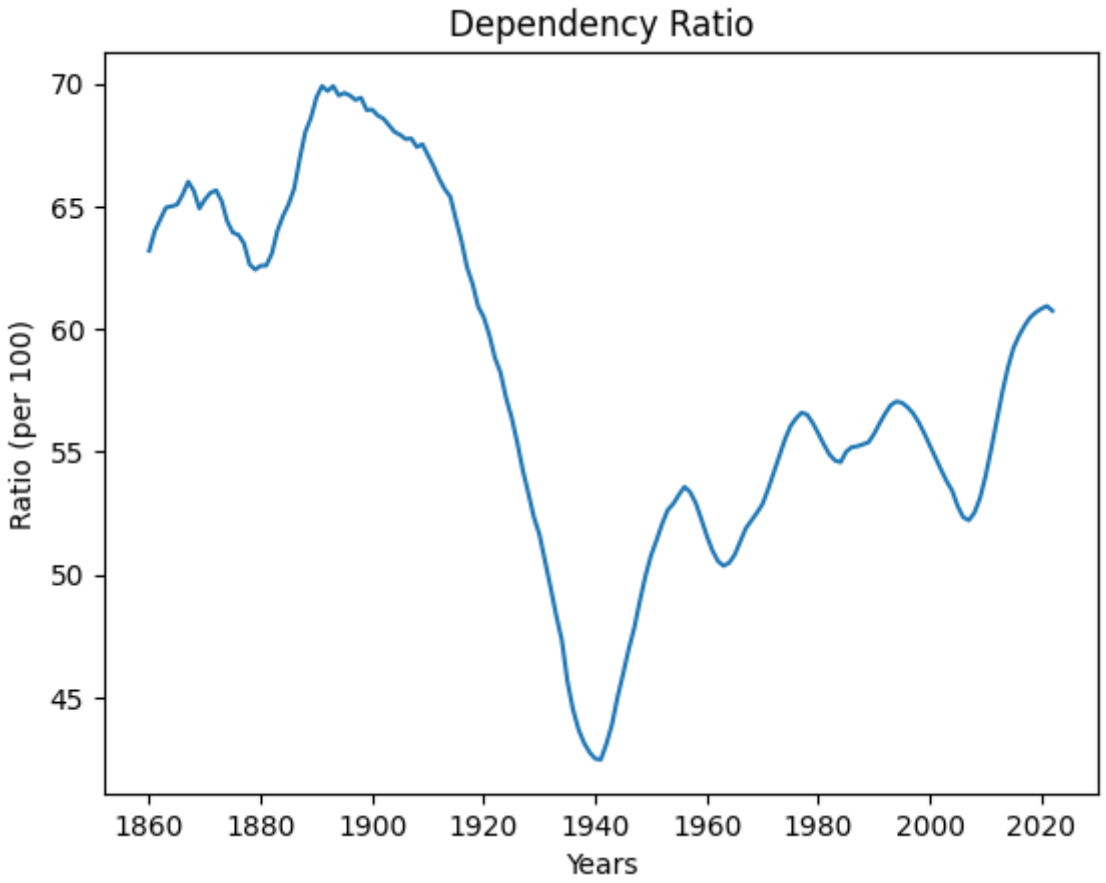
The dependent population includes children (aged 0 to 14) and elderly (aged 65 or older). These age groups are generally more reliant on the working-age population for support, as they either have not yet entered the workforce or have retired from it. The dependency of these groups places unique demands on social services, healthcare, and economic resources.

By examining changes in these population segments over time, we can better understand the shifts in Sweden's demographic structure. This dataset allows us to visualize these demographic trends, helping to reveal how Sweden's labor force has evolved in response to historical events, economic shifts, and advancements in healthcare.

1 Plot the figure of the dependency ratio

In the first section of the assignment, we used the dataset to plot a figure of the dependency ratio over time. The dependency ratio is a demographic measure that represents the proportion of the dependent population (children aged 0-14 and elderly aged 65+).

To conduct this analysis, we set up a work environment using Jupyter Notebook and Anaconda. The figure can be found below.



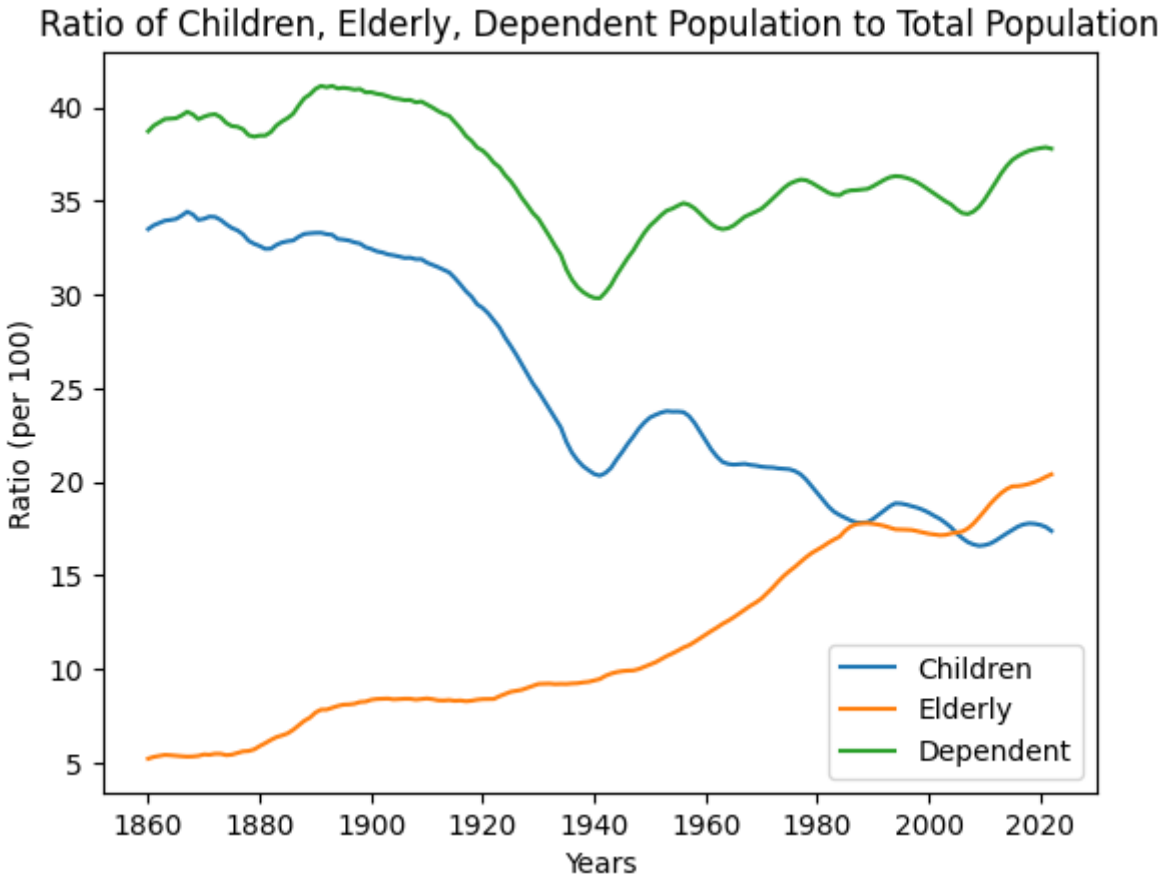
The data analyzed reveals fluctuations in Sweden's labor force over the years, particularly a significant decline from the 1910s through the 1940s. This period was marked by a wave of emigration, as many Swedes left for countries like the United States in search of new opportunities. Between 1850 and 1930 alone, an estimated 1.1 million Swedes emigrated to the U.S., largely seeking work and a better life. This mass departure of workers undoubtedly impacted Sweden's domestic labor market. Additionally, the 1930s saw the Great Depression in the United States, followed closely by the disruptions of World War II, which likely compounded these shifts in Sweden's workforce. Sweden suffered severely during the early years of the Great Depression. In the early 1930s unemployment rose, and reductions in wages caused a series of harsh labor conflicts [2].

The 1930s brought additional challenges. As the Great Depression unfolded, Sweden experienced a sharp rise in unemployment and wage reductions, which led to a series of intense labor conflicts. This global economic downturn, followed by the disruptions of World War II likely compounded the challenge within Sweden's workforce [1]. Although Sweden remained neutral during the war, industrial production slowed due to shortages in imports and exports. However, this focus on self-sufficiency and domestic production may have laid a foundation for a rapid economic recovery after the war, helping to stabilize the labor

force. Sweden's capacity to harness its own resources during the war likely contributed to its post-war economic expansion, enabling more Swedes to enter the workforce, start families, and contribute to population growth [5].

During the period between the 1960s to 2000s, immigration to Sweden became a significant factor in its population dynamics. Refugees from regions like Iran and the former Yugoslavia, fleeing conflict in their home countries, arrived in increasing numbers. The new residents contributed to Sweden's population growth, as immigrants settled and began to start families. This expansion of the labor force through both natural population growth and immigration supported Sweden's economic development during this period [3].

2 Include the fraction of children and elderly



In this figure, we can see that the dependent population and the children population follow a similar pattern, decreasing in the 1940s and going up afterwards. With the elderly population increasing steadily, the dependent population goes up. As discussed in the previous question, the child population increasing during the 1940s could be attributed to the economic stability in

Sweden after World War II. Sweden choosing to remain neutral and mobilizing their domestic production created better conditions for the job market, allowing more people to start families.

The elderly population increasing could correlate with the life expectancy overall increasing. With healthcare improving each year, more people will live longer than their previous generations. During 1860 in Sweden, the average life expectancy for a newborn baby was roughly 43 years. This number has gradually increased throughout the years, and as of 2022, the average life expectancy of a newborn baby in Sweden is roughly 81 years. With this change, it is no surprise that the number of elderly population has increased [4].

3 Assignment code

Here is our code that shows how we plotted the first figure that displayed the dependency ratio.

```
import pandas as pd
import matplotlib.pyplot as plt
raw_data = pd.read_csv("swedish_population_by_year_and_sex_1860-2022.csv")
raw_data
work_data = raw_data.melt(
    id_vars=["age", "sex"], var_name="year", value_name="population"
)
work_data
work_data["age"] = work_data["age"].replace(r"\D+", "", regex=True).
astype(int)
work_data["year"] = work_data["year"].astype(int)
work_data.dtypes
work_data["age_group"] = pd.cut(
    x=work_data["age"], bins=[-1, 14, 64, 110],
    labels=["children", "labour", "elderly"]
)
work_data
# add as_index=False or .reset_index(name="population_sum")
# without this the columns name is missing
groupby_y_a_s = work_data.groupby(
    by=["year", "age_group"], observed=False, as_index=False
)["population"].sum()
groupby_y_a_s
# use pivot to transform age group row to columns
# .reset_index() adjust the columns, return dataframe object
pivot_df = groupby_y_a_s.pivot(
    index="year", columns="age_group", values="population"
).reset_index()
pivot_df
# dependency ratio
pivot_df["dp_ratio"] = 100 * (
    (pivot_df["children"] + pivot_df["elderly"]) / pivot_df["labour"]
)
```

```
34
35 # children to total populatioin
36 pivot_df["child_to_total"] = 100 * (
37     pivot_df["children"]
38     / (pivot_df["children"] + pivot_df["elderly"] + pivot_df["labour"])
39 )
40
41 # elderly to total populatioin
42 pivot_df["elderly_to_total"] = 100 * (
43     pivot_df["elderly"]
44     / (pivot_df["children"] + pivot_df["elderly"] + pivot_df["labour"])
45 )
46
47 # dependant population to total population
48 pivot_df["dependant_to_total"] = 100 * (
49     (pivot_df["elderly"] + pivot_df["children"])
50     / (pivot_df["children"] + pivot_df["elderly"] + pivot_df["labour"])
51 )
52
53 pivot_df
54 # Dependency ratio plot
55 fig, ax = plt.subplots()
56 x = pivot_df["year"]
57 y = pivot_df["dp_ratio"]
58 ax.plot(x, y)
59
60 plt.title("Dependency␣Ratio")
61 plt.xlabel("Years")
62 plt.ylabel("Ratio␣(per␣100)")
63
64 plt.show()
```

References

[1] Lennart T. Norman Jörgen Weibull. "Sweden in the 20th century". In: *Britannica* (2024). URL: https://www.britannica.com/topic/history-of-Sweden/Sweden-in-the-20th-century.

[2] Jonas Helgertz Martin Dribe Björn Eriksson. "From Sweden to America: migrant selection in the transatlantic migration, 1890–1910". In: *European Review of Economic History* (2023). URL: https://academic.oup.com/ereh/article/27/1/24/6644609.

[3] Unknown. "Immigration and emigration 1970-2023 and projection 2024-2070". In: *Statistics Sweden* (2024). URL: https://www.scb.se/en/finding-statistics/statistics-by-subject-area/population/population-projections/population-projections/pong/tables-and-graphs/immigration-and-emigration-by-sex-and-country-of-birth-and-projection/.

[4] Unknown. “Life expectancy 1751-2023”. In: *Statistics Sweden* (2024). URL: https://www.scb.se/en/finding-statistics/statistics-by-subject-area/population/population-composition/population-statistics/pong/tables-and-graphs/births-and-deaths/life-expectancy-17512023/.

[5] Katarina Wagman. “From war to the Swedish model”. In: *Ekonomi Fakta* (2024). URL: https://www.ekonomifakta.se/en/swedish-economic-history/from-war-to-the-swedish-model_1209204.html.