Master of Science in Applied Information and DataScience

Group work - Tableau Dashboard

Emigration Analysis of Switzerland

CSC01 - Computer Science Concepts for Data Scientists

Students: Eduardo Pacheco

Giancarlo Meccariello

Philipp Bächler

Lecturers: Dr. Halldor Janetzko

Prof. Dr. Martin Zimmermann

Date of Submission: 16. January 2022

Table of content

1. Motivation

The source of inspiration for this topic has been the growth of population in Switzerland. The overall population of Switzerland has enjoyed a steady and relatively consistent growth throughout the country's history. Throughout the 19th and 20th centuries it is claimed that the growth levels did not vary from figures of 0.7% to 0.8% but over the decades, that led to a doubling of the population.

There have been some anomalies in those figures such as a 1.1% rise in numbers in 2007 and the reasons for these are put firmly at the door of increased immigration. That growth level continues to this day at a point where the Swiss population of 2014 stands at around the eight million figures.

Switzerland Population Projections:
The population change in Switzerland has been declining since 1995 and that is expected to continue. It is predicted that the rate of growth will reach a temporary high in 2020 at 0.835 before slowly decreasing towards stagnation. The annual growth rate is expected to get as low as **0.28%** by 2050. During this same period, projections say that the population of Switzerland will be 8'670'535 in 2020, 9'203'908 by 2030, 9'586'938 by 2040, and 9'879'901 by 2050. [1]

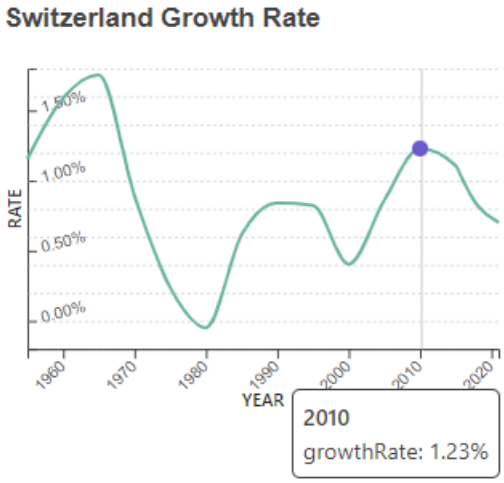


Figure 1. worlpopulationreview.com

The relationship between population growth and growth of economic output has been studied extensively (Heady & Hodge, 2009). Many analysts believe that economic growth in high-income countries is likely to be relatively slow in coming years in part because population growth in these countries is predicted to slow considerably (Baker, Delong, & Krugman, 2005). In this context we would be approaching the emigration phenomenon in Switzerland by trying to quantify it and get some interesting insight hidden in the chosen data sets.

2. Data Source

A data set containing information of the past decade (2011 – 2020) will be analysed. The name of the data set is "Emigration of the permanent resident population by canton, citizenship (category), country of destination, sex and age class". The data source provider is the Swiss Federal Statistical Office (FSO). [2]

Another data set used to conduct more specific analysis was taken from the FSO called "Demographic balance by age and canton". [3]

3. Process

To tackle the challenges of a big data set, the methodology known as CRISP-DM was applied. This methodology is a cross-industry standard process for data mining and predictive analytics.

The process of developing visualizations is broken down into six major phases which will be discussed in greater detail:

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Outcome
5. Evaluation

3.1. Business Understanding (use case/usage)

Before the start of creating a dashboard, it is essential to know the target group. This consideration is therefore important to understand the needs of the target group.

Since the dashboard is going to display information about emigrations of Switzerland, the dashboard could be useful for any emigration institutes of Switzerland. It could be some specific cantonal institutes which is interested in growths, or destinations of emigrations. Specifically, for people without any statistic knowledge which are looking for insight information.

3.2. Data Understanding

After having defined the use case of the dashboard, the data set has to be understood. The used data set includes, the year, canton of origin, citizenship, country of destination ID, the name of the country of destination, the gender and age divided into sets. The age starts at 10 years and a group always includes four years. In this case 10-14 years, 15-19 years and so on.

Therefore, in the corresponding cell is written the amount of people which are emigrating to the specific country.

With these attributes it is possible to create different kind of visualisation and answering a range of specific questions about the emigration behaviour of Swiss inhabitants.

3.3. Data Preparation

The names of the canton were written in the different languages spoken in that canton. For example, in Bern there are two languages spoken: German and French. The names are separated by a "/". For example, Bern was written as "Bern / Berne". Tableau was not able to assign these fields to the correct canton. The shell tool "sed" can be uses to rename the fields. To check, if everywhere "Bern / Berne" was used, one can use "grep".

```
planck@planck-MS-7B49:~/ownCloud/projects/csc$ grep "Bern / Berne" Emigration_Swiss_Last10Years.csv
2011,Bern / Berne,Switzerland,Albania,Male,0,0,0,0,0,0,0,0,0,0,0
2011,Bern / Berne,Switzerland,Albania,Female,0,0,0,0,0,0,0,0,0,0,0
2011,Bern / Berne,Switzerland,Andorra,Male,0,0,0,1,0,0,0,0,0,0,0
2011,Bern / Berne,Switzerland,Andorra,Female,0,0,0,0,0,0,0,0,0,0,0
2011,Bern / Berne,Switzerland,Belgium,Male,0,0,3,2,1,0,0,1,1,0,1
2011,Bern / Berne,Switzerland,Belgium,Female,0,0,1,0,2,3,0,0,0,1,0
2011,Bern / Berne,Switzerland,Bulgaria,Male,0,0,0,0,0,0,0,0,0,1,0
2011,Bern / Berne,Switzerland,Bulgaria,Female,0,0,1,0,0,0,0,0,0,0,0
2011,Bern / Berne,Switzerland,Denmark,Male,0,0,0,0,0,0,0,0,0,0,0
```

By piping "grep" with "wc -l" it is possible to check how many times Bern occured in the data set.

```
planck@planck-MS-7B49:~/ownCloud/projects/csc$ grep "Bern" Emigration_Swiss_Last10Years.csv | wc -l
7920
```

Following, "sed" was applied with the parameters:

1. -i: in place
2. -e: editing
3. 's/Bern \/ Berne/Bern/g': search pattern / new name
4. Emigration_Swiss_Last10Years.csv: target file

```
planck@planck-MS-7B49:~/ownCloud/projects/csc$ sed -i -e 's/Bern \/ Berne/Bern/g' Emigration_Swiss_Last10Years.csv
planck@planck-MS-7B49:~/ownCloud/projects/csc$ grep "Bern" Emigration_Swiss_Last10Years.csv
2011,Bern,Switzerland,Albania,Male,0,0,0,0,0,0,0,0,0,0,0
2011,Bern,Switzerland,Albania,Female,0,0,0,0,0,0,0,0,0,0,0
2011,Bern,Switzerland,Andorra,Male,0,0,0,1,0,0,0,0,0,0,0
2011,Bern,Switzerland,Andorra,Female,0,0,0,0,0,0,0,0,0,0,0
2011,Bern,Switzerland,Belgium,Male,0,0,3,2,1,0,0,1,1,0,1
2011,Bern,Switzerland,Belgium,Female,0,0,1,0,2,3,0,0,0,1,0
2011,Bern,Switzerland,Bulgaria,Male,0,0,0,0,0,0,0,0,0,1,0
2011,Bern,Switzerland,Bulgaria,Female,0,0,1,0,0,0,0,0,0,0,0
2011,Bern,Switzerland,Denmark,Male,0,0,0,0,0,0,0,0,0,0,0
2011,Bern,Switzerland,Denmark,Female,0,0,3,2,1,0,0,0,0,0,0
2011,Bern,Switzerland,Germany,Male,1,4,20,26,18,6,12,11,4,3,2,4
2011,Bern,Switzerland,Germany,Female,1,7,28,30,26,10,11,6,6,2,5,5
```

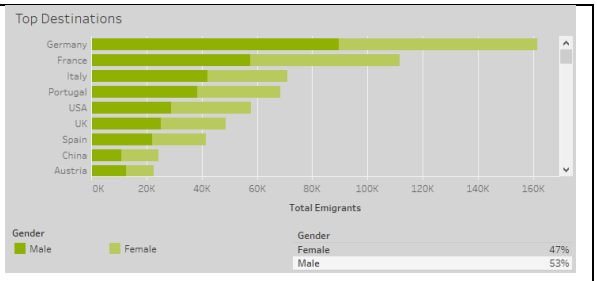
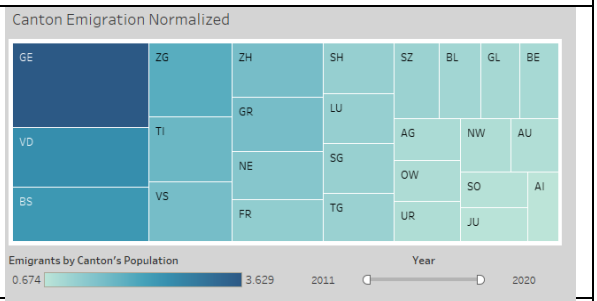
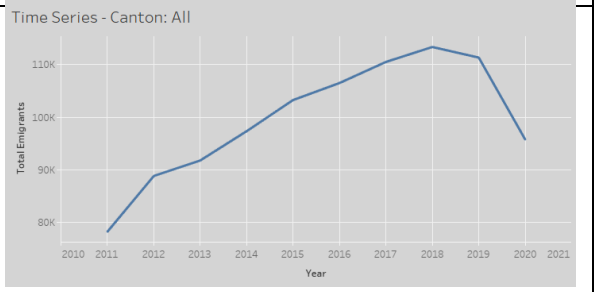
In the search pattern (3.) an escape character "\" was used. This was necessary to include the "/" in the search pattern. The tool "grep" was used to check if "sed" worked correctly.

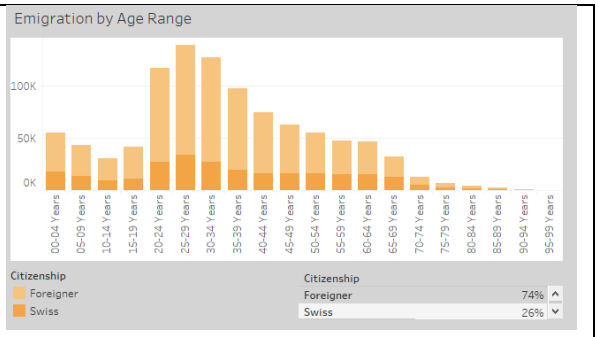
This procedure was repeated four times with following corrections:

1. Bern / Berne -> Bern
2. Fribourg / Freiburg -> Fribourg
3. Graubünden / Grigioni / Grischun -> Graubünden
4. Valais / Wallis -> Valais

4. Outcome

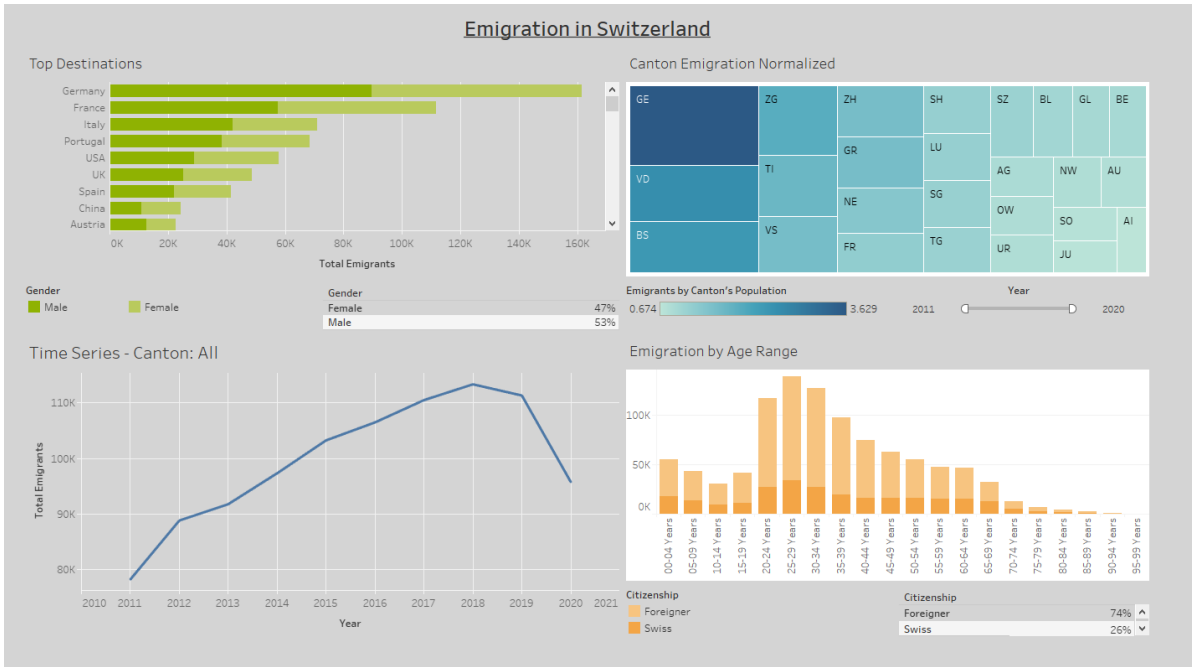
The created dashboard consists of four individual visualisations. By setting filters do the visualisations show the requested insights.

Top Destination The bar chart shows the countries which the most emigrations in an ascending order. No doubt the neighbour countries are in the top of the list, as is also related to the official speaking languages of Switzerland. The fact that Switzerland has a large community of Spaniards and Portuguese, most likely, these persons would like to get retired in their home countries. China stands out as is so far away from Switzerland, in means of distance and culture. Perhaps, the fast-growing economy may play an important role in attracting investment and people to settle down in China.	
Canton Emigration Normalized This visualisation shows the most emigrations per canton. Notable is the data was normalized per 100'000 inhabitants. This prevents wrong interpretation of the data. The canton Zurich has the most inhabitants which would simply lead to the most emigrations in absolute numbers.	
Time Series The goal of a time series analysis is to visualize pattern and trends, how the emigration changed overtime. As it may be seen extreme the growing rate of emigration, it might have a proportional and direct connection to the Swiss population growth. An interesting point to mention is the year 2020. During the year 2020 roughly 15% less people left the country which might be a result of the corona pandemic.	

Emigration by Age This vertical bar charts shows the emigration patterns by age groups. Here one can see an effect of the mandatory attendance to school in Switzerland. As student must attend school until approximately 15 years old. As people grow old the emigration rate decreases. The reason for this might be the quality of the health system, people settling down or just getting too old to travel and take risks. It is also good visible, that much more foreigners leave the country. One must keep in mind, that about 25% of the population is considered as foreigner.	

5. Evaluation

These four plots are combined to one single dashboard. The Plots can be used as filters which is useful to explore the data even further. If e.g., one is interested in the canton of Geneva, he could simply select this canton in the treemap. Then one can see in the "Top Destinations" graphic where the population of Geneva tends to emigrate. Or it is possible to see the annual changes and the age groups. Another filter could be the age groups. If one is keen on finding out how the top destinations for the age groups 25-29 and 60-64 differ – he selects these groups in the "Emigration by Age Groups" boxplot.



6. References

[1] https://worldpopulationreview.com/countries/switzerland-population

[2] https://www.pxweb.bfs.admin.ch/pxweb/en/px-x-0103020300_104/-/px-x-0103020300_104.px/

[3] https://www.bfs.admin.ch/asset/en/px-x-0102020000_104