

# WDB Mini-Challenge

Pascal Trösch

November 2025

## 1 Introduction

The Web is a huge open source database. The data stored on different websites can be exploited for all kinds of purposes. By using programming languages such as Python, one can access this data in a very efficient and easy way to inspect the data for different kinds of goals. The aim of this project was to extract data from job ads posted on [www.jobscout24.ch](http://www.jobscout24.ch). I am especially interested in the metadata of the job ads, especially the date when job ads were uploaded, the quota of jobs and the regions of where the jobs are located. This data can be used by researchers and decision makers to observe labour market development in quasi real-time.

## 2 Scope and Relevance of the Project

The labour market is a very important feature of capitalist economies. If people have no labour, they have to rely on government subsidies. To mitigate a crisis in labour markets such as mass unemployment, governments can implement different measures to counter factors that create unemployment. On the other hand, the economy needs sufficient qualified personnel to fill vacancies in companies. If qualified workers are not available in a sufficient manner, production shrinks and a shortage in goods and services may result.

However, real time data on the labour market is often hard to get. One could rely on stock prices or stock indexes, but consequences of rising or falling stock prices on employment are hard to predict. Labour market data often lags several weeks or even months, as it is often gathered from new inscriptions at job centers or large scale surveys at companies (such as the Economic Barometer from the Swiss Economic Institute (KOF)).

This is where data from job ad websites such as [www.jobs.ch](http://www.jobs.ch) or [www.jobscout24.ch](http://www.jobscout24.ch) can be very useful. Companies do have an incentive to publish open jobs as fast as possible to the public, to encourage workers to apply and to fill the vacancy in a short period of time. On one hand, the sheer number of job ads published can be an indicator on the situation of the labour market. The more open job ads are posted, the higher is demand for labour force among companies and therefore the better is the general state of the economy. But I am also particularly interested in the duration of publication of job ads. I argue that longer periods of publication show difficulties for companies to fill their vacancies with qualified personnel. Contrary to that, shorter periods show that more people are out looking for new jobs and therefore there is a higher risk of unemployment. Furthermore, I am interested in published quotas of jobs, as I argue that more jobs with higher quotas represent a higher demand for labour force than more jobs with lower quotas.

I therefore aim to collect job ad data from the website [www.jobscout24.ch](http://www.jobscout24.ch), which is one of the biggest job ad platforms of Switzerland with ads from all regions and all industries of Switzerland.

Such data is very interesting for researchers and decision makers, as it represents real time data from the labour market which can be collected at very low cost and 24/7.

## 3 Technical Procedure

For the extraction of the data, I use the Selenium Package for Python which is useful to extract data from publicly available data from websites. I start scraping at the base page of [www.jobscout24.ch](http://www.jobscout24.ch) without taking

further measurements. Therefore, the most relevant job ads are shown. For reasons of feasibility and efficiency, I only scraped the first 100 pages of job ads on the website, which results in a total of 2401 job ads (for the time of the conduction of the scraping, this number may change to a small extent depending on the moment of execution of the programm). To test the clean execution of the programm, I conductet the following 4 different tests:

- Scraper intialization: Does the scraper and the corresponding web driver initailize correctly and can they access the web page?
- Title extraction: Is the extraction of the title of the job ad successful and can the scraper access the correct field on the web page?
- Handling of missing informaiton: Can the scraper handle empty fields or missing information in job ads and continue to the next relevant fields?
- Clean quota input: Does the scraper recognize quota numbers in a correct matter?

These tests ensure that the scraper works in the correct manner, in order to only start scraping when the programm works successful.

## 4 Structure of the Data

The data collected by the programm is stored to a csv-file which can then be used for further analysis. Table 1 in the appendix shows the collected attributes by the programm. I deliberately relinquished from scraping the main content of the ads, as for the purpose of my analysis, this is not relevant. The most important data points for the following analysis are especially the date of publication, the quota of the job and the municipality or location of the job. Date and quota inputs had to be changed to a readable format, as they were stored as character value by the programm. Furthermore, quotas had to be separated, as many of them were indicated in a manner such as 80-100%. To enable analysis by cantons, I merged the dataset with a dataset from the federal office of statistics, taking the municipality as main indicator variable for merging. Unfortunately, not all municipalities could be assigned to a canton, which created several NA's in the canton variable. Table 2 shows the structure of the data after treatment and creation of new variables.

## 5 Data Analysis

As I already outlined, I am particularly interested in publication dates and quotas. I will therefore in the following outline some results of my data analysis from the data obtained through the web scraping process.

Figure 1 shows a box-plot indicating the publication dates for all job ads collected. The vast majority of all ads was published in November (The data collection was executed on 23.11.2025). There are only very few job ads posted before October 15th. This leads to the conclusion, that most of the open positions are filled fast. Particularly interesting would be a comparison bewtween multiple points in time to observe, whether the median or the the percentiles shifted more towards the actual date of further away from it. By looking at this graph, we can conclude that collected job ads do not show a pattern of particular difficulties for employers to fill their vacancies.

Distribution of publishing dates for job ads

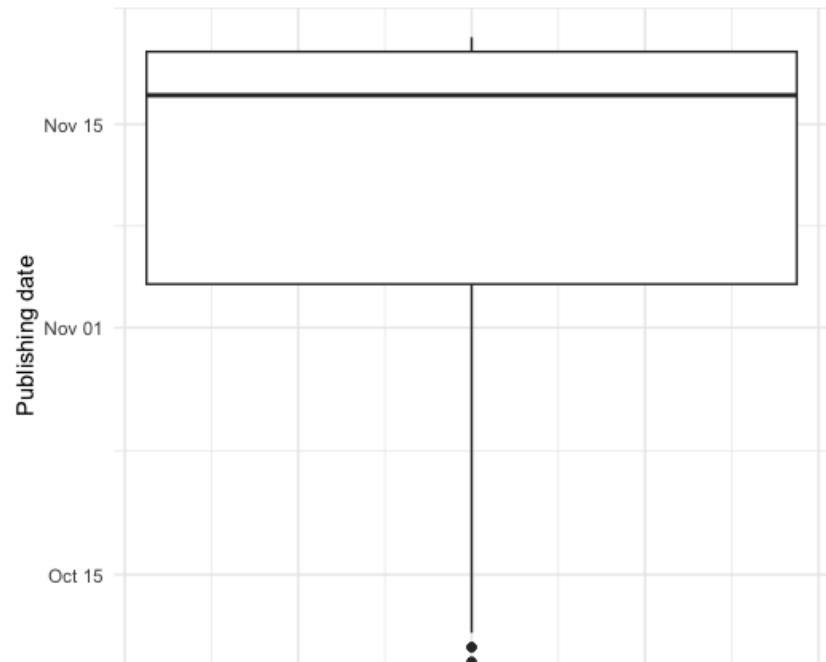


Figure 1: Publishing dates for all job ads collected

Figure 2 also shows the distribution of publication dates of job ads. For the majority of cantons, the big part of job ads were published in November. The canton of Nidwalden immediately catches the attention, as it figures as an outlier among the cantons, having a median date of publication around october 15th. One hypothesis that came to my mind that this may be due to the Pilatus Airplane Factory, which is in the canton of Nidwalden, which heavily relies on specially educated and trained staff, that may be hard to get on the labour market, as Nidwalden is a very small canton.

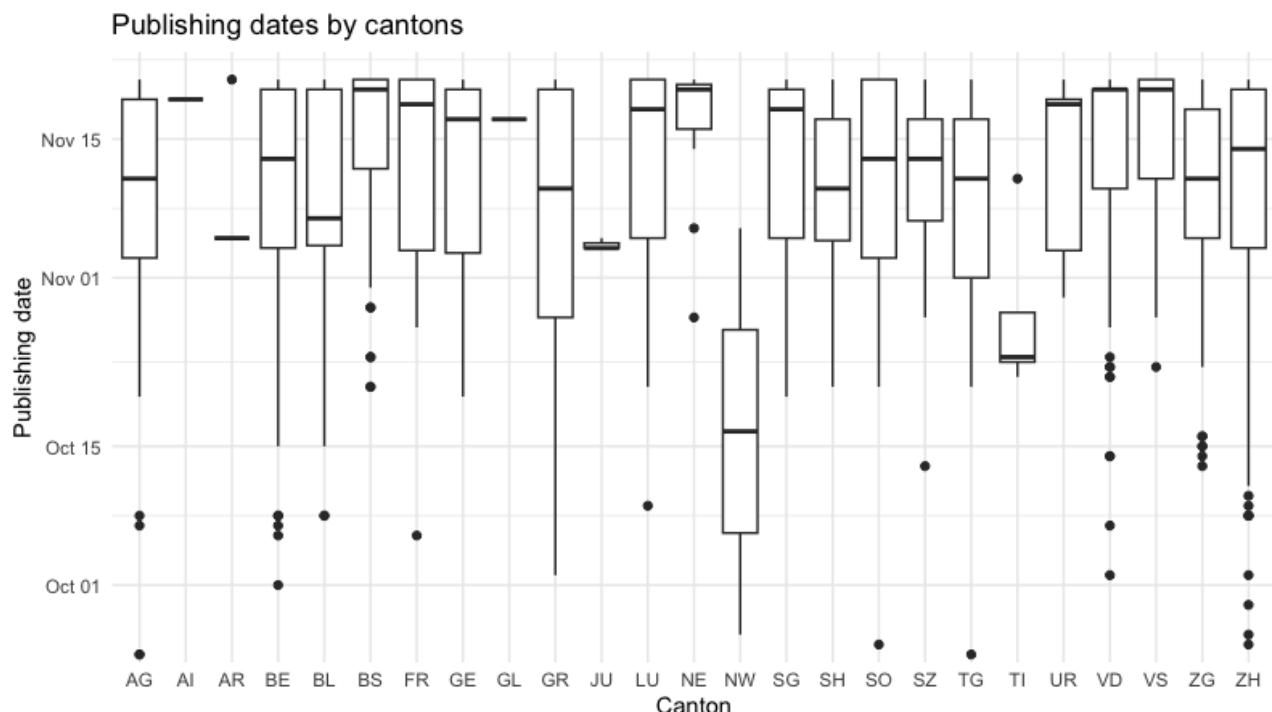


Figure 2: Publishing dates by cantons

Finally, Figure 3 shows the distribution of quotas in job ads. One can clearly see, that most job ads were issued for jobs with higher quotas such as from 80 to 100%. Quotas lower than 80% are still the exception, even if part time working is becoming more attractive and common for quite some people. As for the other two plots, it would also be interesting to observe if there are changes over time. It remains to say that, when looking at quotas, employers seem to have a high need for employment, as part time jobs are scarce.

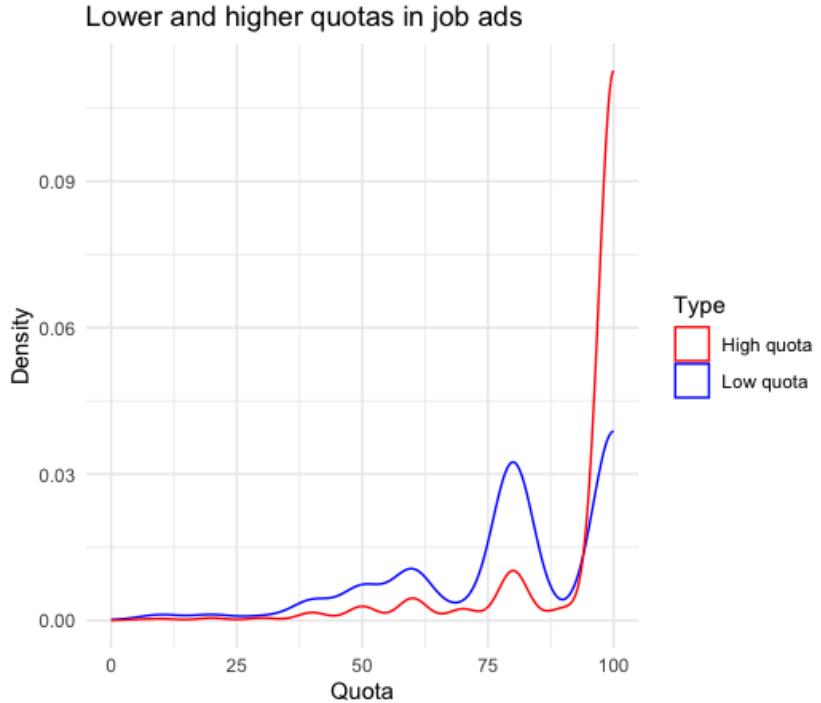


Figure 3: Density of lower and higher quotas in job ads

## 6 Conclusion and possible Extensions

Using a web scraping tool for collecting job ad data can have the big advantage of getting real time data about the labour market. This has a broad range of possible applications, such as research or database for policy makers. However, my application of the tool also shows, that one would also have to look further at some other aspects of job ad data. For example, a comparison over different time periods would be interesting. This could be done quite easily, using cronjobs for example and scraping data weekly or monthly to observe differences in time periods. Another interesting feature would be to compare different industries, to see, if unemployment or skilled labour shortage is specially present in some industries. For this, one would need to reprogramm the scraping programm, as the job ads usually do not list industries. A further limitation is the number of job ads I was able to extract from the webpage. To create really solid analysis results, one would need to extract all public job ads on the page. This would require more powerful hardware than I had access to in order to extract data efficiently from the website.

## 7 Appendix

Table 1: Structure of Data before treatment (Rows 1-5 shown)

title	company	publishing_date	quota	location	url	job_id
Diagnosetechniker	Garage Wismer AG	2025-11-20T14:45:07+01:00:00	100%	rotkreuz	<a href="https://www.jobscout24.ch">https://www.jobscout24.ch</a>	9, 951, 992
Pizzaiolo	TCS Touring Club Schweiz	2025-11-21T15:02:11+01:00:00	50% - 80%	flaach	<a href="https://www.jobscout24.ch">https://www.jobscout24.ch</a>	9, 948, 232
Verantwortliche	TCS Touring Club Schweiz	2025-11-21T15:02:04+01:00:00	50%	ostermundigen	<a href="https://www.jobscout24.ch">https://www.jobscout24.ch</a>	9, 950, 469
Un.e aide infirmier.ère	EHNV	2025-11-21T15:02:06+01:00:00	80% - 100%	site de chamblon	<a href="https://www.jobscout24.ch">https://www.jobscout24.ch</a>	9, 954, 226
Buchhalter	Weisser Immobilien AG	2025-11-21T15:21:09+01:00:00	80% - 100%	zürich	<a href="https://www.jobscout24.ch">https://www.jobscout24.ch</a>	9, 949, 163

For reasons of readability, job titles and URL's have been shortened

Table 2: Structure of Data after treatment (Rows 1-5 shown)

title	company	quota	location	url	job_id	publishing_date	publishing_time	canton	quota_clean	quota_low	quota_high
Diagnosetechniker In	Garage Wismer AG	100%	rotkreuz	<a href="https://www.jobscout24.ch">https://www.jobscout24.ch</a>	9, 951, 992	2025-11-20	2025-11-20 13:45:07		100	100	100
Pizzaiolo	TCS Touring Club Schweiz	50% - 80%	flaach	<a href="https://www.jobscout24.ch">https://www.jobscout24.ch</a>	9, 948, 232	2025-11-21	2025-11-21 14:02:11	ZH	50-80	50	80
Verantwortliche	TCS Touring Club Schweiz	50%	ostermundigen	<a href="https://www.jobscout24.ch">https://www.jobscout24.ch</a>	9, 950, 469	2025-11-21	2025-11-21 14:02:04	BE	50	50	50
Un.e aide infirmier.ère	EHNV	80% - 100%	site de chamblon	<a href="https://www.jobscout24.ch">https://www.jobscout24.ch</a>	9, 954, 226	2025-11-21	2025-11-21 14:02:06		80-100	80	100
Buchhalter	Weisser Immobilien AG	80% - 100%	zürich	<a href="https://www.jobscout24.ch">https://www.jobscout24.ch</a>	9, 949, 163	2025-11-21	2025-11-21 14:21:09	ZH	80-100	80	100

For reasons of readability, job titles and URL's have been shortened