Hungry for Data: What 2 Years of Scraping Cafeteria Data Taught Me

Paul Gies · 18. Mai 2026

Über Politik aus der Wissenschaft

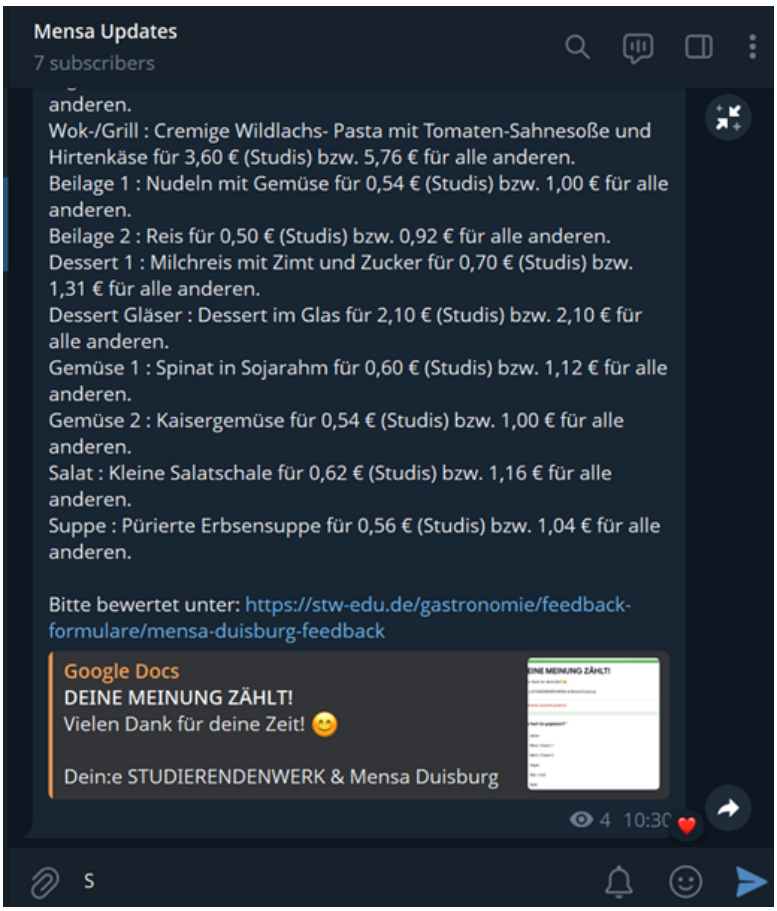
Usually, we talk serious business on this blog: the rise of the far right, war, political violence, pandemics, distrust; all topics hardly palatable in today's world of constant bad news and unpleasant social media feeds.

Therefore, I decided to bring our readers a slightly less serious post today. In this blog, I will talk about how a Python course by my colleagues (instructors at the time) during my political science master's program, inspired me to create a hobby project that scrapes our university's cafeteria homepage and sends daily updates to my phone. Moreover, I will do some data analysis on the 2 years' worth of food data I collected in the process. I have shared this dataset on my Github and hope that some of you will do data analysis of your own with it, as there are many more fun possibilities I did not engage with. If you discover cool findings, please mail them to me at paul.gies@uni-due.de.

But enough with the prerequisites, let's talk food: Back in 2023, my university offered a Python course for beginners in the social sciences. I had some experience with data analysis in R and Stata but had not really done anything beyond that. In this great course, we learned the basics of python, some webscraping, natural language processing workflows and a little data analysis. I was hooked. Really hooked. So, from that course on, I began tinkering with some hobby projects in my free time. One of my first, and longest running, of these hobby projects was the "Mensabot": a little webapp that scrapes our university's cafeteria website once a week, saves the menu as a .csv file and pushes daily updates to my phone in the morning. While the code is undoubtedly janky, it still works, two years after its first deployment. As they say "Never change a running system" (one week after writing that sentence the cafeteria rolled out a new website so I

had to rewrite it).

Figure 1: The daily updates via telegram



Besides being well informed on my daily dietary options, this project amassed the most comprehensive dataset on cafeteria food in the campus Duisburg Mensa I know of to date. For a long time I thought I should do something with this data. So, without further ado, let me present to you some interesting findings about our cafeteria!

OK, first of all let’s look into what data I actually have. Below, you can see the days per week with missing data. Since the cafeteria is closed on Saturday and Sunday, 2 is our baseline here. You might notice a few 3s. This is due to the fortunate number of holidays sprinkled across the German calendar year. Then you will likely notice the five weeks of fully missing data. The first one (and the 5-day peak shortly after) were caused by the bot being somewhat unreliable in the beginning, which I eventually got around to fixing. Since then, the bot has been running like a clockwork. Well, to some extent… Two larger periods of missing data can be explained by the holidays between years. Neither mine,

nor the bot's fault. Then there are two more periods, one around July of last year and one at the end of October. The latter one was caused by an outage of the cafeteria's homepage. I did not notice soon enough. Regarding the former one, well, I embarrassingly forgot to top up the prepaid budget of the cloud provider I used back then for saving the data. It took me two weeks to figure out why the bot was broken all of a sudden. So, to anyone interested in a longtime data collection project, I cannot stress two important lessons enough: a) set up notifications for failures and b) pay your bills.

Figure 2: Days with missing data per week

OK, so now you know what the dataset is and what it's lacking. Let's dive into the food itself:

Figure 3: Wordcloud generated using the names of main dishes



As is quite clear, the Mensa is a consistent provider of fries ("pommes frites"), vegetables ("gemüse", "paprika", "tomaten", "salat" and "zwiebeln") and pan-fried dishes ("gebraten"). Let's take a closer look at what this entails. Figure 4 shows the top 3 dishes

per category in the four main dish categories (I used the OpenAI API to unify minor changes, such as when the cafeteria misspelled a dish, i.e. ‚Kartoffel-Endivien-Stampf mit 3 Green Mounten Balls‘ to ‚Kartoffel-Endivien-Stampf mit 3 Green Mountain Balls‘. I also used the API to auto-translate labels into English for the blogpost).

Figure 4: Top 3 main dishes per category

By far the most common dish appears to be the vegan lentil lasagna, with 35 occurrences during the 2 years. All other dishes appear to be quite close to each other, hovering around 15-20 servings during the data collection. This bears the question: do vegans get less variety in their diet options (of course they do, because they only get to pick from one category, so more precisely: is the vegan category less varied than the other ones)?

The opposite is actually true. Besides the Wok-/Grill category (which is supposed to serve seasonal special dishes), the vegan category has the largest number of unique dishes. The cafeteria’s “vegan friendly”-badge awarded by PETA is seemingly deserved.

Figure 5: Number of unique main dishes per category

So far, we have omitted the most important factor from discussions around a student cafeteria: the prices. In a world of economic crisis and high inflation, how does our cafeteria fare?

Figure 6: Prices across dish categories and status groups

The median prices are lowest for the non-veggie “Essen 1” category. The cheapest (based on student prices) main dish in the dataset were the “Farmer’s potatoes with egg and bacon” on November 18th of last year, while the most expensive was the Duckleg and -breast served 7 days prior. What this plot additionally shows is that it really does pay to be a student. The average difference in prices ranges from 1.94€ (“Essen 1”) to 2.33€ (“Essen 2”). Assuming 5 meals a week during the 16 weeks per semester, that equates to a discount of around 160€, a good deal compared to 125€ paid to the Studierendenwerk from the tuition fees. Mind you, this does not even include savings on side dishes or desserts!

However, while students arguably get a good deal, inflation has clearly gotten hold of the cafeteria as well. Both for faculty and students alike, prices have increased quite sharply over the past 2 years (except for students' Wok-Grill dishes, but this category is somewhat special since dishes are rotated so often). Overall, the mean price for a main dish for students over all categories increased from 2.82€ in 2024 to 3.06€ in 2026 (so far). For faculties, the mean prices increased from 4.82€ to 5.24€. While this may not seem like a large increase, calculated against the historical German inflation rate between 2024 to 2026 (2.23% in 2024 and 2.18% in 2025), this increase seems rather high. Of course, the German inflation rate is based on other products than just food, so this comparison might be a little unfair. What stands however, is that prices have been rising in the cafeteria over the past 2.5 years.

Figure 7: Monthly average prices for student's main dishes over time

Figure 8: Monthly average prices for faculty's main dishes over time

So, to reminisce, this project was (and still is) a fun side project. While deeply unserious, it nevertheless is a nice showcase ofhow skills from a political science program can transfer to highly unrelated fields. If anyone wants to give "Mensa-Analysis" their own shot, I provide the data on my Github. This blog only provided some very shallow analysis I ran in my spare time. Proper digging in the data likely reveals a lot of more interesting finds. If you manage to dig them up, hit me up at paul.gies@uni-due.de. I did dabble in trying to forecast Mensa food, however no classifier I used was able to predict even a single dish correctly. So if you are able to build a food forecasting algorithm, hit me up!

Stay up to date: Join our mailing list and get a short email whenever a new post goes live, usually every three weeks. No ads, and you can unsubscribe at any time. Sign up here

---

Paul Gies (2026): Hungry for Data: What 2 Years of Scraping Cafeteria Data Taught Me. Empirische Politik UDE.
https://www.politik-wissenschaft.org/2026/05/18/hungry-for-data/
Lesefassung aus dem Onlinebeitrag. Interaktive Inhalte sind in der Onlinefassung verfügbar.