

Data Stewardship – Assignment 1:

Part 1 - data science use case

Experiment overview

Given to the huge amount of data available, I picked a topic first and then looked for some data then. I focused on financial data and found some interesting numbers and will look deeper into it. The goal is to visualize the data and find some dependencies between the data sets afterwards. Later, if not satisfied, we can add more data to our experiment.

- Net international investment position - quarterly data, % of GDP
 - <https://data.europa.eu/euodp/en/data/dataset/Zaa87C9KaW0Oc7V1BeXRA>
- c - quarterly data, million units of national currency
 - <https://data.europa.eu/euodp/en/data/dataset/XPGtzRQl1bar3lO50yMYGg>
- Direct investment - quarterly data, million units of national currency
 - <https://data.europa.eu/euodp/en/data/dataset/Yu9NbJr3KNmCNuAcrrzWfQ>
- General government gross debt (EDP concept), consolidated - quarterly data
 - <https://data.europa.eu/euodp/en/data/dataset/uEVcriXpO2FOSnwlSkt4Q>

First I download the data and import into a Microsoft SQL Server to do further transformations. This will result in 4 tables.

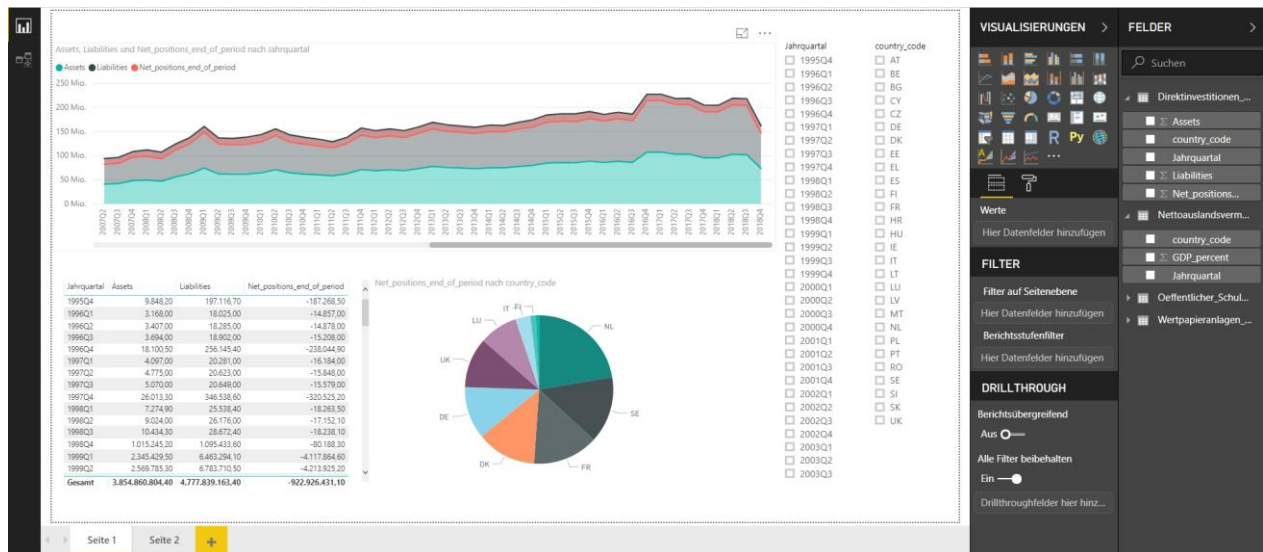
Imported data:

	currency_bop_item_sector10.sectpart.stk_flow.partner.geo time	1995Q4	1996Q1	1996Q2	1996Q3
22	MIO_NAC.FA__D__F.S1.S1.A.LE.WRL_REST.PL	:	:	:	:
23	MIO_NAC.FA__D__F.S1.S1.A.LE.WRL_REST.PT	:	3168.0	3407.0	3694.0
24	MIO_NAC.FA__D__F.S1.S1.A.LE.WRL_REST.RO	:	:	:	:
25	MIO_NAC.FA__D__F.S1.S1.A.LE.WRL_REST.SE	:	:	:	:
26	MIO_NAC.FA__D__F.S1.S1.A.LE.WRL_REST.SI	658.2	:	:	:
27	MIO_NAC.FA__D__F.S1.S1.A.LE.WRL_REST.SK	:	:	:	:
28	MIO_NAC.FA__D__F.S1.S1.A.LE.WRL_REST.UK	:	:	:	:
29	MIO_NAC.FA__D__F.S1.S1.L.LE.WRL_REST.AT	:	:	:	:
30	MIO_NAC.FA__D__F.S1.S1.L.LE.WRL_REST.BE	:	:	:	:
31	MIO_NAC.FA__D__F.S1.S1.L.LE.WRL_REST.BG	:	:	:	:
32	MIO_NAC.FA__D__F.S1.S1.L.LE.WRL_REST.CY	:	:	:	:

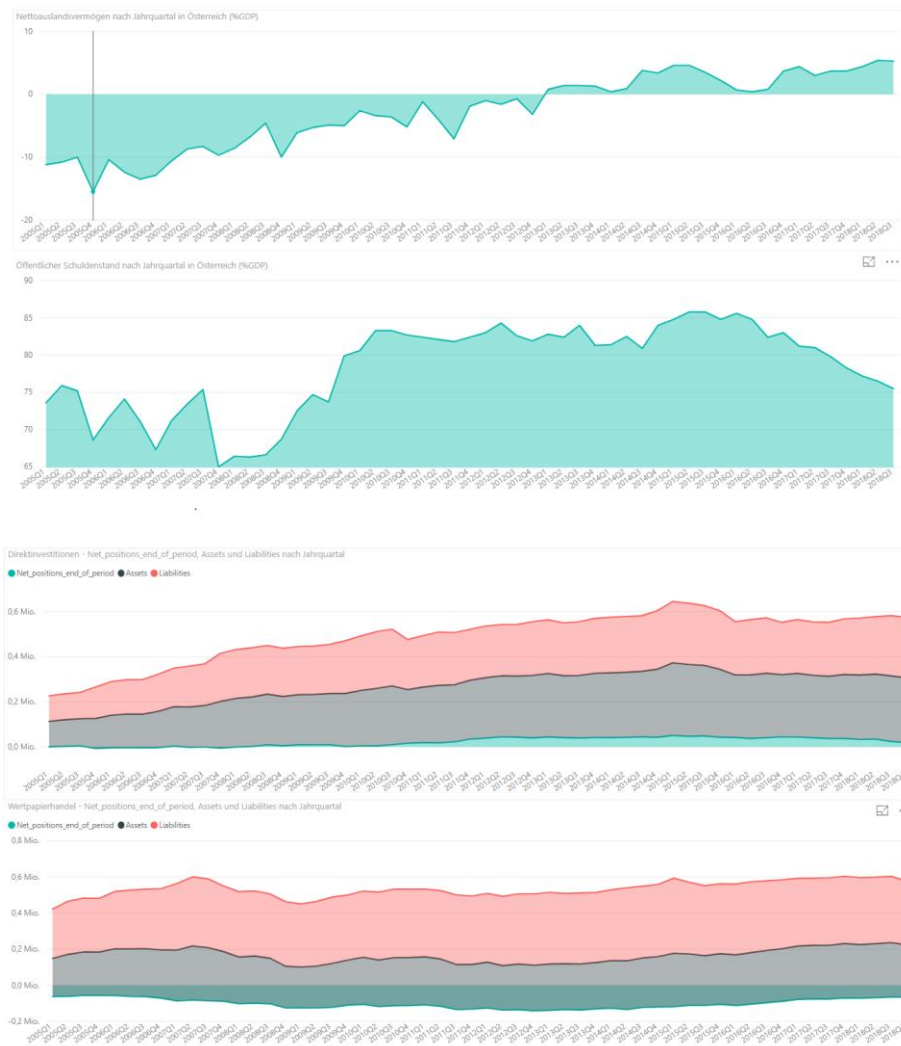
Transformed data:

	Jahrquartal	country_code	Net_positions_end_of_period	Assets	Liabilities
10	1998Q1	AT	NULL	NULL	NULL
11	1998Q2	AT	NULL	NULL	NULL
12	1998Q3	AT	NULL	NULL	NULL
13	1998Q4	AT	NULL	NULL	NULL
14	1999Q1	AT	-3922.00	18678.00	22600.00
15	1999Q2	AT	-4012.00	19266.00	23278.00
16	1999Q3	AT	-3520.00	20023.00	23543.00
17	1999Q4	AT	-3518.00	22679.00	26197.00
18	2000Q1	AT	-3813.00	23257.00	27070.00
19	2000Q2	AT	-2617.00	23897.00	26514.00
20	2000Q3	AT	-1507.00	25983.00	27490.00
21	2000Q4	AT	-5594.00	29773.00	35367.00

After we have transformed the data we can use Microsoft Power BI Desktop to visualize the data. We create a Power BI page for each dataset and use some interactive tools to get an idea of the content.



We can compare the international investments in percent of the GDP and the general government gross debt in percent of the GDP in Austria. (data used from 2005 to now)



Part 2 – data management

Github:

https://github.com/Darigas1/data_stewardship_01127076

<https://doi.org/10.5281/zenodo.2648777>

Raw data DOI:

<https://doi.org/10.5281/zenodo.2648201>

DMP:

<https://doi.org/10.5281/zenodo.2648724>

Part 3 – compare DMPs

DMP:

<https://doi.org/10.5281/zenodo.2648781>

Level of details and scope of DMPs

I think dmponline gives a deeper level of details because you can write some text and ds-wizzard is more like a checklist to know what to do.

Quality and information provided in generated DMPs

I think both is good, there are some strange questions at the ds-wizzard and there is not as much information that dmponline provides.

Guidance provided by tools

Both are easy to use but I think they do not have the same goals.

Time needed to create a DMP

Ds-wizzard is faster because you do not have to write any text. The tools help a lot to save some time but if you want a professional dmp, it will take a lot of it.

Complexity of the DMP creation process

Pretty simple.

Functionality of tools

Both work well. It's easy to use and can be exported as well.

Which tool would you use for your future DMPs and why?

I would use dmponline for future dmps because it contains much more information. But maybe I would use ds-wizzard for developing the dmp like a checklist. To keep in mind what the dmp is about and check out if all needs are met.

Whom would you recommend to use DMP Online and whom the other tool?

For beginner I would recommend the ds-wizzard, you can get an idea of the process without a lot of trouble. But for real use I always would recommend dmponline, it just gives you more possibilities.

Other comments/remarks/differences

I think the questions of the ds-wizzard questionnaire doesn't really fit in every project. Dmponline is more flexible.s