

Using R to extract and structure portfolio data

R Meetup Luzern

Thomas Spycher
31.03.2021



Supported by:

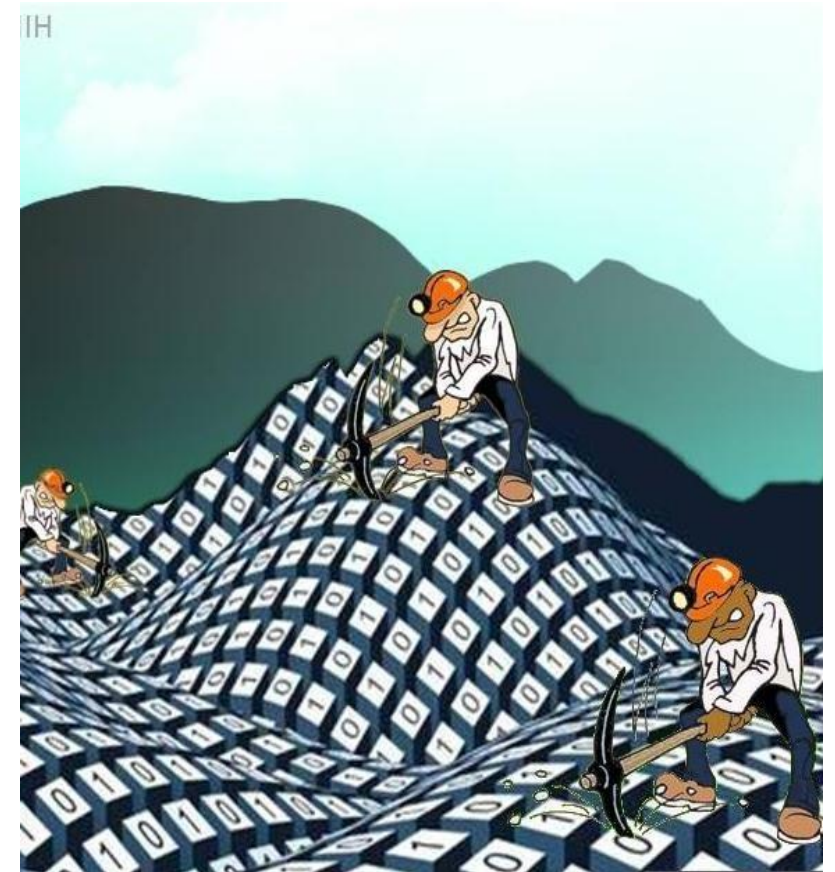


Stiftung für technologische Innovation
Fondation pour l'innovation technologique
Foundation for technological innovation



Background

- Studies in **Economics** with focus on **empirical analysis**
- First experiences in **econ/finance & data**
- PhD with focus on household financial decisions & applied econometrics
- Entrepreneurship: Founded DataCareer (datacareer.de) – career platform for data science
- Left academia in 2018; co-founder of Novalytica AG



Our team combines passion and expertise in data science with a strong understanding for economics and finance

- Data science startup in the area of data collection, data visualization and data modelling
- Team of 8 people + Freelancers
- Clients in the area of real estate, finance, marketing/campaign planning and the public sector – often work with georeferenced data
- Full range of services from data collection through analysis and modelling to visualization and BI



ALPHAPROP

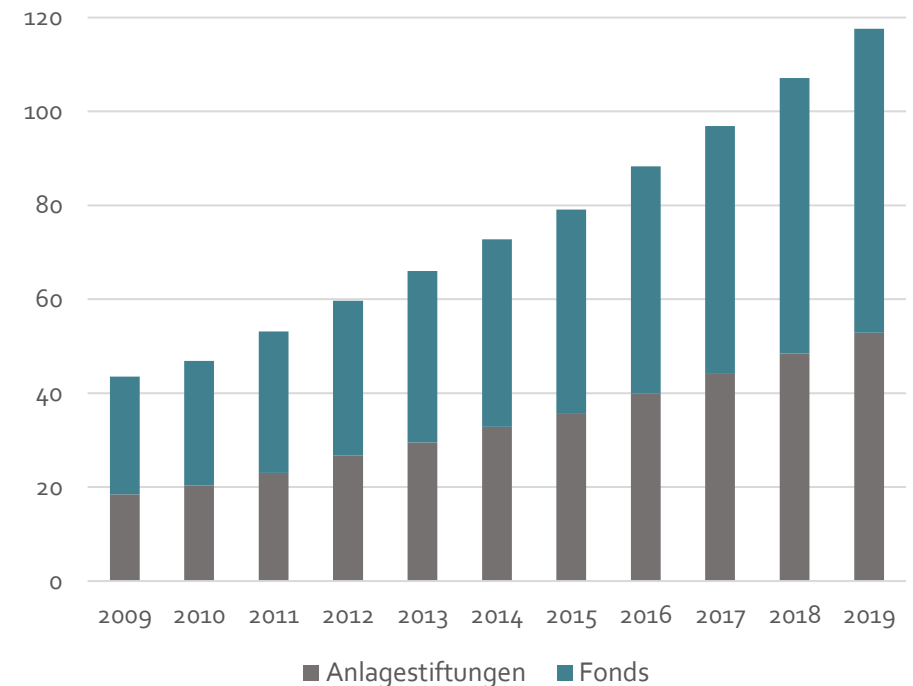
Indirect real estate investments made transparent

The market for indirect (Swiss) real estate investments

- Approx. CHF 150bn in Assets in real estate stocks, funds and investment foundations
- As comparison: US REITs market is approx. 1.3tr
- Swiss Pension funds held in 2018 20.3% of their total assets of CHF 876bn in real estate
- Swiss Pension funds hold about half of RE holdings in indirect assets → approx. CHF 87 bn.

Indirect Swiss Real Estate Investment

Property value in bn CHF



The market for indirect (Swiss) real estate investments

- Approx. CHF 150bn in Assets in real estate stocks, funds and investment foundations
- As comparison: US REITs market is approx. 1.3tr
- Swiss Pension funds held in 2018 20.3% of their total assets of CHF 876bn in real estate
- Swiss Pension funds hold about half of RE holdings in indirect assets → approx. CHF 87 bn.

- For asset managers, that is a tiny market



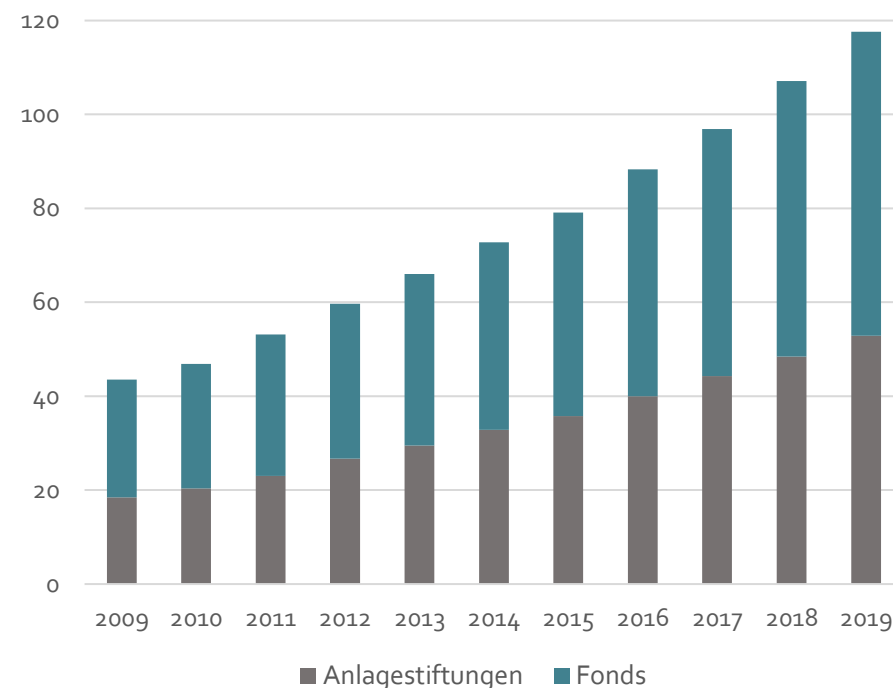
2 trillion in
Market Cap



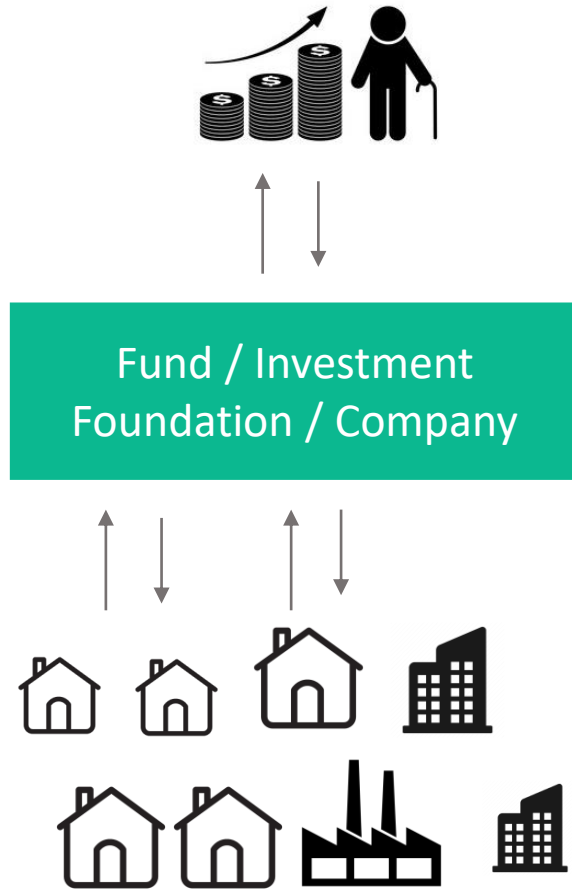
SMI > 1 trillion
in Market Cap

Indirect Swiss Real Estate Investment

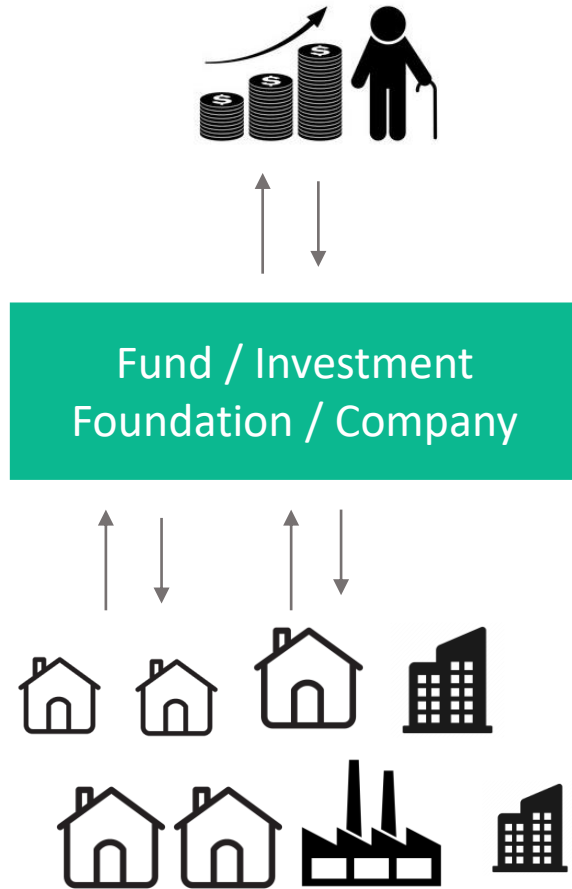
Property value in bn CHF



A very simplified image of indirect real estate investments



A very simplified image of indirect real estate investments



So far, the analysis was focused on the layer in between.

What kind of information is available?

- Prices and performance
- Annual/Other reports



Lack of data: Aside from price and performance data, there is no structured data source available.

Missing overview: It is extremely hard and cumbersome for investors to obtain a reliable big picture and compare to benchmarks

Missing Optimization: No reliable data bases for rule-based asset management or quantitative models.

Data driven decisions for investments in indirect real estate



Data collection & cleaning

Moving from unstructured to structured and clean information



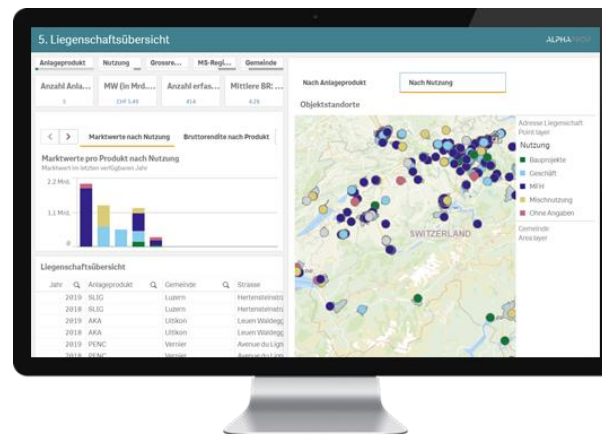
Move to Smart Data

Visualisations, Benchmarking and Reports give new insights.



Connect data and modelling

Connecting location information and information on economic activity.



Product selection

Portfolio analysis

Portfolio-Simulation

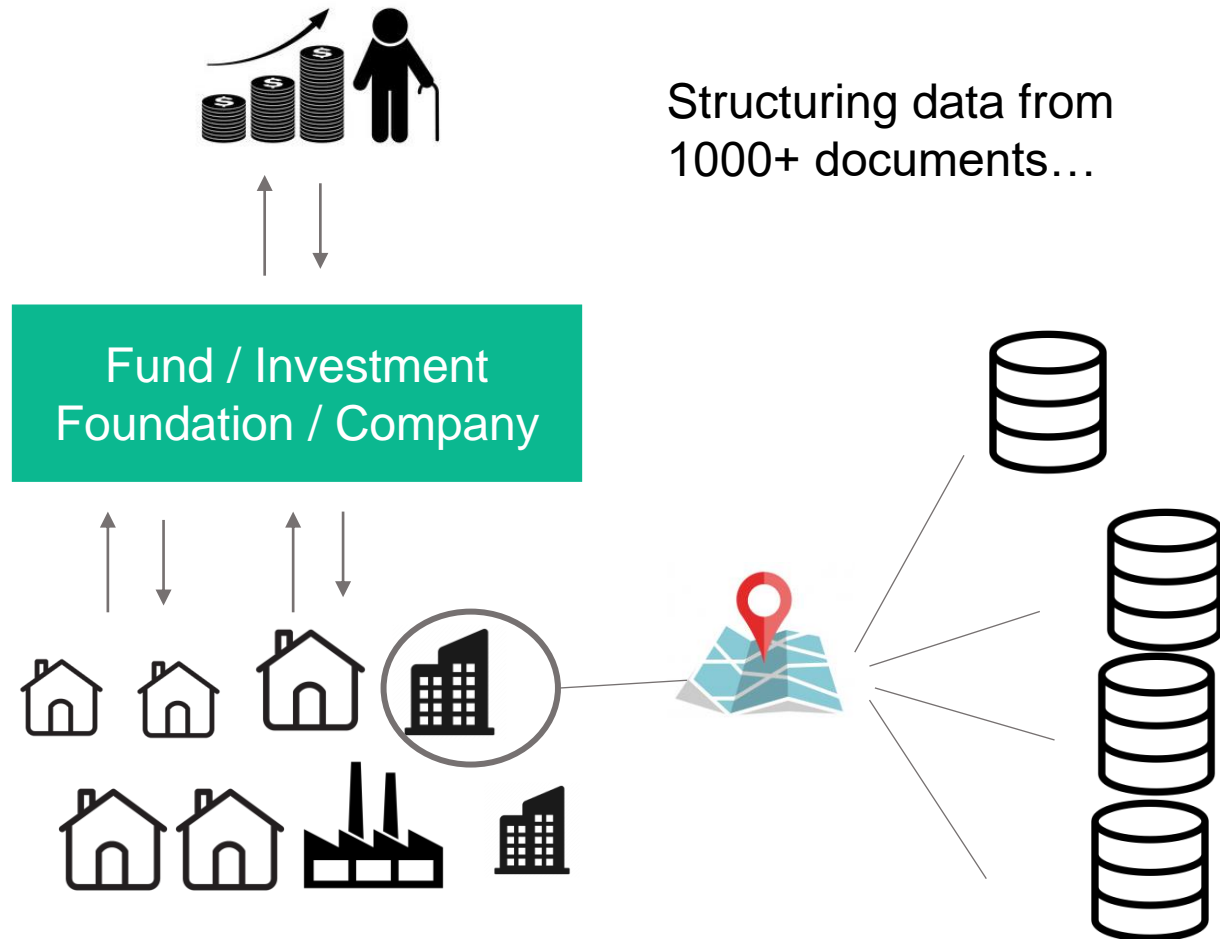
Alphasense Analysis-Tool



Qlik Sense



A very simplified image of indirect real estate investments



INVESTMENT PROPERTIES AS AT 31 DECEMBER 2020

Acquired in current year	Canton	Place	Address	Market value (CHFm)	Annualised full occupancy property rent (CHFm)	Net annualised property rent (CHFm)	Occupancy rate	Site area (sqm)	Ownership type ¹⁾																																																																																																																																																																																																																																																																																													
Residential properties																																																																																																																																																																																																																																																																																																						
GE	Geneva		Rue du Môle 5	15.7	0.6	0.6	100.0%	277	S																																																																																																																																																																																																																																																																																													
GE	Geneva		Rue de la Servette 23	16.2	0.7	0.7	100.0%	421	S																																																																																																																																																																																																																																																																																													
GE	Geneva		Rue Charles Cusin 10	13.9	0.5	0.4	87.9%	279	S																																																																																																																																																																																																																																																																																													
GE	Geneva		Rue de Bâle 28/30	19.8	0.7	0.5	80.5%	1,230	S																																																																																																																																																																																																																																																																																													
GE	Geneva		Rue des Asters 8	7.0	0.3	0.3	100.0%	302	S																																																																																																																																																																																																																																																																																													
GE	Geneva	F F																																																																																																																																																																																																																																																																																																				
<table><tr><th></th><th></th><th>Grundstückfläche in m²</th><th>1- und 1½- Zimmer-Wohnungen</th><th>2- und 2½- Zimmer-Wohnungen</th><th>3- und 3½- Zimmer-Wohnungen</th><th>4- und 4½- Zimmer-Wohnungen</th><th>5- und mehr Zimmer-Wohnungen</th><th>Parking</th><th>Diverse</th><th>Kommerzielle Nutzungsflächen in m²</th><th>Buchwert CHF</th><th>Versicherungswert CHF</th><th>Marktwert CHF</th><th>Mietetrag netto CHF</th></tr><tr><td>Basel</td><td>Altkircherstrasse 15</td><td>366</td><td>3</td><td></td><td>7</td><td>1</td><td>1</td><td></td><td></td><td></td><td>4'182'400</td><td>2'703'000</td><td>5'161'000</td><td>174'554</td></tr><tr><td>Basel</td><td>Hechtliacker 44</td><td>BR</td><td></td><td>32</td><td>32</td><td>16</td><td>1</td><td>89</td><td></td><td>176</td><td>24'787'000</td><td>19'628'000</td><td>29'610'000</td><td>1'391'774</td></tr><tr><td>Basel</td><td>Herrengabenweg 14</td><td>384</td><td></td><td>3</td><td>8</td><td></td><td></td><td>1</td><td></td><td>35</td><td>4'092'400</td><td>2'534'000</td><td>5'152'000</td><td>173'602</td></tr><tr><td>Basel</td><td>Pilgerstrasse 8-16</td><td>2'239</td><td>7</td><td>12</td><td>18</td><td></td><td></td><td>34</td><td></td><td>40</td><td>13'317'900</td><td>11'614'000</td><td>17'160'000</td><td>595'148</td></tr><tr><td>Bern</td><td>von-Gunten-Strasse 7</td><td>BR</td><td></td><td>5</td><td>5</td><td>2</td><td>1</td><td>11</td><td></td><td></td><td>6'077'100</td><td>3'430'000</td><td>6'430'000</td><td>251'294</td></tr><tr><td>Biel</td><td>Hans-Hugi-Strasse 5, 7, Silbergasse 2, 4</td><td>1'293</td><td>29</td><td>15</td><td>9</td><td>14</td><td></td><td>42</td><td></td><td>1'198</td><td>21'858'600</td><td>20'900'000</td><td>24'850'000</td><td>1'076'818</td></tr><tr><td>Chiasso</td><td>Via B. Fontana 6, 8, 10</td><td>1'146</td><td></td><td>12</td><td>18</td><td></td><td>6</td><td>33</td><td>3</td><td>370</td><td>6'616'300</td><td>8'519'000</td><td>6'930'000</td><td>353'442</td></tr><tr><td>Dübendorf</td><td>Am Wasser 6, 15</td><td>3'191</td><td>8</td><td></td><td>8</td><td>16</td><td>10</td><td>20</td><td>4</td><td>826</td><td>28'484'800</td><td>17'846'000</td><td>38'510'000</td><td>1'198'790</td></tr><tr><td>Ecublens</td><td>Avenue du Tir Fédéral 38, 40</td><td>5'448</td><td>2</td><td>32</td><td>30</td><td>16</td><td></td><td>85</td><td></td><td></td><td>16'923'400</td><td>14'227'417</td><td>22'070'000</td><td>931'895</td></tr><tr><td>Kerzers</td><td>Widacker 1, 3</td><td>1'983</td><td></td><td>8</td><td>6</td><td>6</td><td></td><td>29</td><td></td><td></td><td>4'032'200</td><td>5'878'100</td><td>4'690'000</td><td>208'613</td></tr><tr><td>La Tour-de-Peilz</td><td>Chemin des Pléiades 8</td><td>2'135</td><td></td><td>6</td><td>6</td><td></td><td>3</td><td>17</td><td></td><td>214</td><td>4'688'400</td><td>4'462'344</td><td>6'359'000</td><td>264'327</td></tr><tr><td>Lyss</td><td>Knospenweg 2, 4, 4a, 6, 6a, 8</td><td>3'731</td><td></td><td>3</td><td>23</td><td>14</td><td>1</td><td>33</td><td></td><td>10</td><td>9'743'300</td><td>9'936'400</td><td>12'560'000</td><td>527'088</td></tr><tr><td>Nidau</td><td>Lyss-Strasse 65, 67, 69, 71</td><td>4'448</td><td></td><td>16</td><td>22</td><td>22</td><td></td><td>57</td><td></td><td></td><td>14'705'500</td><td>16'110'707</td><td>15'850'000</td><td>720'494</td></tr><tr><td>Onex</td><td>Route de Chancy 130-134, Chemin de l'Auberge 2-4, Route de Loëx 3C-E, 5, 5bis, 5ter</td><td>5'366</td><td></td><td>6</td><td>6</td><td>19</td><td>12</td><td>22</td><td></td><td>774</td><td>25'834'900</td><td>11'987'700</td><td>23'010'000</td><td>229'672</td></tr><tr><td>Riehen</td><td>Wasserstelzenweg 60</td><td>1'287</td><td></td><td>4</td><td>4</td><td>4</td><td></td><td>8</td><td>1</td><td></td><td>4'942'900</td><td>3'386'000</td><td>6'080'000</td><td>190'754</td></tr><tr><td>Schlieren</td><td>Badenerstrasse 96, 98, 106, 108</td><td>2'454</td><td>7</td><td>2</td><td>22</td><td>28</td><td>0</td><td>53</td><td>4</td><td>235</td><td>36'500'000</td><td>29'017'000</td><td>41'870'000</td><td>223'034</td></tr><tr><td>Winterthur</td><td>Dättnerstrasse 131-163</td><td>11'558</td><td>9</td><td>6</td><td>41</td><td>31</td><td>7</td><td>80</td><td>2</td><td></td><td>39'350'600</td><td>24'209'935</td><td>51'750'000</td><td>1'776'816</td></tr><tr><td colspan="2">Total fertige Bauten</td><td>47'029</td><td>65</td><td>162</td><td>265</td><td>189</td><td>42</td><td>614</td><td>14</td><td>3'878</td><td>266'137'700</td><td>206'389'643</td><td>318'042'000</td><td>10'288'115</td></tr></table>												Grundstückfläche in m²	1- und 1½- Zimmer-Wohnungen	2- und 2½- Zimmer-Wohnungen	3- und 3½- Zimmer-Wohnungen	4- und 4½- Zimmer-Wohnungen	5- und mehr Zimmer-Wohnungen	Parking	Diverse	Kommerzielle Nutzungsflächen in m²	Buchwert CHF	Versicherungswert CHF	Marktwert CHF	Mietetrag netto CHF	Basel	Altkircherstrasse 15	366	3		7	1	1				4'182'400	2'703'000	5'161'000	174'554	Basel	Hechtliacker 44	BR		32	32	16	1	89		176	24'787'000	19'628'000	29'610'000	1'391'774	Basel	Herrengabenweg 14	384		3	8			1		35	4'092'400	2'534'000	5'152'000	173'602	Basel	Pilgerstrasse 8-16	2'239	7	12	18			34		40	13'317'900	11'614'000	17'160'000	595'148	Bern	von-Gunten-Strasse 7	BR		5	5	2	1	11			6'077'100	3'430'000	6'430'000	251'294	Biel	Hans-Hugi-Strasse 5, 7, Silbergasse 2, 4	1'293	29	15	9	14		42		1'198	21'858'600	20'900'000	24'850'000	1'076'818	Chiasso	Via B. Fontana 6, 8, 10	1'146		12	18		6	33	3	370	6'616'300	8'519'000	6'930'000	353'442	Dübendorf	Am Wasser 6, 15	3'191	8		8	16	10	20	4	826	28'484'800	17'846'000	38'510'000	1'198'790	Ecublens	Avenue du Tir Fédéral 38, 40	5'448	2	32	30	16		85			16'923'400	14'227'417	22'070'000	931'895	Kerzers	Widacker 1, 3	1'983		8	6	6		29			4'032'200	5'878'100	4'690'000	208'613	La Tour-de-Peilz	Chemin des Pléiades 8	2'135		6	6		3	17		214	4'688'400	4'462'344	6'359'000	264'327	Lyss	Knospenweg 2, 4, 4a, 6, 6a, 8	3'731		3	23	14	1	33		10	9'743'300	9'936'400	12'560'000	527'088	Nidau	Lyss-Strasse 65, 67, 69, 71	4'448		16	22	22		57			14'705'500	16'110'707	15'850'000	720'494	Onex	Route de Chancy 130-134, Chemin de l'Auberge 2-4, Route de Loëx 3C-E, 5, 5bis, 5ter	5'366		6	6	19	12	22		774	25'834'900	11'987'700	23'010'000	229'672	Riehen	Wasserstelzenweg 60	1'287		4	4	4		8	1		4'942'900	3'386'000	6'080'000	190'754	Schlieren	Badenerstrasse 96, 98, 106, 108	2'454	7	2	22	28	0	53	4	235	36'500'000	29'017'000	41'870'000	223'034	Winterthur	Dättnerstrasse 131-163	11'558	9	6	41	31	7	80	2		39'350'600	24'209'935	51'750'000	1'776'816	Total fertige Bauten		47'029	65	162	265	189	42	614	14	3'878	266'137'700	206'389'643	318'042'000	10'288'115
		Grundstückfläche in m²	1- und 1½- Zimmer-Wohnungen	2- und 2½- Zimmer-Wohnungen	3- und 3½- Zimmer-Wohnungen	4- und 4½- Zimmer-Wohnungen	5- und mehr Zimmer-Wohnungen	Parking	Diverse	Kommerzielle Nutzungsflächen in m²	Buchwert CHF	Versicherungswert CHF	Marktwert CHF	Mietetrag netto CHF																																																																																																																																																																																																																																																																																								
Basel	Altkircherstrasse 15	366	3		7	1	1				4'182'400	2'703'000	5'161'000	174'554																																																																																																																																																																																																																																																																																								
Basel	Hechtliacker 44	BR		32	32	16	1	89		176	24'787'000	19'628'000	29'610'000	1'391'774																																																																																																																																																																																																																																																																																								
Basel	Herrengabenweg 14	384		3	8			1		35	4'092'400	2'534'000	5'152'000	173'602																																																																																																																																																																																																																																																																																								
Basel	Pilgerstrasse 8-16	2'239	7	12	18			34		40	13'317'900	11'614'000	17'160'000	595'148																																																																																																																																																																																																																																																																																								
Bern	von-Gunten-Strasse 7	BR		5	5	2	1	11			6'077'100	3'430'000	6'430'000	251'294																																																																																																																																																																																																																																																																																								
Biel	Hans-Hugi-Strasse 5, 7, Silbergasse 2, 4	1'293	29	15	9	14		42		1'198	21'858'600	20'900'000	24'850'000	1'076'818																																																																																																																																																																																																																																																																																								
Chiasso	Via B. Fontana 6, 8, 10	1'146		12	18		6	33	3	370	6'616'300	8'519'000	6'930'000	353'442																																																																																																																																																																																																																																																																																								
Dübendorf	Am Wasser 6, 15	3'191	8		8	16	10	20	4	826	28'484'800	17'846'000	38'510'000	1'198'790																																																																																																																																																																																																																																																																																								
Ecublens	Avenue du Tir Fédéral 38, 40	5'448	2	32	30	16		85			16'923'400	14'227'417	22'070'000	931'895																																																																																																																																																																																																																																																																																								
Kerzers	Widacker 1, 3	1'983		8	6	6		29			4'032'200	5'878'100	4'690'000	208'613																																																																																																																																																																																																																																																																																								
La Tour-de-Peilz	Chemin des Pléiades 8	2'135		6	6		3	17		214	4'688'400	4'462'344	6'359'000	264'327																																																																																																																																																																																																																																																																																								
Lyss	Knospenweg 2, 4, 4a, 6, 6a, 8	3'731		3	23	14	1	33		10	9'743'300	9'936'400	12'560'000	527'088																																																																																																																																																																																																																																																																																								
Nidau	Lyss-Strasse 65, 67, 69, 71	4'448		16	22	22		57			14'705'500	16'110'707	15'850'000	720'494																																																																																																																																																																																																																																																																																								
Onex	Route de Chancy 130-134, Chemin de l'Auberge 2-4, Route de Loëx 3C-E, 5, 5bis, 5ter	5'366		6	6	19	12	22		774	25'834'900	11'987'700	23'010'000	229'672																																																																																																																																																																																																																																																																																								
Riehen	Wasserstelzenweg 60	1'287		4	4	4		8	1		4'942'900	3'386'000	6'080'000	190'754																																																																																																																																																																																																																																																																																								
Schlieren	Badenerstrasse 96, 98, 106, 108	2'454	7	2	22	28	0	53	4	235	36'500'000	29'017'000	41'870'000	223'034																																																																																																																																																																																																																																																																																								
Winterthur	Dättnerstrasse 131-163	11'558	9	6	41	31	7	80	2		39'350'600	24'209'935	51'750'000	1'776'816																																																																																																																																																																																																																																																																																								
Total fertige Bauten		47'029	65	162	265	189	42	614	14	3'878	266'137'700	206'389'643	318'042'000	10'288'115																																																																																																																																																																																																																																																																																								

Tabulizer to extract tables

- Starting from tabula
- <https://github.com/tabulapdf/tabula>

Tabulizer offering (semi)-automation in R

- `library("tabulizer")`

```
library("tabulizer")
f <- system.file("examples", "data.pdf", package = "tabulizer")
# extract table from first page of example PDF
tab <- extract_tables(f, pages = 1) head(tab[[1]])
```

Tabula



Tabula is a tool for liberating data tables locked inside PDF files.

[View the Project on GitHub](https://github.com/tabulapdf/tabula)
tabulapdf/tabula

Download for
Windows

Download for
Mac

View source on
GitHub

Property level: Property level data (1/2) – data extraction

Semi-automated data extraction

Indirect real estate investments (especially funds and investment foundations) offer property level data of their respective real estate portfolio. We extract these data points with a R-script and followup with proper data pre-processing to provide readily accessible and cleaned data tables for our clients.

Ort, Adresse	Gesteuerungs-kosten		Versicherungs-wert		Marktwert		Mietzeins-einnahmen
	CHF	%	CHF	%	CHF	%	CHF %
Kommerziell genutzte Liegenschaften							
Aarau, Rohrerstrasse 100/102	31 757 557	10.06	29 987 000	9.39	34 120 000	11.12	1 505 834 9.29
Aclens, Chemin du Coteau 23	11 339 796	3.59	7 053 400	2.21	11 510 000	3.75	586 901 3.62
Alschwil, Gewerbestrasse 25	10 270 593	3.25	10 100 000	3.16	8 940 000	2.91	596 913 3.68
Bubendorf, Grüngartenstrasse 17/17a	8 456 843	2.68	8 488 000	2.66	8 260 000	2.69	408 131 2.52
Buchs, Amsterdamerweg 6	30 250 627	9.58	41 456 000	12.98	26 940 000	8.78	2 186 863 13.49
Castione, Via alle Cave 20	6 310 893	2.00	4 000 000	1.25	5 830 000	1.90	335 116 2.07
Castione, Via San Gottardo 18a/18d	24 604 773	7.80	18 000 000	5.64	24 440 000	7.97	968 090 5.97
Dietikon, Lerzenstrasse 10	40 863 083	12.94	53 796 000	16.85	34 260 000	11.17	872 485 5.38
Grancia Collina d'Oro, Via al Molino 49	31 776 430	10.07	26 650 000	8.35	33 740 000	11.00	2 085 625 12.87
Lamone TI, Via Industria 16	15 775 227	5.00	8 226 000	2.58	15 600 000	5.08	881 258 5.44
Langenthal, Gaswerkstrasse 33/35	21 238 882	6.73	32 282 600	10.11	19 030 000	6.20	1 300 967 8.03
Lyss, Industriehof 17	18 396 118	5.83	18 150 000	5.68	19 620 000	6.39	1 200 000 7.40
Mestraschweiden, Industriestrasse 6	6 338 634	2.01	5 780 000	1.81	6 580 000	2.14	259 191 1.60
Mühlin, Industriestrasse 3/5	11 079 517	3.51	8 534 000	2.67	10 390 000	3.39	598 717 3.69
Spreitenbach, Limalstrasse 8-12	22 309 223	7.07	20 851 000	6.53	26 120 000	8.51	1 458 010 8.99
Staad, Hauptstrasse 104	16 462 006	5.22	16 086 000	5.04	14 430 000	4.70	389 026 2.40
Weinfelden, Weststrasse 15	8 388 436	2.66	9 907 000	3.10	7 010 000	2.28	469 895 2.90
Total kommerziell genutzte Liegen-schaften	315 608 637		319 347 000		306 820 000		16 103 051

INVESTMENT PROPERTIES AS AT 31 DECEMBER 2020

Acquired in current year	Canton	Place	Address	Market value (CHFm)	Annualised full occupancy property rent (CHFm)	Net annualised property rent (CHFm)	Occupancy rate	Site area (sqm)	Ownership type 1)
Residential properties									
GE	Geneva		Rue du Môle 5	15.7	0.6	0.6	100.0%	277	S
GE	Geneva		Rue de la Servette 23	16.2	0.7	0.7	100.0%	421	S
GE	Geneva		Rue Charles Cusin 10	13.9	0.5	0.4	87.9%	279	S
GE	Geneva		Rue de Bâle 28/30	19.8	0.7	0.5	80.5%	1,230	S
GE	Geneva		Rue des Asters 8	7.0	0.3	0.3	100.0%	302	S
GE	Geneva		Rue Antoine-Carteret 5	25.7	0.9	0.9	99.7%	1,242	S
			Rue du Colombier 11/13						

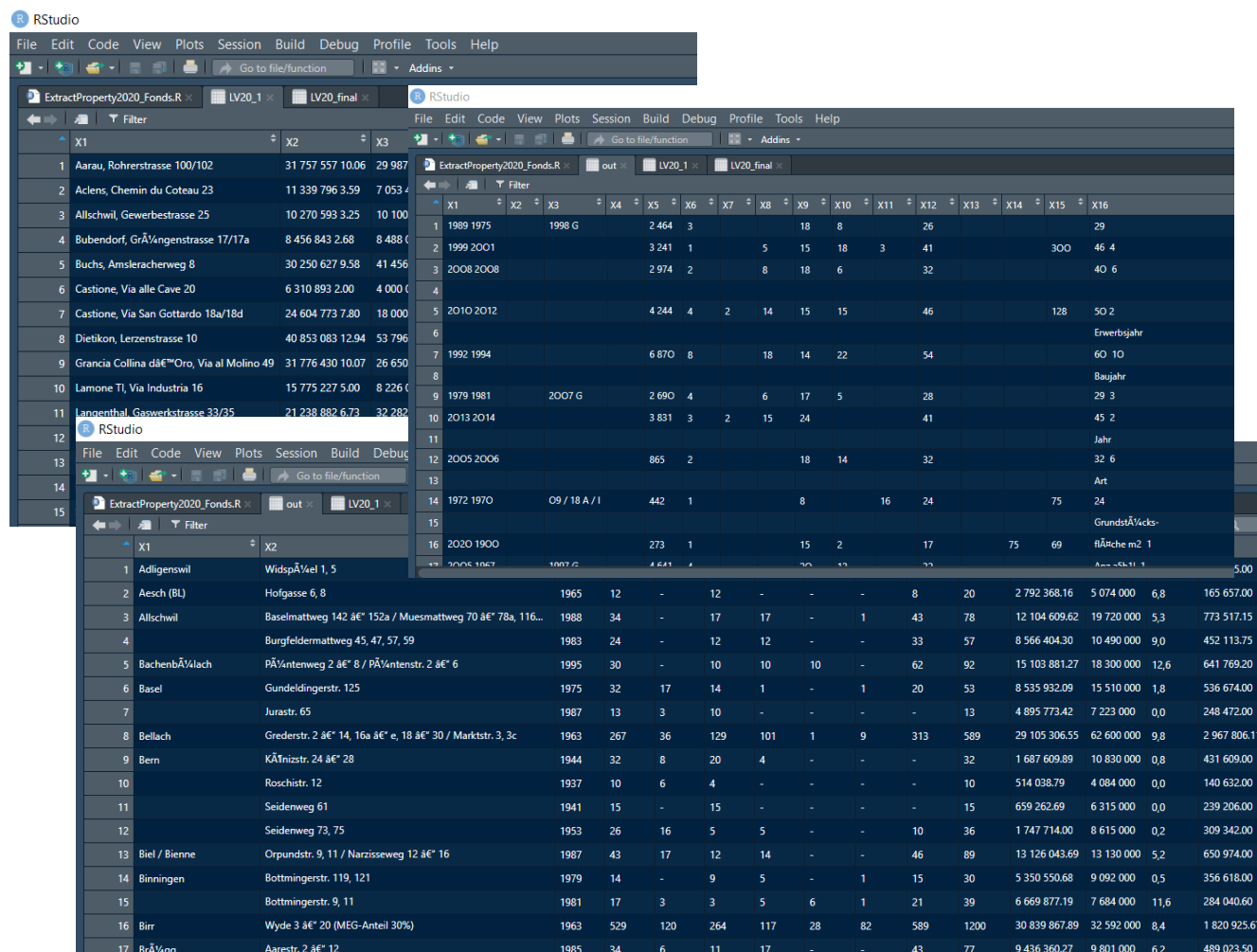
			Grundstückfläche in m²	1. und 1½-Zimmer-Wohnungen	2. und 2½-Zimmer-Wohnungen	3. und 3½-Zimmer-Wohnungen	4. und 4½-Zimmer-Wohnungen	5. und mehr Zimmer-Wohnungen	Parking	Diverse	Kommunale Nutzflächen in m²	Buchwert CHF	Versicherungswert CHF	Marktwert CHF	Mietrend netto CHF
Basel	Altkircherstrasse 15		366	3		7	1	1				4'182'400	2'703'000	5'161'000	174'554
Basel	Hechtliacker 44	BR			32	32	16	1	89		176	24'787'000	19'628'000	29'610'000	1'391'774
Basel	Herrengrabenweg 14		384		3	8			1		35	4'092'400	2'534'000	5'152'000	173'602
Basel	Pfligerstrasse 8-16		2'239	7	12	18			34		40	13'317'900	11'614'000	17'160'000	595'148
Bern	von-Gunten-Strasse 7	BR			5	5	2	1	11			6'077'100	3'430'000	6'430'000	251'294
Biel	Hans-Hugi-Strasse 5, 7, Silbergasse 2, 4		1'293	29	15	9	14		42		1'198	21'858'600	20'900'000	24'850'000	1'076'818
Chiasso	Via B. Fontana 6, 8, 10		1'146		12	18		6	33	3	370	6'618'300	8'519'000	6'930'000	353'442
Dübendorf	Am Wasser 6, 15		3'191	8		8	16	10	20	4	826	28'484'800	17'846'000	38'510'000	1'198'790
Ecublens	Avenue du Tir Fédéral 38, 40		5'448	2	32	30	16		85			16'923'400	14'227'417	22'070'000	931'895
Kerzers	Widacker 1, 3		1'983		8	6	6		29			4'032'200	5'878'100	4'690'000	208'613
La Tour-de-Peilz	Chemin des Pélades 8		2'135		6	6		3	17		214	4'688'400	4'462'344	6'359'000	264'327
Lyss	Knospenweg 2, 4, 4a, 6, 6a, 8		3'731		3	23	14	1	33		10	9'743'300	9'936'400	12'560'000	527'088
Nidau	Lyss-Strasse 65, 67, 69, 71		4'448		16	22	22		57			14'705'500	16'110'707	15'850'000	720'494
Onex	Route de Chancy 130-134, Chemin de l'Auberge 2-4, Route de Loix 3C-E, 5, 5bis, 5ter		5'366		6	6	19	12	22		774	25'834'900	11'987'700	23'010'000	229'672
Riehen	Wasserstelzenweg 60		1'287		4	4	4		8	1		4'942'900	3'386'000	6'080'000	190'754
Schlieren	Badenerstrasse 96, 98, 106, 108		2'454	7	2	22	28	0	53	4	235	36'500'000	29'017'000	41'870'000	223'034
Winterthur	Dättnerstrasse 131-163		11'568	9	6	41	31	7	80	2		39'350'600	24'209'935	51'750'000	1'776'816
Total fertige Bauten			47'029	65	162	265	189	42	614	14	3'878	266'137'700	206'389'643	318'042'000	10'288'115

Property level: Property level data (1/2) – data extraction

Issues

We are mainly working with the tabulizer package which works well to extract the data but due to improper formatting of the report, manual adjustments still often have to be made within the code, e.g. for issues like:

- Two addresses on the same or different lines for the same property.
- Addresses with spelling mistakes
- Figures which are not in the original unit
- Multiple columns extracted as one without separation
- Encoding issues e.g. with inconvenient fonts



The screenshot displays two RStudio windows. The top window shows a table with columns X1, X2, and X3. The bottom window shows a table with columns X1, X2, and X3, and a large number of additional columns (X4 to X16). The data in the bottom table is more complex, including years, numerical values, and some text labels.

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16
1 Aarau, Rohrerstrasse 100/102	31 757 557 10.06	29 987													
2 Aclens, Chemin du Coteau 23	11 339 796 3.59	7 053													
3 Allschwil, Gewerbestrasse 25	10 270 593 3.25	10 100													
4 Bubendorf, Gräbenstrasse 17/17a	8 456 843 2.68	8 488													
5 Buchs, Amsleracherweg 8	30 250 627 9.58	41 456													
6 Castione, Via San Gottardo 18a/18d	6 310 893 2.00	4 000													
7 Castione, Via San Gottardo 18a/18d	24 604 773 7.80	18 000													
8 Dietikon, Lerzenstrasse 10	40 853 083 12.94	53 796													
9 Grancia Collina d'oro, Via al Molino 49	31 776 430 10.07	26 650													
10 Lamone TI, Via Industria 16	15 775 227 5.00	8 226													
11 Lenggathal, Gewerbestrasse 33/35	21 238 882 6.73	32 282													
12															
13															
14															
15															
1 Adligenswil	Widspäuel 1, 5														
2 Aesch (BL)	Hofgasse 6, 8														
3 Allschwil	Baselmattweg 142 à 152a / Muesmattweg 70 à 78a, 116...														
4	Burgfeldermattweg 45, 47, 57, 59														
5 Bachenbühlach	Pfaffenweg 2 à 8 / Pfaffenstr. 2 à 6														
6 Basel	Gundeldingerstr. 125														
7	Jurastr. 65														
8 Bellach	Gredenstr. 2 à 14, 16a à 30 / Marktstr. 3, 3c														
9 Bern	Kätnizstr. 24 à 28														
10	Roschiistr. 12														
11	Seidenweg 61														
12	Seidenweg 73, 75														
13 Biel / Bienne	Orpundstr. 9, 11 / Narzisseweg 12 à 16														
14 Binningen	Bottmingerstr. 119, 121														
15	Bottmingerstr. 9, 11														
16 Birr	Wyde 3 à 20 (MEG-Anteil 30%)														
17 Brühl	Aarestr. 2 à 12														

Cleaning: Property level data (2/3) – data cleaning

Cleaning Process

Pre-defined cleaning functions do most of the work
– but some issue remain...

- besides the data extraction, the data needs to be formatted so it matches the existing database
- Data quality checks



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

ExtractProperty2020_Fonds.R LV20_1 LV20_final

Filter

	X1	X2	X3	X4	X5	X6
1	Aarau, Rohrerstrasse 100/102	31 757 557 10.06	29 987 000 9.39		34 120 000 11.12	1 505 834 9.29
2	Aclens, Chemin du Coteau 23	11 339 796 3.59	7 053 400 2.21		11 510 000 3.75	586 931 3.62
3	Allschwil, Gewerbestrasse 25	10 270 593 3.25	10 100 000 3.16		8 940 000 2.91	596 913 3.68
4	Bubendorf, Gräbigenstrasse 17/17a	8 456 843 2.68	8 488 000 2.66		8 260 000 2.69	408 131 2.52
5	Buchs, Amsleracherweg 8	30 250 627 9.58	41 456 000 12.98		26 940 000 8.78	2 186 863 13.49
6	Castione, Via alle Cave 20	6 310 893 2.00	4 000 000 1.25		5 830 000 1.90	335 116 2.07
7	Castione, Via San Gottardo 18a/18d	24 604 773 7.80	18 000 000 5.64		24 440 000 7.97	968 090 5.97
8	Dietikon, Lerzenstrasse 10	40 853 083 12.94	53 796 000 16.85		34 260 000 11.17	872 485 5.38
9	Grancia Collina d'Orò, Via al Molino 49	31 776 430 10.07	26 650 000 8.35		33 740 000 11.00	2 085 625 12.87
10	Lamone TI, Via Industria 16	15 775 227 5.00	8 226 000 2.58		15 600 000 5.08	881 258 5.44
11	Langenthal, Gaswerkstrasse 33/35	21 238 882 6.73	32 282 600			
12	Lyss, Industriering 17	18 396 118 5.83	18 150 000			
13	Meisterschwanden, Industriestrasse 6	6 338 634 2.01	5 780 000			
14	Möhlthlin, Industriestrasse 3/5	11 079 517 3.51	8 534 000			
15	Spreitenbach, Limmatsstrasse 8-12	22 309 223 7.07	20 850 000			



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

ExtractProperty2020_Fonds.R LV20_final

Filter

	Gestehungskosten	Versicherungswert	Marktwert	Sollmietetrag	Typ	Jahr	Sub_Key	GDE	Strasse
1	31757557	29987000	34120000	1505834	Geschäft	2020	SFC	Aarau	Rohrerstrasse 100/102
2	11339796	7053400	11510000	586931	Geschäft	2020	SFC	Aclens	Chemin du Coteau 23
3	10270593	10100000	8940000	596913	Geschäft	2020	SFC	Allschwil	Gewerbestrasse 25
4	8456843	8488000	8260000	408131	Geschäft	2020	SFC	Bubendorf	Grünenstrasse 17/17a
5	30250627	41456000	26940000	2186863	Geschäft	2020	SFC	Buchs	Amsleracherweg 8
6	6310893	4000000	5830000	335116	Geschäft	2020	SFC	Castione	Via alle Cave 20
7	24604773	18000000	24440000	968090	Geschäft	2020	SFC	Castione	Via San Gottardo 18a/18d
8	40853083	53796000	34260000	872485	Geschäft	2020	SFC	Dietikon	Lerzenstrasse 10
9	31776430	26650000	33740000	2085625	Geschäft	2020	SFC	Grancia Collina d'Orò	Via al Molino 49
10	15775227	8226000	15600000	881258	Geschäft	2020	SFC	Lamone TI	Via Industria 16
11	21238882	32282600	19030000	1300967	Geschäft	2020	SFC	Langenthal	Gaswerkstrasse 33/35
12	18396118	18150000	19620000	1200000	Geschäft	2020	SFC	Lyss	Industriering 17
13	6338634	5780000	6580000	259191	Geschäft	2020	SFC	Meisterschwanden	Industriestrasse 6
14	11079517	8534000	10390000	598717	Geschäft	2020	SFC	Möhlthlin	Industriestrasse 3/5
15	22309223	20851000	26120000	1458010	Geschäft	2020	SFC	Spreitenbach	Limmatsstrasse 8-12
16	16462006	16086000	14430000	389026	Geschäft	2020	SFC	Staad	Hauptstrasse 104
17	8388436	9907000	7010000	469895	Geschäft	2020	SFC	Weinfelden	Weststrasse 15

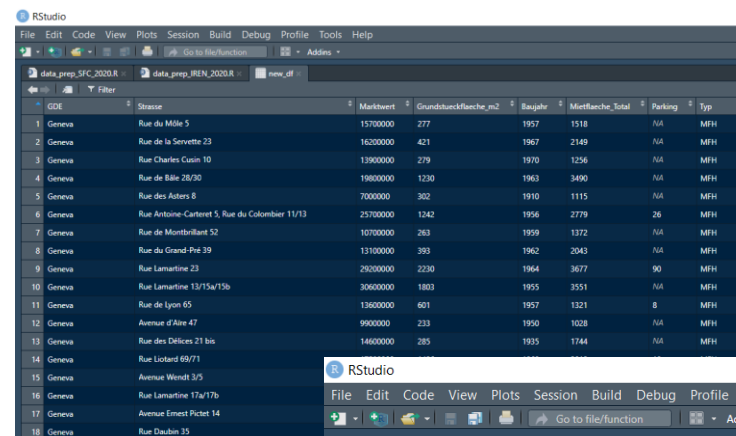
Showing 1 to 17 of 17 entries, 9 total columns

Geocoding: Property level data (3/3) – geocoding

Geocoding

Address data often contain errors such as the assignment of the wrong municipality.

Several automated checks required till we can merge properties with additional information based on the plot



GID	Adresse	Marktwert	Grundstueckflaeche_m2	Baugjahr	Mietflaeche_Total	Parking	Typ
1	Geneva Rue du Môle 5	15700000	277	1957	1518	NA	MFH
2	Geneva Rue de la Servette 23	16200000	421	1967	2149	NA	MFH
3	Geneva Rue Charles Cusin 10	13900000	279	1970	1256	NA	MFH
4	Geneva Rue de Bâle 26/30	19800000	1230	1963	3490	NA	MFH
5	Geneva Rue des Asters 8	7000000	302	1910	1115	NA	MFH
6	Geneva Rue Antoine-Carrellet 5, Rue du Colombier 11/13	25700000	1242	1956	2779	26	MFH
7	Geneva Rue de Montbailant 52	10700000	263	1959	1372	NA	MFH
8	Geneva Rue du Grand-Prie 39	13700000	393	1962	2043	NA	MFH
9	Geneva Rue Lamartine 23	25000000	2230	1964	3677	90	MFH
10	Geneva Rue Lamartine 13/15a/15b	30600000	1803	1955	3551	NA	MFH
11	Geneva Rue de Lyon 65	13600000	601	1957	1321	8	MFH
12	Geneva Avenue d'Alte 47	9900000	233	1950	1028	NA	MFH
13	Geneva Rue des Délices 21 bis	14600000	285	1935	1744	NA	MFH
14	Geneva Rue Lottard 69/71						
15	Geneva Avenue Windt 3/5						
16	Geneva Rue Lamartine 17a/17b						
17	Geneva Avenue Ernest Pictet 14						
18	Geneva Rue Daubin 35						



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Addins

data_prep_SFC_2020.Rdata_prep_IREN_2020.R*out4out3new_df

Filter

	GMDNR	HouseKey	Latitude	Longitude	GMDNAME	MS-Name	Kanton-Name	Grossraum-Name
1	843	IREN, Saanen, gstaadstr68	46.48839	7.264610	Saanen	Saanen-Obersimmental	Bern	Espace Mittelland
2	5516	IREN, Cugy (VD), chdespetitserts1	46.58897	6.643027	Cugy (VD)	Lausanne	Vaud	Région lémanique
3	5586	IREN, Lausanne, avalexandreinet39	46.52663	6.627812	Lausanne	Lausanne	Vaud	Région lémanique
4	5586	IREN, Lausanne, rtealoyfauquez60	46.53506	6.633609	Lausanne	Lausanne	Vaud	Région lémanique
5	5586	IREN, Lausanne, chducloset46810	46.51453	6.632475	Lausanne	Lausanne	Vaud	Région lémanique
6	5586	IREN, Lausanne, avdechallens8789	46.52861	6.614734	Lausanne	Lausanne	Vaud	Région lémanique
7	5586	IREN, Lausanne, avdouchy7274	46.50749	6.626808	Lausanne	Lausanne	Vaud	Région lémanique
8	5586	IREN, Lausanne, chdeslys14	46.52897	6.647353	Lausanne	Lausanne	Vaud	Région lémanique
9	5586	IREN, Lausanne, rtealoyfauquez122124	46.54042	6.633786	Lausanne	Lausanne	Vaud	Région lémanique
10	5586	IREN, Lausanne, chdelachapelle2	46.57895	6.604951	Lausanne	Lausanne	Vaud	Région lémanique
11	5586	IREN, Lausanne, chdemontmeillan1921	46.52832	6.638513	Lausanne	Lausanne	Vaud	Région lémanique
12	5586	IREN, Lausanne, ruedumapas61636567	46.52583	6.623763	Lausanne	Lausanne	Vaud	Région lémanique
13	5586	IREN, Lausanne, placeduvallon1	46.52684	6.638561	Lausanne	Lausanne	Vaud	Région lémanique
14	5586	IREN, Lausanne, placedelanavigation46	46.50709	6.626866	Lausanne	Lausanne	Vaud	Région lémanique
15	5586	IREN, Lausanne, avdesoiseaux1517	46.53225	6.629004	Lausanne	Lausanne	Vaud	Région lémanique
16	5586	IREN, Lausanne, avvictoruffy33	46.52774	6.648201	Lausanne	Lausanne	Vaud	Région lémanique
17	5588	IREN, Paudex, rtedelabernadaz1	46.50791	6.671861	Paudex	Lausanne	Vaud	Région lémanique
18	5589	IREN, Prilly, ruedelacombette2224	46.53004	6.603333	Prilly	Lausanne	Vaud	Région lémanique

Feeding data into the application...

RStudio

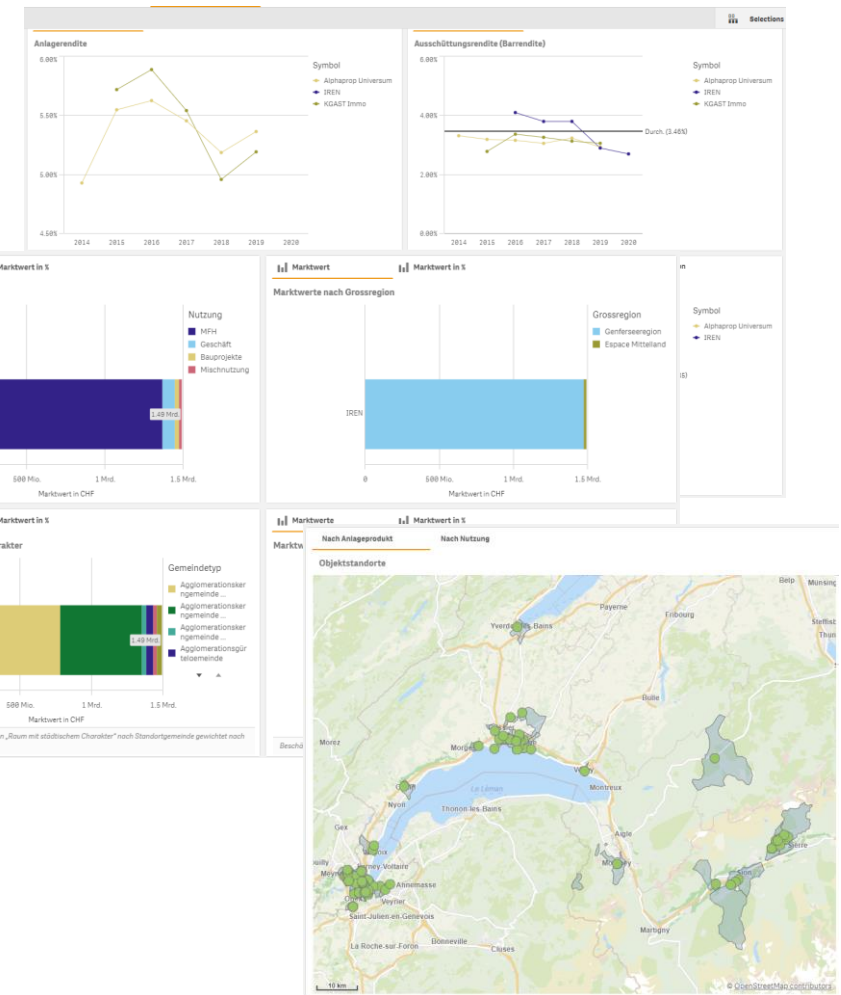
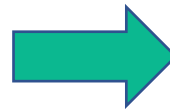
File Edit Code View Plots Session Build Debug Profile Tools Help

data_prep_SFC_2020.R data_prep_IREN_2020.R out4 out3 new_df

Filter

	GMDNR	HouseKey	Latitude	Longitude	GMDNAME	MS-Name	Kanton-Name	Grossraum-Name
1	843	IREN, Saanen, gstaadstr68	46.48839	7.264610	Saanen	Saanen-Obersimmental	Bern	Espace Mittelland
2	5516	IREN, Cugy (VD), chdespetitsesserts1	46.58897	6.643027	Cugy (VD)	Lausanne	Vaud	Région lémanique
3	5586	IREN, Lausanne, avalexandrevinet39	46.52663	6.627812	Lausanne	Lausanne	Vaud	Région lémanique
4	5586	IREN, Lausanne, rtealoyfauquez60	46.53506	6.633609	Lausanne	Lausanne	Vaud	Région lémanique
5	5586	IREN, Lausanne, chducloset46810	46.51453	6.632475	Lausanne	Lausanne	Vaud	Région lémanique
6	5586	IREN, Lausanne, avdechallenc8789	46.52861	6.614734	Lausanne	Lausanne	Vaud	Région lémanique
7	5586	IREN, Lausanne, avdouchy7274	46.50749	6.626808	Lausanne	Lausanne	Vaud	Région lémanique
8	5586	IREN, Lausanne, chdeslys14	46.52897	6.647353	Lausanne	Lausanne	Vaud	Région lémanique
9	5586	IREN, Lausanne, rtealoyfauquez122124	46.54042	6.633786	Lausanne	Lausanne	Vaud	Région lémanique
10	5586	IREN, Lausanne, chdelachapelle2	46.57895	6.604951	Lausanne	Lausanne	Vaud	Région lémanique
11	5586	IREN, Lausanne, chdemontmeillan1921	46.52832	6.638513	Lausanne	Lausanne	Vaud	Région lémanique
12	5586	IREN, Lausanne, ruedumaupas61636567	46.52583	6.623763	Lausanne	Lausanne	Vaud	Région lémanique
13	5586	IREN, Lausanne, placeduvallon1	46.52684	6.638561	Lausanne	Lausanne	Vaud	Région lémanique
14	5586	IREN, Lausanne, placedelanavigation46	46.50709	6.626866	Lausanne	Lausanne	Vaud	Région lémanique
15	5586	IREN, Lausanne, avdesoiseaux1517	46.53225	6.629004	Lausanne	Lausanne	Vaud	Région lémanique
16	5586	IREN, Lausanne, avictoruffy33	46.52774	6.648201	Lausanne	Lausanne	Vaud	Région lémanique
17	5588	IREN, Paudex, rtedelabernadaz1	46.50791	6.671861	Paudex	Lausanne	Vaud	Région lémanique
18	5589	IREN, Prilly, ruedelacombette2224	46.53004	6.603333	Prilly	Lausanne	Vaud	Région lémanique

Qlik



If you have any questions or suggestions – please reach out!

thomas.spycher@novalytica.com

