R-Bootcamp - Analysis of Building Projects in Zurich
MSC.IDS - HSLU Lucerne

Keith Lawless, Daniel Leibrock

2022-08-29

Contents

Introduction to the Analysis

Selection of the dataset

The dataset we want to analyze in the cause of this project, is the "**Neubauwohnungen nach Bauprojektstatus, Eigentumsart und statistischer Zone seit 2009**" dataset from the opendata.swiss website. The dataset contains information about housing projects in the city of Zurich since 2009 and holds information such as project status, ownership and location of the project. Furthermore, we add enriching data regarding real estate prices as well as adress data from the City of Zurich Open Data Portal to enrich our main dataset.

Goals of the Analysis

The goal of the Analysis is to get an insight into the development of the real estate market in Zurich throughout the past decade. We want to offer a diverse view on several aspects of the market, including the development of different types of projects, prices, owners and locations. Through this, we want to identify trends and offer an overview of the market structure.

Research questions

Following the previous introduction and the goals of our analysis, we aim to answer the following research questions:

1) How did prices develop, differentiated by the type of ownership?
2) How did prices develop, differentiated by location?
3) How did the number of overall building projects develop, differentiated by location?
4) Who are the main owners / drivers of building projects?
5) Does the share of different building size groups change over time? ?
6) Is there a relation between price development and building project development?

Data Preparation

Data preparation consists of several steps. These include:

- Deletion of obsolete columns from the different data sets
- Renaming of several columns
- Grouping and summarising of several columns
- Merging of the different datasets

Exploratory Data Analysis

The summary of the final dataframe reveals several insights about the general structure of the data as well as the real estate market itself.

Regarding the general structure of the data, we can see that in total **27950 different datapoints** spread over a period of 13 years from **2009 until 2021**. In total, projects with a volume of **51016 apartments received approval** while **69194 appartments started construction** in Zurich during this period. *It is not clear, if individual apartments can appear in both categories at different points of time.*

The projects are located within **12 "Kreise" / regions**, split into **34 "Quartiere" / districts** which are then again differentiated into **215 statistical zones**, the most granular geo-statistical level for the city of Zurich. While data about the number of apartments and geographic coordinates is available on this level, price data is only available on Quartier-level. Furthermore, the projects are assigned to **5 different types of ownership**.

Visual Data Analysis

In this chapter, we want to visually analyze the structure of the real estate market.

```
## 'summarise()' has grouped output by 'Jahr'. You can override using the
## '.groups' argument.
```

Price development and differences

Average price per year & sq.m. of condominium ownership and ground (CHF) Average price development per sq.m. of condominium ownership and ground (CHF): As a first visual analysis of the data we will look at price developments on different levels. We want to get an overview of the development of yearly-averaged prices over time, split in the different categories available. These are CHF / m^2 Boden

(Ganze Liegenschaft), CHF / m^2 Boden (Stockwerkbesitz), CHF / m^2 Boden (Kombiniert) as well as CHF / m^2 Wohnfläche (Stockwerkbesitz).

The following table and plot show information about real estate prices averaged over the whole city of Zurich for each year in CHF:

Jahr

CHF / m^2 Boden (Ganze Liegensch.)

CHF / m^2 Boden (Stockwerkb.)

CHF / m^2 Boden (Kombiniert)

CHF / m^2 Wohnfläche (Stockwerkb.)

2009

10901.27

12796.64

11935.92

8593.750

2010

12021.01

15576.15

13844.78

8992.903

2011

13333.41

15603.49

15023.58

9940.323

2012

14382.35

16967.21

17080.93

10363.448

2013

14220.12

16107.28
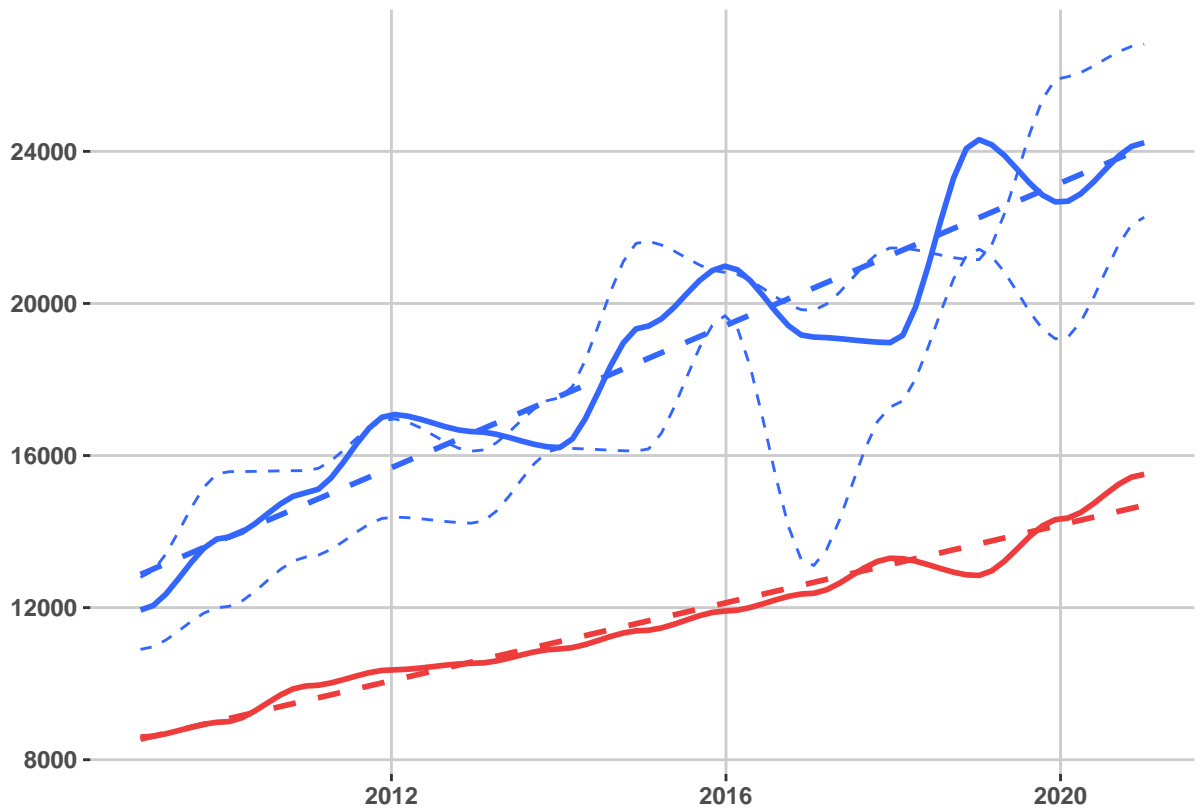
16627.28

10536.129

2014

16192.51

17519.03

16209.24

10913.667

2015

16116.10

21647.96

19382.91

11397.419

2016

19681.47

20815.68

20981.48

11909.688

2017

13076.26

19807.66

19117.49

12371.000

2018

17291.06

21467.14

18969.10

13303.226

2019

21432.53

21143.72

24311.94

12843.667

2020

19043.26

25945.96

22653.35

14330.667

2021

22267.91

26815.59

24229.06

15505.484

As we can see, the overall price trend is clearly upwards. The red curve, which represents the price for living space / Wohnfläche in condominium ownership / Stockwerkbesitz, is showing a continuous and constant **growth of about 575 CHF per year** over the observed timeline. The thick blue curve, which represents the price for ground / Boden, combined for both Stockwerkbesitz (upper dashed, blue line) and whole buildings / ganze Liegenschaft (lower dashed, blue line), reveals a yearly average **price increase of about 1000 CHF**.

Total number of projected flats per year and location   Jahr

KreisLang

WhgSum

2009

Kreis 1

4

2009

Kreis 10

622

2009

Kreis 11

1971

2009

Kreis 12

75

2009

Kreis 2

375

2009

Kreis 3

841

2009

Kreis 4

226

2009

Kreis 5

991

2009

Kreis 6

201

2009

Kreis 7

269

2009

Kreis 8

230

2009

Kreis 9

1274

2010

Kreis 1

0

2010

Kreis 10

626

2010

Kreis 11

1824

2010

Kreis 12

543

2010

Kreis 2

493

2010

Kreis 3

847

2010

Kreis 4

251

2010

Kreis 5

956

2010

Kreis 6

242

2010

Kreis 7

271

2010

Kreis 8

249

2010

Kreis 9

1507

2011

Kreis 1

0

2011

Kreis 10

663

2011

Kreis 11

1717

2011

Kreis 12

1246

2011

Kreis 2

605

2011

Kreis 3

826

2011

Kreis 4

546

2011

Kreis 5

1048

2011

Kreis 6

257

2011

Kreis 7

232

2011

Kreis 8

222

2011

Kreis 9

1133

2012

Kreis 1

2

2012

Kreis 10

830

2012

Kreis 11

1591

2012

Kreis 12

1543

2012

Kreis 2

660

2012

Kreis 3

585

2012

Kreis 4

432

2012

Kreis 5

1078

2012

Kreis 6

148

2012

Kreis 7

349

2012

Kreis 8

175

2012

Kreis 9

1748

2013

Kreis 1

3

2013

Kreis 10

1271

2013

Kreis 11

1162

2013

Kreis 12

2037

2013

Kreis 2

708

2013

Kreis 3

463

2013

Kreis 4

439

2013

Kreis 5

803

2013

Kreis 6

388

2013

Kreis 7

552

2013

Kreis 8

242

2013

Kreis 9

2290

2014

Kreis 1

2

2014

Kreis 10

1250

2014

Kreis 11

972

2014

Kreis 12

1840

2014

Kreis 2

1057

2014

Kreis 3

579

2014

Kreis 4

483

2014

Kreis 5

596

2014

Kreis 6

393

2014

Kreis 7

581

2014

Kreis 8

267

2014

Kreis 9

3069

2015

Kreis 1

2

2015

Kreis 10

1143

2015

Kreis 11

1282

2015

Kreis 12

1182

2015

Kreis 2

1160

2015

Kreis 3

646

2015

Kreis 4

434

2015

Kreis 5

153

2015

Kreis 6

602

2015

Kreis 7

487

2015

Kreis 8

279

2015

Kreis 9

2869

2016

Kreis 1

0

2016

Kreis 10

496

2016

Kreis 11

1159

2016

Kreis 12

1248

2016

Kreis 2

1334

2016

Kreis 3

920

2016

Kreis 4

596

2016

Kreis 5

147

2016

Kreis 6

891

2016

Kreis 7

510

2016

Kreis 8

428

2016

Kreis 9

1808

2017

Kreis 1

0

2017

Kreis 10

160

2017

Kreis 11

1622

2017

Kreis 12

852

2017

Kreis 2

1414

2017

Kreis 3

712

2017

Kreis 4

495

2017

Kreis 5

132

2017

Kreis 6

842

2017

Kreis 7

647

2017

Kreis 8

309

2017

Kreis 9

1934

2018

Kreis 1

0

2018

Kreis 10

146

2018

Kreis 11

1799

2018

Kreis 12

608

2018

Kreis 2

1000

2018

Kreis 3

449

2018

Kreis 4

584

2018

Kreis 5

120

2018

Kreis 6

573

2018

Kreis 7

463

2018

Kreis 8

311

2018

Kreis 9

1552

2019

Kreis 1

0

2019

Kreis 10

260

2019

Kreis 11

2387

2019

Kreis 12

612

2019

Kreis 2

1358

2019

Kreis 3

327

2019

Kreis 4

606

2019

Kreis 5

109

2019

Kreis 6

749

2019

Kreis 7

529

2019

Kreis 8

249

2019

Kreis 9

1582

2020

Kreis 1

0

2020

Kreis 10

400

2020

Kreis 11

2756

2020

Kreis 12

681

2020

Kreis 2

1547

2020

Kreis 3

1140

2020

Kreis 4

654

2020

Kreis 5

308

2020

Kreis 6

607

2020

Kreis 7

695

2020

Kreis 8

273

2020

Kreis 9

2200

2021

Kreis 1

0

2021

Kreis 10

295

2021

Kreis 11

2694

2021

Kreis 12

602

2021

Kreis 2

1151

2021

Kreis 3

687

2021

Kreis 4

648

2021

Kreis 5
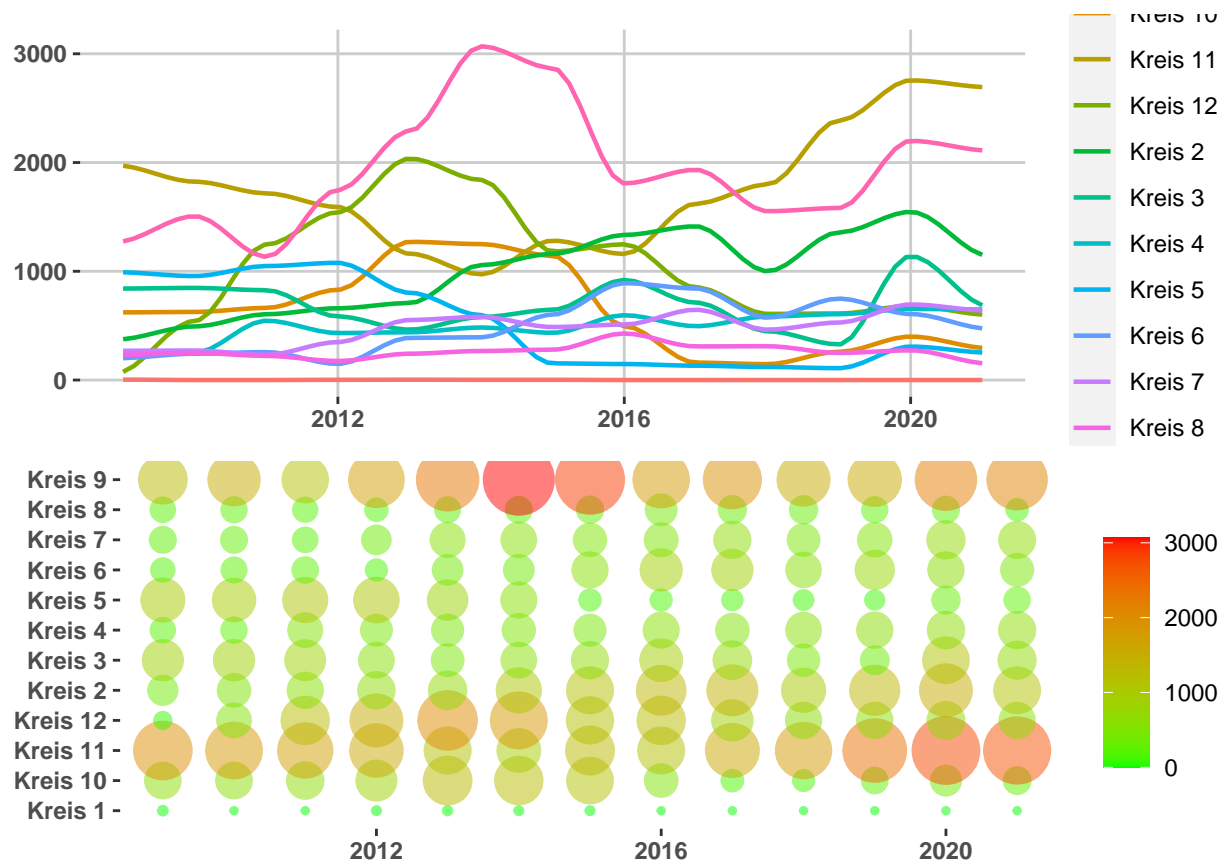
253

2021

Kreis 6

475

2021

Kreis 7

639

2021

Kreis 8

154

2021

Kreis 9

2112

Average price per sq.m. of condominium ownership and ground (CHF) per Kreis   Kreis

CHF / m^2 Boden (Ganze Liegensch.)

CHF / m^2 Boden (Stockwerkb.)

CHF / m^2 Boden (Kombiniert)

CHF / m^2 Wohnfläche (Stockwerkb.)

Kreis 1

66331.745

54126.194

60143.031

15309.600

Kreis 10

6715.157

11173.218

9471.065

10699.615

Kreis 11

6066.582

10193.949

8760.521

8791.795

Kreis 12

3603.991

7848.958

5259.269

8128.611

Kreis 2

9517.523

15630.428

12573.564

11234.615

Kreis 3

10022.699

16611.077

11702.246

10591.667

Kreis 4

19754.423

28229.548

23896.731

9976.452

Kreis 5

12651.683

30571.173

21600.013
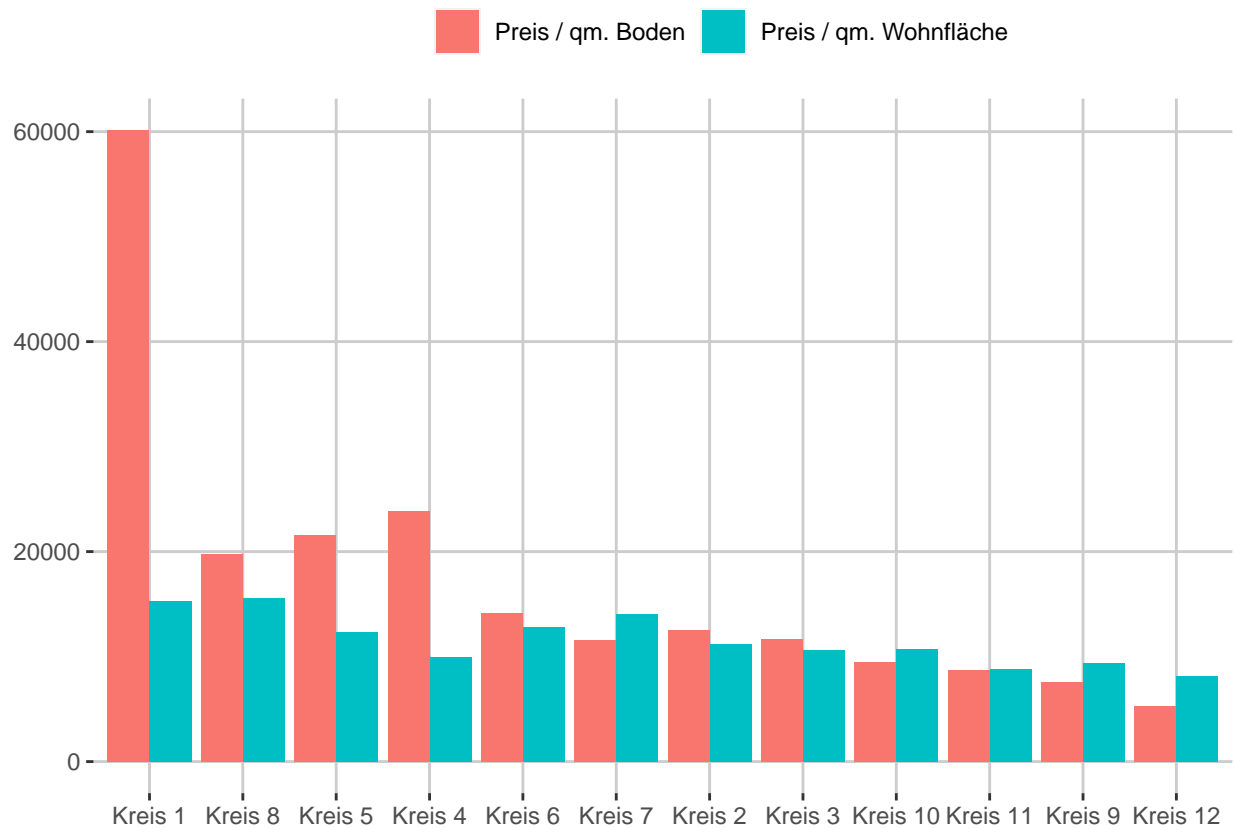
12295.600

Kreis 6

10491.287

17741.990

14187.053

12800.769

Kreis 7

8311.202

13618.579

11569.249

14019.615

Kreis 8

16643.761

24388.579

19718.626

15552.368

Kreis 9

5677.249

9945.179

7550.902

9422.308

Special Topic: Leaflet with R/Shiny

For our chapter of choice, we chose leaflet and integrate it with a simple Shiny App. A preview of the app
is shown below. Alternatively, the app is available online by clicking here.

```
knitr::include_app("https://vordaten.shinyapps.io/StadtZurich/", height = "600px")
```

PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is installed, please

Experience with Leaflet and Shiny Integration Our main goal with leaflet was to provide a rich interactive experience for users of our app, such that they could intuitively explore the cleaned dataset. Initially, there were some challenges to overcome and ultimately, we were not entirely satisfied with the final outcome. However, in development of the app, some of the challenges we encountered forced us to change our design thinking on the overall role that maps can play as part of a rich, interactive visual that can tell a useful story which is in itself a great thing.

Leaflet Pros Firstly, the range of base map options that are available with leaflet are stunning. In the opinion of the authors, this option alone makes leaflet the only option when it comes to implementing maps ahead of commercial tools such as Tableau and PowerBI (PowerBI in particular is limited here).

Secondly, the code to quickly develop a map is straightforward. Add your data, chose a basemap, select the marker types you would like and customize.

Next - the options to fix the boundaries, zoom level and focus point are a big advantage. Again, options like these are absent from tools like PowerBI. In these tools, it is entirely possible to mistakenly zoom out of the map or move it to another area with no way to reset them. With leaflet, you can help the user by ensuring they do not stray from the area of the map containing the interesting data points.

Leaflet Cons As mentioned, we experienced a couple of challenges. Firstly performance. We would have liked to pass the entire data set to Leaflet and for it to summarize as needed in the background. However, when we did this, R-Studio frequently crashed. As a result, we condensed the data set down into an aggregated format, such that the map would display quickly and comfortably.

Secondly, and somewhat related, it was frustrating not to be able to use the map as a filter by clicking on one of the polygons. The package does indeed contain functionality for generating map click events, however, to pass these events to other elements in our app would have meant that we also pre-aggregated the data inputs supporting this element which was not possible for all elements.

Leaflet Conclusions Overall we would recommend leaflet as a map visualization. As mentioned, due to some of the challenges we would recommend to our users that they think of how we use maps somewhat differently. A personal frustration of the author is when searching an address on Google maps, it zooms right into the lowest level of the map. Therefore, a wider context is missing. Extra steps are required in order for me to zoom out and get a general idea of where the location is (in relation to the city center, nearest train station, motorway, etc.).

I would like to make decisions based on that context (is it better to drive or get the train) and I can worry about the exact location later when I am approaching.

Similarly with leaflet, I think it is more interesting for users to click on a data point and then see that update in the context of the fixed map.

Additionally, performance remains a concern. I would recommend to our client the use of SQL inside the app could be used to take click events and perform the map filtering in the data set in the database, before returning the results of that aggregated query to the map or other element for the visualization.

Modelling and Prediction

Model XYZ

Prediction XYZ

Summary and Conclusion