

R-Bootcamp - Analysis of Building Projects in Zurich

MSC.IDS - HSLU Lucerne

Keith Lawless, Daniel Leibrock

2022-08-29

Contents

Introduction to the Analysis	1
Selection of the dataset	1
Goals of the Analysis	2
Research questions	2
Data Preparation	2
Exploratory Data Analysis	2
Visual Data Analysis	4
Basic visual Analysis / ggplot2	4
Special Topic: Mapping / Shiny	6
Modelling and Prediction	6
Model XYZ	6
Prediction XYZ	6
Summary and Conclusion	6

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 the number of overall building projects develop over time in different statistical zones / quartiers?
- 2) Who are the main owners / drivers of building projects?
- 3) Does the share of different building size groups change over time?
- 4) How did prices in the different parts of the city (quartiers) change over time?
- 5) 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 apartments 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**.

The following table shows information about real estate prices averaged 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

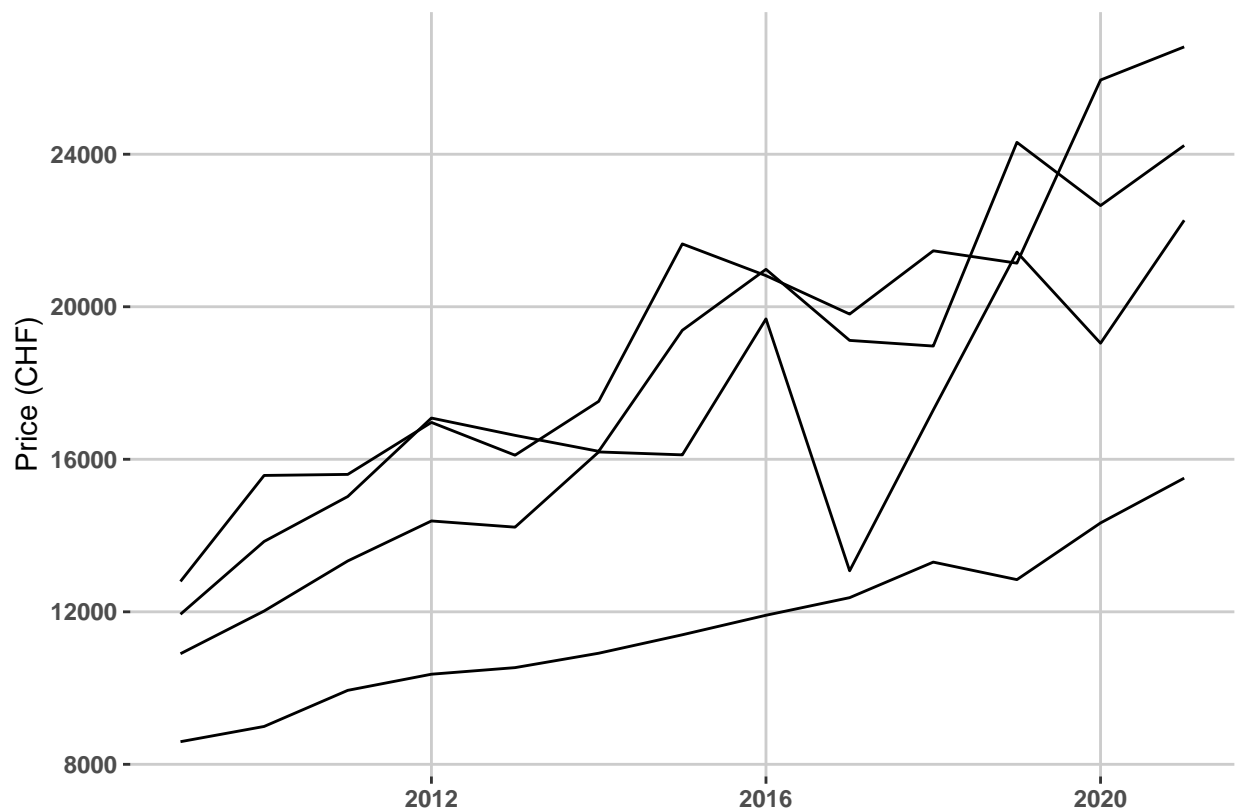
11397.419

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

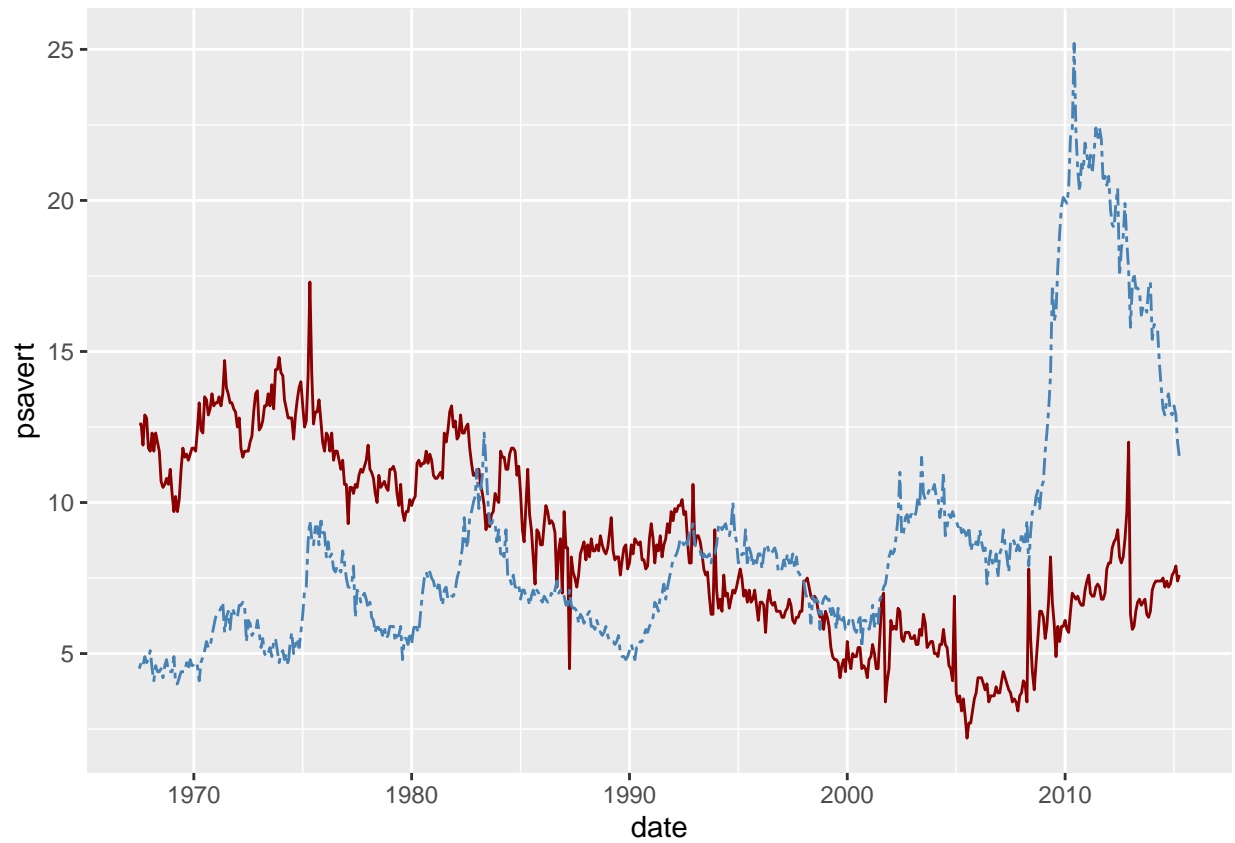
Visual Data Analysis

Basic visual Analysis / ggplot2

```
ggplot(yearprices, aes(x = Jahr), colour = price_cols_renamed) +
  geom_line(aes(y = `CHF / $m^{2}$ Wohnfläche (Stockwerkb.)`), size = 0.5) +
  geom_line(aes(y = `CHF / $m^{2}$ Boden (Stockwerkb.)`), size = 0.5) +
  geom_line(aes(y = `CHF / $m^{2}$ Boden (Ganze Liegensch.)`), size = 0.5) +
  geom_line(aes(y = `CHF / $m^{2}$ Boden (Kombiniert)`), size = 0.5) +
  theme(
    panel.grid.major = element_line(colour = "gray80"),
    panel.grid.minor = element_line(colour = NA),
    axis.text = element_text(family = "sans",
                             face = "bold"),
    panel.background = element_rect(fill = "white")
  ) +
  ylab("Price (CHF)") +
  xlab("")
```



```
ggplot(economics, aes(x=date)) +
  geom_line(aes(y = psavert), color = "darkred") +
  geom_line(aes(y = uempmed), color="steelblue", linetype="twodash")
```



Special Topic: Mapping / Shiny

Modelling and Prediction

Model XYZ

Prediction XYZ

Summary and Conclusion