

R-Bootcamp report

Bernardo Freire Barboza da Cruz & Leonid Gavriluk

01 February, 2022

Contents

1	Introduction	2
1.1	Which is the most activate Kreis?	2
1.2	TBD	2
1.3	TBD	2
2	Extract and Transform	2
2.1	Data Import	2
2.2	Data Exploration	4
2.3	Merging Data	4
2.4	Data Visualization	4
2.5	Fit Model	4
2.6	Chapter of Choice TBD	4
3	References	4

1 Introduction

Describe the use case....

1.1 Which is the most activate Kreis?

1.2 TBD

1.3 TBD

2 Extract and Transform

2.1 Data Import

The data is imported from *Stadt Zürich Open Data*.

The data looks as follows:

X...Jahr	Partei	ParteiNr	Wahlkreis	WahlkreisSort	Stimmenanteil
1913	SP	1	Stadt Zrich	0	39.1
1913	BGB/SVP	2	Stadt Zrich	0	NA
1913	FDP	3	Stadt Zrich	0	38.6
1913	GPS	4	Stadt Zrich	0	NA
1913	GLP	5	Stadt Zrich	0	NA
1913	CVP	6	Stadt Zrich	0	7.9

X... Stich	Sag-Datg	Film-Mo	Datum	M-Sag-DCS	SexLang	Alter	Vor-Nr	Hilf-Kunz	Kreis	Cheis-Langr	Qdar	Datens-tant	Gst-andl	Baug-Wir
1998	1	Januar	13910	1	mnnlich	1	0-19	1	Schweizer/inKreis 11	Rathau-	Ws	Verffentlidt	14	
1998	1	Januar	13910	1	mnnlich	1	0-19	1	Schweizer/inKreis 12	Hochschu-	len	Verffentlicht	24	
1998	1	Januar	13910	1	mnnlich	1	0-19	1	Schweizer/inKreis 13	Lindenhof	-	Verffentlicht	37	
1998	1	Januar	13910	1	mnnlich	1	0-19	1	Schweizer/inKreis 14	City V	-	Verffentlicht	48	
1998	1	Januar	13910	1	mnnlich	1	0-19	1	Schweizer/inKreis 21	Wollishofen	-	Verffentlicht	1950	
1998	1	Januar	13910	1	mnnlich	1	0-19	1	Schweizer/inKreis 23	Leimbach	-	Verffentlicht	1361	

X... Jahr	Alter	Geschlecht	Anteil
2019	18	Männer	22.6
2019	18	Frauen	25.1
2019	19	Männer	19.1
2019	19	Frauen	21.4
2019	20	Männer	21.4
2019	20	Frauen	20.6

X...Jahr	QNr	Qname	Sberechtigte	Nteilnehmende	teilnehmende	Beteiligung
2006	11	Rathaus (Kreis 1)	1974	1186	788	39.9
2006	12	Hochschulen (Kreis 1)	377	232	145	38.5
2006	13	Lindenhof (Kreis 1)	1335	962	373	27.9
2006	14	City (Kreis 1)	597	440	157	26.3
2006	21	Wollishofen (Kreis 2)	10115	6168	3947	39.0
2006	23	Leimbach (Kreis 2)	3123	1997	1126	36.1

and the column types look as follows:

```
## 'data.frame': 4437 obs. of 6 variables:
## $ X...Jahr : int 1913 1913 1913 1913 1913 1913 1913 1913 1913 1913 1916 ...
## $ Partei : chr "SP" "BGB/SVP" "FDP" "GPS" ...
## $ ParteiNr : int 1 2 3 4 5 6 7 8 9 1 ...
## $ Wahlkreis : chr "Stadt Z\303\274rich" "Stadt Z\303\274rich" "Stadt Z\303\274rich" "Stadt Z\303\274rich" ...
## $ WahlkreisSort: chr "0" "0" "0" "0" ...
## $ Stimmenanteil: num 39.1 NA 38.6 NA NA 7.9 NA NA 14.4 43.9 ...

## 'data.frame': 194897 obs. of 17 variables:
## $ X...StichtagDatJahr: int 1998 1998 1998 1998 1998 1998 1998 1998 1998 1998 1998 1998 ...
## $ StichtagDatMM : int 1 1 1 1 1 1 1 1 1 1 ...
## $ StichtagDatMonat : chr "Januar" "Januar" "Januar" "Januar" ...
## $ StichtagDat : int 13910 13910 13910 13910 13910 13910 13910 13910 13910 13910 13910 13910 ...
## $ SexCd : int 1 1 1 1 1 1 1 1 1 1 ...
## $ SexLang : chr "m\303\244nnlich" "m\303\244nnlich" "m\303\244nnlich" "m\303\244nnlich" ...
## $ AlterV20Sort : int 1 1 1 1 1 1 1 1 1 1 ...
## $ AlterV20Kurz : chr "0-19" "0-19" "0-19" "0-19" ...
## $ HerkunftCd : int 1 1 1 1 1 1 1 1 1 1 ...
## $ HerkunftLang : chr "Schweizer/in" "Schweizer/in" "Schweizer/in" "Schweizer/in" ...
## $ KreisCd : int 1 1 1 1 2 2 2 3 3 3 ...
## $ KreisLang : chr "Kreis 1" "Kreis 1" "Kreis 1" "Kreis 1" ...
## $ QuarCd : int 11 12 13 14 21 23 24 31 33 34 ...
## $ QuarLang : chr "Rathaus" "Hochschulen" "Lindenhof" "City" ...
## $ DatenstandCd : chr "V" "V" "V" "V" ...
## $ DatenstandLang : chr "Ver\303\266ffentlich" "Ver\303\266ffentlich" "Ver\303\266ffentlich" ...
## $ AnzBestWir : int 114 24 37 48 950 361 444 592 961 721 ...

## 'data.frame': 312 obs. of 4 variables:
## $ X...Jahr : int 2019 2019 2019 2019 2019 2019 2019 2019 2019 2019 2019 ...
## $ Alter : int 18 18 19 19 20 20 21 21 22 22 ...
## $ Geschlecht: chr "M\303\244nner" "Frauen" "M\303\244nner" "Frauen" ...
## $ Anteil : num 22.6 25.1 19.1 21.4 21.4 20.6 20.1 19.4 19.8 22.1 ...

## 'data.frame': 136 obs. of 7 variables:
## $ X...Jahr : int 2006 2006 2006 2006 2006 2006 2006 2006 2006 2006 2006 ...
## $ QNr : int 11 12 13 14 21 23 24 31 33 34 ...
## $ Qname : chr "Rathaus (Kreis 1)" "Hochschulen (Kreis 1)" "Lindenhof (Kreis 1)" "City (Kreis 1)" ...
## $ Sberechtigte : int 1974 377 1335 597 10115 3123 5093 8358 6005 10821 ...
## $ Nteilnehmende: int 1186 232 962 440 6168 1997 3209 5555 3804 7643 ...
## $ teilnehmende : int 788 145 373 157 3947 1126 1884 2803 2201 3178 ...
## $ Beteiligung : num 39.9 38.5 27.9 26.3 39 36.1 37 33.5 36.7 29.4 ...
```

- 2.2 Data Exploration
- 2.3 Merging Data
- 2.4 Data Visualization
- 2.5 Fit Model
- 2.6 Chapter of Choice TBD
- 3 References