

# **Analyse und Dekomposition von Zeitreihen am Beispiel von Sales Zahlen**

**Deskriptive Statistik**

Jürgen Mayer

2024-04-15

# Inhaltsverzeichnis

<b>1</b>	<b>Information zu dem Datensatz</b>	<b>3</b>
1.1	Details . . . . .	3
<b>2</b>	<b>Deskriptive Statistik</b>	<b>5</b>
2.1	Gleitender Durchschnitt . . . . .	6
2.2	Verteilung und Ausreißer . . . . .	7
<b>3</b>	<b>Decomposition</b>	<b>9</b>
3.1	Autocorrelation . . . . .	12
3.2	Details . . . . .	12
<b>4</b>	<b>Export</b>	<b>14</b>

Dies ist ein Beispiel der Zeitreihenanalyse mit dem PrediFo Paket. Zu Beginn der Analyse und des Forecastings steht die deskriptive Statistik, um einen Gesamteindruck der Daten zu bekommen.

Die Zahlen sind reale neutralisierte Umsatzzahlen, die hier aus einer Azure SQL-Datenbank eingelesen werden.

Wichtig ist hier auch die Interpretation der Daten und nicht nur eine Präsentation der Grafiken.

# 1 Information zu dem Datensatz

## Datenübersicht pro Kategorie

category	Werte		
	min_date	max_date	Anzahl_Monate
10	1 Februar 2010	1 November 2021	142
20	1 Februar 2010	1 November 2021	142
30	1 Februar 2010	1 November 2021	142
40	1 Februar 2010	1 November 2021	142
50	1 Februar 2010	1 November 2021	142
60	1 Februar 2010	1 November 2021	142
70	1 Februar 2010	1 November 2021	142

## 1.1 Details

Klassische statistische Werte können relativ leicht aufgelistet werden.

## Statistische Werte für 2026 und 2027 pro Kategorie

category	Werte					Anzahl	sd
	min	max	median	mean	Summe		
2016							
10	17349	40135	20821	23587	283039	12	6544
20	1783	4378	2593	2872	34470	12	731
30	1410	3702	2150	2287	27444	12	765
40	1563	4591	3411	3291	39489	12	946
50	1057	7487	2122	2439	29264	12	1656
60	440	1008	834	782	9390	12	207
70	6505	14804	7954	9076	108913	12	2774
2017							
10	15588	29993	25235	24658	295891	12	4018
20	1765	3437	2665	2664	31974	12	538

Header text

---

30	1107	2097	1630	1556	18674	12	289
40	1753	5536	3685	3647	43764	12	1218
50	1495	7206	2413	3082	36978	12	1838
60	479	3031	969	1084	13007	12	661
70	6705	13265	9646	9725	116704	12	2211

---

Tabelle 1.1: Durchschnittswert

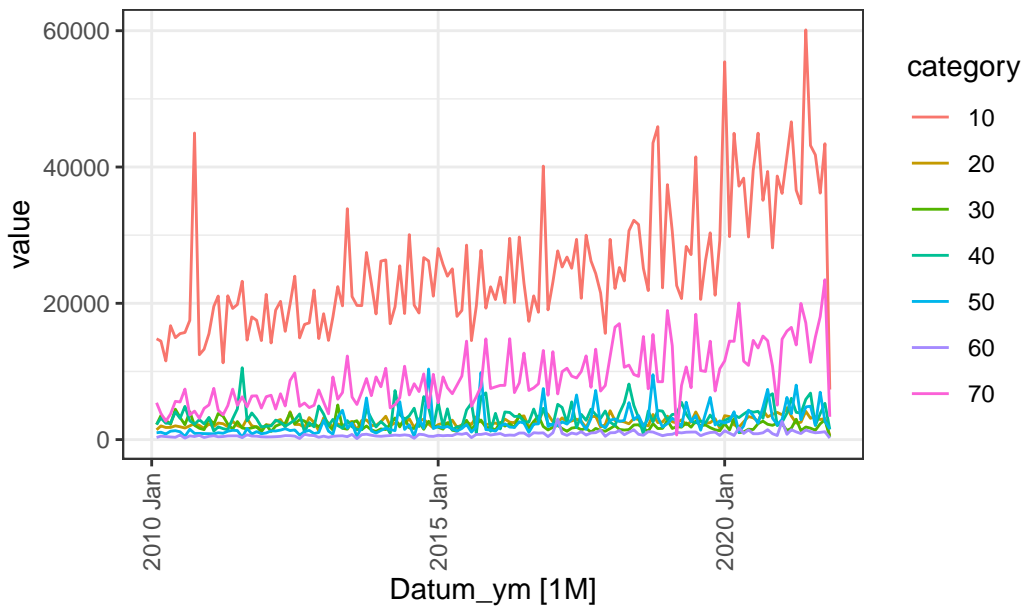
category	mean
60	767
30	2078
50	2532
20	2590
40	3358
70	9103
10	25267

Tabelle 1.2: Quantile

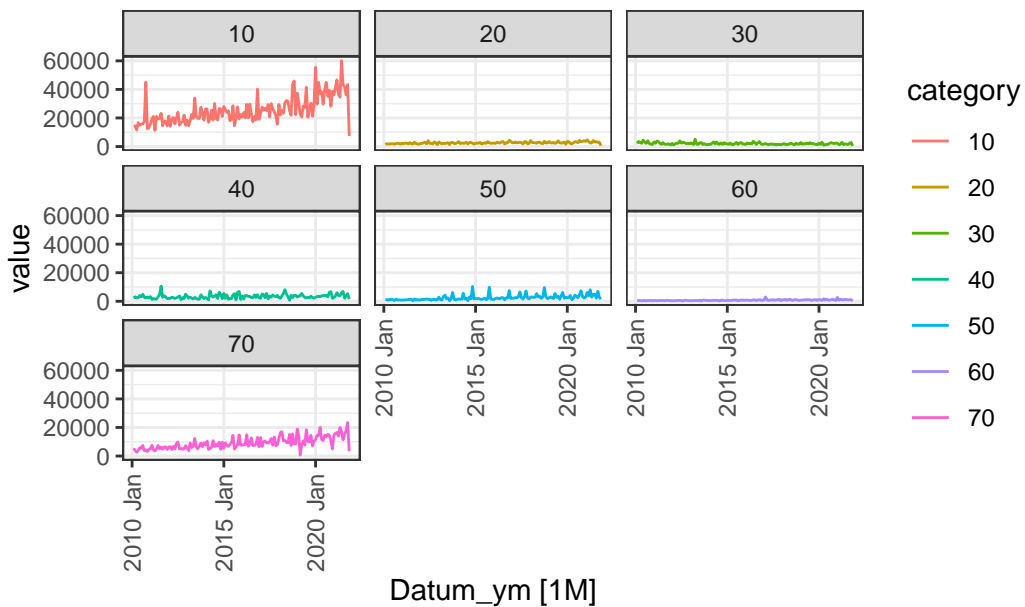
category	0%	25%	50%	75%	100%
10	7345	18966	22765	29679	60121
20	598	2081	2508	3012	4680
30	394	1528	1910	2398	5078
40	790	2118	3080	4177	10543
50	333	1294	1930	2942	10362
60	141	516	668	965	3031
70	676	6157	7946	11470	23454

## 2 Deskriptive Statistik

Darstellung aller Zeitreihen, Periode = month



Darstellung aller Zeitreihen, aufgeteilt, Periode = month



### 2.0.1 Interpretation:

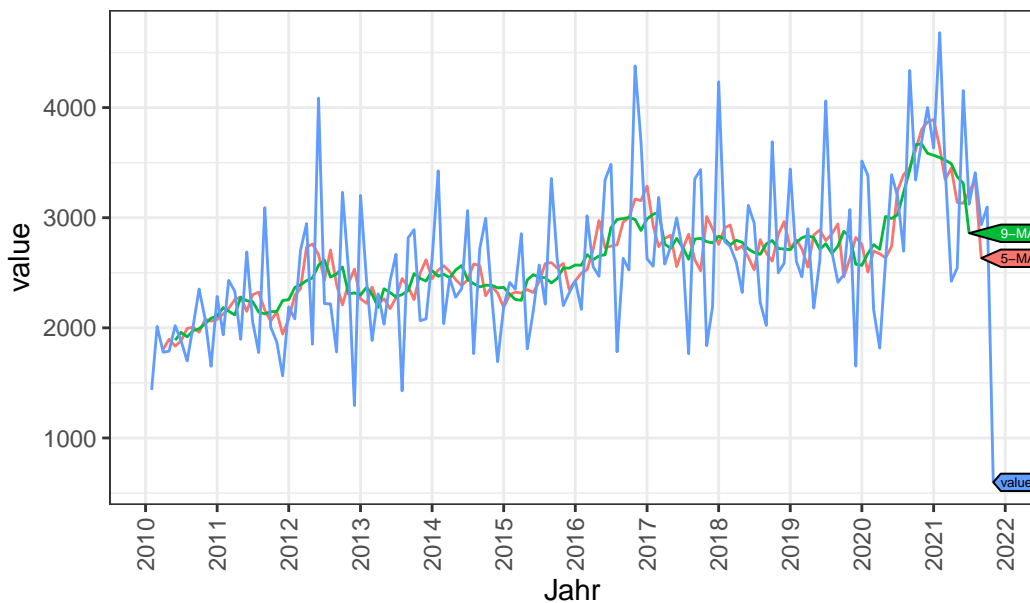
- Kategorie 10 hat deutlich höhere Werte
- Alle Kategorien zeigen eine hohe Streuung.
- Das Absinken der Werte am Ende liegt an der nicht kompletten Periode beim abspeichern der Daten. Dies muss z.B. beim Forecast berücksichtigt werden.

## 2.1 Gleitender Durchschnitt

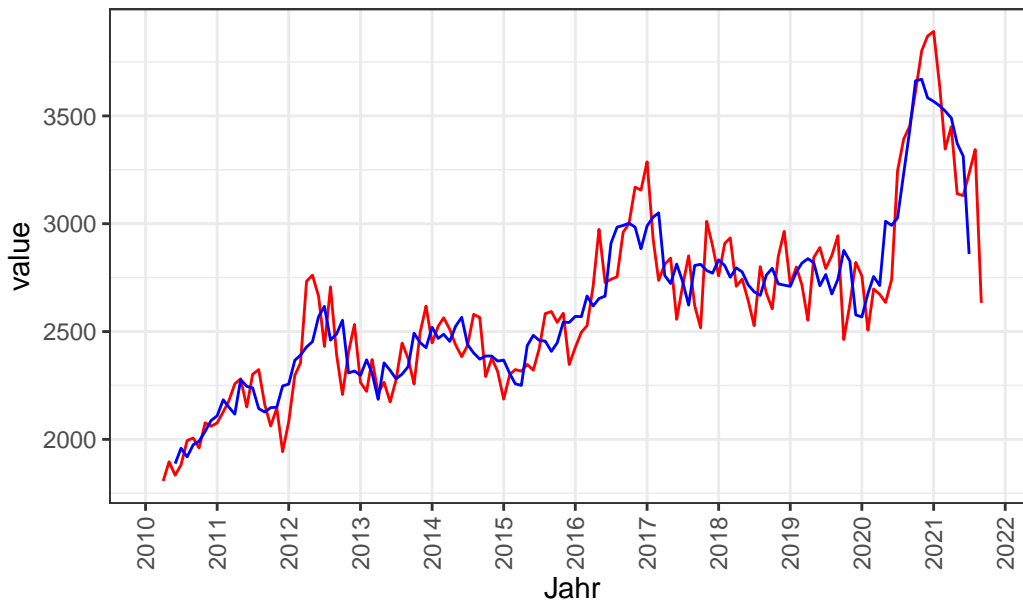
Hier wird ein gleitender Durchschnitt berechnet, wobei der jeweilige Wert der Mittelwert aus den Werten vor und nach dem jeweiligen Einzelwert ist. Somit können Gewichtungen realisiert werden.

Im Beispiel wird ein 5 Monate und 9 Monate moving average (MA) dargestellt.

5 und 9 Monate gleitender Durchschnitt, Kategorie: 20



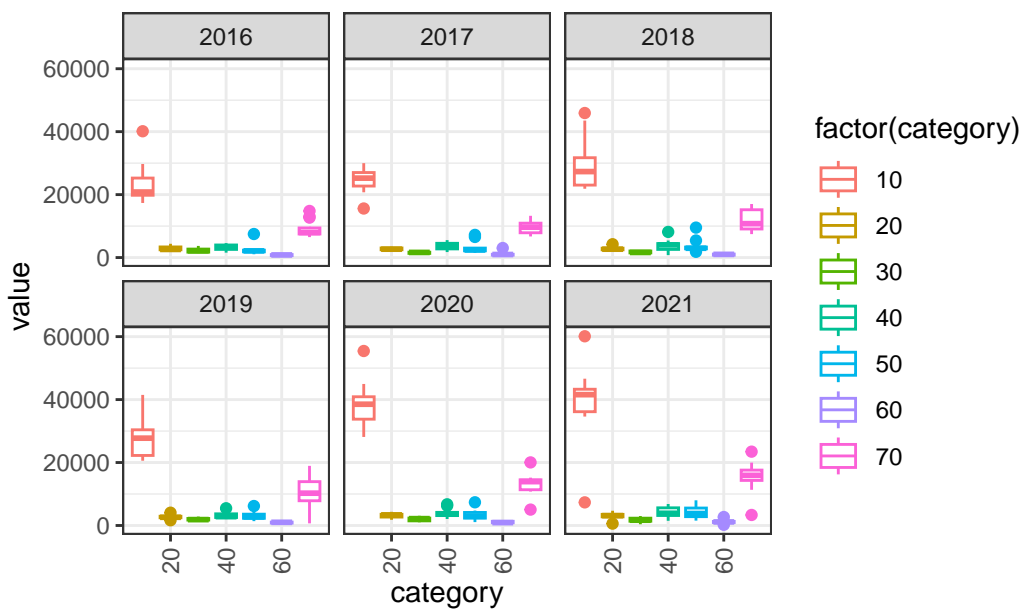
## 5 und 9 Monate gleitender Durchschnitt, Kategorie: 20



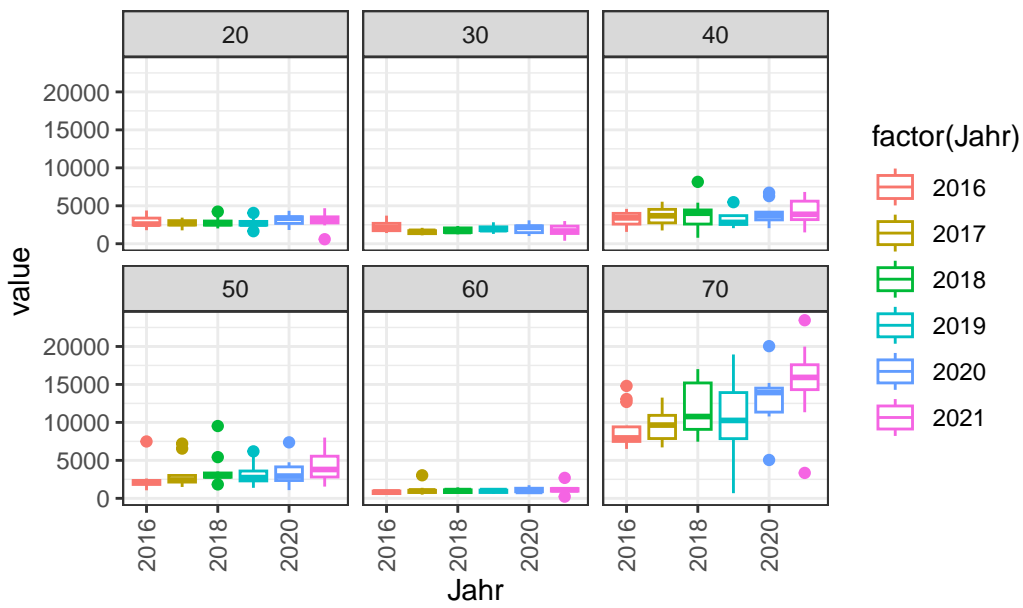
## 2.2 Verteilung und Ausreißer

Mittels Boxplots können die Verteilung und Ausreißer erkannt werden. Für die Prognoseberechnung können die Ausreißer dann eliminiert werden.

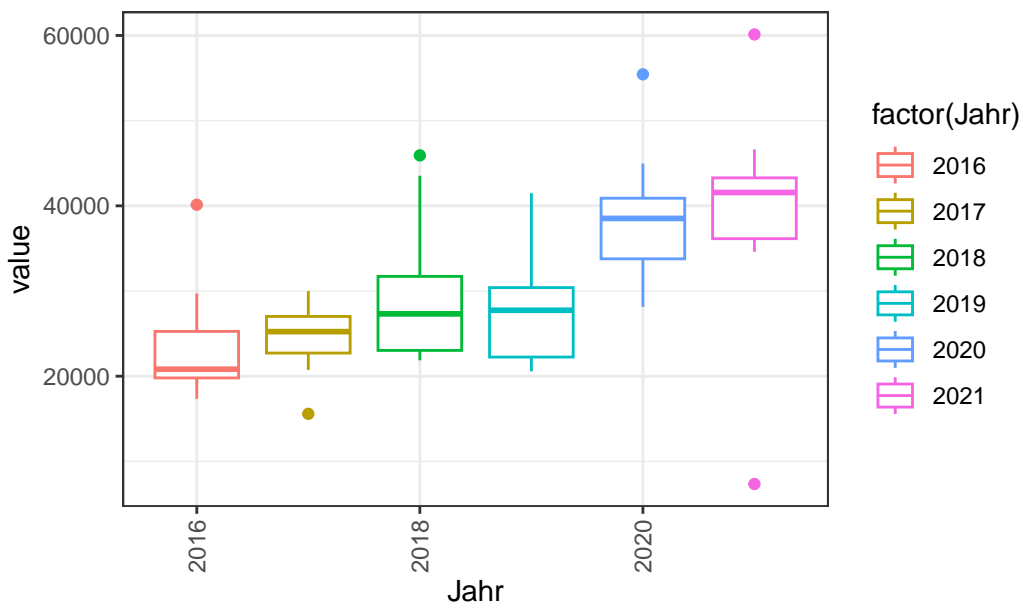
### Verteilung pro Jahr



## Verteilung pro Kategorie



## Verteilung Details, Kategorie = 10

**2.2.1 Interpretation:**

- Speziell Kategorie 10 und 70 zeigen eine hohe Bandbreite der Werte.



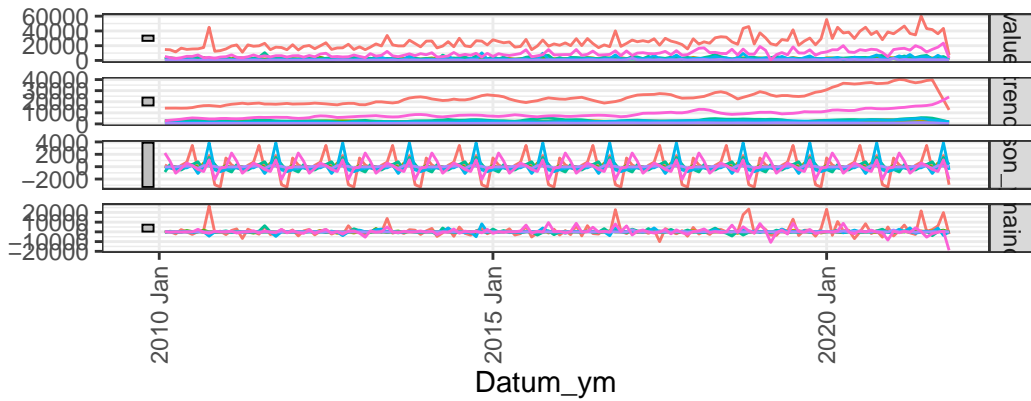
# 3 Decomposition

Dies ist eine Beschreibung des Decomposition-Moduls von PrediFo.

Decomposition ist die Zerlegung der Zeitreihendaten (Time Series Analysis).

Die aufbereiteten Daten können dann analysiert werden. Die decomposition erfolgt mit dem STL Model: STL = "Seasonal and Trend decomposition using Loess"

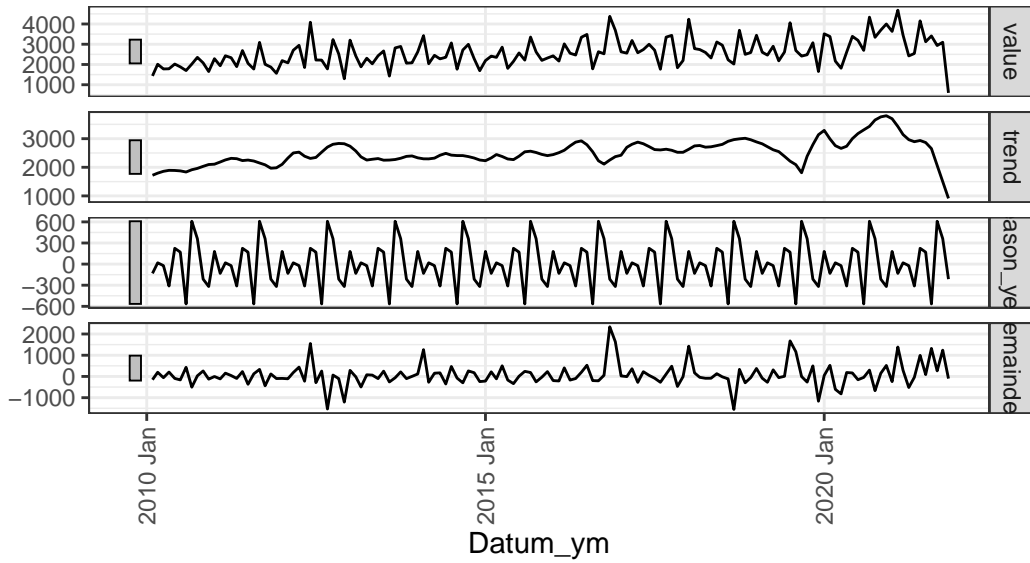
STL Decomposition, trend: 7 , season: periodic robust: TR  
 value = trend + season\_year + remainder



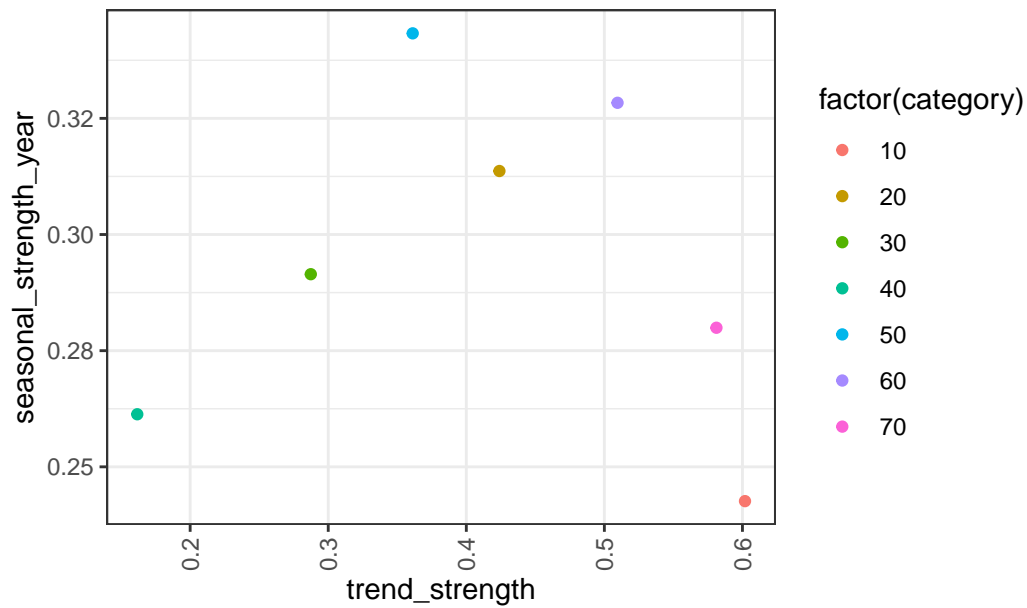
rend(window = STL\_trend) + season(window = STL\_season),  
 robust)  
 rend(window = STL\_trend) + season(window = STL\_season),  
 robust)

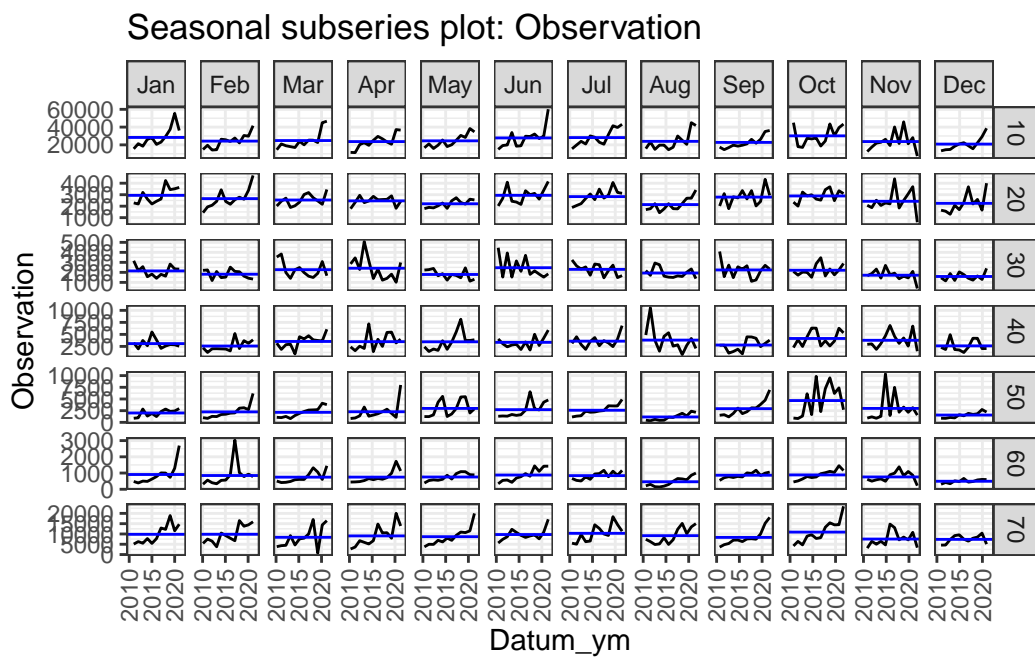
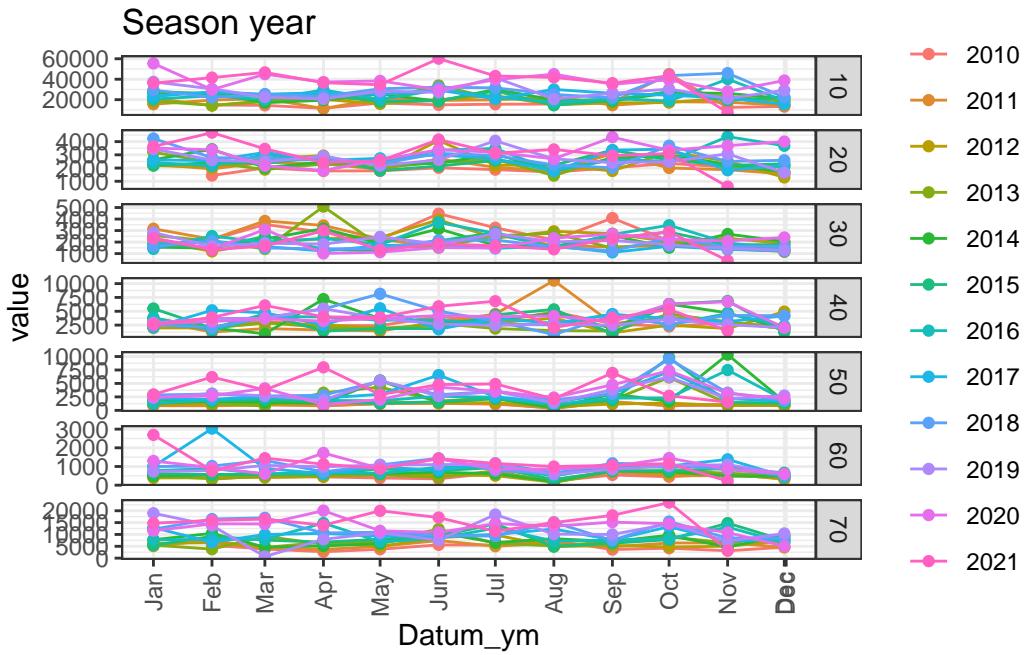
— 50/STL(value ~ trend(window =  
 robust = STL\_robust)  
 — 60/STL(value ~ trend(window =  
 robust = STL\_robust)

STL Decomposition, Kategorie: 20, trend: 7 , season: periodi  
 value = trend + season\_year + remainder

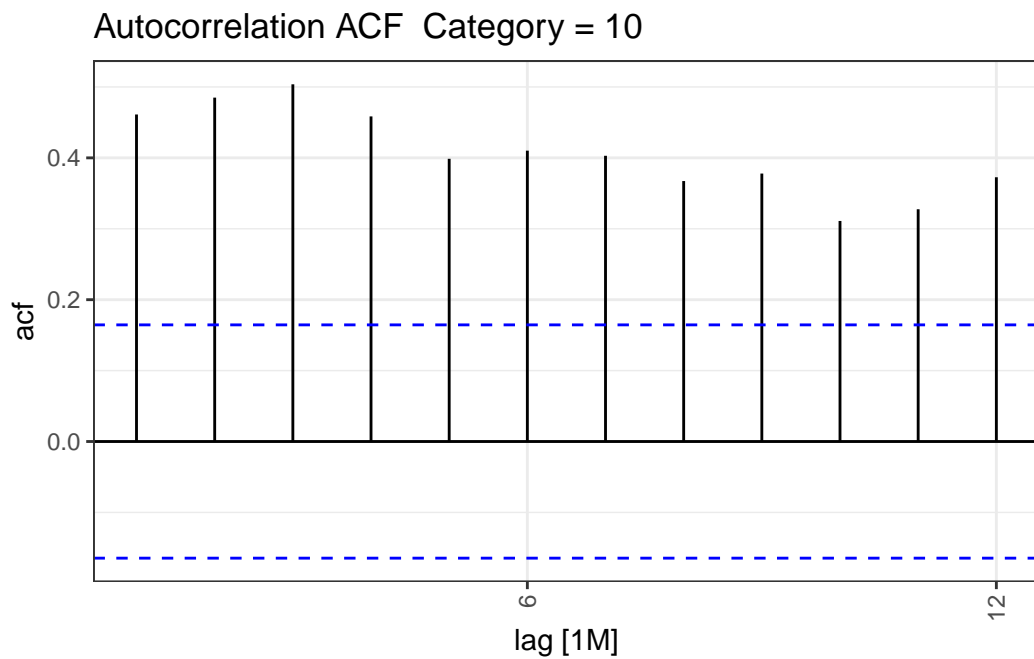
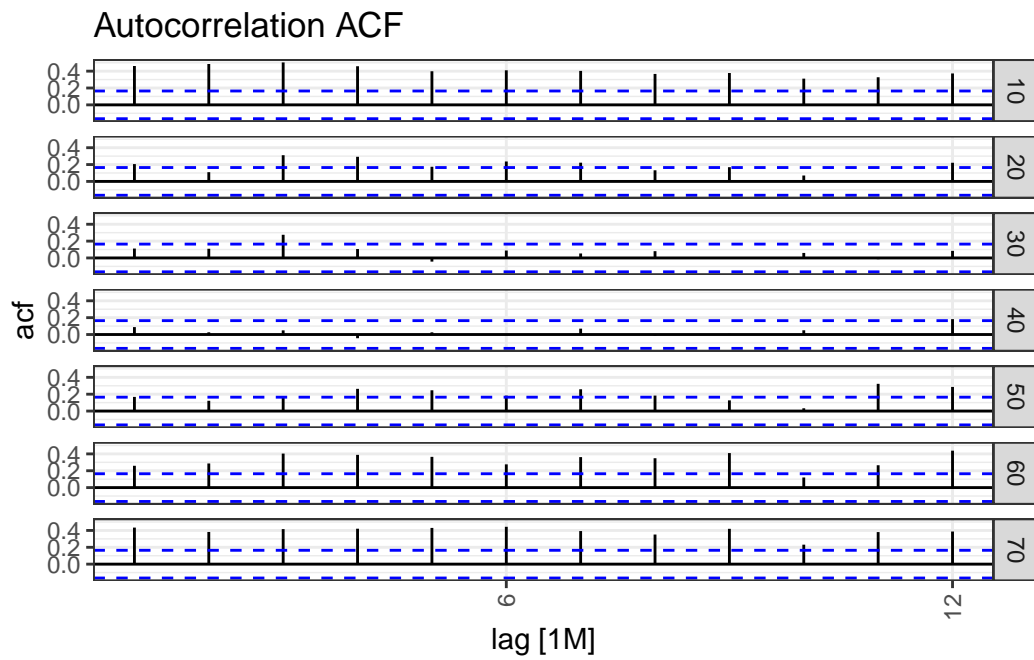


STL features





### 3.1 Autocorrelation



### 3.2 Details

Klassische statistische Werte können relativ leicht aufgelistet werden.

Tabelle 3.1: Durchschnittswert

category	mean
60	767
30	2078
50	2532
20	2590
40	3358
70	9103
10	25267

Tabelle 3.2: Quantile

category	0%	25%	50%	75%	100%
10	7345	18966	22765	29679	60121
20	598	2081	2508	3012	4680
30	394	1528	1910	2398	5078
40	790	2118	3080	4177	10543
50	333	1294	1930	2942	10362
60	141	516	668	965	3031
70	676	6157	7946	11470	23454

Tabelle 3.3: ACF

category	acf1	acf10	diff1_acf1	diff1_acf10	diff2_acf1	diff2_acf10	season_acf1
10	0.46	1.78	-0.49	0.26	-0.64	0.44	0.37
20	0.21	0.42	-0.41	0.27	-0.56	0.39	0.22
30	0.11	0.13	-0.48	0.32	-0.61	0.45	0.08
40	0.09	0.02	-0.46	0.24	-0.63	0.45	0.18
50	0.17	0.35	-0.47	0.31	-0.65	0.54	0.29
60	0.26	1.11	-0.51	0.43	-0.64	0.58	0.44
70	0.43	1.57	-0.42	0.34	-0.58	0.58	0.39

## 4 Export

Alle aufbereiteten Daten können exportiert werden. Aktuelle Möglichkeiten:

- Azure SQL Datenbank
- SQL Datenbank
- CSV Datei
- Excel Datei