

Ein Kochrezept für die Controlling-Praxis am Beispiel der Besucherprognose für ein Sommerbad der Berliner Bäder-Betriebe

Jonas Krembsler, Sandra Spiegelberg, Nicki Lena Kämpf

12.11.2021



Inhaltsverzeichnis

- 1 Einführung
 - Ausgangslage
 - Erwartungshaltung
- 2 Datenfindung - Zutaten zusammenstellen
 - Identifizierung von Einflussfaktoren
 - Datenfindung
- 3 Datenaufbereitung - Zutaten abwiegen
 - Daten der Bäderbetriebe
 - Aufbereitung der Exogenen Variablen
- 4 Nutzung von Streamlit - Rühren und Kochen/Backen
 - Vorstellung des Tools
- 5 Möglichkeiten und Grenzen von Predictive Analytics - Kosten und Genießen
 - Potentielle Probleme
 - Wann funktioniert Predictive Analytics und wann nicht?
- 6 Fazit und Vergabe der Sterne

Überblick

① Einführung

Ausgangslage

Erwartungshaltung

② Datenfindung - Zutaten zusammenstellen

Identifizierung von Einflussfaktoren

Datenfindung

③ Datenaufbereitung - Zutaten abwiegen

Daten der Bäderbetriebe

Aufbereitung der Exogenen Variablen

④ Nutzung von Streamlit - Rühren und Kochen/Backen

Vorstellung des Tools

⑤ Möglichkeiten und Grenzen von Predictive Analytics -

Kosten und Genießen

Potentielle Probleme

Wann funktioniert Predictive Analytics und wann nicht?

⑥ Fazit und Vergabe der Sterne

Ausgangslage

Problem:

Frau Bäderspaß sitzt freitags vor dem Wochenende im Büro und fragt sich, wie die Besucherzahlen eines Sommerbads in der nächsten Woche aussehen werden.

Warum?

Durch eine gute Prognose der Besucherzahlen, kann der Chloreinsatz und eventueller zusätzlicher Personalaufwand durch externe Dienstleister frühzeitig geplant und bestenfalls optimiert werden.

Lösung:

Eine automatisierte Prognose der Besucherzahlen basierend auf exogenen Einflüssen (Treiber), welche die Besucherzahlen beeinflussen.

Bonus:

Monatsprognose der Besucher für das nächste Geschäftsjahr

Erwartungshaltung

Beispiel BBB:

- Durch die kurze Vorbereitung ist keine optimale Prognose zu erwarten
- Aktuell kein Drill-Down möglich
- Prognose steht und fällt mit der Qualität der Daten

Prognosetool:

- Übernimmt nicht die Aufgabe der Datenaufbereitung
- Je nach Fragestellung müssen andere Daten aufbereitet und eingelesen werden

Überblick

- 1 Einführung
 - Ausgangslage
 - Erwartungshaltung
- 2 Datenfindung - Zutaten zusammenstellen
 - Identifizierung von Einflussfaktoren
 - Datenfindung
- 3 Datenaufbereitung - Zutaten abwägen
 - Daten der Bäderbetriebe
 - Aufbereitung der Exogenen Variablen
- 4 Nutzung von Streamlit - Rühren und Kochen/Backen
 - Vorstellung des Tools
- 5 Möglichkeiten und Grenzen von Predictive Analytics - Kosten und Genießen
 - Potentielle Probleme
 - Wann funktioniert Predictive Analytics und wann nicht?
- 6 Fazit und Vergabe der Sterne

Identifizierung von Einflussfaktoren

Zum Start eines Prognoseprojekts, welches über autoregressive Modelle hinaus gehen soll, müssen Treiber/Einflüsse/exogene Variablen identifiziert werden, die die Zielgröße der Prognose beeinflussen (können):

Mögliches Vorgehen:

- Experteneinschätzung innerhalb der Unternehmen
- Externe Expertisen
- Gemeinsames Brainstorming/Mindmap erstellen mit „gesundem Menschenverstand“

Selbstreflexion:

- Welche Daten haben wir im Unternehmen?
- Was ist extern verfügbar (kostenlos oder erwerbbar)?
- Welche Daten sollten wir in Zukunft innerhalb des Unternehmens erheben?
- Welche Faktoren haben einen Einfluss, sind aber nur als Annäherung oder auf eigene Annahmen basiert, zu generieren?

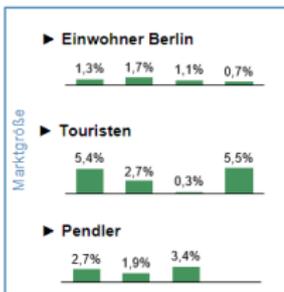
Beispiel BVG

Treiber Nachfrage

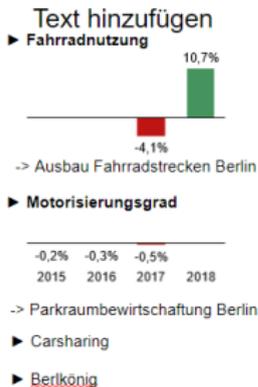
Delta zum Vorjahr in %



Marktgröße



Alternative Mobilität



Vertrieb/Marketing



Angebot

- ▶ Angebotsveränderung
- ▶ Ausfälle Leistung
- ▶ Auslastung

*) Berlin triEAV + Sozialticket

Beispiel BBB

Mögliche Treiber	Spezifizierung
Wetter	<ul style="list-style-type: none"> • Niederschlagsmenge • Sonnenscheindauer • Durchschnittstemperatur • Maximaltemperatur • Minimaltemperatur • ...
Kalendereffekte	<ul style="list-style-type: none"> • Wochentag • Feiertag • Ferientag • Monat • Jahre
...	...

Verfügbarkeit relevanter, unternehmensinterner Daten

Im Unternehmen in den unterschiedlichen Abteilungen auf die Suche nach Daten gehen, welche entweder als Zielgröße relevant sind oder auch als Kennzahl für eine Einflussgröße verwendet werden können:

Beispiele:

- **Mögliche Zielgrößen:**
Verkaufszahlen, Umsatzerlöse, **Kundenanzahl**, Liquidität, Aufträge...
- **Mögliche Einflussfaktoren:**
Angebot, Preis, Kundenzufriedenheit, Reichweite Marketingmaßnahmen, Verweildauer von Kunden in Geschäft, Variable Kosten, Eingesetzte Arbeitskräfte, Aufwand...

→ Müssen in der Regel noch aufbereitet werden.

Amtliche/behördliche Daten

Beispiele:

- **Statistisches Bundesamt:**
Bevölkerung, Tourismuszahlen, Verbraucherpreise, Einkommens, Steuern, Öffentliche Finanzen,...
- **Statistische Ämter der Bundesländer:**
ähnlich wie statistisches Bundesamt, aber landesspezifischer (mitunter Überschneidungen)
- **Bundesagentur für Arbeit:**
Arbeitsmarktspezifische Zahlen (mitunter Überschneidungen mit Genesis): Pendler, Beschäftigung, Erwerbslosigkeit,...
- **Senatsverwaltungen:**
Verkehrszählstellen, Bevölkerungsprognosen,...
- **Deutscher Wetterdienst:**
Diverse Wetterkennzahlen

Daten von privaten Anbietern

Beispiele:

- **Google Trends:**
Suchhäufigkeit von Schlagwörtern bei Google
- **Google Cloud Plattform:**
Sammlung von diversen Datasets (privat und öffentlich)
- **Statista:**
Sammlung diverser Datasets. Teils kostenlos, teils über Pro-Abo erwerbbar
- **Diverse Anbieter, welche Daten verkaufen:**
Bewegungsdaten, Online-Einkaufsdaten, Appdaten,...

→ Nicht immer kostenlos

Selbstständige Erhebung

Falls die Daten dennoch nicht verfügbar sind und trotzdem für unabdingbar betrachtet werden, kann eine eigene Erhebung in Betracht gezogen werden.

Beispiele:

- Auswertung eigener App-Daten
- Auswertung Website Traffic (Google Analytics o.ä.)
- Kundenumfragen, sonstige Umfragen, Marktforschung
- IT-Systeme anpassen und gewünschte Kennzahlen erfassen lassen
- Externe Unternehmen mit Erhebung beauftragen

→ Mit größerem Aufwand und/oder Kosten verbunden. Der Datenschutz sollte gerade bei der Auswertung eigener App-Daten und des Website Traffic nicht verletzt werden

Beispiel Bäderbetriebe

Potenziale beim Datenmanagement:

- Von einer Person abhängig, die Daten exportieren kann. Was passiert wenn im Urlaub oder krank?
- Daten für andere Jahre?
- Umstellung des Systems: Konsistenz
- Data Scientist?
- Zugang für Alle?
- Zentrales Data Warehouse?

→ Datenmanagement genauso wichtig wie Datenqualität!

Überblick

- 1 Einführung
 - Ausgangslage
 - Erwartungshaltung
- 2 Datenfindung - Zutaten zusammenstellen
 - Identifizierung von Einflussfaktoren
 - Datenfindung
- 3 **Datenaufbereitung - Zutaten abwiegen**
 - Daten der Bäderbetriebe
 - Aufbereitung der Exogenen Variablen
- 4 Nutzung von Streamlit - Rühren und Kochen/Backen
 - Vorstellung des Tools
- 5 Möglichkeiten und Grenzen von Predictive Analytics - Kosten und Genießen
 - Potentielle Probleme
 - Wann funktioniert Predictive Analytics und wann nicht?
- 6 Fazit und Vergabe der Sterne

Übermittelte Daten

- pro Jahr 1 Datei *Besuche Durchgang JJJJ.csv* (2019: 300.000 Zeilen)
- pro Jahr 1 Datei *Besuche Kartenausgabe JJJJ.csv* (2019: 200.000 Zeilen)
- eine Datei mit dem Namen *Artikel-Nr. und Bezeichnungen.csv*

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1	Kontrolldaten - Journal																					
2																						
3																						
4	Selektionskriterien:																					
5																						
6	Datum:				01.01.2017 - 31.12.2017																	
7	Zeitraum:				-- keine Einschränkung --																	
8	Zeit:				-- keine Einschränkung --																	
9	Wochentage:				-- keine Einschränkung --																	
10	Betreiber:				-- keine Einschränkung --																	
11	Tarifgruppen:				Bad-Sammelkarten (A1), Bad-Sammelkarten erm? Igt (A2), Dauerkarten (A3), Dauerkarten erm? Igt (A4)																	
12	Kartensorten:				-- keine Einschränkung --																	
13	Mitglieder:				0 - 127-9999999																	
14	Durchgangsort:				Eingänge																	
15																						
16																						
17			Zeit					Berechtigung			Barcode		Medium			Passage			Durchgang			
18																						
19	Datum	Uhrzeit		BBB/Artikel	BarCode	BarCode																
20	29.04.2017	06:59:35		10400 / 1	5,04E+23	400000000000078CE320A																
21																						
22	29.04.2017	06:59:36		10400 / 1	5,08E+23	400000000000005C905F3C																
23																						
24	29.04.2017	06:59:41		10400 / 417	5,00E+23												1				19	
25																						
26	29.04.2017	06:59:43		10400 / 1	5,04E+23	400000000000029CC829A																
27																						
28	29.04.2017	06:59:44		10400 / 417	5,05E+23												1				19	
29																						
30	29.04.2017	06:59:45		10400 / 417	5,08E+23												1				19	
31																						
32	29.04.2017	06:59:46		10400 / 2	5,03E+23	400000000000022E8B2CC																
33																						
34	29.04.2017	06:59:47		10400 / 417	5,09E+23												1				19	
35																						
36	29.04.2017	06:59:49		10400 / 1	5,02E+23	400000000000029CE000A																
37																						

Fragen zu den Daten

- **Soll die Prognose für die gesamten Besucherzahlen oder für eines bestimmten Artikels erfolgen?** Gesamt (must have), Artikel (nice to have)
- **Soll die Prognose für jede Stunde des Tages erfolgen oder für den gesamten Tag?** - Tag (must have), Stunde (nice to have)
- **Bei den Besuche Durchgang gab es Einschränkungen bei den Tarifgruppen. Ist diese Einschränkung so gewollt?** - Ja
- **Welche der Daten sollen verwendet werden? Durchgänge oder Kartenausgabe?** - Beide müssen addiert werden
- **Ist jeder Timestamp ein Gast?** - Ja
- **Können die Inkonsistenzen und Exportprints (Seitenanzahlen) beseitigt werden?** - Nein

Inkonsistenzen innerhalb der Dateien (Bsp. Durchgänge)

- Spalten unterschiedlich bezeichnet/unterschiedlicher Inhalt
- Print des Exportprogramms + Seitenzahl (rot)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
16																				
17									Berechtigung	Barcode	Medium					Passage		Durchgang		
18																				
19	Datum	Uhrzeit	BBB/Artikelnummer	BarCode	BarCode															
20	07.05.2016	06:58:38	10400 / 1	5,08E+23	s00000000000029BEOBAA				176 >>>											
21																				
22	07.05.2016	06:58:43	10400 / 3	5,00E+23	s00000000000029D6914A				176 >>>											
23																				
24	07.05.2016	06:58:53	10400 / 415	5,07E+23					176 >>>							1				19
25																				
26	07.05.2016	06:58:57	10400 / 1	5,01E+23	s0000000000007BCA048A				176 >>>											
27																				
28	07.05.2016	06:58:59	10400 / 1	5,07E+23	s0000000000007BCD690A				176 >>>											
29																				
47																				
48	07.05.2016	06:59:28	10400 / 415	5,02E+23					176 >>>							1				19
49																				
62	07.05.2016	07:07:50	10400 / 1	5,03E+23	s0000000000007BC5441A				176 >>>											
63																				
64	07.05.2016	07:09:06	10400 / 415	5,08E+23					176 >>>								1			18
65																				
66	<u>#10400.102 Berliner B7der Betriebe, Access Office 4.00 Build 350</u>															<u>Seite: 1 / 2874</u>				
67																				
68		Zeit				Berechtigung	Barcode	Medium	Passage	Durchgang	Wertekonto	Gel								
69																				
70																				
71	07.05.2016	07:12:23	10400 / 415	5,06E+23					176 >>>							1				19
72																				
73	07.05.2016	07:16:27	10400 / 415	5,10E+23					176 >>>							1				19
74																				
75	07.05.2016	07:16:34	10400 / 1	5,04E+23	s0000000000007BC49DFA				176 >>>											
76																				

Inkonsistenzen zwischen den Dateien

- Tabelle liefert andere Daten
- Tabelle anders aufgebaut

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
19																		
20																		
21																		
22	Datum	Uhrzeit		Artikelnum	Artikel-Bezeichnung													
23																		
34	07.05.2016	06:59:24	17/ 117- 2	115	Einzelkarte Haupttarif Sommerbad													
36	07.05.2016	08:00:36	17/ 117- 2	115	Einzelkarte Haupttarif Sommerbad													
37	07.05.2016	08:32:02	17/ 117- 2	115	Einzelkarte Haupttarif Sommerbad													
39	07.05.2016	08:32:55	17/ 117- 2	115	Einzelkarte Haupttarif Sommerbad													
45	07.05.2016	08:51:13	17/ 117- 2	115	Einzelkarte Haupttarif Sommerbad													
46	07.05.2016	09:02:11	17/ 117- 2	115	Einzelkarte Haupttarif Sommerbad													
47	07.05.2016	09:02:21	17/ 117- 2	115	Einzelkarte Haupttarif Sommerbad													
48	07.05.2016	09:06:37	17/ 117- 2	115	Einzelkarte Haupttarif Sommerbad													
49	07.05.2016	09:07:17	17/ 117- 2	115	Einzelkarte Haupttarif Sommerbad													
50	07.05.2016	09:11:40	17/ 117- 2	115	Einzelkarte Haupttarif Sommerbad													
51	07.05.2016	09:12:54	17/ 117- 2	115	Einzelkarte Haupttarif Sommerbad													
52																		
53	#10400.102 Berliner B?der Betriebe, Access Office 4.00 Buld 350												Seite: 1 / 2723					
54																		
55		Zeit			Verkaufsstelle									Preis/Stk		Steuer		Gesamt
56																		
57														Brutto		Gesamt		Brutto
58																		
59																		
60	07.05.2016	09:22:18	17/ 117- 2	115	Einzelkarte Haupttarif Sommerbad													
61	07.05.2016	09:32:29	17/ 117- 2	125	Einzelkarte erm? igt Haupttarif Sommerbad													
62	07.05.2016	09:32:51	17/ 117- 2	248	Nullk EF Begleitperson													
63	07.05.2016	09:35:05	17/ 117- 2	150	Familienkarte - 1.Erwachsene													
64	07.05.2016	09:35:06	17/ 117- 2	152	Badespa - Kind													

Lösung

Bei grob geschätzten 1,5 Mio. - 2 Mio. Zeilen insgesamt (nicht Besuchern) ist eine händische Bereinigung offensichtlich unmöglich.

→ **Automatisierung**

Skizzierte Schritte der Automatisierung in Python:

- Alle Dateien untereinander hängen
- Nur die Zeilen behalten, welche in Spalte A ein Datum enthalten
- Nach dem Datum gruppieren und sortieren
- Alle Einträge je Tag aufaddieren
- Laufzeit des Programms (knapp 3h)

→ Alle weiteren Informationen (Ticketart, Barcode, usw.) sind nicht mehr enthalten.

Aufbereitet

- Links: Tägliche Besucher. Rechts: Monatliche Besucher pro Tag
- Warum nicht Gesamtbesucher pro Monat? - Durch die unterschiedlichen Öffnungsdaten können mal mehr oder weniger Öffnungstage im April/Mai und September liegen.

	A	B
1		Besucher
2	2016-05-07 00:00:00	1191
3	2016-05-08 00:00:00	1478
4	2016-05-09 00:00:00	956
5	2016-05-10 00:00:00	1501
6	2016-05-11 00:00:00	1521
7	2016-05-12 00:00:00	1144
8	2016-05-13 00:00:00	756
9	2016-05-14 00:00:00	261
10	2016-05-15 00:00:00	201
11	2016-05-16 00:00:00	241
12	2016-05-17 00:00:00	161
13	2016-05-18 00:00:00	406
14	2016-05-19 00:00:00	531
15	2016-05-20 00:00:00	710
16	2016-05-21 00:00:00	849
17	2016-05-22 00:00:00	3773
18	2016-05-23 00:00:00	1287
19	2016-05-24 00:00:00	556
20	2016-05-25 00:00:00	356
21	2016-05-26 00:00:00	329

	A	B	P
1		Besucher pro Tag	
2	2016-05-31 00:00:00	971,28	
3	2016-06-30 00:00:00	1580,8	
4	2016-07-31 00:00:00	1550	
5	2016-08-31 00:00:00	1716,806452	
6	2016-09-30 00:00:00	2636	
7	2017-04-30 00:00:00	294,5	
8	2017-05-31 00:00:00	1312,096774	
9	2017-06-30 00:00:00	1430,733333	
10	2017-07-31 00:00:00	1307,419355	
11	2017-08-31 00:00:00	1859,935484	
12	2017-09-30 00:00:00	601,4705882	
13	2018-04-30 00:00:00	761	
14	2018-05-31 00:00:00	2101,709677	
15	2018-06-30 00:00:00	1737,033333	
16	2018-07-31 00:00:00	3245,870968	
17	2018-08-31 00:00:00	2722,064516	
18	2018-09-30 00:00:00	1482,142857	
19	2019-04-30 00:00:00	777,8333333	
20	2019-05-31 00:00:00	606,4193548	
21	2019-06-30 00:00:00	3844,933333	

Was nehmen wir mit?

- **zwei verschiedene Tabellen pro Jahr müssen erst zusammengefasst werden.**
- **wenn möglich: Inkonsistenz innerhalb einer Datei vermeiden**
- **wenn möglich: Inkonsistenz zwischen den Dateien vermeiden**
- **Rücksprachen mit Experten notwendig**

Was wäre für uns optimal gewesen?

Daten an Fragestellung angepasst in einer konsistenten Datei übergeben: Besucher pro Tag (oder Monat)

Wetterdaten

Schritte

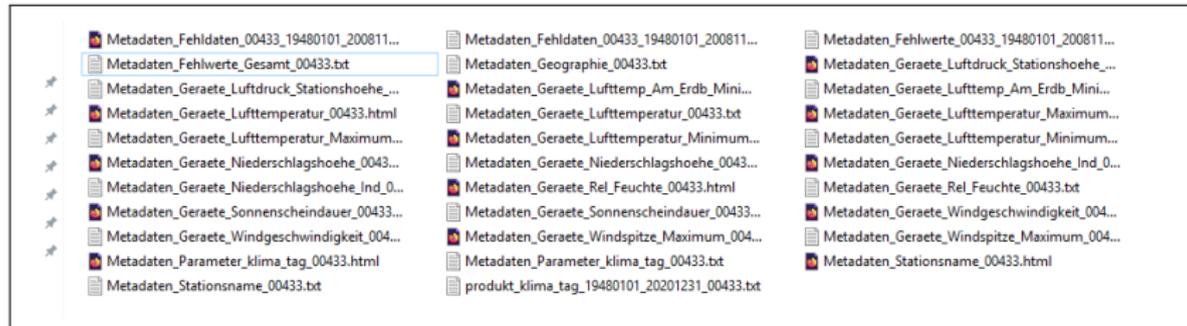
- Auf der Seite des DW zurecht finden
- Passende Daten für meine Problemstellung finden
- Richtige Wetterstation finden

Quelle: Deutscher Wetterdienst

Wetterdaten

Schritte

- Richtige Datei identifizieren



Wetterdaten

Schritte

- Überblick über die Daten verschaffen (Datensatzbeschreibung)

produkt_klima_tag_19480101_20201231_00433.txt - Editor

Datei Bearbeiten Format Ansicht Hilfe

STATIONS_ID;MESS_DATUM;QN_3; FX; FM;QN_4; RSK;RSKF; SDK;SHK_TAG; NM; VPM; PM; TMK; UPM; TXK; TNK; TGK;eor
433;19480101;-999;-999;-999; 5; 5.7; 8;-999; 4; 5.0; 4.9; 1010.70; -1.8; 89.00; 0.8; -5.0; -4.7;eor
433;19480102;-999;-999;-999; 5; 3.3; 1;-999; 8; 8.0; 7.9; 999.70; 4.4; 95.00; 6.8; -2.7; -2.5;eor
433;19480103;-999;-999;-999; 5; 4.0; 1;-999; 0; 5.3; 9.8; 1001.70; 8.5; 86.00; 10.6; 5.2; 4.6;eor
433;19480104;-999;-999;-999; 5; 0.2; 1;-999; 0; 8.0; 10.2; 999.80; 10.2; 83.00; 11.5; 6.6; 4.7;eor
433;19480105;-999;-999;-999; 5; 2.0; 1;-999; 0; 7.7; 9.8; 987.80; 10.5; 78.00; 12.0; 8.0; 5.9;eor
433;19480106;-999;-999;-999; 5; 0.0; 0;-999; 0; 4.7; 6.0; 993.60; 3.3; 75.00; 10.8; 1.8; 2.9;eor
433;19480107;-999;-999;-999; 5; 1.5; 1;-999; 0; 8.0; 6.9; 980.90; 2.9; 96.00; 5.2; -0.9; -2.0;eor
433;19480108;-999;-999;-999; 5; 7.8; 8;-999; 0; 5.3; 7.3; 978.40; 6.2; 74.00; 9.5; 4.7; 4.0;eor
433;19480109;-999;-999;-999; 5; 2.1; 7;-999; 0; 6.0; 5.5; 997.50; 0.4; 86.00; 4.7; -0.2; -2.2;eor
433;19480110;-999;-999;-999; 5; 8.4; 8;-999; 0; 7.3; 6.1; 992.50; 0.7; 92.00; 3.0; -1.2; -3.0;eor
433;19480111;-999;-999;-999; 5; 0.0; 7;-999; 1; 5.0; 3.8; 1011.30; -3.0; 79.00; -0.1; -4.9; -6.1;eor
433;19480112;-999;-999;-999; 5; 8.5; 1;-999; 1; 7.0; 6.8; 1000.40; 3.8; 85.00; 7.0; -2.8; -4.0;eor
433;19480113;-999;-999;-999; 5; 33.4; 1;-999; 0; 8.0; 10.3; 988.10; 9.3; 92.00; 11.4; 4.2; 4.1;eor
433;19480114;-999;-999;-999; 5; 3.8; 1;-999; 0; 7.3; 7.5; 991.80; 5.7; 80.00; 11.9; 4.2; 4.8;eor
433;19480115;-999;-999;-999; 5; 0.4; 8;-999; 0; 8.0; 5.7; 997.40; 0.5; 87.00; 4.6; -0.9; 1.0;eor
433;19480116;-999;-999;-999; 5; 0.7; 7;-999; 1; 8.0; 4.1; 1003.30; -1.7; 74.00; 0.2; -2.3; -3.2;eor
433;19480117;-999;-999;-999; 5; 0.0; 0;-999; 1; 7.0; 4.4; 1005.00; -2.8; 90.00; -1.6; -4.7; -5.7;eor
433;19480118;-999;-999;-999; 5; 3.9; 7;-999; 2; 4.7; 5.6; 991.60; 0.7; 88.00; 2.5; -2.5; -5.0;eor
433;19480119;-999;-999;-999; 5; 0.0; 0;-999; 0; 2.7; 5.7; 988.90; 0.0; 92.00; 2.7; -0.7; -3.6;eor
433;19480120;-999;-999;-999; 5; 0.0; 0;-999; 0; 7.3; 5.4; 990.10; -0.7; 92.00; 3.4; -5.0; -5.1;eor
433;19480121;-999;-999;-999; 5; 0.0; 0;-999; 0; 7.0; 4.6; 998.00; -0.4; 78.00; 0.7; -1.7; -3.4;eor
433;19480122;-999;-999;-999; 5; 0.0; 7;-999; 0; 7.7; 4.4; 994.80; -1.0; 78.00; 1.4; -2.6; -5.0;eor
433;19480123;-999;-999;-999; 5; 0.0; 7;-999; 2; 8.0; 5.6; 998.00; 0.4; 88.00; 2.1; -1.9; -4.7;eor
433;19480124;-999;-999;-999; 5; 0.0; 0;-999; 2; 8.0; 5.2; 1001.70; 0.0; 86.00; 1.0; -1.3; -4.0;eor
433;19480125;-999;-999;-999; 5; 0.0; 0;-999; 1; 5.3; 5.8; 1000.20; 0.3; 93.00; 2.3; -1.4; -3.0;eor
433;19480126;-999;-999;-999; 5; 0.0; 0;-999; 0; 1.7; 6.5; 994.80; 2.4; 89.00; 7.0; -1.5; -3.2;eor
433;19480127;-999;-999;-999; 5; 0.0; 0;-999; 0; 4.3; 6.7; 992.20; 2.2; 93.00; 4.8; -0.2; -2.7;eor
433;19480128;-999;-999;-999; 5; 3.9; 1;-999; 0; 8.0; 7.3; 992.40; 3.1; 97.00; 4.3; 1.1; -1.8;eor
433;19480129;-999;-999;-999; 5; 0.1; 1;-999; 0; 5.0; 7.0; 996.50; 3.0; 92.00; 5.5; -0.5; -2.4;eor

Wetterdaten

Schritte

- Daten für meine Fragestellung und meinen Zeithorizont aufbereiten
- Umgang mit fehlenden Werten

1		Niederschlagshöhe	Sonnenscheindauer	Tagesmittel der Temperatur	Tagesmaximum der Temperatur	Tagesminimum der Temperatur
2	2016-05-07 00:00:00	0	13,8	18,3	24,7	10,1
3	2016-05-08 00:00:00	0	12,8	18,4	24,2	10,8
4	2016-05-09 00:00:00	0	13,6	19,1	24,9	11,3
5	2016-05-10 00:00:00	0	13,6	18,5	24,8	11,9
6	2016-05-11 00:00:00	0	11,6	19,1	25,3	11,4
7	2016-05-12 00:00:00	2,2	9,7	18,4	24,5	13,5
8	2016-05-13 00:00:00	0	11,7	16,4	22,5	9,7
9	2016-05-14 00:00:00	0,9	6,6	10,9	15	7
10	2016-05-15 00:00:00	0,2	5,7	9	13,5	6,2
11	2016-05-16 00:00:00	0,3	3,7	9,9	15	6,4
12	2016-05-17 00:00:00	0,5	0,3	10	12,9	7,2
13	2016-05-18 00:00:00	0	5,4	15	20,6	8
14	2016-05-19 00:00:00	0	5,3	17,6	22,7	12,7
15	2016-05-20 00:00:00	0	5,6	17,8	23,4	13,1
16	2016-05-21 00:00:00	0	6,6	19,5	24,3	14,9
17	2016-05-22 00:00:00	0	10,8	22,4	29,4	14,2
18	2016-05-23 00:00:00	15,2	9,9	21	26,3	15,1
19	2016-05-24 00:00:00	0	1	16,3	20,2	13,2
20	2016-05-25 00:00:00	3,6	0	13,9	15,5	12,6
21	2016-05-26 00:00:00	0	0	14,2	16	11,4
22	2016-05-27 00:00:00	0	5,8	16,4	21,6	9,4
23	2016-05-28 00:00:00	0	7,6	18,4	23,9	10,7
24	2016-05-29 00:00:00	0,2	10,2	21,4	28,1	14,9
25	2016-05-30 00:00:00	18,7	8,3	21	27,4	15,2
26	2016-05-31 00:00:00	0,8	5,2	19,9	25,5	14,1
27	2016-06-01 00:00:00	13,2	3,3	19,3	22,8	17,4

Wetterdaten (Monat)

Schritte

- Daten für meine Fragestellung und meinen Zeithorizont aufarbeiten (Monatsdaten)

	A	J	K	L	M	N	P
1		Tage mit Regen >3	Median Sonnenscheindauer	Median Tagestemperatur	Median Höchsttemperatur	Median Tiefsttemperatur	
2	2016-05-31 00:00:00	3	6,6	18,3	23,9	11,4	
3	2016-06-30 00:00:00	4	8,15	19,45	24,2	13,85	
4	2016-07-31 00:00:00	4	6,3	20,2	25,8	16,6	
5	2016-08-31 00:00:00	5	6,5	19,1	23	14,4	
6	2016-09-30 00:00:00	1	10,9	21,1	27,2	15,4	
7	2017-04-30 00:00:00	0	8,8	7,7	13,2	0,95	
8	2017-05-31 00:00:00	2	7,133	15,7	19,8	9,7	
9	2017-06-30 00:00:00	9	6,6335	18,45	23,3	13,5	
10	2017-07-31 00:00:00	10	6,367	19	23,7	14,7	
11	2017-08-31 00:00:00	5	7,233	19,1	24	14,2	
12	2017-09-30 00:00:00	2	4,483	14,8	19,1	10,4	
13	2018-04-30 00:00:00	0	10,183	17,3	23,9	11,6	
14	2018-05-31 00:00:00	2	12,383	17,8	24	11,2	
15	2018-06-30 00:00:00	3	7,4585	20,05	25,05	14,35	
16	2018-07-31 00:00:00	3	12	22	27,8	15,7	
17	2018-08-31 00:00:00	0	9,55	21,6	27,6	15,6	
18	2018-09-30 00:00:00	0	7,633	19,6	26,3	13,6	
19	2019-04-30 00:00:00	1	11,075	13,8	19,95	8,3	
20	2019-05-31 00:00:00	3	6,133	13,1	17,6	8,1	

Kalenderdaten

Schritte

- Identifizierung der Kalendereffekte die für meine Prognose wichtig sind: Wochentag, Feiertag, Ferientag, Monat
- Wochentag, Monat, gesetzliche Feiertage automatisierbar (Python, Makro etc.)
- Sonderfeiertage (z.B. Tag der Befreiung, Weltfrauentag (neu) müssen händisch ergänzt werden)
- Ferientage müssen recherchiert werden und bestenfalls anschließend automatisieren
- Daten ins gewünschte Format bringen (Vorwissen gefragt)

Kalenderdaten täglich

- Umgang mit kategorialen Variablen

1	Jahr	Monat	Wochentag	Ferientag	April	August	July	June	May	September	Friday	Monday	Saturday	Sunday/Holiday	Thursday	Tuesday	Wednesday
2	2016-05-07 00:00:00	2016 May	Saturday	FALSCH	0	0	0	0	1	0	0	0	1	0	0	0	0
3	2016-05-08 00:00:00	2016 May	Sunday/Holiday	FALSCH	0	0	0	0	1	0	0	0	0	1	0	0	0
4	2016-05-09 00:00:00	2016 May	Monday	FALSCH	0	0	0	0	1	0	0	1	0	0	0	0	0
5	2016-05-10 00:00:00	2016 May	Tuesday	FALSCH	0	0	0	0	1	0	0	0	0	0	0	0	1
6	2016-05-11 00:00:00	2016 May	Wednesday	FALSCH	0	0	0	0	1	0	0	0	0	0	0	0	1
7	2016-05-12 00:00:00	2016 May	Thursday	FALSCH	0	0	0	0	1	0	0	0	0	0	1	0	0
8	2016-05-13 00:00:00	2016 May	Friday	FALSCH	0	0	0	0	1	0	1	0	0	0	0	0	0
9	2016-05-14 00:00:00	2016 May	Saturday	FALSCH	0	0	0	0	1	0	0	0	1	0	0	0	0
10	2016-05-15 00:00:00	2016 May	Sunday/Holiday	FALSCH	0	0	0	0	1	0	0	0	0	1	0	0	0
11	2016-05-16 00:00:00	2016 May	Sunday/Holiday	FALSCH	0	0	0	0	1	0	0	0	0	1	0	0	0
12	2016-05-17 00:00:00	2016 May	Tuesday	WAHR	0	0	0	0	1	0	0	0	0	0	0	1	0
13	2016-05-18 00:00:00	2016 May	Wednesday	WAHR	0	0	0	0	1	0	0	0	0	0	0	0	1
14	2016-05-19 00:00:00	2016 May	Thursday	FALSCH	0	0	0	0	1	0	0	0	0	0	1	0	0
15	2016-05-20 00:00:00	2016 May	Friday	FALSCH	0	0	0	0	1	0	1	0	0	0	0	0	0
16	2016-05-21 00:00:00	2016 May	Saturday	FALSCH	0	0	0	0	1	0	0	1	0	0	0	0	0
17	2016-05-22 00:00:00	2016 May	Sunday/Holiday	FALSCH	0	0	0	0	1	0	0	0	0	1	0	0	0
18	2016-05-23 00:00:00	2016 May	Monday	FALSCH	0	0	0	0	1	0	0	1	0	0	0	0	0
19	2016-05-24 00:00:00	2016 May	Tuesday	FALSCH	0	0	0	0	1	0	0	0	0	0	0	1	0
20	2016-05-25 00:00:00	2016 May	Wednesday	FALSCH	0	0	0	0	1	0	0	0	0	0	0	0	1
21	2016-05-26 00:00:00	2016 May	Thursday	FALSCH	0	0	0	0	1	0	0	0	0	0	1	0	0
22	2016-05-27 00:00:00	2016 May	Friday	FALSCH	0	0	0	0	1	0	1	0	0	0	0	0	0
23	2016-05-28 00:00:00	2016 May	Saturday	FALSCH	0	0	0	0	1	0	0	0	1	0	0	0	0
24	2016-05-29 00:00:00	2016 May	Sunday/Holiday	FALSCH	0	0	0	0	1	0	0	0	0	1	0	0	0
25	2016-05-30 00:00:00	2016 May	Monday	FALSCH	0	0	0	0	1	0	0	1	0	0	0	0	0
26	2016-05-31 00:00:00	2016 May	Tuesday	FALSCH	0	0	0	0	1	0	0	0	0	0	0	0	1
27	2016-06-01 00:00:00	2016 June	Wednesday	FALSCH	0	0	0	1	0	0	0	0	0	0	0	0	1
28	2016-06-02 00:00:00	2016 June	Thursday	FALSCH	0	0	0	1	0	0	0	0	0	0	1	0	0
29	2016-06-03 00:00:00	2016 June	Friday	FALSCH	0	0	0	1	0	0	1	0	0	0	0	0	0
30	2016-06-04 00:00:00	2016 June	Saturday	FALSCH	0	0	0	1	0	0	0	0	1	0	0	0	0

Kalenderdaten Monatlich

	A	C	D	E	F	G	H	I	O	P
1		Montage	Dienstage	Mittwochs	Donnerstage	Freitage	Samstage	Sonn-/Feiertag	Monat	
2	2016-05-31 00:00:00	3	4	3	3	3	4	5	May	
3	2016-06-30 00:00:00	4	4	5	5	4	4	4	June	
4	2016-07-31 00:00:00	4	4	4	4	5	5	5	July	
5	2016-08-31 00:00:00	5	5	5	4	4	4	4	August	
6	2016-09-30 00:00:00	2	2	2	3	2	2	2	September	
7	2017-04-30 00:00:00	0	0	0	0	0	1	1	April	
8	2017-05-31 00:00:00	4	5	5	3	4	4	4	May	
9	2017-06-30 00:00:00	3	4	4	5	5	4	4	June	
10	2017-07-31 00:00:00	5	4	4	4	4	5	5	July	
11	2017-08-31 00:00:00	4	5	5	5	4	4	4	August	
12	2017-09-30 00:00:00	2	2	2	2	3	3	3	September	
13	2018-04-30 00:00:00	1	0	0	0	0	1	1	April	
14	2018-05-31 00:00:00	3	4	5	4	4	4	4	May	
15	2018-06-30 00:00:00	4	4	4	4	5	5	4	June	
16	2018-07-31 00:00:00	5	5	4	4	4	4	4	July	
17	2018-08-31 00:00:00	4	4	5	5	5	4	4	August	
18	2018-09-30 00:00:00	3	3	3	3	3	3	3	September	
19	2019-04-30 00:00:00	1	2	1	1	1	2	2	April	
20	2019-05-31 00:00:00	4	5	5	4	4	4	4	May	

Daten für Streamlit

	A	B	C	D	E	F	Z	AA	AB	AC	AD
1		Besucher	Jahr	Monat	Wochentag	Ferientag	Niederschlagshöhe	Sonnenscheindauer	Tagesmittel der Temperatur	Tagesmaximum der Temperatur	Tagesminimum der Temperatur
2	2016-05-07 00:00:00	1191	2016	May	Saturday	0	0	13,8	18,3	24,7	10,1
3	2016-05-08 00:00:00	1478	2016	May	Sunday/Holiday	0	0	12,8	18,4	24,2	10,8
4	2016-05-09 00:00:00	956	2016	May	Monday	0	0	13,6	19,1	24,9	11,3
5	2016-05-10 00:00:00	1501	2016	May	Tuesday	0	0	13,6	18,5	24,8	11,9
6	2016-05-11 00:00:00	1521	2016	May	Wednesday	0	0	11,6	19,1	25,3	11,4
7	2016-05-12 00:00:00	1144	2016	May	Thursday	0	2,2	9,7	18,4	24,5	13,5
8	2016-05-13 00:00:00	756	2016	May	Friday	0	0	11,7	16,4	22,5	9,7
9	2016-05-14 00:00:00	261	2016	May	Saturday	0	0,9	6,6	10,9	15	7
10	2016-05-15 00:00:00	201	2016	May	Sunday/Holiday	0	0,2	5,7	9	13,5	6,2
11	2016-05-16 00:00:00	241	2016	May	Sunday/Holiday	0	0,3	3,7	9,9	15	6,4
12	2016-05-17 00:00:00	161	2016	May	Tuesday	1	0,5	0,3	10	12,9	7,2
13	2016-05-18 00:00:00	406	2016	May	Wednesday	1	0	5,4	15	20,6	8
14	2016-05-19 00:00:00	531	2016	May	Thursday	0	0	5,3	17,6	22,7	12,7
15	2016-05-20 00:00:00	710	2016	May	Friday	0	0	5,6	17,8	23,4	13,1
16	2016-05-21 00:00:00	849	2016	May	Saturday	0	0	6,6	19,5	24,3	14,9
17	2016-05-22 00:00:00	3773	2016	May	Sunday/Holiday	0	0	10,8	22,4	29,4	14,2
18	2016-05-23 00:00:00	1287	2016	May	Monday	0	15,2	9,9	21	26,3	15,1
19	2016-05-24 00:00:00	556	2016	May	Tuesday	0	0	1	16,3	20,2	13,2
20	2016-05-25 00:00:00	356	2016	May	Wednesday	0	3,6	0	13,9	15,5	12,6
21	2016-05-26 00:00:00	329	2016	May	Thursday	0	0	0	14,2	16	11,4
22	2016-05-27 00:00:00	678	2016	May	Friday	0	0	5,8	16,4	21,6	9,4
23	2016-05-28 00:00:00	1141	2016	May	Saturday	0	0	7,6	18,4	23,9	10,7
24	2016-05-29 00:00:00	2269	2016	May	Sunday/Holiday	0	0,2	10,2	21,4	28,1	14,9
25	2016-05-30 00:00:00	1086	2016	May	Monday	0	18,7	8,3	21	27,4	15,2
26	2016-05-31 00:00:00	900	2016	May	Tuesday	0	0,8	5,2	19,9	25,5	14,1
27	2016-06-01 00:00:00	511	2016	June	Wednesday	0	13,2	3,3	19,3	22,8	17,4
28	2016-06-02 00:00:00	1051	2016	June	Thursday	0	2,9	6,6	19,9	25,1	14,5
29	2016-06-03 00:00:00	1837	2016	June	Friday	0	0	13,4	21,8	25,8	16,8

Überblick

- 1 Einführung
 - Ausgangslage
 - Erwartungshaltung
- 2 Datenfindung - Zutaten zusammenstellen
 - Identifizierung von Einflussfaktoren
 - Datenfindung
- 3 Datenaufbereitung - Zutaten abwiegen
 - Daten der Bäderbetriebe
 - Aufbereitung der Exogenen Variablen
- 4 Nutzung von Streamlit - Rühren und Kochen/Backen**
 - Vorstellung des Tools**
- 5 Möglichkeiten und Grenzen von Predictive Analytics - Kosten und Genießen
 - Potentielle Probleme
 - Wann funktioniert Predictive Analytics und wann nicht?
- 6 Fazit und Vergabe der Sterne

Vorstellung des Tools

Streamlit:

- Streamlit ermöglicht es, Web-Apps basierend auf Python-Code zu erstellen
- kostenlose Software, die auch kommerziell genutzt werden kann
- App wird auf dem lokalen Rechner installiert und ausgeführt

Vision:

- automatisierte Prognose für verschiedene Zeitreihen (Erträge, Einnahmen, ...) für beliebige Branchen
- Vergleich von verschiedenen Methoden für die Prognose einer Zeitreihe
- Modellparameter sollen ohne Zutun des Anwenders gefunden werden

Überblick

- 1 Einführung
 - Ausgangslage
 - Erwartungshaltung
- 2 Datenfindung - Zutaten zusammenstellen
 - Identifizierung von Einflussfaktoren
 - Datenfindung
- 3 Datenaufbereitung - Zutaten abwiegen
 - Daten der Bäderbetriebe
 - Aufbereitung der Exogenen Variablen
- 4 Nutzung von Streamlit - Rühren und Kochen/Backen
 - Vorstellung des Tools
- 5 Möglichkeiten und Grenzen von Predictive Analytics -
Kosten und Genießen
 - Potentielle Probleme
 - Wann funktioniert Predictive Analytics und wann nicht?
- 6 Fazit und Vergabe der Sterne

Potentielle Probleme

Mögliche fehlende Informationen

- Exogene Einflüsse die über unser Wetter hinausgehen, können die Besucherzahl beeinflussen (Bsp. WM-Fußballspiel, usw.)
- Interne Einflüsse (Reparaturarbeiten, Aktionen, usw.)

Sommerbetriebproblematik

- Das Sommerbad hat nur von April/Mai bis September geöffnet.
→ Es gibt einen Sprung von September zum April/Mai, welche autoregressive Modelle falsch lernen lässt!
- Das Sommerbad öffnet und schließt an unterschiedlichen Daten.
→ Falls das Bad im Trainingsdatensatz nicht im April geöffnet hat, ist auch eine Prognose für April nicht möglich!
- Das Sommerbad öffnet und schließt an unterschiedlichen Wochentagen.
→ Zeitreihenmodelle lernen aus Zyklen (Wochenzyklus). Wenn das Bad an einem Sonntag schließt und an einem Freitag öffnet, ist der Zyklus ein anderer.

Wann funktioniert Predictive Analytics und wann nicht?

- Sehr gute Datenqualität:
Keine Systemänderungen im Reporting, Vollständigkeit, Plausibilität,...
- Datenverfügbarkeit (auch ausreichende Verfügbarkeit der Historie oder Granularität)
- Keine unerwarteten Buchungsverschiebungen/Abrechnungsverschiebungen
- Keine gravierende Änderung der Ausgangssituation (innerhalb oder außerhalb des Unternehmens):
Schocks, neue Markteintritte/Konkurrenten, rechtliche/politische Veränderungen, Produktveränderungen,...
- Modell korrekt spezifiziert:
Kausaler Zusammenhang zwischen Treibern und Zielgröße gegeben, Modellparameter korrekt bestimmt,...

Fazit und Vergabe der Sterne
Wie hat es Ihnen geschmeckt?

Fragen

- **Wie würde ein konkreter Anwendungsfall in Ihrem Unternehmen aussehen? Wofür wird die Prognose verwendet?**
 - Klare Zielsetzung
 - Zeitreihenproblem
 - Big-Data
 - ...
- **Was wäre der Mehrwert dieses Anwendungsfalls?**
 - Zeitliche Ersparnis
 - Monetär
 - ...
- **Sind Voraussetzungen für Predictive Analytics gegeben?**
 - Daten (Qualität, Management, Verfügbarkeit, ...)
 - Personal (Zeit, Finanzen, ...)
 - ...

Kontakte

Sandra Spiegelberg

Wissenschaftliche Mitarbeiterin
Berliner Hochschule für Technik
Email: Sandra.Spiegelberg@bht-berlin.de

Nicki Lena Kämpf

Wissenschaftliche Mitarbeiterin
Berliner Hochschule für Technik
Email: NickiLena.Kaempf@bht-berlin.de

Jonas Krembsler

Wissenschaftlicher Mitarbeiter
Hochschule für Wirtschaft und Recht
Email: Jonas.Krembsler@hwr-berlin.de
Tel: +49 (0) 30-30877-2618