







The image features a close-up of a wooden surface with a vertical grain. The wood is painted with the colors of the Venezuelan flag: yellow at the top, blue in the middle, and red at the bottom. White stars are visible on the blue band. A large, dark blue diagonal shape cuts across the right side of the image, serving as a background for the text.

Venezuela – Flucht oder Kampf?

Lisa Yerebakan, Sandro Rüttimann,
Calebe Künzle, Silvan Saxer
& Pascal Kohler

Präsentation: 21.05.2019



Calebe Künzle

Journalismus



Lisa Yerebakan

Mathematik



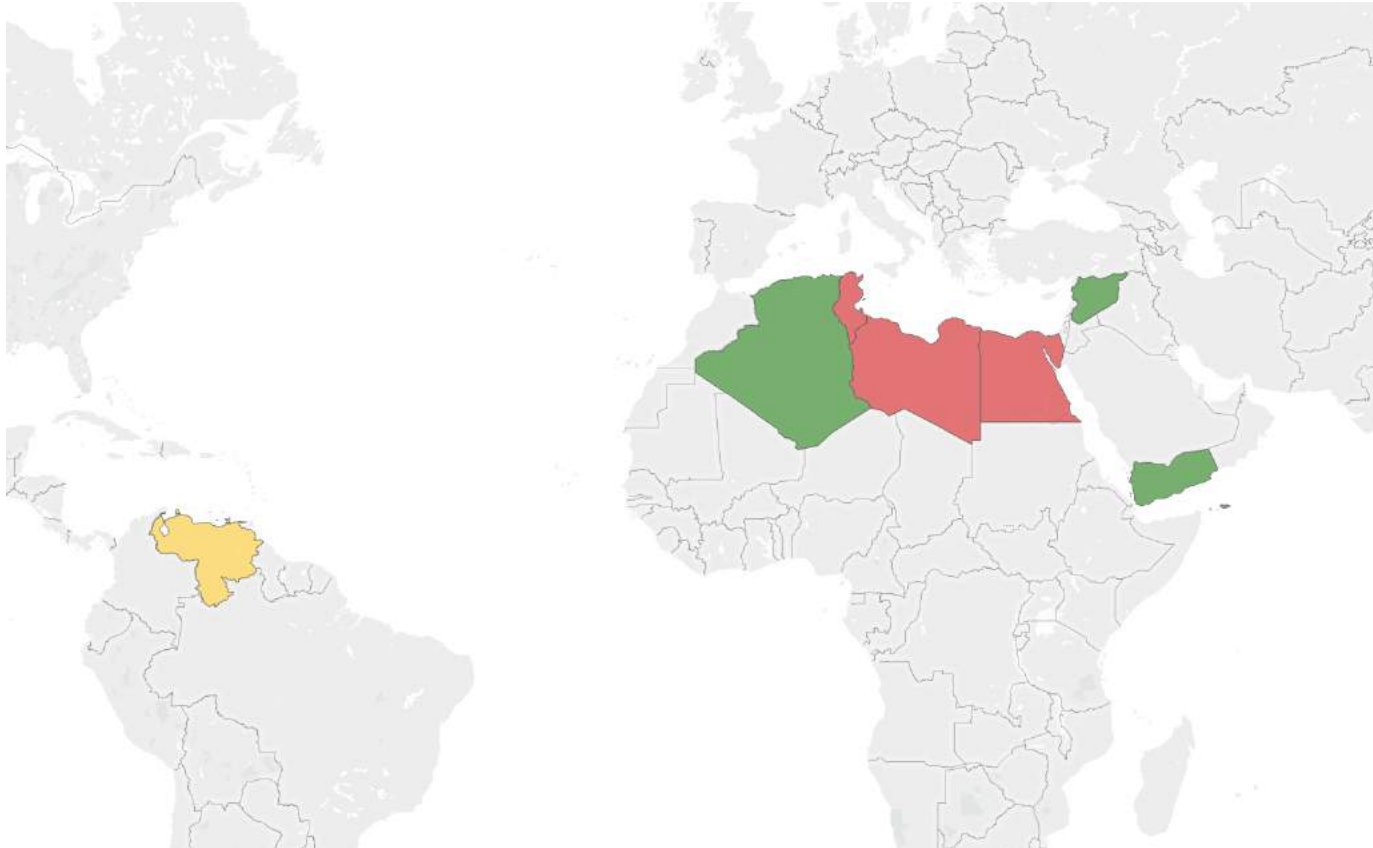
Silvan Saxer

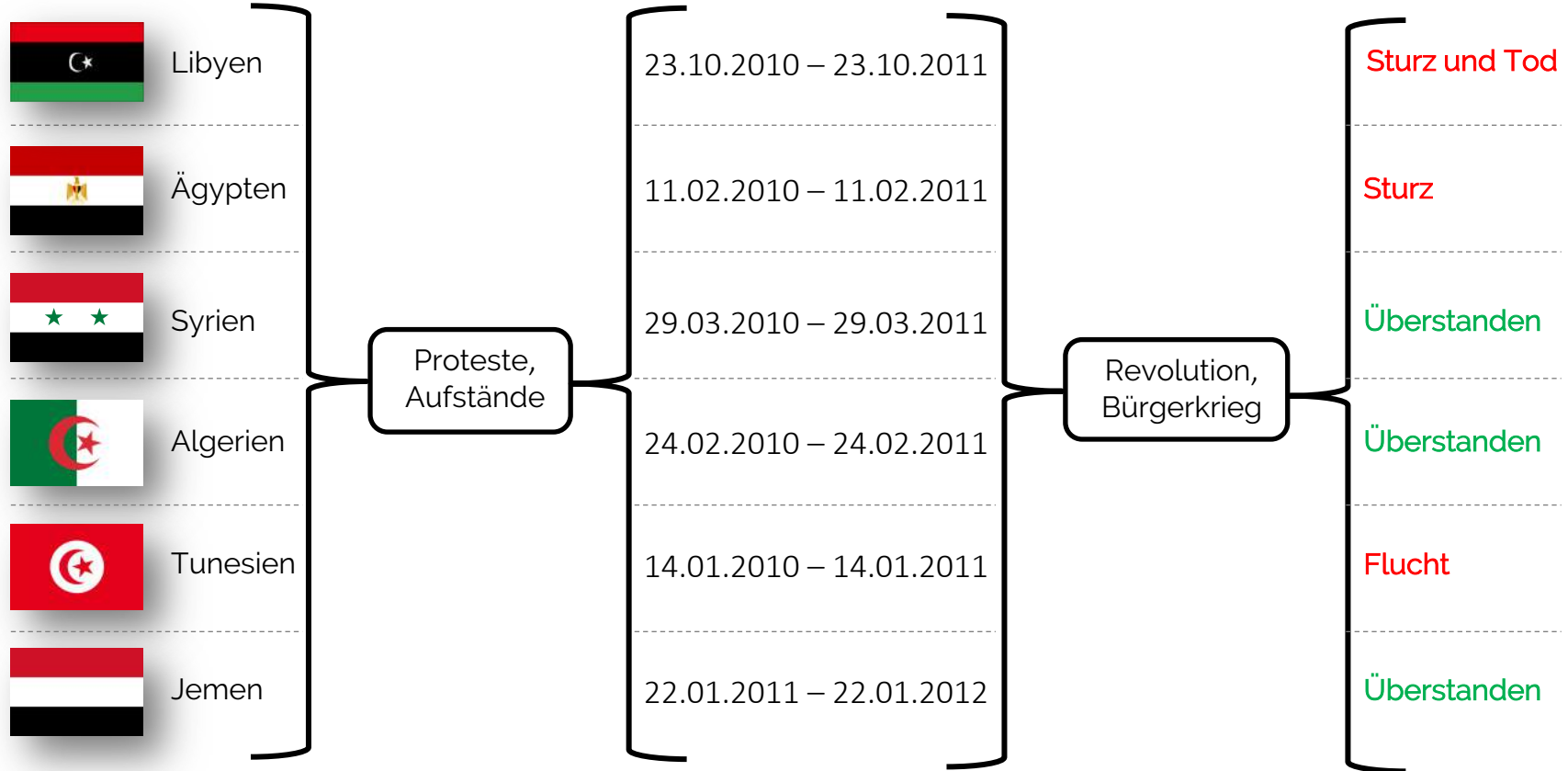
Makroökonomie



Sandro Rüttimann

Informatik







Ölpreis

Preisentwicklung des
Rohöls



New York Times

Anzahl Artikel mit
einem
länderspezifischen Titel



BIP

Bruttoinlandprodukt der
einzelnen Länder



Al Jazeera

Anzahl Tweets, welche
das Land erwähnen



Wiki-Log

Anzahl Änderungen von
Wikipedia-Einträgen



Korruptions-Index

Messung des Grades
der Korruption



Jugendarbeitslosigkeit

Jugendarbeitslosenkeitsrate
(15-24 Jahre alt)



Inflation

Länderspezifische
Entwicklung der
Inflationsrate



Ölpreis

Preisentwicklung des
Rohöls



New York Times

Anzahl Artikel mit
einem
länderspezifischen Titel



BIP

Bruttoinlandprodukt der
einzelnen Länder



Al Jazeera

Anzahl Tweets welche
das Land erwähnen



Wiki-Log

Anzahl Änderungen von
Wikipedia-Einträgen



Korruptions-Index

Messung des Grades
der Korruption



Jugendarbeitslosigkeit

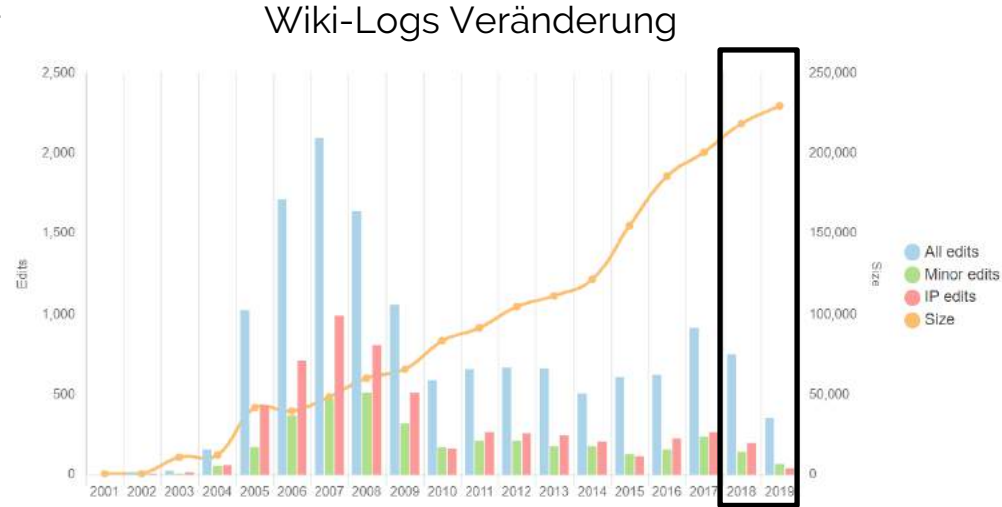
Jugendarbeitslosenkeitsrate
(15-24 Jahre alt)

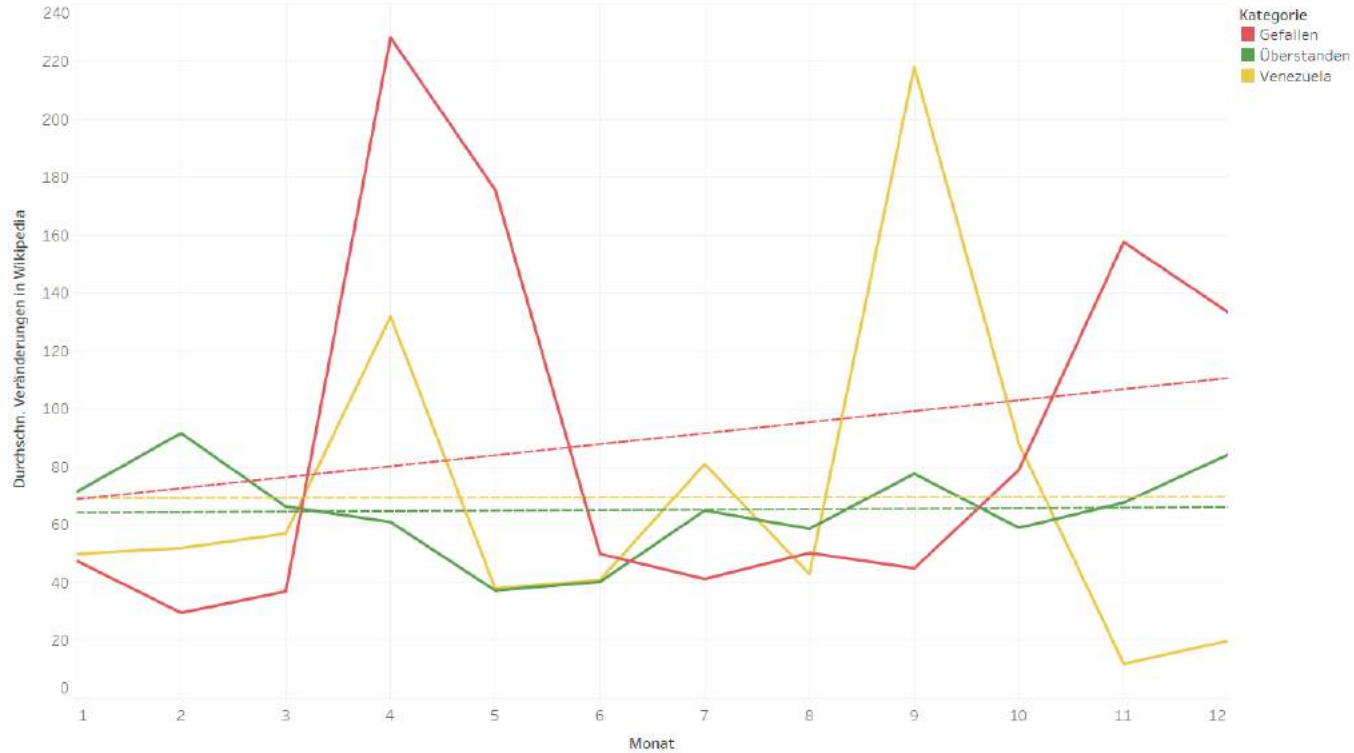


Inflation

Länderspezifische
Entwicklung der
Inflationsrate

- Veränderungen im Wikipedia Artikel der einzelnen Länder in der Krisenperiode
- Wikipedia hat bei den Veränderungen des Artikels einen externen Partner: xtools
- Dieses Tool lässt nur 6 Aufrufe in 5 Minuten zu und sperrt die API für Scrapy
- Manuelle Berechnung





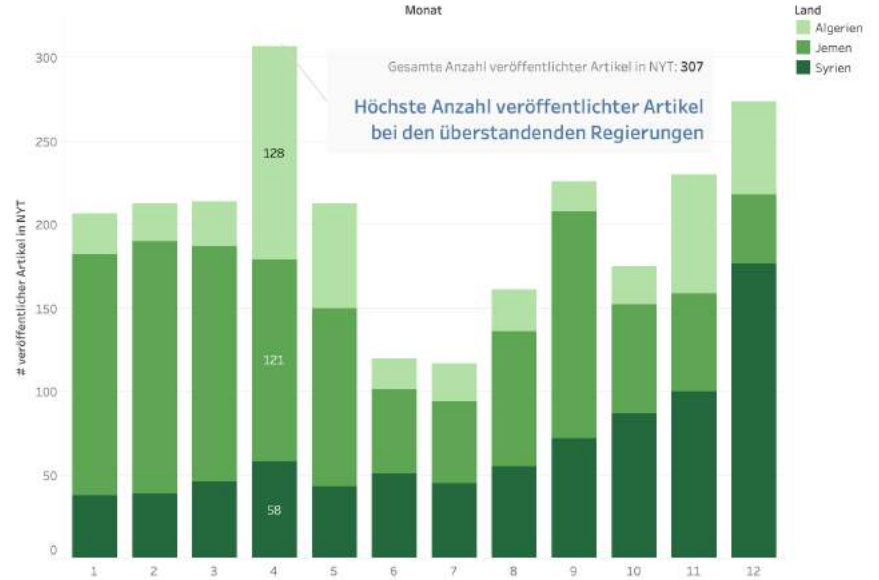
- NYT Archiv API benutzt
- Anzahl veröffentlichter NYT Artikeln mit Erwähnung der jeweiligen Länder (e.g 'Venezuela', 'Libya', 'Egypt') abgerufen
- **9'743 Artikel**
- Herausforderungen:
 - Eingrenzung der Requests auf bestimmte Zeitperiode → durch Schleife und Anpassung der URL gelöst
 - Nicht mehr als 10 Requests pro Minute → durch time function sleep() gelöst

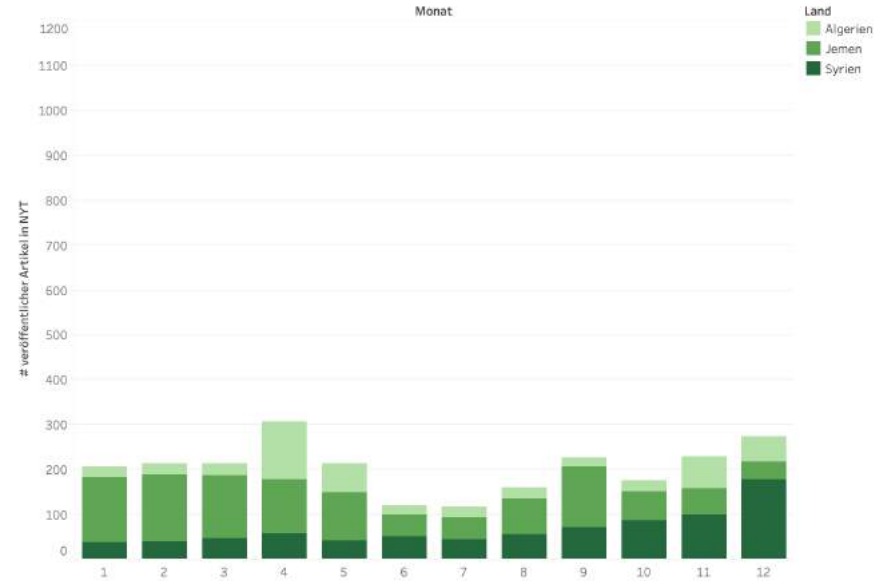
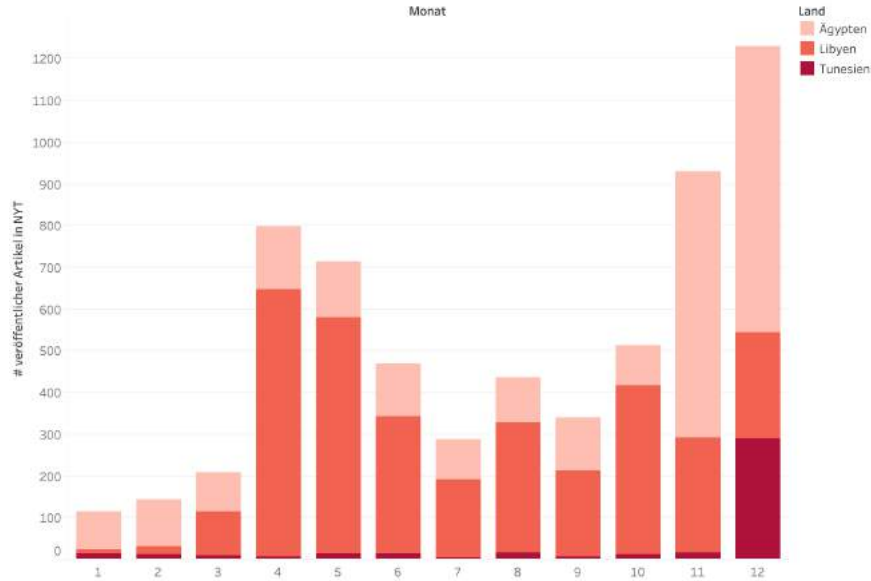
```
app_id = 'ebd056b2-12ba-4233-9673-b35102f9a4dc'  
api_key = 'yLQx1LlB0u5YAqjus1vYhBDa6ILAX0bk'  
keyword = 'venezuela'
```

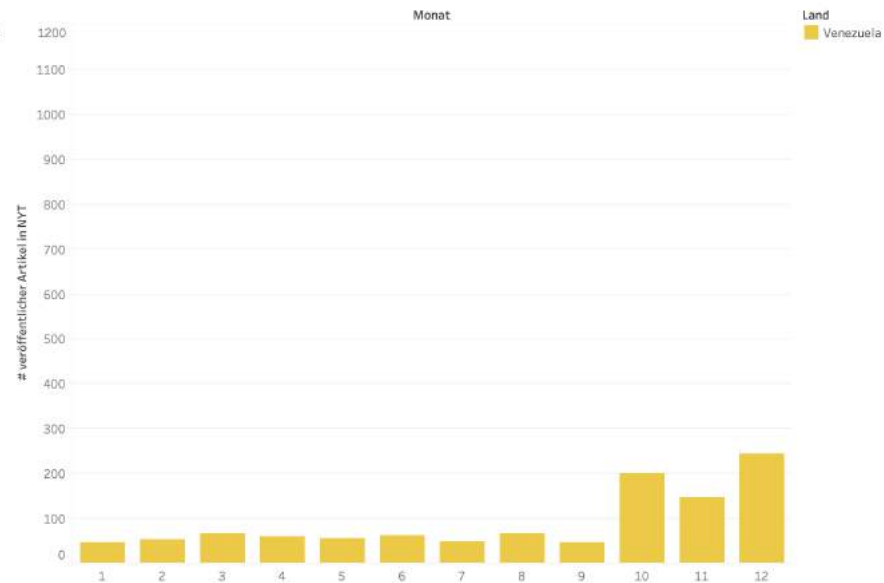
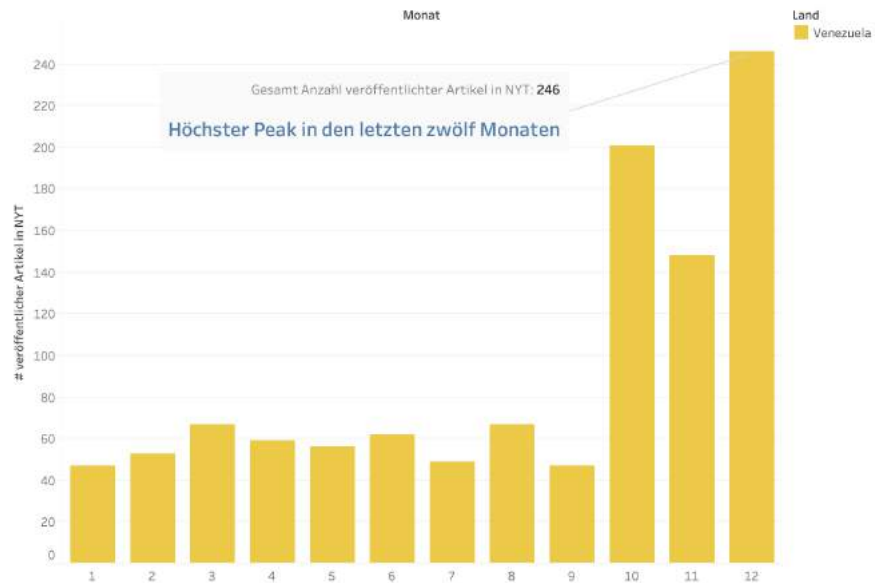
```
begin_date=20180916&end_date=20181016
```

```
if (p-1)%10==0:  
    time.sleep(60)
```

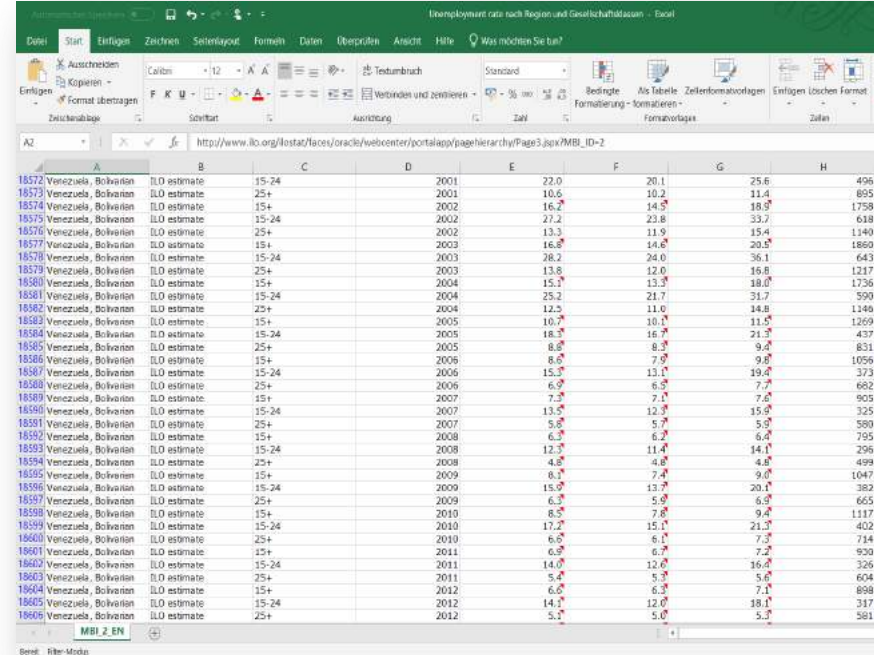






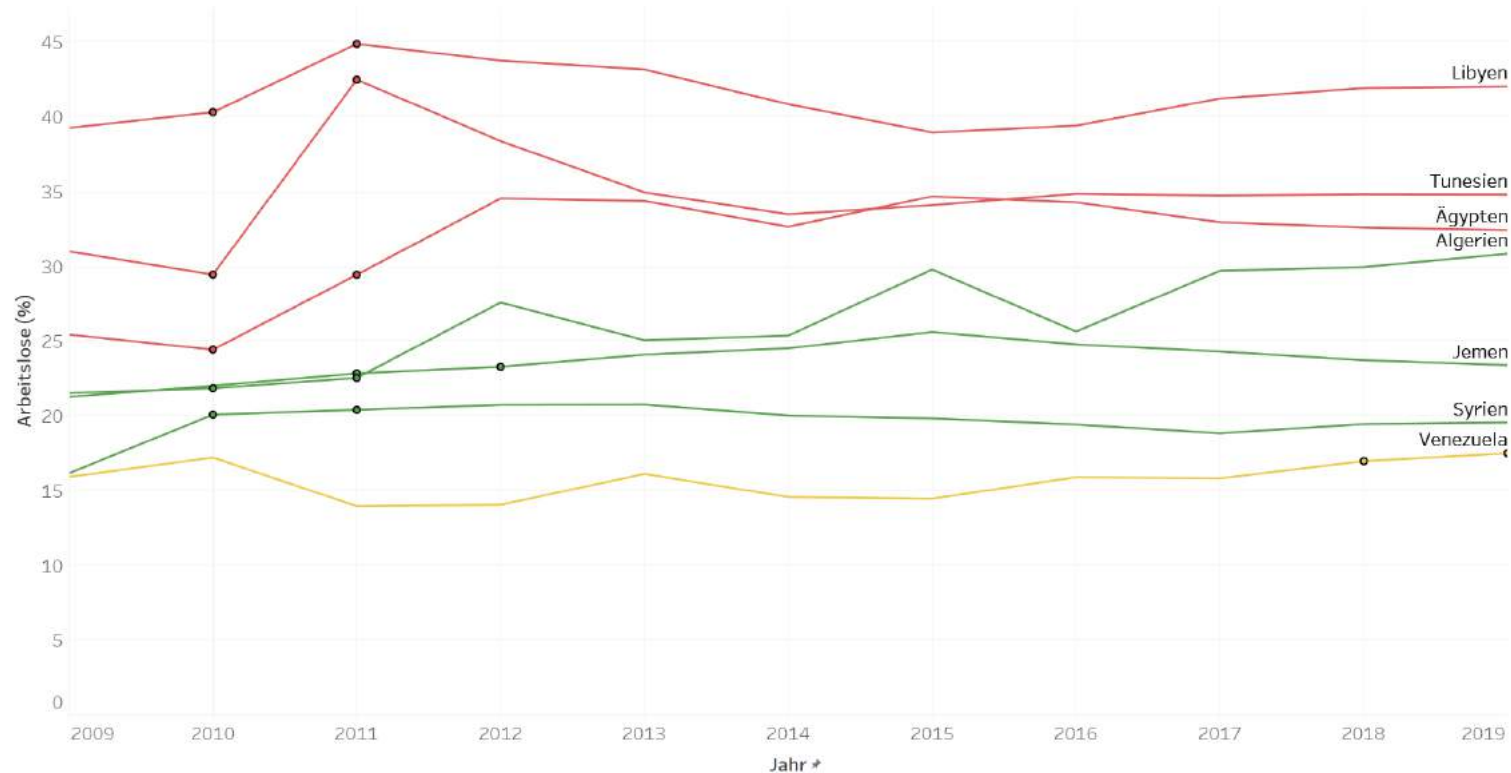


- Datenset von ILOSTAT – International Labour Organization (18'981 Datensätze)
- Jugendarbeitslosigkeit der untersuchten Länder (15 bis 24 Jahre)
- Prozentuale Veränderung manuell berechnet



	A	B	C	D	E	F	G	H
18572	Venezuela, Bolivien	ILO estimate	15-24	2001	22.0	20.1	25.6	496
18573	Venezuela, Bolivien	ILO estimate	25+	2001	10.6	10.2	11.4	895
18574	Venezuela, Bolivien	ILO estimate	15+	2002	16.2	14.5	18.9	1758
18575	Venezuela, Bolivien	ILO estimate	15-24	2002	27.2	23.8	33.7	638
18576	Venezuela, Bolivien	ILO estimate	25+	2002	13.3	11.9	15.4	1140
18577	Venezuela, Bolivien	ILO estimate	15+	2003	16.8	14.6	20.5	1860
18578	Venezuela, Bolivien	ILO estimate	15-24	2003	28.2	24.0	36.1	643
18579	Venezuela, Bolivien	ILO estimate	25+	2003	13.8	12.0	16.8	1217
18580	Venezuela, Bolivien	ILO estimate	15+	2004	15.1	13.3	18.0	1736
18581	Venezuela, Bolivien	ILO estimate	15-24	2004	25.2	21.7	31.7	590
18582	Venezuela, Bolivien	ILO estimate	25+	2004	12.5	11.0	14.8	1146
18583	Venezuela, Bolivien	ILO estimate	15+	2005	10.7	10.1	11.5	1269
18584	Venezuela, Bolivien	ILO estimate	15-24	2005	18.3	16.7	21.3	437
18585	Venezuela, Bolivien	ILO estimate	25+	2005	8.8	8.3	9.4	831
18586	Venezuela, Bolivien	ILO estimate	15+	2006	8.5	7.9	9.8	1056
18587	Venezuela, Bolivien	ILO estimate	15-24	2006	15.3	13.1	19.4	373
18588	Venezuela, Bolivien	ILO estimate	25+	2006	6.9	6.5	7.7	682
18589	Venezuela, Bolivien	ILO estimate	15+	2007	7.3	7.1	7.6	905
18590	Venezuela, Bolivien	ILO estimate	15-24	2007	13.5	12.3	15.9	325
18591	Venezuela, Bolivien	ILO estimate	25+	2007	5.9	5.7	5.8	580
18592	Venezuela, Bolivien	ILO estimate	15+	2008	6.3	6.2	6.4	795
18593	Venezuela, Bolivien	ILO estimate	15-24	2008	12.3	11.4	14.1	296
18594	Venezuela, Bolivien	ILO estimate	25+	2008	4.8	4.6	4.8	499
18595	Venezuela, Bolivien	ILO estimate	15+	2009	6.1	7.4	9.0	1047
18596	Venezuela, Bolivien	ILO estimate	15-24	2009	13.9	13.7	20.1	382
18597	Venezuela, Bolivien	ILO estimate	25+	2009	6.3	5.0	6.9	665
18598	Venezuela, Bolivien	ILO estimate	15+	2010	8.3	7.8	9.4	1117
18599	Venezuela, Bolivien	ILO estimate	15-24	2010	17.2	15.1	21.3	402
18600	Venezuela, Bolivien	ILO estimate	25+	2010	6.5	6.1	7.3	714
18601	Venezuela, Bolivien	ILO estimate	15+	2011	6.8	6.7	7.2	930
18602	Venezuela, Bolivien	ILO estimate	15-24	2011	14.0	12.0	16.4	326
18603	Venezuela, Bolivien	ILO estimate	25+	2011	5.4	5.3	5.6	604
18604	Venezuela, Bolivien	ILO estimate	15+	2012	6.0	6.3	7.1	898
18605	Venezuela, Bolivien	ILO estimate	15-24	2012	14.1	12.0	18.1	317
18606	Venezuela, Bolivien	ILO estimate	25+	2012	5.1	5.0	5.3	581





- Analyse aller geposteten Tweets von «Al Jazeera English»
- Twitter Scraper
- Herausforderungen
 - Rechenleistung (>4h)
 - Dateigrösse
- Code Funktion:
 - Öffnen von Twitter mittels Webdriver & Selenium
 - Abfrage jedes einzelnen Tages (sequentiell)
 - Datenablage in .json-Datei

Datensetgrösse: 3.12 Mio. Zeilen

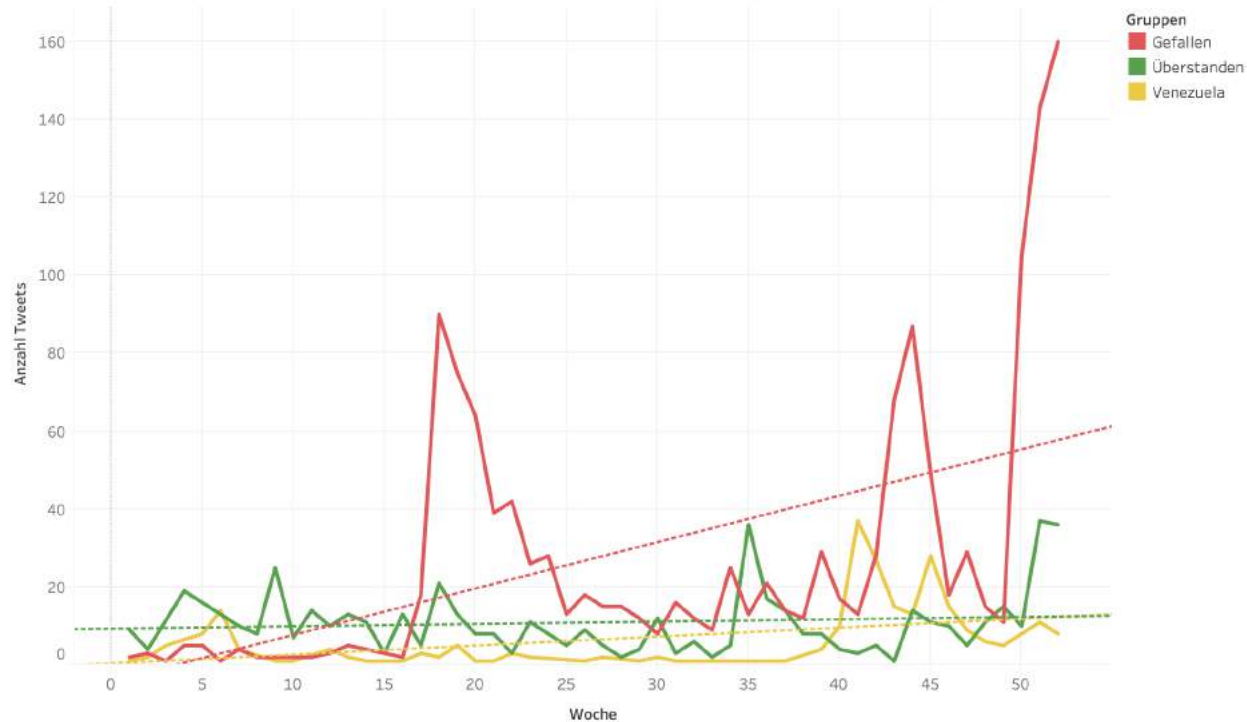
```
3122044  "profile_image_url": "http://pbs.twimg.com/profile_images/875638617606987776/Y8OKib96_normal.jpg
3122045  "profile_image_url_https": "https://pbs.twimg.com/profile_images/875638617606987776/Y8OKib96_nor
3122046  "profile_banner_url": "https://pbs.twimg.com/profile_banners/4970411/1554222591",
3122047  "profile_link_color": "23408C",
3122048  "profile_sidebar_border_color": "FFFFFF",
3122049  "profile_sidebar_fill_color": "000000",
3122050  "profile_text_color": "000000",
3122051  "profile_use_background_image": false,
3122052  "has_extended_profile": false,
3122053  "default_profile": false,
3122054  "default_profile_image": false,
3122055  "following": false,
3122056  "follow_request_sent": false,
3122057  "notifications": false,
3122058  "translator_type": "none"
3122059  },
3122060  "geo": null,
3122061  "coordinates": null,
3122062  "place": null,
3122063  "contributors": null,
3122064  "is_quote_status": false,
3122065  "retweet_count": 6,
3122066  "favorite_count": 0,
3122067  "favorited": false,
3122068  "retweeted": false,
3122069  "lang": "en"
3122070  }
3122071  }
```

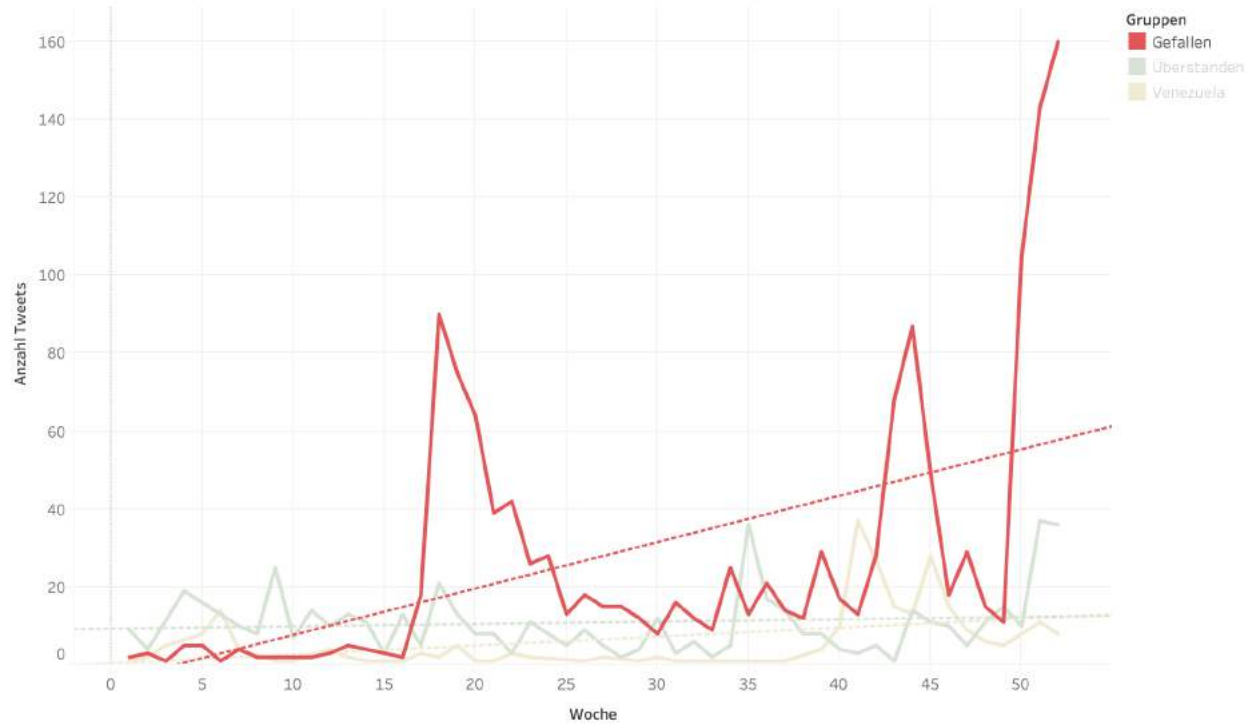


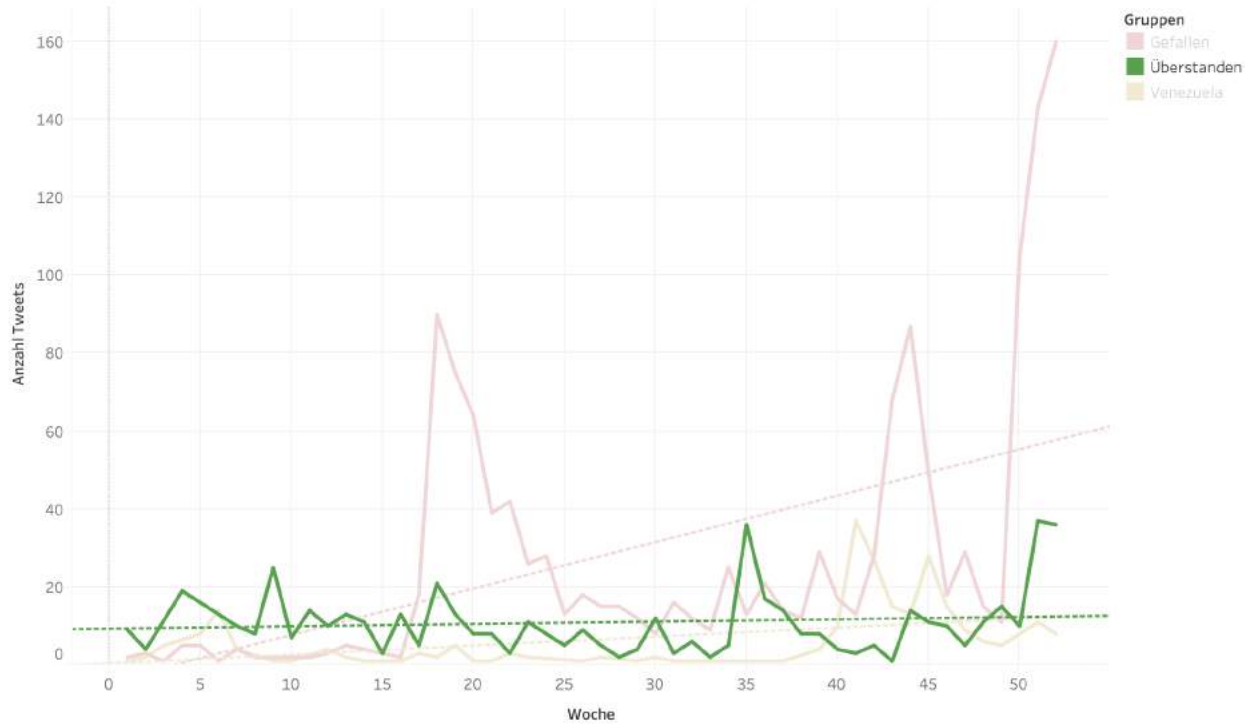
- Analyse der Daten aus .json-Datei
- Code-Funktion
 - Alles Kleinbuchstaben
 - Suche nach Land (hier «yemen»)
 - Wenn Land gefunden, dann «created at» herauslesen

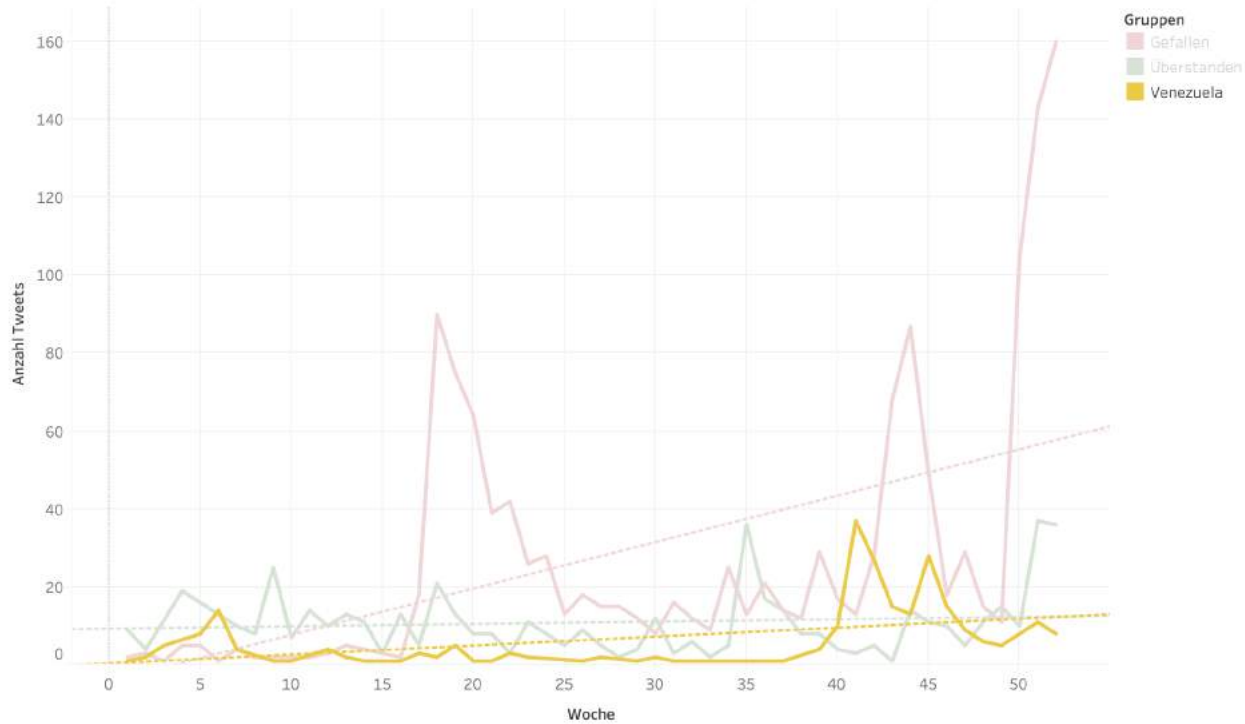
```
1 import json
2
3 f = open('ajenglish_Arabien.json')
4 data = json.load(f)
5 l = []
6 for d in data:
7     if 'yemen' in d['text'].lower():
8         l.append(d)
9
10 dates = []
11 for d in l:
12     dates.append(d['created_at'])
13
14 with open('dates.txt', 'w') as f:
15     for d in dates:
16         f.write('%s\n' %d)
```

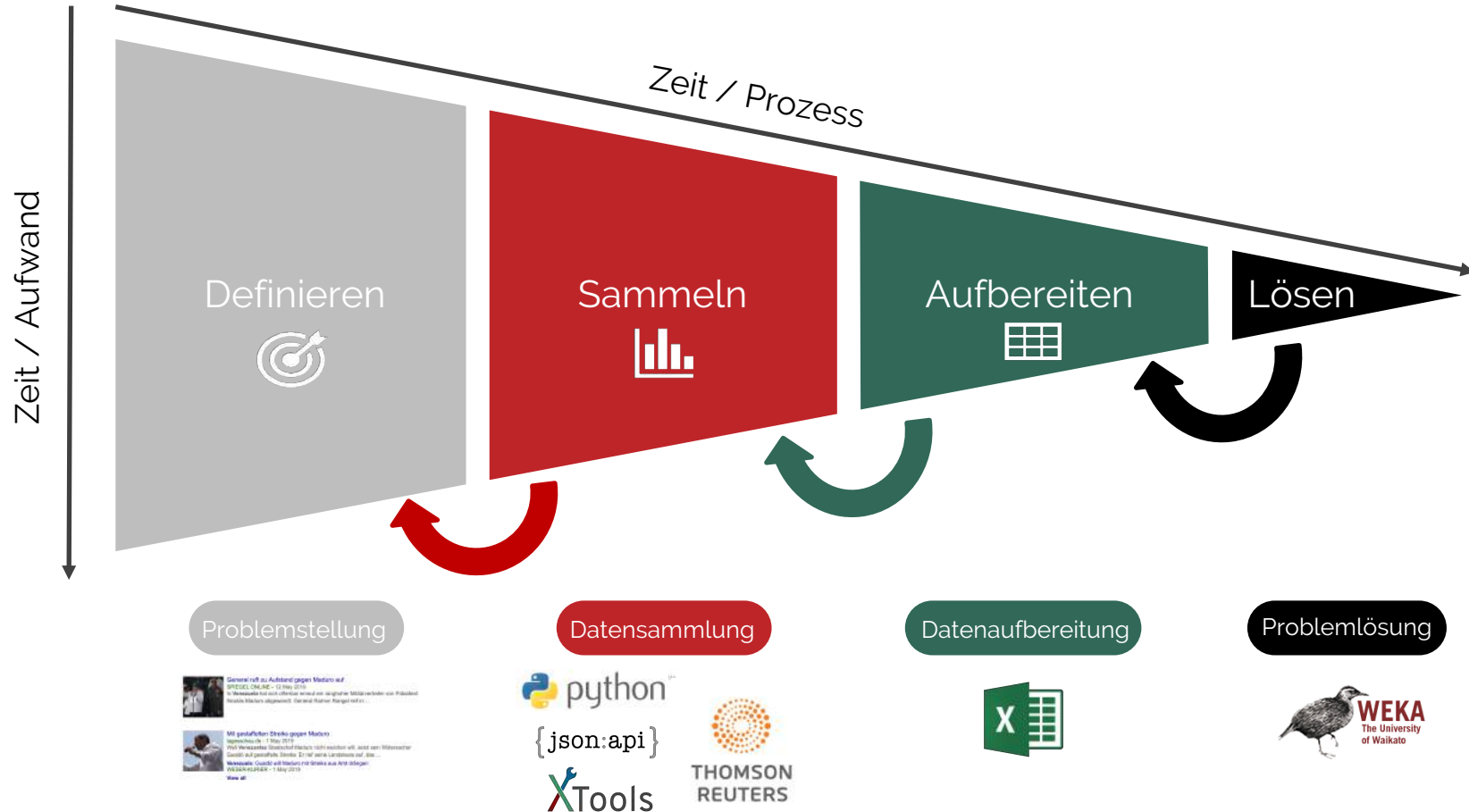






































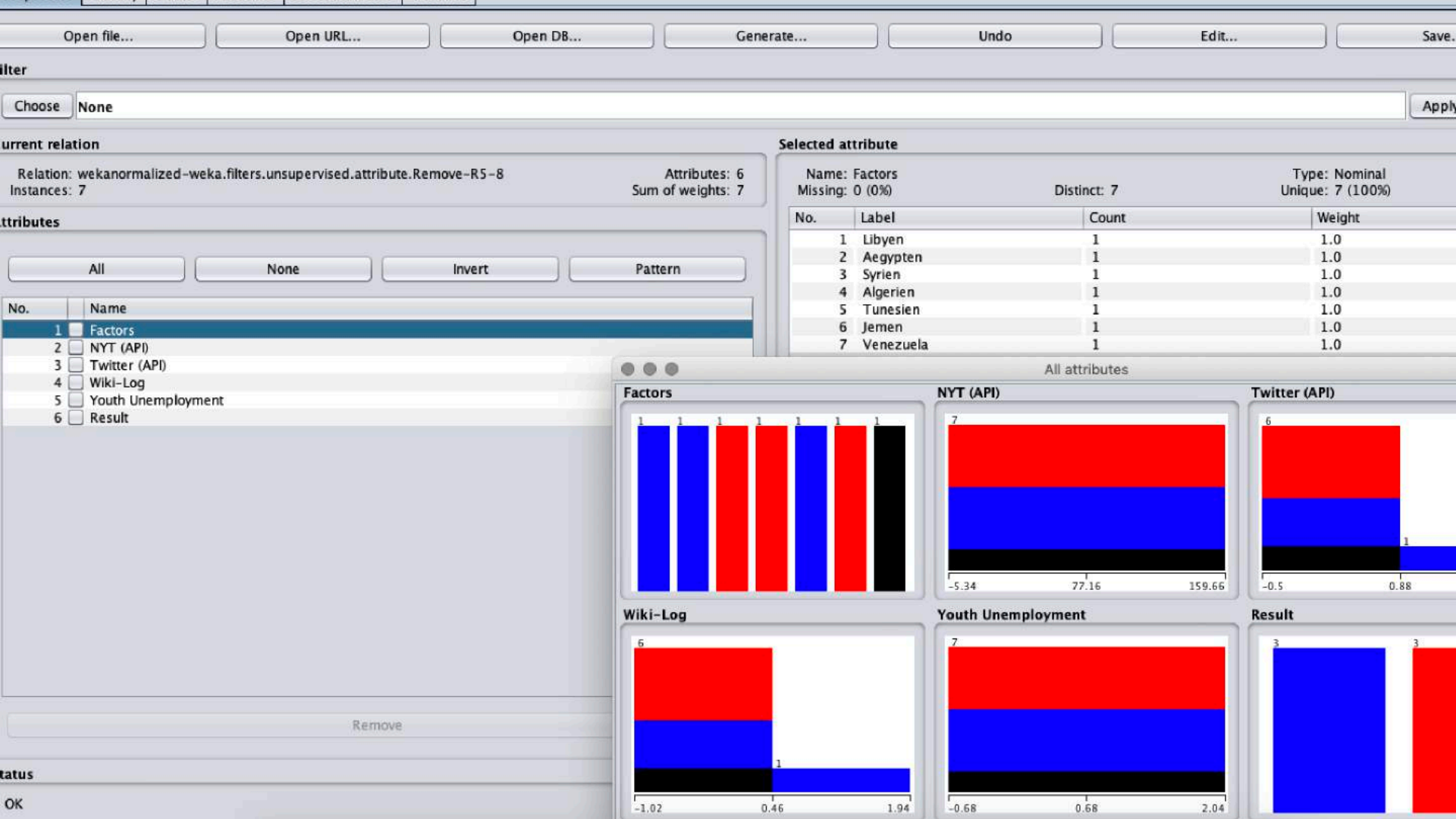


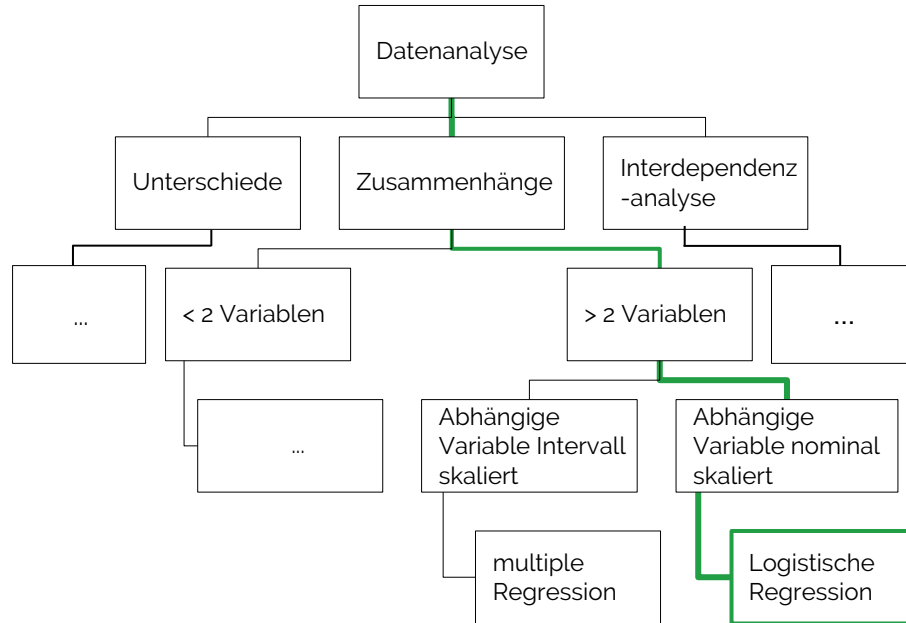




Prozentuale normalisierte Veränderung der Faktoren										
Libyen		159.66459	2.255357362	1.937493317	-1.93930092	-0.904917248	-0.037986859	-0.377620572	-0.05730276	Ja
Ägypten		48.36352935	-0.465078524	-0.804642033	0.502937456	0.737970308	-0.037986859	-0.378095323	0.522091812	Ja
Syrien		28.87422401	-0.464144694	-0.211747903	-0.057902301	1.233786511	-1.101618906	-0.378151507	-0.681266145	Nein
Algerien		13.47888252	-0.220716554	0.274095897	1.147306539	0.452620709	1.557461212	-0.378011048	-0.579394572	Nein
Tunesien		122.7361447	-0.498466567	0.372911586	0.351391394	0.393054689	-1.101618906	-0.378151507	2.037431463	Ja
Jemen		-5.335523644	-0.29217616	-0.549368172	0.518847804	-0.27578662	-0.303894871	-0.378109369	-0.655798252	Nein
Venezuela		32.54160065	-0.314774866	-1.018742691	-0.523279972	-1.636728349	1.025645188	-0.37803914	-0.585761546	?

Prozentuale normalisierte Veränderung der Faktoren										
Libyen		159.66459	2.255357362	1.937493317	-1.93930092	-0.904917248	-0.037986859	-0.377620572	-0.05730276	Ja
Ägypten		48.36352935	-0.465078524	-0.804642033	0.502937456	0.737970308	-0.037986859	-0.378095323	0.522091812	Ja
Syrien		28.87422401	-0.464144694	-0.211747903	-0.057902301	1.233786511	-1.101618906	-0.378151507	-0.681266145	Nein
Algerien		13.47888252	-0.220716554	0.274095897	1.147306539	0.452620709	1.557461212	-0.378011048	-0.579394572	Nein
Tunesien		122.7361447	-0.498466567	0.372911586	0.351391394	0.393054689	-1.101618906	-0.378151507	2.037431463	Ja
Jemen		-5.335523644	-0.29217616	-0.549368172	0.518847804	-0.27578662	-0.303894871	-0.378109369	-0.655798252	Nein
Venezuela		32.54160065	-0.314774866	-1.018742691	-0.523279972	-1.636728349	1.025645188	-0.37803914	-0.585761546	?



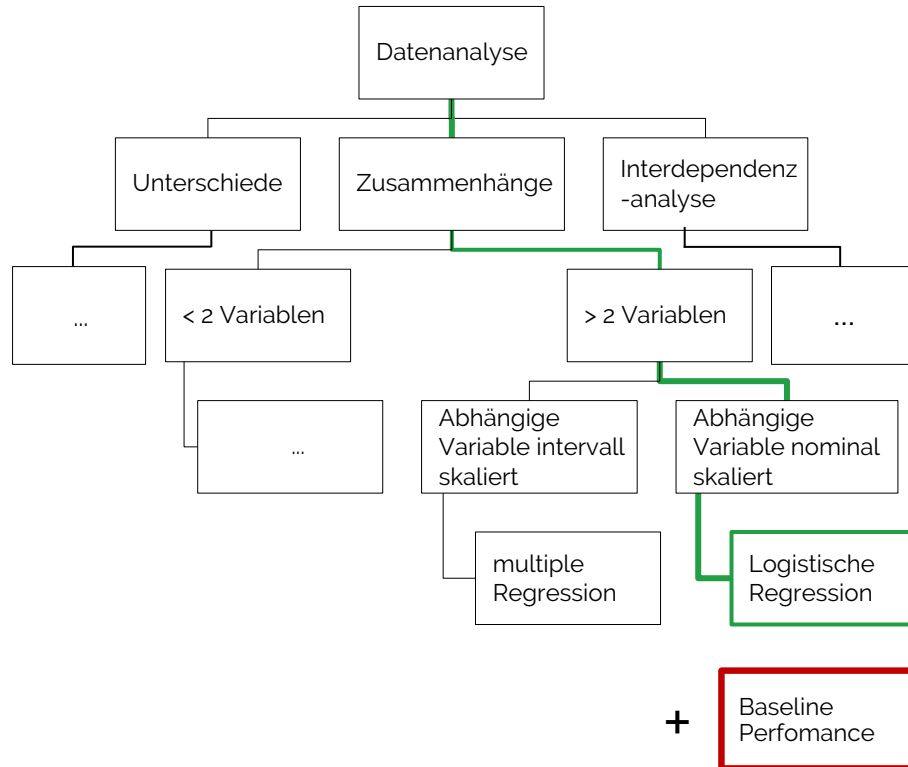


Logistische Regression:

“Haben die unabhängigen Variablen einen Einfluss auf die Wahrscheinlichkeit, dass die abhängige Variable den Wert 1 annimmt? Wie stark ist deren Einfluss?”

Voraussetzungen:

- Abhängige Variable ist binär (0-1) codiert
- Unabhängige Variablen metrisch codiert
- Die unabhängigen Variablen sind untereinander nicht hoch korreliert



Logistische Regression:

"Haben die unabhängigen Variablen einen Einfluss auf die Wahrscheinlichkeit, dass die abhängige Variable den Wert 1 annimmt? Wie stark ist deren Einfluss?"

Voraussetzungen:

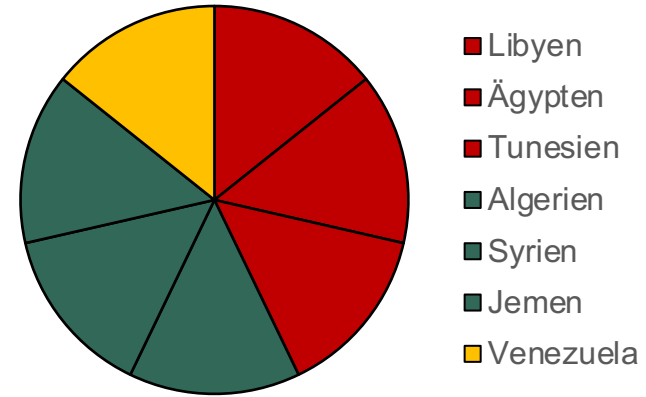
- Abhängige Variable ist binär (0-1) codiert
- Unabhängige Variablen metrisch codiert
- Die unabhängigen Variablen sind untereinander nicht hoch korreliert

Baseline Performance für Regressionsprobleme:

- Es ist nicht bekannt, welcher Algorithmus am besten performt
- Stellt Referenzpunkt dar

7-fold Cross Validation

- Systemische Methode «Percentage Split»
- Teilen des Datensets in 7 Teile
- Pro Durchlauf: 6 Teile Training & 1 Teil Testing
- 7 durchschnittliche Resultate
- 8. Durchgang ist der finale Algorithmen-Durchgang, um das Resultat zu finden





=== Run information ===

Scheme: weka.classifiers.rules.ZeroR

Instances: 7

Factors:

NYT (API)

Twitter (API)

Wiki-Log

Youth Unemployment

=== Predictions on test data ===

inst#	actual	predicted	error
-------	--------	-----------	-------

1	1:Yes	1:Yes	
---	-------	-------	--

2	1:Yes	1:Yes	
---	-------	-------	--

3	2:No	1:Yes	+
---	------	-------	---

4	2:No	1:Yes	+
---	------	-------	---

5	1:Yes	1:Yes	
---	-------	-------	--

6	2:No	1:Yes	+
---	------	-------	---

7	1:?	1:Yes	
---	-----	-------	--

Correctly Classified Instances	3	50	%
--------------------------------	---	----	---

Incorrectly Classified Instances	3	50	%
----------------------------------	---	----	---

Kappa statistic	0		
-----------------	---	--	--

Mean absolute error	0.5		
---------------------	-----	--	--

Classifier

Choose

Logistic -R 1.0E-8 -M -1 -num-decimal-places 4

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation

☐ Percentage split

Set...

Folds 7

% 66

More options...

(Nom) Result

Start

Stop

Result list (right-click for options)

12:12:01 - rules.ZeroR

12:12:10 - rules.ZeroR

12:19:01 - functions.Logistic

12:19:07 - functions.Logistic

Classifier output

Time taken to build model: 0 seconds

=== Predictions on test data ===

inst#	actual	predicted	error	prediction
1	2:No	2:No		1
1	2:No	2:No		1
1	2:No	2:No		1
1	1:Yes	2:No	+	0.999
1	1:Yes	2:No	+	1
1	1:Yes	1:Yes		1
1	1:?	2:No		0.993

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	4	66.6667 %
Incorrectly Classified Instances	2	33.3333 %
Kappa statistic	0.3333	
Mean absolute error	0.3331	
Root mean squared error	0.577	
Relative absolute error	58.2991 %	
Root relative squared error	100.9767 %	
Total Number of Instances	6	
Ignored Class Unknown Instances	1	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.333	0.000	1.000	0.333	0.500	0.447	0.583	0.698	Yes
	1.000	0.667	0.600	1.000	0.750	0.447	0.750	0.639	No
Weighted Avg.	0.667	0.333	0.800	0.667	0.625	0.447	0.667	0.669	

=== Confusion Matrix ===

a b <-- classified as



=== Run information ===

Scheme:

weka.classifiers.functions.Logistic -R
1.0E-8 -M -1 -num-decimal-places 4

Instances: 7

Attributes: 6

Factors: NYT (API)

Twitter (API)

Wiki-Log

Youth Unemployment

Test mode: 7-fold cross-validation

=== Predictions on test data ===

inst#	actual	predicted	error
1	2:No	2:No	
1	2:No	2:No	
1	2:No	2:No	
1	1:Yes	2:No	+
1	1:Yes	2:No	+
1	1:Yes	1:Yes	
1	1:?	2:No	

Correctly Classified Instances	4	66.6667 %
Incorrectly Classified Instances	2	33.3333 %
Kappa statistic	0.3333	
Mean absolute error	0.3331	



Logistische Regression ist das richtige Vorhersagemodell



Genauigkeit der Vorhersage liegt bei **66.67%**



Präsident Maduro wird **nicht** fallen (in absehbarer Zeit)



Flucht muss ergriffen werden



Geringe Anzahl betrachteter Länder



Herausforderung der Festlegung des länderspezifischen Betrachtungszeitraums



Technologische Entwicklung zwischen den einzelnen Betrachtungszeiträumen



Einschränkung auf englischsprachige Quellen (Twitter & Co.)

Gracias
Preguntas (Fragen)?

Backup



=== Run information ===

Scheme:

weka.classifiers.functions.Logistic -R
1.0E-8 - M -1 -num-decimal-places 4

Instances: 7

Factors: NYT (API)

Twitter (API)

Wiki-Log

GDP

Oil Prices

Corruption

Inflation

Youth Unemployment

Result

Test mode: 7-fold cross-validation

=== Predictions on test data ===

inst#	actual	predicted	error
1	2:No	1:Yes	+
1	2:No	2:No	
1	2:No	1:Yes	+
1	1:Yes	1:Yes	
1	1:Yes	2:No	+
1	1:Yes	1:Yes	
1	1:?	1:Yes	

Correctly Classified Instances 3 50 %

Incorrectly Classified Instances 3 50 %

Kappa statistic 0

Mean absolute error 0.5003



Anhang

ARD (2011). Assad entlässt syrische Regierung. Abgerufen von: <https://web.archive.org/web/20110401103829/http://www.tagesschau.de:80/ausland/syrien180.html>

Asseburg, M. (2011, Juli). Der arabische Frühling. Herausforderung und Chance für die deutsche und europäische Politik. *Stiftung Wissenschaft und Politik. Deutsches Institut für internationale Politik und Sicherheit*. Abgerufen von https://www.swp-berlin.org/fileadmin/contents/products/studien/2011_S17_ass_ks.pdf

Asseburg, M. (2011, 11. Oktober). Die historische Zäsur des arabischen Frühlings. *Bundeszentrale für politische Bildung* [bpb]. Abgerufen von <http://www.bpb.de/internationales/afrika/arabischer-fruehling/52389/einfuehrung?p=all>

Bpb27. (2017, 17. Januar). *Grab all a user's tweets (and get past 3200 limit)*. Abgerufen von https://github.com/bpb27/twitter_scraping

Brownlee, Jason (2018). *How To Estimate A Baseline Performance For Your Machine Learning Models in Weka*. Machine Learning Mastery. Abgerufen von: <https://machinelearningmastery.com/estimate-baseline-performance-machine-learning-models-weka/>

Burgmer, C. (2012, 06. April). Wenig gewonnen, viel erreicht: Die politische Zukunft des «arabischen Frühlings». Deutschlandfunk. Abgerufen von https://www.deutschlandfunk.de/wenig-gewonnen-viel-erreicht-die-politische-zukunft-des.1184.de.html?dram:article_id=185495

Futurelearn (2019). Cross Validation. Abgerufen von <https://www.futurelearn.com/courses/data-mining-with-weka/0/steps/25384>

Frankfurter Allgemeine (2011a). Ausnahmezustand offiziell aufgehoben. Abgerufen von: <https://www.faz.net/aktuell/politik/algerien-ausnahmezustand-offiziell-aufgehoben-1596133.html>

Frankfurter Allgemeine (2011b). Internationaler Haftbefehl gegen Ben Ali. Abgerufen von: <https://www.faz.net/aktuell/politik/ausland/tunesien-internationaler-haftbefehl-gegen-ben-ali-1578818.html>

Helm, M. & Witzenberger, B. (2017, 31. Juli). Hungern im ölreichsten Land der Welt. Süddeutsche Zeitung. <https://www.sueddeutsche.de/politik/venezuela-hungern-im-oelreichsten-land-der-welt-1.3611061>

International Labour Organization (2019). Youth NEET rate. Abgerufen von: http://www.ilo.org/ilostat/faces/oracle/webcenter/portalapp/pagehierarchy/Page3.jspx?MBI_ID=20

Kim, S. K., Park, M. J., & Rho, J. J. (2015). *Effect of the Government's Use of Social Media on the Reliability of the Government: Focus on Twitter*. *Public Management Review*, 17(3), 328–355. <https://doi.org/10.1080/14719037.2013.822530>

Korruption in der arabischen Welt. (2010, 22. Januar). Deutsche Welle [DW]. Abgerufen von <https://www.dw.com/de/korruption-in-der-arabischen-welt/a-5157865>

NTV (2011). Libyen erklärt sich für befreit. Abgerufen von: <https://www.n-tv.de/politik/Libyen-erklaert-sich-fuer-befreit-article4597551.html>

Pfaffenberger, F. (2016). *Twitter als Basis für wissenschaftliche Studien. Eine Bewertung gängiger Erhebungs- und Analysemethoden der Twitter-Forschung*. Wiesbaden: Springer.

Pfaffmann, C., & Meckel, M. (2011). Social Media and Democratic Utopia. Is Cyberspace Transforming the Game of Power Politics? Caroline Pfaffmann. Retrieved from <https://search.ebscohost.com/login.aspx?direct=true&db=cat00327a&AN=stgal.000545968&lang=de&site=eds-live>

Pond, P., & Lewis, J. (2019). Riots and Twitter: connective politics, social media and framing discourses in the digital public sphere. *Information, Communication & Society*, 22(2), 213–231. <https://doi.org/10.1080/1369118X.2017.1366539>

PYE4 (2019). *Phyton for Everybody*. Abgerufen von: <https://www.py4e.com/lessons>

Statista (2019). *Venezuela: Gross domestic product (GDP) in current prices from 2013 to 2023 (in billion U.S. dollars)*. Abgerufen von: <https://www.statista.com/statistics/370937/gross-domestic-product-gdp-in-venezuela/>

Tableau (2019). Tutorial: *Get Started with Tableau Desktop*. Abgerufen von: <https://onlinehelp.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-home.htm>

Tages Anzeiger (2012). Saleh bittet sein Volk um Verzeihung – und geht. Abgerufen von: <https://www.tagesanzeiger.ch/ausland/naher-osten-und-afrika/Saleh-bittet-sein-Volk-um-Verzeihung--und-geht/story/20788112?track>

TAZ (2011). Mubarak verantwortlich für 846 Tote. Abgerufen von: <http://www.taz.de/!5122286/>

The World Bank (2019). *GDP (current US\$) World Bank national accounts data, and OECD National Accounts data files.*. Abgerufen von: <https://data.worldbank.org/indicator/ny.gdp.mktc>

Transparency International (2018). *Corruption Perceptions Index*. Abgerufen von: <https://www.transparency.org/research/cpi/overview>

Universität Zürich (n.d.). *Datenanalyse mit SPSS*. Methodenberatung. Abgerufen von: https://www.methodenberatung.uzh.ch/de/datenanalyse_spss.html

Wikipedia (2019). Venezuela. Abgerufen von: <https://xtools.wmflabs.org/articleinfo/en.wikipedia.org/Venezuela>