

Fallstudie 2



```
# tweepy ist eine API (Application Programming Interface= Programmierschnittstelle) und ermöglicht die Kommunikation zwischen meinem Programm und Twitter
# Mit dem Framework tweepy können Phyton und Twitter kommunizieren und interagieren
```

```
import tweepy
```

```
# Wir haben ein Script geschrieben, um unsere Keys und Tokens zu speichern
key = 'keys.py'
```

```
%%writefile {key}
```

```
consumer_key= "06S9Dqz59HcsTCmWqGt2I2UK0"
consumer_secret= "P9sXknnuA2qVM1jh38asDlQZadnaBX1xh2Bc2ptQDXQhzpGXUL"
access_token= "1457401307304706065-txNKnVEyeTJ0wPkNzMLRcWwNEPx8FX"
access_token_secret= "04Kz7tEGWIcCDtuvUJbwBOT0C3HznQjwAkWxcmWmRoZ2R"
bearer_token = "AAAAAAAAAAAAAAAAAAAAJuHWgEAAAAAZZzGNSewF0bbzQNwrjaFck36sAg%3DodSUJhWL4dV1kfQnu45S5QYyH08Txq0z15Ttc2PjMHuvSP00G8"
```

Writing keys.py

```
from keys import *
```

```
# Diese Keys und Tokens brauchen wir, um unser Authentifizieren zu können.
# Ohne diese Authentifizierung ist es uns nicht möglich die API zu nutzen. Der Client überprüft unsere Keys und Tokens
```

```
import requests
```

```
client = tweepy.Client( bearer_token=bearer_token,
                        consumer_key=consumer_key,
                        consumer_secret=consumer_secret,
                        access_token=access_token,
                        access_token_secret=access_token_secret,
                        return_type = requests.Response,
                        wait_on_rate_limit=True)
```

```
# Define query: Wir spezifizieren die Daten, welche wir herausziehen wollen:
# In meinem Beispiel analysieren wir die letzten 10 Retweets von LeBron James
query = 'from:KingJames -is:retweet'
```

```
# Wir können auch mehrere Felder spezifizieren. In unserem Fall haben wir auch die Author_Id und das Datum mit der Uhrzeit, wann der retweet erstellt worden ist.
# Weitere Felder welche wir nutzen könnten wären z.B. public_metrics, wie retweet_count, reply_count, like_count, quote_count ; Unser_Id, User_name
```

```
tweets = client.search_recent_tweets(query= 'from:KingJames -is:retweet',
                                     tweet_fields=['author_id', 'created_at'],
                                     max_results=10)
```

1.Schritt

Import von tweepy

Request, um zu

authentifizieren, dass wir

Twitter Developer sind

Query erstellen, um Daten

von Twitter zu holen

```
# Save data as dictionary: Das machen wir um Schlüssel-Wert Paare zu speichern
```

```
tweets_dict = tweets.json()
```

```
# Extract "data" value from dictionary
```

```
tweets_data = tweets_dict['data']
```

```
# Transform to pandas Dataframe
```

```
df = pd.json_normalize(tweets_data)
```

```
df
```

	text	id	author_id	created_at
0	Jumping on someone podcast soon. Maybe my own 🤔	1533632259755896832	23083404	2022-06-06T02:09:58.000Z
1	It's so funny to me how many basketball expert...	1533632033972310016	23083404	2022-06-06T02:09:05.000Z
2	We need to come together and help do whatever ...	1533599792701378560	23083404	2022-06-06T00:00:58.000Z
3	Our condolences goes out to the famlly who los...	1532754469728837633	23083404	2022-06-03T16:01:57.000Z
4	YES WE DID!!! https://t.co/OlWm2GEt18	1532225483596738560	23083404	2022-06-02T04:59:57.000Z
5	🤔 "In the face of Impossible odds, people who...	1531837452591042561	23083404	2022-06-01T03:18:03.000Z
6	Congratulations brother!!! 🍌🍌🍌🍌🍌🍌🍌🍌🍌htt...	1531784889279795201	23083404	2022-05-31T23:49:11.000Z
7	Ayyyyyyyyeeee!!!! 🤔🤔🤔🤔🤔 https://t.co/gxJ6zf0dHh	1531260582657138688	23083404	2022-05-30T13:05:47.000Z

```
# Exportieren der Daten als CSV: Dies machen wir damit wir die Daten in Schritt 2 Textmining betreiben können.
```

```
# Textmining = Informationen aus einem Text herausziehen
```

```
df.to_csv("tweets_Fallstudie.csv")
```

2. Schritt

CSV erstellen um mit Daten arbeiten zu können

✓ 2.9s

Unnamed: 0		text	id	author_id	created_at
0	0	Jumping on someone podcast soon. Maybe my own 🤔	1533632259755896832	23083404	2022-06-06T02:09:58.000Z
1	1	It's so funny to me how many basketball expert...	1533632033972310016	23083404	2022-06-06T02:09:05.000Z
2	2	We need to come together and help do whatever ...	1533599792701378560	23083404	2022-06-06T00:00:58.000Z
3	3	Our condolences goes out to the family who los...	1532754469728837633	23083404	2022-06-03T16:01:57.000Z
4	4	YES WE DID!!! https://t.co/OlWm2Gt18	1532225483596738560	23083404	2022-06-02T04:59:57.000Z
5	5	🤔 "In the face of Impossible odds, people who...	1531837452591042561	23083404	2022-06-01T03:18:03.000Z
6	6	Congratulations brother!!! 🍌🍌🍌🍌🍌🍌🍌htt...	1531784889279795201	23083404	2022-05-31T23:49:11.000Z
7	7	Ayyyyyyyye!!!! 🤔🤔🤔 https://t.co/gxj6zf0dHh	1531260582657138688	23083404	2022-05-30T13:05:47.000Z

✓ 0.1s

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8 entries, 0 to 7
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Unnamed: 0   8 non-null      int64
1   text         8 non-null      object
2   id           8 non-null      int64
3   author_id    8 non-null      int64
4   created_at   8 non-null      object
dtypes: int64(3), object(2)
memory usage: 448.0+ bytes
```

✓ 0.7s

df

3. Schritt

Nun können wir mit dem Textmining starten:
Zuerst lesen wir DataFrame ein und schauen uns die Daten an.

**Wir sehen, dass die Daten
Tweets sehr viele Emojis
haben, was uns stört. Diese
möchten wir entfernen.**


```
# wir importieren das natural language toolkit der Stanford University, um Textmining zu betreiben und die Erkenntnisse zu visualisieren.
import nltk
nltk.download('stopwords')
```

✓

0.3s

```
[nltk_data] Downloading package stopwords to
[nltk_data]   /Users/rafaela/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

True

```
# wir importieren das Paket "stopwords", welches wichtig ist, um Wörter entfernen zu können. Das Paket beinhaltet typische Wörter wie "the", "of", "to".
import nltk
from nltk.corpus import stopwords
stopwords = nltk.corpus.stopwords.words("english")

# Wir erstellen eine Liste mit den Wörtern die wir entfernen wollen:
my_stopwords = ['https', 't', 'co', '1ac2oyfv6m', '2chzznfrjx', 'oiwm2get18', 'bpq5qsonfo', '2gww3ff81p', 'gxj6zf0dhh']

stopwords.extend(my_stopwords)
```

✓

0.1s

Unnamed: 0	text	id	author_id	created_at	text_token
0	0 jumping on someone podcast soon. maybe my own 🤔	1533632259755896832	23083404	2022-06-06T02:09:58.000Z	[jumping, on, someone, podcast, soon, maybe, m...
1	1 it's so funny to me how many basketball expert...	1533632033972310016	23083404	2022-06-06T02:09:05.000Z	[it, s, so, funny, to, me, how, many, basketba...
2	2 we need to come together and help do whatever ...	1533599792701378560	23083404	2022-06-06T00:00:58.000Z	[we, need, to, come, together, and, help, do, ...
3	3 our condolences goes out to the family who los...	1532754469728837633	23083404	2022-06-03T16:01:57.000Z	[our, condolences, goes, out, to, the, family,...
4	4 yes we did!!! https://t.co/olwm2get18	1532225483596738560	23083404	2022-06-02T04:59:57.000Z	[yes, we, did, https, t, co, olwm2get18]
5	5 🤔 "In the face of Impossible odds, people who...	1531837452591042561	23083404	2022-06-01T03:18:03.000Z	[in, the, face, of, Impossible, odds, people, ...
6	6 congratulations brother!!! 🍌🍌🍌🍌🙏🔥🇸🇰 htt...	1531784889279795201	23083404	2022-05-31T23:49:11.000Z	[congratulations, brother, https, t, co, 1ac2o...
7	7 ayyyyyeeeeee!!!! 🍌🍌🍌🍌 https://t.co/gxj6zf0dhh	1531260582657138688	23083404	2022-05-30T13:05:47.000Z	[ayyyyyyyyy, https, t, co, gxj6zf0dhh]

```
# Entfernt die Stopwords
df['text_token_stopwords'] = df['text_token'].apply(lambda x: [item for item in x if item not in stopwords])

df
```

✓

0.2s

Unnamed: 0	text	id	author_id	created_at	text_token	text_token_stopwords
0	0 jumping on someone podcast soon. maybe my own 🤔	1533632259755896832	23083404	2022-06-06T02:09:58.000Z	[jumping, on, someone, podcast, soon, maybe, m...	[jumping, someone, podcast, soon, maybe]
1	1 it's so funny to me how many basketball expert...	1533632033972310016	23083404	2022-06-06T02:09:05.000Z	[it, s, so, funny, to, me, how, many, basketba...	[funny, many, basketball, experts, app, everyb...
2	2 we need to come together and help do whatever ...	1533599792701378560	23083404	2022-06-06T00:00:58.000Z	[we, need, to, come, together, and, help, do, ...	[need, come, together, help, whatever, possibl...
3	3 our condolences goes out to the family who los...	1532754469728837633	23083404	2022-06-03T16:01:57.000Z	[our, condolences, goes, out, to, the, family,...	[condolences, goes, family, lost, loved, one, ...
4	4 yes we did!!! https://t.co/olwm2get18	1532225483596738560	23083404	2022-06-02T04:59:57.000Z	[yes, we, did, https, t, co, olwm2get18]	[yes]
5	5 🤔 "In the face of Impossible odds, people who...	1531837452591042561	23083404	2022-06-01T03:18:03.000Z	[in, the, face, of, Impossible, odds, people, ...	[face, Impossible, odds, people, love, country...
6	6 congratulations brother!!! 🍌🍌🍌🍌🙏🔥🇸🇰 htt...	1531784889279795201	23083404	2022-05-31T23:49:11.000Z	[congratulations, brother, https, t, co, 1ac2o...	[congratulations, brother]
7	7 ayyyyyeeeeee!!!! 🍌🍌🍌🍌 https://t.co/gxj6zf0dhh	1531260582657138688	23083404	2022-05-30T13:05:47.000Z	[ayyyyyyyyy, https, t, co, gxj6zf0dhh]	[ayyyyyyyyy]

4. Schritt

Um bestimmte Wörter entfernen zu können, brauchen wir das Paket "stopwords".

```
# Alle Wörter werden zu einem String

df['text_string'] = df['text_token_stopwords'].apply(lambda x: ' '.join([item for item in x if len(item)>1]))

all_words = ''.join([word for word in df["text_string"]])

all_words
```

[28] ✓ 0.1s Python

... 'jumping someone podcast soon maybefunny many basketball experts app everybody dr james naismith fascinatingneed come together help whatever possibly bring bg home quickly safely voice athletes stronger together uninterrupted wearebgcondolences goes family lost loved one heavens watch tragedy pray communityyesface impossible odds people love country change barack obamacongratulations brotherayyyyyyyyy'

```
# wir machen aus dem String eine Liste
tokenized_words = nltk.tokenize.word_tokenize(all_words)
```

[29] ✓ 0.9s Python

```
# mit FreqDist erstellen wir eine Häufikeitsverteilung; Wir zählen wie oft die Wörter vorkommen
from nltk.probability import FreqDist

fdist = FreqDist(tokenized_words)
fdist
```

[30] ✓ 0.8s Python

... FreqDist({'together': 2, 'jumping': 1, 'someone': 1, 'podcast': 1, 'soon': 1, 'maybefunny': 1, 'many': 1, 'basketball': 1, 'experts': 1, 'app': 1, ...})

```
# Wir filtern nun die Wörter raus, welche mehr als 1 Mal vorkommen
df['text_string_fdist'] = df['text_token_stopwords'].apply(lambda x: ' '.join([item for item in x if fdist[item] >= 1 ]))

df
```

[38] ✓ 0.1s Python

...

	Unnamed: 0	text	id	author_id	created_at	text_token	text_token_stopwords	text_string	text_string_fdist
0	0	jumping on someone podcast soon. maybe my own 🥰	1533632259755896832	23083404	2022-06-06T02:09:58.000Z	[jumping, on, someone, podcast, soon, maybe, m...	[jumping, someone, podcast, soon, maybe]	jumping someone podcast soon maybe	jumping someone podcast soon
1	1	It's so funny to me how many basketball expert...	1533632033972310016	23083404	2022-06-06T02:09:05.000Z	[It, s, so, funny, to, me, how, many, basketba...	[funny, many, basketball, experts, app, everyb...	funny many basketball experts app everybody dr...	many basketball experts app everybody dr james...
2	2	we need to come together and help do whatever ...	1533599792701378560	23083404	2022-06-06T00:00:58.000Z	[we, need, to, come, together, and, help, do, ...	[need, come, together, help, whatever, possibl...	need come together help whatever possibly brin...	come together help whatever possibly bring bg ...
3	3	our condolences goes out to the family who los...	1532754469728837633	23083404	2022-06-03T16:01:57.000Z	[our, condolences, goes, out, to, the, family,...	[condolences, goes, family, lost, loved, one, ...	condolences goes family lost loved one heavens...	goes family lost loved one heavens watch trage...
4	4	yes we dld!!! https://t.co/olwm2get18	1532225483596738560	23083404	2022-06-02T04:59:57.000Z	[yes, we, dld, https, t, co, olwm2get18]	[yes]	yes	
5	5	🥰 *In the face of Impossible odds, people who...	1531837452591042561	23083404	2022-06-01T03:18:03.000Z	[In, the, face, of, Impossible, odds, people, ...	[face, impossible, odds, people, love, country...	face impossible odds people love country chang...	impossible odds people love country change barack
6	6	congratulations brother!!! 🥳🥳🥳🥳🥳🥳htt...	1531784889279795201	23083404	2022-05-31T23:49:11.000Z	[congratulations, brother, https, t, co, 1ac2o...	[congratulations, brother]	congratulations brother	
7	7	ayyyyyyyyy!!!! 🥳🥳🥳🥳🥳https://t.co/gxj6zf0dhh	1531260582657138688	23083404	2022-05-30T13:05:47.000Z	[ayyyyyyyyy, https, t, co, gxj6zf0dhh]	[ayyyyyyyyy]	ayyyyyyyyy	

5. Schritt

In der ersten Zeile kann man, dass die Stopwords entfernt worden sind.

Mit FreqDist können wir nun die Häufigkeit der Wörter zählen

```
# Mit Lemmatizer kann man Wörter welche aus der gleichen Wortfamilie stammen, gruppieren.
from nltk.stem import WordNetLemmatizer
```

```
wordnet_lem = WordNetLemmatizer()
```

```
df['text_string_lem'] = df['text_string_fdist'].apply(wordnet_lem.lemmatize)
```

Python

```
df['is_equal'] = (df['text_string_fdist'] == df['text_string_lem'])
```

```
# Es wurden 8 gleiche Wörter gefunden
df.is_equal.value_counts()
```

Python

8

```
: is_equal, dtype: int64
```

df

✓ 0.8s

Python

Unnamed: 0		text	id	author_id	created_at	text_token	text_token_stopwords	text_string	text_string_fdlist	text_string_lem	is_equal
0	0	jumping on someone podcast soon. maybe my own 🤔	1533632259755896832	23083404	2022-06-06T02:09:58.000Z	[jumping, on, someone, podcast, soon, maybe, m...	[jumping, someone, podcast, soon, maybe]	Jumping someone podcast soon maybe	jumping someone podcast soon	jumping someone podcast soon	True
1	1	It's so funny to me how many basketball expert...	1533632033972310016	23083404	2022-06-06T02:09:05.000Z	[It, s, so, funny, to, me, how, many, basketba...	[funny, many, basketball, experts, app, everyb...	funny many basketball experts app everybody dr...	many basketball experts app everybody dr james...	many basketball experts app everybody dr james...	True
2	2	we need to come together and help do whatever ...	1533599792701378560	23083404	2022-06-06T00:00:58.000Z	[we, need, to, come, together, and, help, do, ...	[need, come, together, help, whatever, possibl...	need come together help whatever possibly brin...	come together help whatever possibly bring bg ...	come together help whatever possibly bring bg ...	True
3	3	our condolences goes out to the family who los...	1532754469728837633	23083404	2022-06-03T16:01:57.000Z	[our, condolences, goes, out, to, the, family,...	[condolences, goes, family, lost, loved, one, ...	condolences goes family lost loved one heavens...	goes family lost loved one heavens watch trage...	goes family lost loved one heavens watch trage...	True
4	4	yes we did!!! https://t.co/olwm2get18	1532225483596738560	23083404	2022-06-02T04:59:57.000Z	[yes, we, did, https, t, co, olwm2get18]	[yes]	yes			True
5	5	🤔 "In the face of Impossible odds, people who...	1531837452591042561	23083404	2022-06-01T03:18:03.000Z	[In, the, face, of, impossible, odds, people, ...	[face, impossible, odds, people, love, country...	face impossible odds people love country chang...	impossible odds people love country change barack	impossible odds people love country change barack	True
6	6	congratulations brother!!! 🎉🎉🎉🎉🎉 👉htt...	1531784889279795201	23083404	2022-05-31T23:49:11.000Z	[congratulations, brother, https, t, co, 1ac2o...	[congratulations, brother]	congratulations brother			True
7	7	ayyyyyyyyy!!!! 🤔🤔🤔 https://t.co/gxj6zf0dhh	1531260582657138688	23083404	2022-05-30T13:05:47.000Z	[ayyyyyyyyy, https, t, co, gxj6zf0dhh]	[ayyyyyyyyy]	ayyyyyyyyy			True

```
# wir schaffen eine Wortcloud:

all_words_lem = ' '.join([word for word in df['text_string_lem']])
```

✓ 0.1s

```
%matplotlib inline
import matplotlib.pyplot as plt
from wordcloud import WordCloud

wordcloud = WordCloud(width=400,
                        height=125,
                        random_state=2,
                        max_font_size=100).generate(all_words_lem)

plt.figure(figsize=(10, 7))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off');
```

✓ 0.6s



```
import numpy as np

x, y = np.ogrid[:300, :300]
mask = (x - 150) ** 2 + (y - 150) ** 2 > 130 ** 2
mask = 255 * mask.astype(int)

wc = WordCloud(background_color="white", repeat=True, mask=mask)
wc.generate(all_words)

plt.axis("off")
plt.imshow(wc, interpolation="bilinear");
```



7. Schritt

Ergebnisse der Wortcloud


```
from nltk.tokenize import word_tokenize
from nltk.probability import FreqDist
```

```
words = nltk.word_tokenize(all_words_lem)
fd = FreqDist(words)
```

✓ 1.4s

```
# wir geben die drei Wörter aus die am meisten vorkommen
fd.most_common(3)
```

✓ 0.1s

```
[('together', 2), ('jumping', 1), ('someone', 1)]
```

```
# wir geben die drei Wörter aus welche am meisten vorkommen in Form einer Tabelle
fd.tabulate(3)
```

✓ 0.1s

together	jumping	someone
2	1	1

```
# Wir geben die drei Wörter aus welche am meisten vorkommen und speichern sie in der Variable top_10
top_10 = fd.most_common(10)
```

```
# Wir erstellen eine Panda-Serie, um einen Plot zu erstellen
fdist = pd.Series(dict(top_10))
```

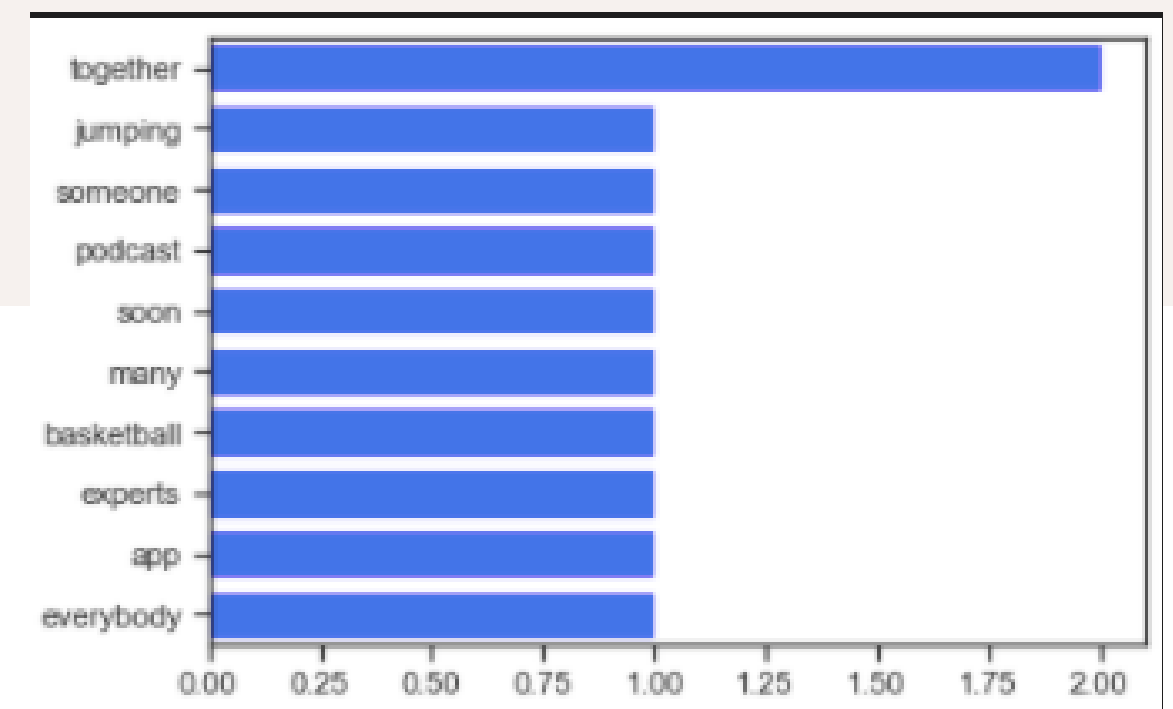
```
top_10
```

✓ 0.2s

```
[('together', 2),
 ('jumping', 1),
 ('someone', 1),
 ('podcast', 1),
 ('soon', 1),
 ('many', 1),
 ('basketball', 1),
 ('experts', 1),
 ('app', 1),
 ('everybody', 1)]
```

8. Schritt

Wir kommen zurück auf die Wörter, welche am meisten vorkommen: In unserem Fall gibt es keine Wörter, die öfters vorkommen, außer das Wort "together"



```
# Vader = Valence Aware Dictionary and sEntiment Reasoner analysiert Stimmungen bzw. die Polaritäten positiv/negativ/neutral
nltk.download('vader_lexicon')

✓ 0.9s Python

[nltk_data] Downloading package vader_lexicon to
[nltk_data] /Users/rafaela/nltk_data...
[nltk_data] Package vader_lexicon is already up-to-date!

True

from nltk.sentiment import SentimentIntensityAnalyzer

analyzer = SentimentIntensityAnalyzer()

✓ 0.1s Python
More...

# wir erzeugen eine neue Spalte, um die Polirität zu sehen

df['polarity'] = df['text_string_lem'].apply(lambda x: analyzer.polarity_scores(x))

df

✓ 1.4s Python
```

Unnamed: 0		text	id	author_id	created_at	text_token	text_token_stopwords	text_string	text_string_fdist	text_string_lem	is_equal	polarity
0	0	jumping on someone podcast soon. maybe my own 🤔	153363225975896832	23083404	2022-06-06T02:09:58.000Z	[jumping, on, someone, podcast, soon, maybe, m...	[jumping, someone, podcast, soon, maybe]	Jumping someone podcast soon maybe	jumping someone podcast soon	jumping someone podcast soon	True	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...
1	1	It's so funny to me how many basketball expert...	1533632033972310016	23083404	2022-06-06T02:09:05.000Z	[It, s, so, funny, to, me, how, many, basketba...	[funny, many, basketball, experts, app, everyb...	funny many basketball experts app everybody dr...	many basketball experts app everybody dr james...	many basketball experts app everybody dr james...	True	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...
2	2	we need to come together and help do whatever ...	1533599792701378560	23083404	2022-06-06T00:00:58.000Z	[we, need, to, come, together, and, help, do, ...	[need, come, together, help, whatever, possibl...	need come together help whatever possibly brin...	come together help whatever possibly bring bg ...	come together help whatever possibly bring bg ...	True	{'neg': 0.0, 'neu': 0.585, 'pos': 0.415, 'comp...
3	3	our condolences goes out to the family who los...	1532754469728837633	23083404	2022-06-03T16:01:57.000Z	[our, condolences, goes, out, to, the, family,...	[condolences, goes, family, lost, loved, one, ...	condolences goes family lost loved one heavens...	goes family lost loved one heavens watch trage...	goes family lost loved one heavens watch trage...	True	{'neg': 0.342, 'neu': 0.204, 'pos': 0.454, 'co...
4	4	yes we did!!! https://t.co/olwm2get18	1532225483596738560	23083404	2022-06-02T04:59:57.000Z	[yes, we, dld, https, t, co, olwm2get18]	[yes]	yes			True	{'neg': 0.0, 'neu': 0.0, 'pos': 0.0, 'compound...
5	5	🤔 "In the face of impossible odds, people who...	1531837452591042561	23083404	2022-06-01T03:18:03.000Z	[In, the, face, of, impossible, odds, people, ...	[face, impossible, odds, people, love, country...	face impossible odds people love country chang...	impossible odds people love country change barack	impossible odds people love country change barack	True	{'neg': 0.0, 'neu': 0.588, 'pos': 0.412, 'comp...
6	6	congratulations brother!!! 🥳🥳🥳 https://t.co/gxj6zf0dhh	1531784889279795201	23083404	2022-05-31T23:49:11.000Z	[congratulations, brother, https, t, co, 1ac2o...	[congratulations, brother]	congratulations brother			True	{'neg': 0.0, 'neu': 0.0, 'pos': 0.0, 'compound...
7	7	ayyyyyyyyye!!!! 🥳🥳🥳 https://t.co/gxj6zf0dhh	1531260582657138688	23083404	2022-05-30T13:05:47.000Z	[ayyyyyyyyye, https, t, co, gxj6zf0dhh]	[ayyyyyyyyye]	ayyyyyyyyye			True	{'neg': 0.0, 'neu': 0.0, 'pos': 0.0, 'compound...

9. Schritt

Nun untersuchen wir die sog. "Polarität". Mit Hilfe dieser können wir analysieren, ob die Tweets positiv, negativ oder neutral sind.

Wir erstellen eine neue Variable "sentiment"

```
df['sentiment'] = df['compound'].apply(lambda x: 'positive' if x >0 else 'negative' if x==0 else 'neutral')
```

df

✓ 0.3s

Python

	text	created_at	text_token	text_token_stopwords	text_string	text_string_fdist	text_string_lem	is_equal	neg	neu	pos	compound	sentiment
0	jumping on someone podcast soon. maybe my own 🤔	2022-06-06T02:09:58.000Z	[jumping, on, someone, podcast, soon, maybe, m...	[jumping, someone, podcast, soon, maybe]	jumping someone podcast soon maybe	jumping someone podcast soon	jumping someone podcast soon	True	0.000	1.000	0.000	0.0000	negative
1	It's so funny to me how many basketball expert...	2022-06-06T02:09:05.000Z	[It, s, so, funny, to, me, how, many, basketba...	[funny, many, basketball, experts, app, everyb...	funny many basketball experts app everybody dr...	many basketball experts app everybody dr James...	many basketball experts app everybody dr James...	True	0.000	1.000	0.000	0.0000	negative
2	we need to come together and help do whatever ...	2022-06-06T00:00:58.000Z	[we, need, to, come, together, and, help, do, ...	[need, come, together, help, whatever, possibl...	need come together help whatever possibly brin...	come together help whatever possibly bring bg ...	come together help whatever possibly bring bg ...	True	0.000	0.585	0.415	0.8176	positive
3	our condolences goes out to the family who los...	2022-06-03T16:01:57.000Z	[our, condolences, goes, out, to, the, family,...	[condolences, goes, family, lost, loved, one, ...	condolences goes family lost loved one heavens...	goes family lost loved one heavens watch trage...	goes family lost loved one heavens watch trage...	True	0.342	0.204	0.454	0.2960	positive
4	yes we did!!! https://t.co/olwm2get18	2022-06-02T04:59:57.000Z	[yes, we, did, https, t, co, olwm2get18]	[yes]	yes			True	0.000	0.000	0.000	0.0000	negative
5	🤔 "In the face of impossible odds, people who...	2022-06-01T03:18:03.000Z	[In, the, face, of, impossible, odds, people, ...	[face, impossible, odds, people, love, country...	face impossible odds people love country chang...	impossible odds people love country change barack	impossible odds people love country change barack	True	0.000	0.588	0.412	0.6369	positive
6	congratulations brother!!! 🍷🍷🍷🍷🍷 ❤️👉htt...	2022-05-31T23:49:11.000Z	[congratulations, brother, https, t, co, 1ac2o...]	[congratulations, brother]	congratulations brother			True	0.000	0.000	0.000	0.0000	negative
7	ayyyyyyyyye!!! 🤔🤔🤔 https://t.co/gxj6zf0dhh	2022-05-30T13:05:47.000Z	[ayyyyyyyyye, https, t, co, gxj6zf0dhh]	[ayyyyyyyyye]	ayyyyyyyyye			True	0.000	0.000	0.000	0.0000	negative

Tweet mit höchstem positivien Sentiment

```
df.loc[df['compound'].idxmax()].values
```

✓ 0.2s

Python

```
array(['we need to come together and help do whatever we possibly can to bring bg home quickly and safely!! 🙏our voice as athletes is stronger together. @uninterrupted 🤔#wearebg https://t.co/2gww3ff81p',  
      '2022-06-06T00:00:58.000Z',  
      list(['we', 'need', 'to', 'come', 'together', 'and', 'help', 'do', 'whatever', 'we', 'possibly', 'can', 'to', 'bring', 'bg', 'home', 'quickly', 'and', 'safely', 'our', 'voice', 'as', 'athletes', 'is', 'stronger', 'together', 'uninterrupted', 'wearebg',  
'https', 't', 'co', '2gww3ff81p']),  
      list(['need', 'come', 'together', 'help', 'whatever', 'possibly', 'bring', 'bg', 'home', 'quickly', 'safely', 'voice', 'athletes', 'stronger', 'together', 'uninterrupted', 'wearebg']),  
      'need come together help whatever possibly bring bg home quickly safely voice athletes stronger together uninterrupted wearebg',  
      'come together help whatever possibly bring bg home quickly safely voice athletes stronger together uninterrupted',  
      'come together help whatever possibly bring bg home quickly safely voice athletes stronger together uninterrupted',  
      True, 0.0, 0.585, 0.415, 0.8176, 'positive'], dtype=object)
```

Tweet mit höchstem negativen Sentiment

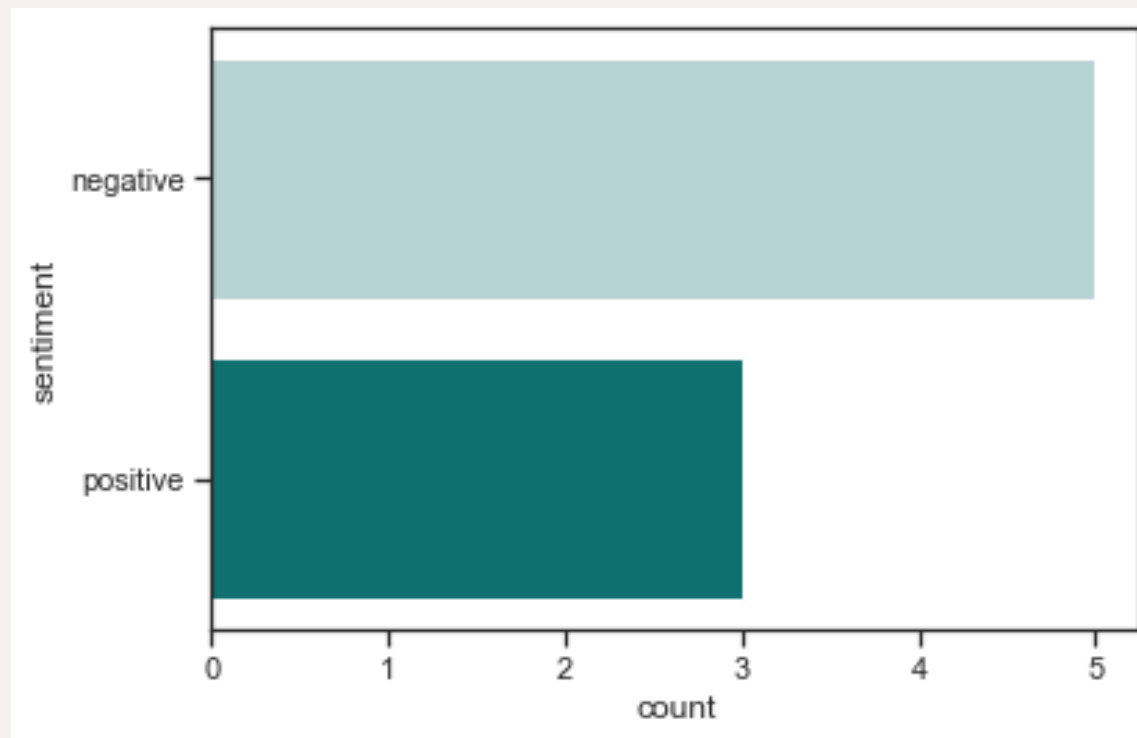
```
df.loc[df['compound'].idxmin()].values
```

✓ 0.1s

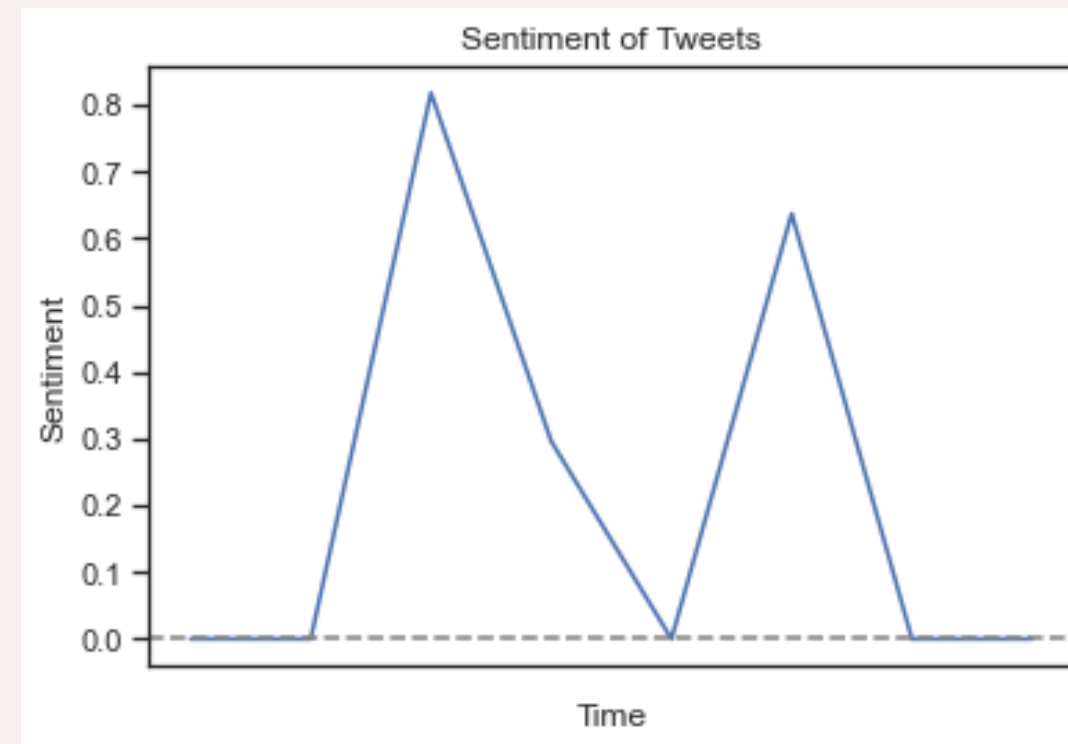
Python

```
array(['jumping on someone podcast soon. maybe my own 🤔',  
      '2022-06-06T02:09:58.000Z',  
      list(['jumping', 'on', 'someone', 'podcast', 'soon', 'maybe', 'my', 'own']),  
      list(['jumping', 'someone', 'podcast', 'soon', 'maybe']),  
      'jumping someone podcast soon maybe',  
      'jumping someone podcast soon', 'jumping someone podcast soon',  
      True, 0.0, 1.0, 0.0, 0.0, 'negative'], dtype=object)
```

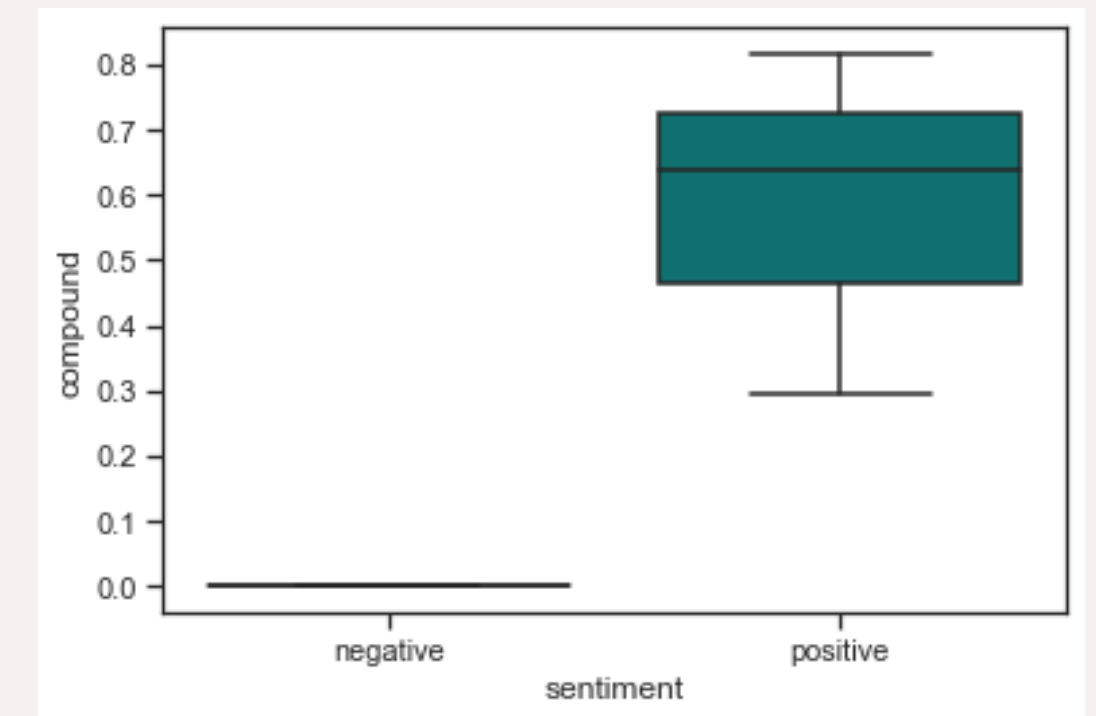
Ergebnisse der Sentiment Analyse



Die meisten Tweets werden als negativ aufgefasst.



Verlauf der Tweets



Im zusammengesetzten Prozentsatz (=Compound) sieht man, dass die positiven Tweets deutlich besser ankommen,

Visualisierung mit Streamlit

```
# SETUP

import streamlit as st
import seaborn as sns
import pandas as pd
import altair as alt
import streamlit as st
import matplotlib.pyplot as plt

# Einlesen der CSV Datei

df = pd.read_csv("Fallstudie2_Daten.csv")

st.title('Herzlich Willkommen auf dem Dashboard von Rafaela Papanikolaou')

st.subheader('Wir analysieren den Twitter Account von LeBron James')

# Erkenntnisse:

st.markdown("Lebrons Aktivität ist um 10% niedriger als im letzten Monat")
st.metric(label="Twitter Aktivität", value="40%", delta="-10%", delta_color="normal")

st.dataframe(df, 1000, 1000)

st.table(df)

st.line_chart(df["sentiment"])

st.bar_chart(df["compound"])

# SelectBox
add_selectbox = st.sidebar.selectbox("Was möchtest du sehen?", ("Datensatz", "Plots", "Lustige Memes"))
```

Lustige Memes

Wir analysieren den Twitter Account von LeBron James

Lebrons Akitvität ist um 10% niedriger als im letzten Monat

Twitter Aktivität

40%

↓ -10%

	Unnamed: 0	text	created_at	text_token
0	0	jumping on someone pod...	2022-06-06T02:09:58.000Z	['jumping', 'on', 'someo
1	1	it's so funny to me how m...	2022-06-06T02:09:05.000Z	['it', 's', 'so', 'funny', 'to',
2	2	we need to come togethe...	2022-06-06T00:00:58.000Z	['we', 'need', 'to', 'come'
3	3	our condolences goes out...	2022-06-03T16:01:57.000Z	['our', 'condolences', 'gc
4	4	yes we did!!! https://t.co/...	2022-06-02T04:59:57.000Z	['yes', 'we', 'did', 'https',
5	5	😬. “in the face of imposs...	2022-06-01T03:18:03.000Z	['in', 'the', 'face', 'of', 'im
6	6	congratulations brother!!...	2022-05-31T23:49:11.000Z	['congratulations', 'brot
7	7	ayyyyyyyyy!!!! 🤔🤔🤔 ...	2022-05-30T13:05:47.000Z	['ayyyyyyyyy', 'https', 't

[illegible]

Was möchtest du sehen?

Lustige Memes

		country can change it."- -barack obama https://t.co/2chzznfrjx	01T03:18:03.000Z	country, can, 'change', 'it', 'barack', 'obama', 'https', 't', 'co', '2chzznfrjx']	'country', 'change', 'barack', 'obama']	love country change barack obama	love country change barack	love country change barack	true	0.0000	0.3660	0.4120	0.6369	positive
6	6	congratulations brother!!! 🙌🙌🙌🙌 🙏👊❤️👑 https://t.co/1ac2oyfv6m	2022-05- 31T23:49:11.000Z	['congratulations', 'brother', 'https', 't', 'co', '1ac2oyfv6m']	['congratulations', 'brother']	congratulations brother	<NA>	<NA>	true	0.0000	0.0000	0.0000	0.0000	negative
7	7	ayyyyyyyyy!!!! 🤔🤔🤔 https://t.co/gxj6zf0dhh	2022-05- 30T13:05:47.000Z	['ayyyyyyyyy', 'https', 't', 'co', 'gxj6zf0dhh']	['ayyyyyyyyy']	ayyyyyyyyy	<NA>	<NA>	true	0.0000	0.0000	0.0000	0.0000	negative

