Welcome!

SENTIMENT ANALYSIS IN PYTHON

Violeta Misheva
Data Scientist

datacamp

What is sentiment analysis?

> **Sentiment analysis is the process of understanding the opinion of an author about a subject.**

What goes into a sentiment analysis system?

First element: Opinion/emotion

- Opinion (polarity): pos, neutral, neg

- Emotion

What goes into a sentiment analysis system?

Second element: subject

- Subject of discussion: **What is being talked about** ?

> _The camera on this phone is great but its battery life is rather disappointing. _

Third element: opinion holder

- Opinion holder (entity): **By whom**?

Why sentiment analysis?

- Social media monitoring
  - Not only what people are talking about but HOW they are talking about it
  - Sentiment can be found also in forums, blogs, news
- Brand monitoring
- Customer service
- Product analytics
- Market research and analysis

Let's look at movie reviews!

```
data.head()
```

	review	label
0	This short spoof can be found on Elite's Mille...	0
1	A singularly unfunny musical comedy that artif...	0
2	An excellent series, masterfully acted and dir...	1
3	The master of movie spectacle Cecil B. De Mill...	1
4	I was gifted with this movie as it had such a ...	0

How many positive and negative reviews?

```
data.label.value_counts()
```

```
0    3782
1    3719
Name: label, dtype: int64
```

Percentage of positive and negative reviews

```
data.label.value_counts() / len(data)
```

```
0    0.504199
1    0.495801
Name: label, dtype: float64
```

How long is the longest review?

```
length_reviews = data.review.str.len()
```

```
type(length_reviews)
pandas.core.series.Series
```

```
# Finding the review with max length
max(length_reviews)
```

```
0    667
1    2982
2    669
3    1087
....
```

How long is the shortest review?

```
length_reviews = data.review.str.len()
```

```
# Finding the review with min length
min(length_reviews)
```

```
0    667
1    2982
2    669
3    1087
4    724
.....
```

Let's practice!

SENTIMENT ANALYSIS IN PYTHON

Sentiment analysis types and approaches

SENTIMENT ANALYSIS IN PYTHON

Violeta Misheva
Data Scientist

datacamp

Levels of granularity

1. Document level
2. Sentence level
3. Aspect level

> The camera in this phone is pretty good but the battery life is disappointing.

Type of sentiment analysis algorithms

- Rule/lexicon-based

nice:+2, good:+1, terrible: -3 ...

Today was a good day.

```
Today: 0, was:0, a:0, good:+1, day:0
Total valence: +1
```

- Automatic/ Machine learning

What is the valence of a sentence?

```
text = "Today was a good day."
```

```
from textblob import TextBlob


my_valence = TextBlob(text)
my_valence.sentiment
```

```
Sentiment(polarity=0.7, subjectivity=0.6000000000000001)
```

Automated or rule-based?

Automated/Machine learning

- Rely on having labelled historical data
- Might take a while to train
- Latest machine learning models can be quite powerful

Rule/lexicon-based

- Rely on manually crafted valence scores
- Different words might have different polarity in different contexts
- Can be quite fast

Let's practice!

SENTIMENT ANALYSIS IN PYTHON

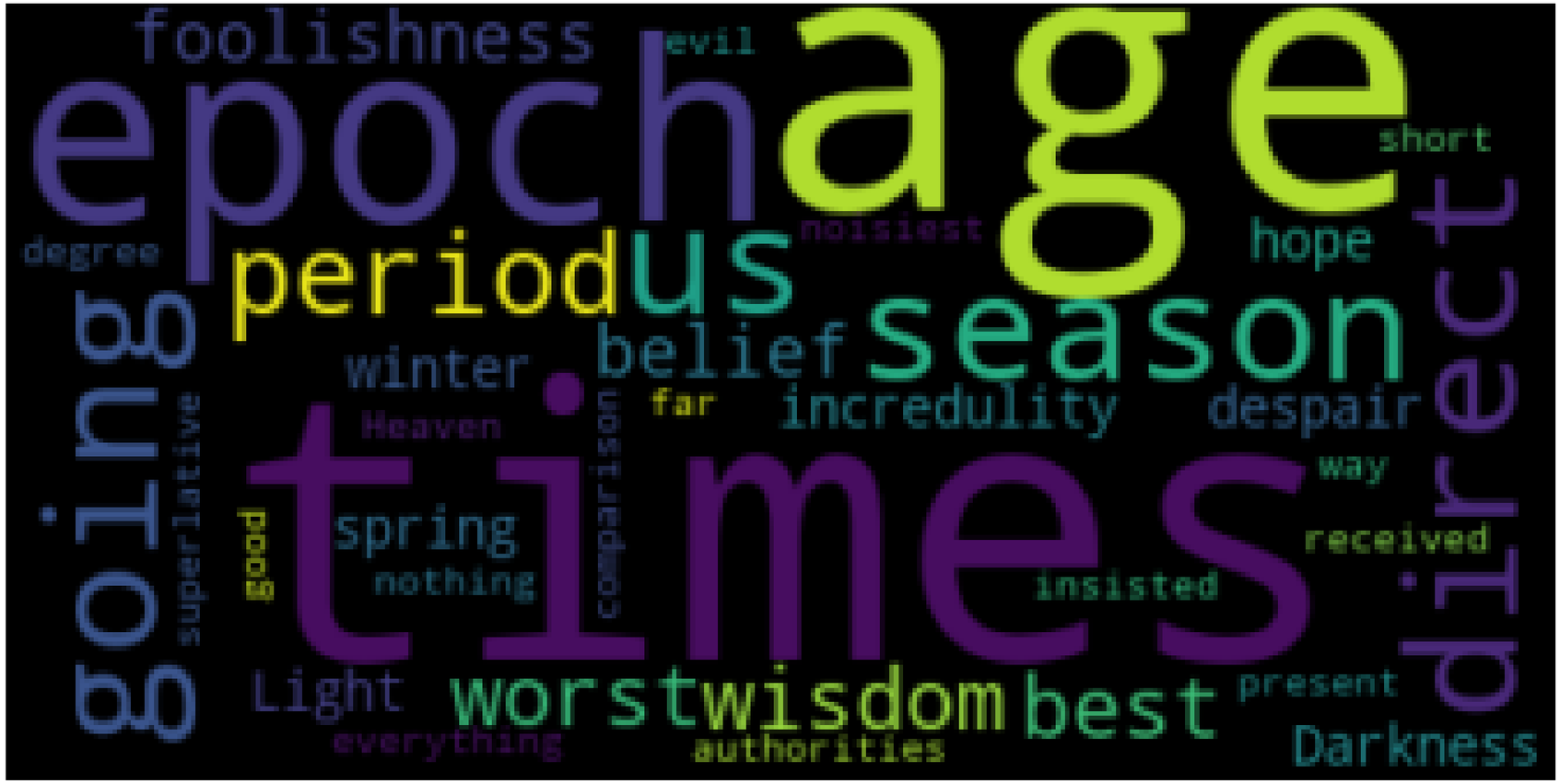
Let's build a word cloud!

SENTIMENT ANALYSIS IN PYTHON

Violeta Misheva
Data Scientist

datacamp

Word cloud example



How do word clouds work?

> The more frequent a word is, the BIGGER and **bolder** it will appear on the word cloud.

- Word cloud generated by one of the longest reviews

Why word clouds?

Pros

- Can reveal the essential
- Provide an overall sense of the text
- Easy to grasp and engaging

Cons

- Sometimes confusing and uninformative
- With larger text, require more work

Let's build a word cloud in Python!

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

```
two_cities = "It was the best of times, it was the worst of times,
        it was the age of wisdom, it was the age of foolishness,
        it was the epoch of belief, it was the epoch of incredulity,
        it was the season of Light, it was the season of Darkness,
        it was the spring of hope, it was the winter of despair,
        we had everything before us, we had nothing before us,
        we were all going direct to Heaven, we were all going
        direct the other way - in short, the period was so far
        like the present period, that some of its noisiest
        authorities insisted on its being received, for good
        or for evil, in the superlative degree of comparison only."
```

Define the WordCloud object

```
cloud_two_cities = WordCloud().generate(two_cities)
```

```
# To see all arguments of the function
?WordCloud
```

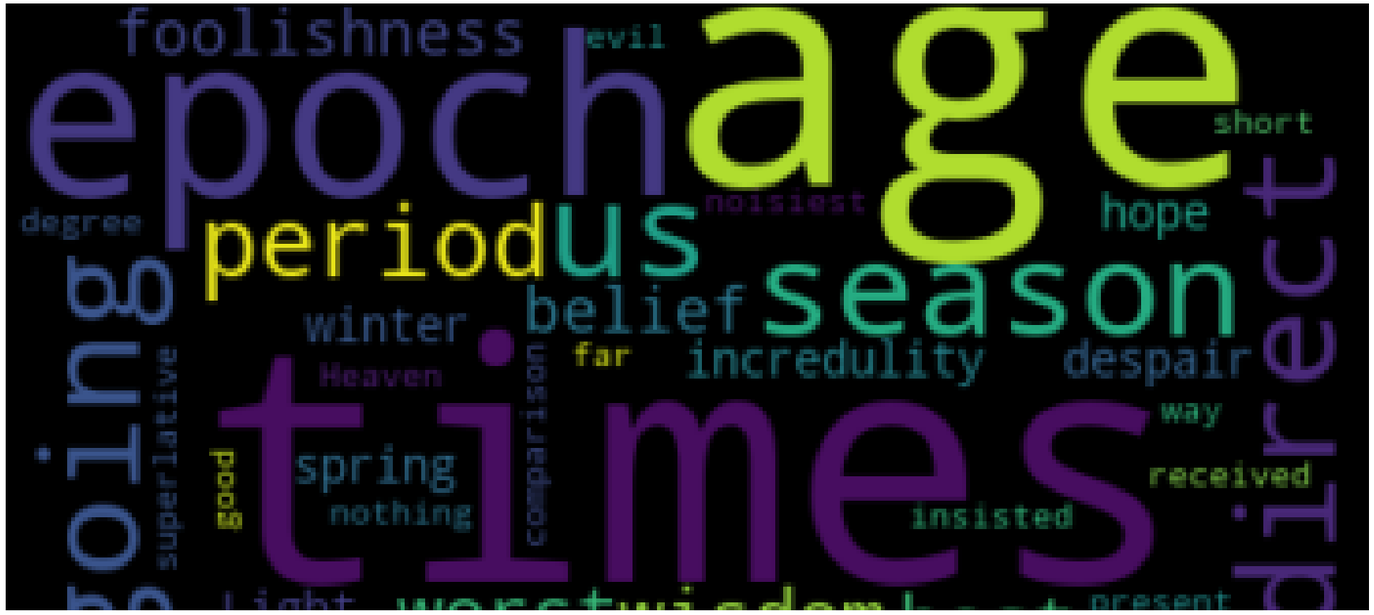
- Background color
- Size and font of the words, scaling
- Stopwords

```
# How does cloud_two_cities look like?
cloud_two_cities
<wordcloud.wordcloud.WordCloud at 0x2585f286d68>
```

Dislaying the word cloud!

```
plt.imshow(cloud_two_cities, interpolation='bilinear')
```

```
plt.axis('off')
plt.show()
```



Let's practice!

SENTIMENT ANALYSIS IN PYTHON