

Twitter Data Analysis

Applied Data Science in Python - Partial Result

Project 5

Sebastian Mohr

Student-ID: 3188691

Wintersemester 2022

1 Loading in of data

The data will be loaded through the Twitter API. The easiest way for this is to use the Python library **Tweepy** that provides different methods to access the Twitter API. The first step here will be the input of a specific events hashtag, for example the current Fifa World Cup 2022 uses the official Hashtag **#FIFAWorldCup**. After the tag has been put into the application, it will request a specific number of tweets for this event from the Twitter API. Currently I go with 100 tweets, but this number can be also set higher or lower, depending on preferences on how detailed the analyzed data should be.

To load the data, a **URL** has to be defined, which accesses the API at a specific endpoint. The URL also defines the query parameters, such as how many tweets are requested, which hashtags should be included in each tweet or which timeframe the tweet has to be created in. In **Tweepy** this process is done by creating a **Tweepy** Client, that includes the Bearer Token, through which the Twitter API can authenticate the current used dev-User. The defined query parameters then get put into the `client.search_recent_tweets(...)` method, that then starts an API call and retrieves all matching tweets. This method also takes in the definition of which fields of the tweets are necessary and will be included in the result.

2 Analyzing the tweets