Sentiment Analysis on Tweets in Switzerland

The goal of this project was to do sentiment analysis on Twitter data geolocated in Switzerland. We created an interactive map to visualize the results.

Challenges

- Multiple languages (French, German, English)
- No labelling for the data
- Learn new tools and libraries
- Rough location of the data (city level at best)
- No service accepts to translate millions of tweets

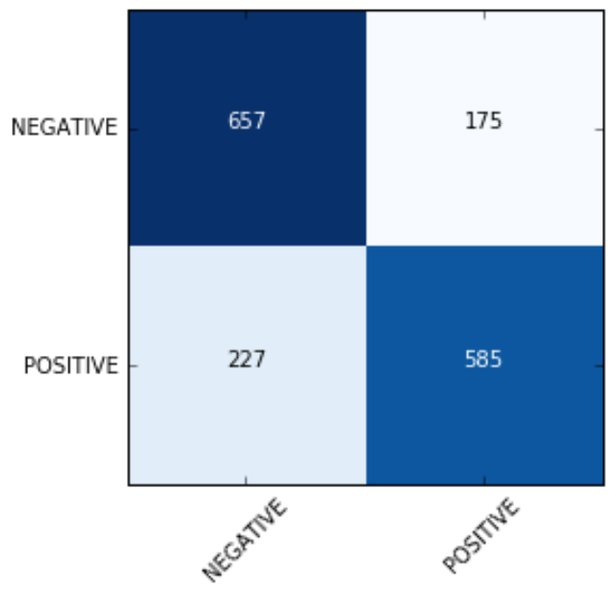
Data sources

- Twitter
  - English tweets (already classified)
  - French/German tweets (reclassified)
- Geolocation using online services

Tweets classification

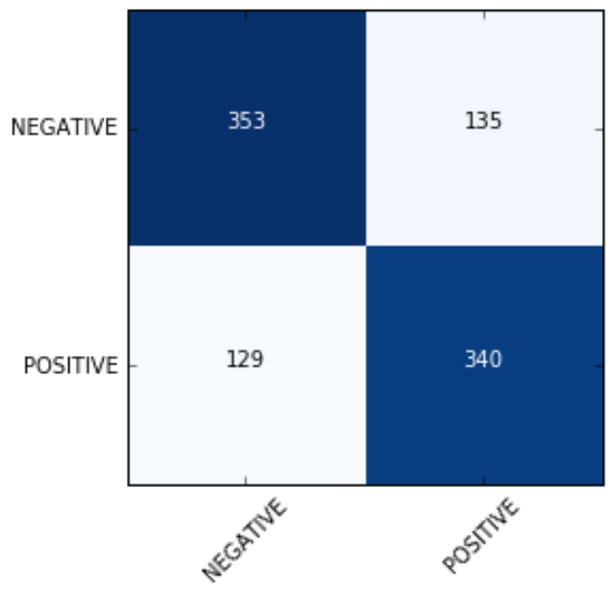
- Use provided sentiments for English tweets
- Reclassify French/German tweets
- Tokenization, stemming: nltk
- Machine learning: scikit-learn

French



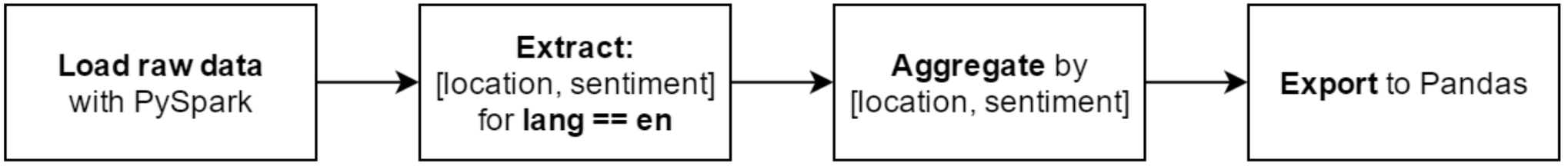
Accuracy: 0.761

German

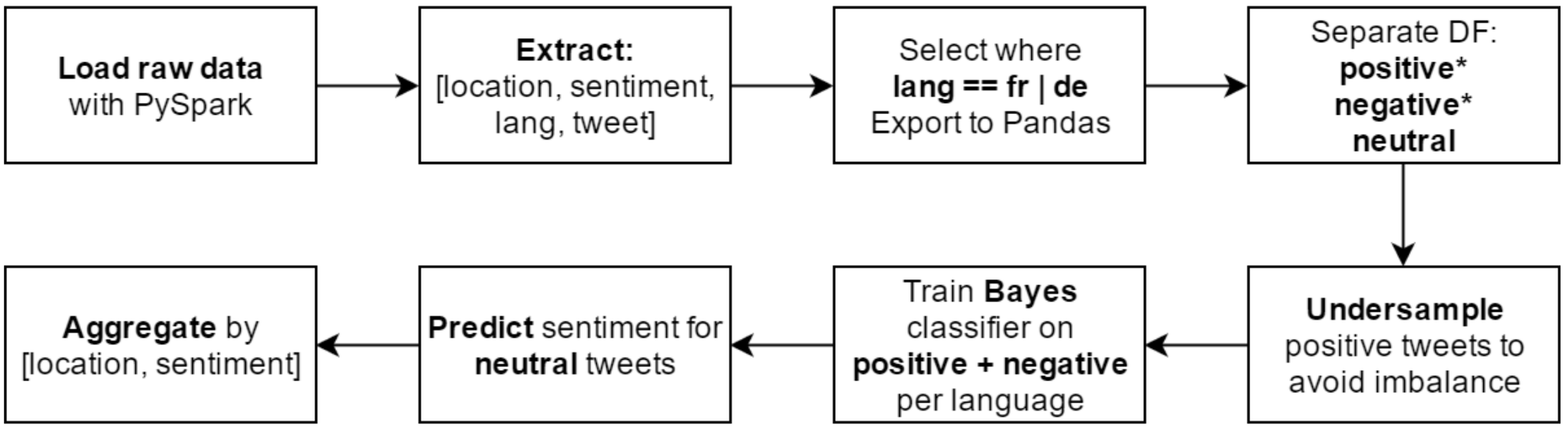


Accuracy: 0.708

Extract english classified tweets:

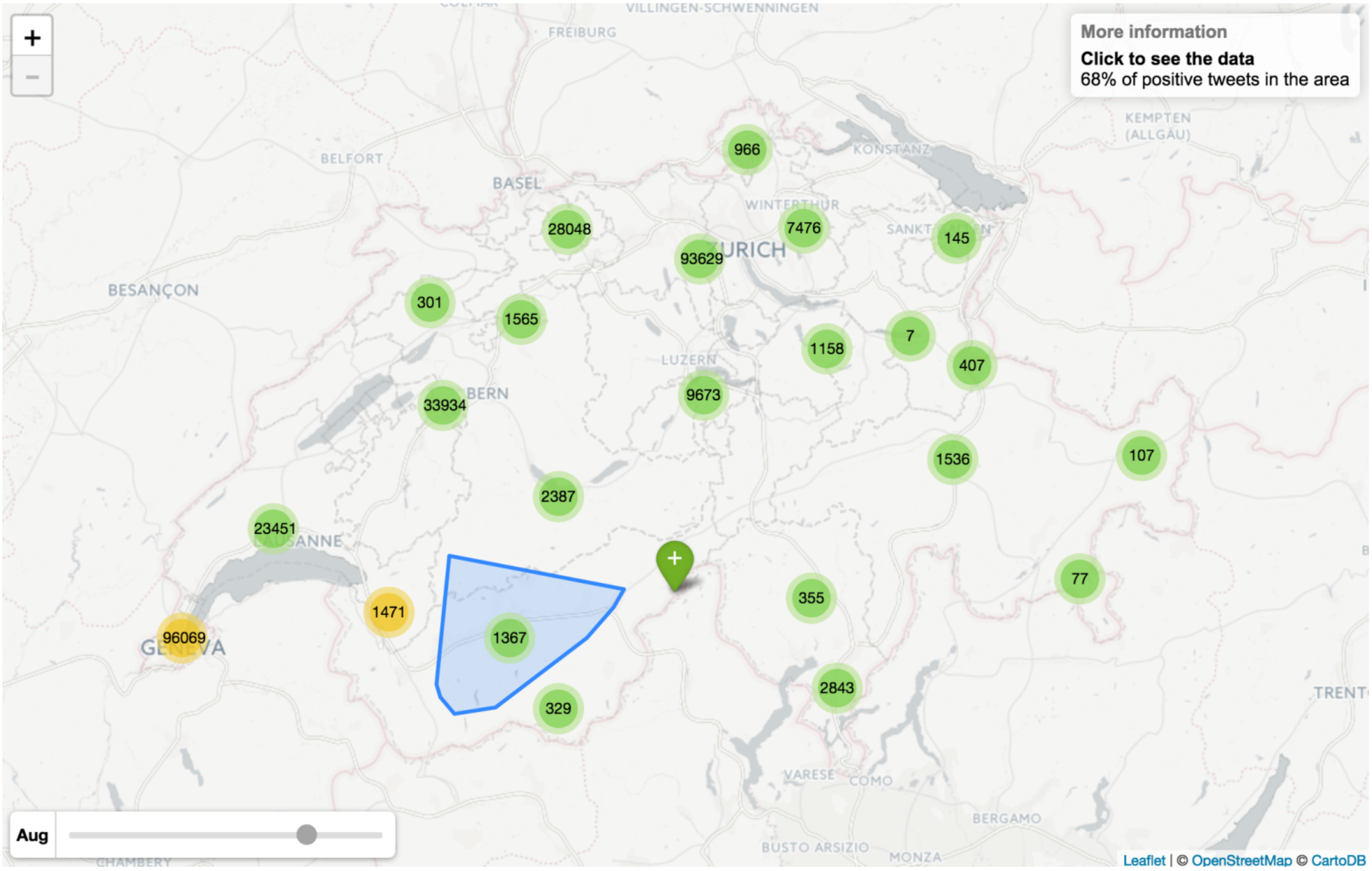


Extract french/german tweets:



Processing the data for visualization

- Merging all the data
- Cleaning up the data
- GeoPy to get coordinates
- Grouping by month, language, location and sentiment
- LeafletJS library for the map



Interactive map

- Each marker is a city
- Markers are clusterized
- Numbers = tweets in the area
- Color = ratio of positive tweets
- Detailed information on mouse over
- Choose the month with the slider