Project Status Report

The disagreeable frogs

1155129475 LUO Xuqi
1155129451 KAI Weiheng
1155132004 LI Ruochong
1155132913 PENG Du
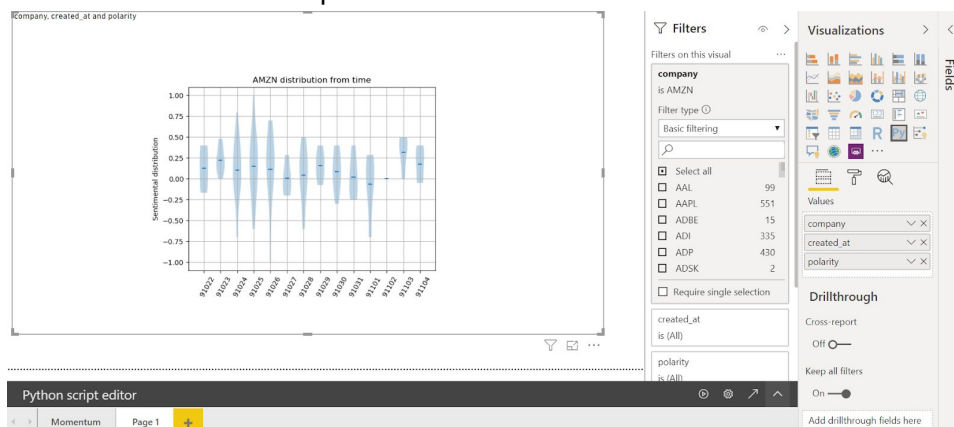
Sentiment Analysis about NASDAQ 100 on Twitter

Timeline Review

Date	Task	Status
10.01	Crawl data from twritter.com	Finished
10.15	Preprocess	Finished
11.01	Apply clustering model	Finished
11.15	Draw the result, visualization	In Progress
11.30	Conclusion	In Progress
12.02	Presentation	In Progress

Implementation List

- The semi-supervised k-mean clustering algorithm has been constructed in the backend.

- The twitter crawler was constructed with Python. We have used it to collect tweets related to NASDAQ 100 companies.

- Data-cleaning has also been implemented.

- The sentiment information is generated.

- Visualization has been implemented.

Realization of 3Vs

- Volume
  The size of the data is large with respective to pure texts.

- Variety
  Twitter feeds are unstructured data that consists of a variety of data types, including words, numbers, hashtags, emoticons, emojis, time and locations.

- Velocity
  The tweets were crawled in real time and being applied to the clustering algorithm.

To-do List

- Applying crawled tweets to the K-mean algorithm.

- Draw conclusions on the result and make possible improvements.

- Front-end implementation of the clustering machine in forms of an online webpage.

- Prepare slides for presentation.