Summary

The twitter provides full archive tweets search and count with relatively high price for commercial usage. However, developers could either apply for academic research access or using outsource scraper package named snscraper to download the historical data, it has no rate and amount limitation, with multi-thread scraping, it is possible to obtain millions of tweets in ten minutes.

Real time streaming is allowed for twitter API, we could obtain the real time tweets that we are interested as our feed data for our model (Still working on AsyncStreamingClient). Using the function provided in the tweepy module, we can stream tweets with strong filtering method as our demands.

For detailed information, please check in the report