1) A short description explaining in English what your interesting analysis is doing

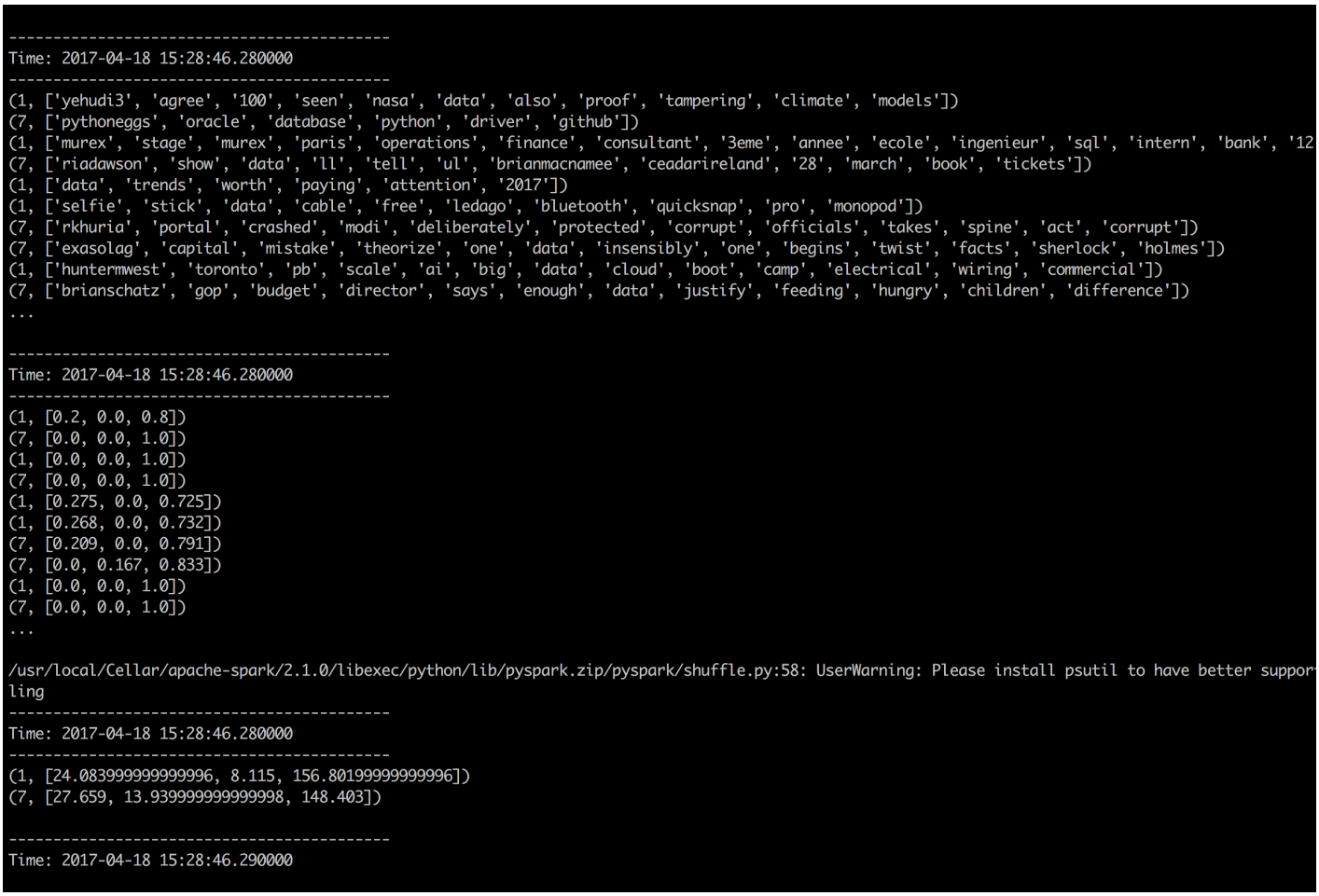
- Our task is to cluster tweets into several clusters based on their semantic content, and to learn the sentiments associated with these different semantic clusters. This lets us understand 2 things –
    - What people are tweeting about (i.e the topics) and what they feel about these topics (the sentiment associated with this set of topics)

- We achieve the clustering using word2vec embeddings that we train using the spark word2vec module on the brown corpus (http://www.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/corpora/list/private/brown/brown.html)

- We then create document embeddings for each tweet using the words that are present in them. These document embeddings are clustered using the Streaming K Means functionality in Spark MLLIB. The corresponding sentiments are computed for each cluster using the NLTK sentiment analyzer, which returns an array with 3 entries – positive sentiment, negative sentiment and neutral sentiment.

- In this screenshot we display some tweets that were assigned to clusters 1 and 7. The corresponding sentiment vectors for these tweets are displayed below the tweets, and the sum of sentiment vectors for cluster 1 and 7 can be seen further below that.

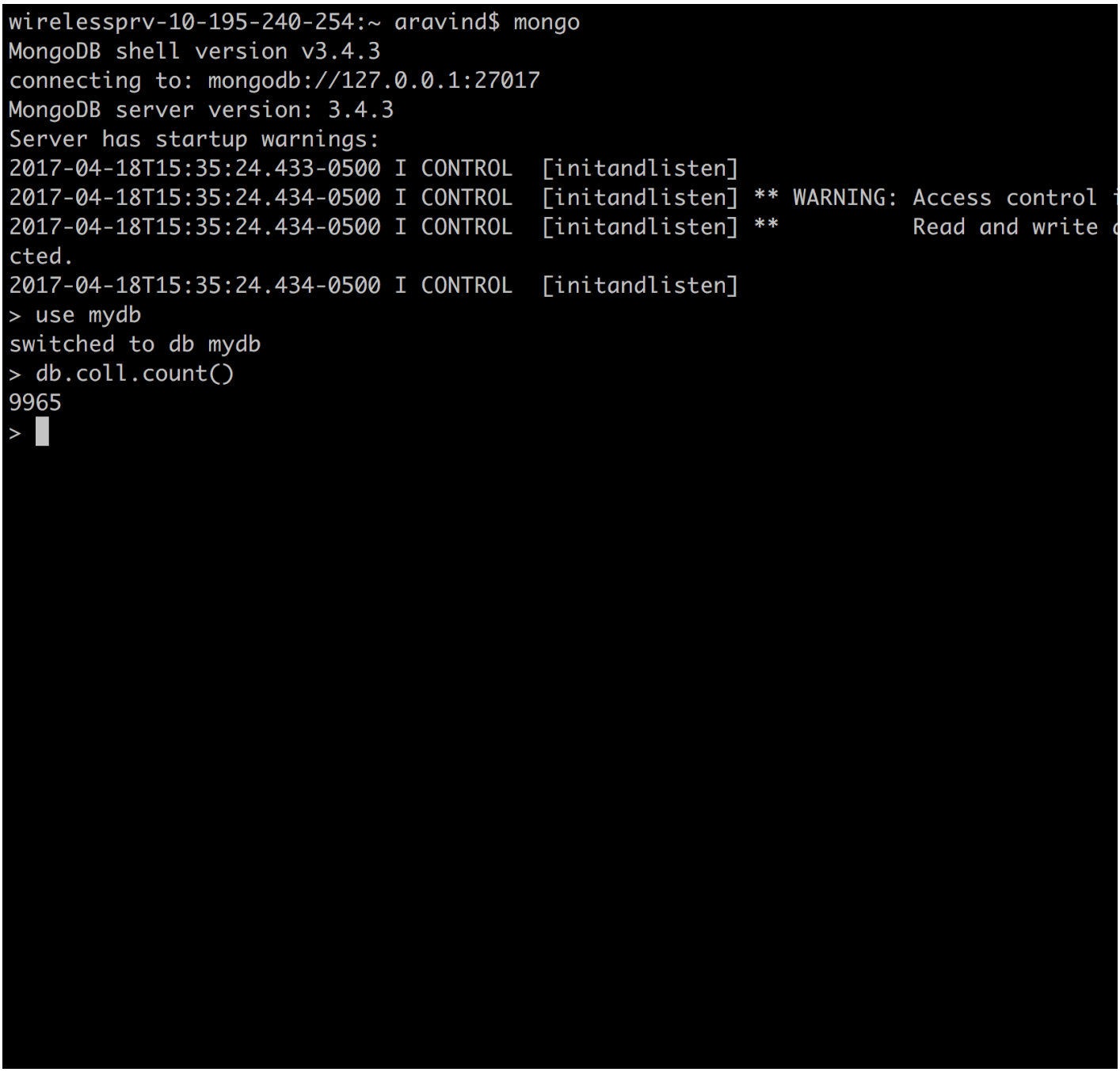
- Cluster 1 has more positive sentiment than cluster 7 as observed in the screenshot, and the neutral sentiment dominates both clusters as expected (since most tweets do not have a strong sentiment associated with them).



Screenshot 1) Results from our clustering and sentiment analysis tasks.

2) The results of running db.yourcollection.count() in MongoDB after running the Kafka and Spark tasks together

Screenshot 2) shows the results of running count on MongoDB after executing the spark task. As expected, there are 9965 tweets in the database, same as the Kafka topic spark_input.



Screenshot 2) Count of MongoDB collection after running Kafka and Spark tasks

3) the results of running db.yourcollection.findOne() in MongoDB after running the Kafka and Spark tasks together

Since the tweet is too large, only a part of it can be seen in the screenshot displayed below (Screenshot 3).


```
> db.coll.findOne()
{
        "_id" : ObjectId("58f413d978b7931539f9c681"),
        "created_at" : "Sat Mar 18 20:43:24 +0000 2017",
        "entities" : {
                "hashtags" : [
                        {
                                "indices" : [
                                        NumberLong(98),
                                        NumberLong(110)
                                ],
                                "text" : "datascience"
                        },
                        {
                                "indices" : [
                                        NumberLong(111),
                                        NumberLong(121)
                                ],
                                "text" : "sqlserver"
                        },
                        {
                                "indices" : [
                                        NumberLong(122),
                                        NumberLong(129)
                                ],
                                "text" : "RStats"
                        }
                ],
                "symbols" : [ ],
                "urls" : [
                        {
                                "display_url" : "bit.ly/2nDdWku",
                                "expanded_url" : "http://bit.ly/2nDdWku",
                                "indices" : [
                                        NumberLong(74),
                                        NumberLong(97)
                                ],
                                "url" : "https://t.co/h5HBbhFOBB"
                        }
                ],
                "user_mentions" : [
                        {
                                "id" : NumberLong(57372793),
                                "id_str" : "57372793",
                                "indices" : [
                                        NumberLong(3),
                                        NumberLong(17)
                                ],
                                "name" : "Niels Berglund",
                                "screen_name" : "nielsberglund"
                        }
                ]
        },
```