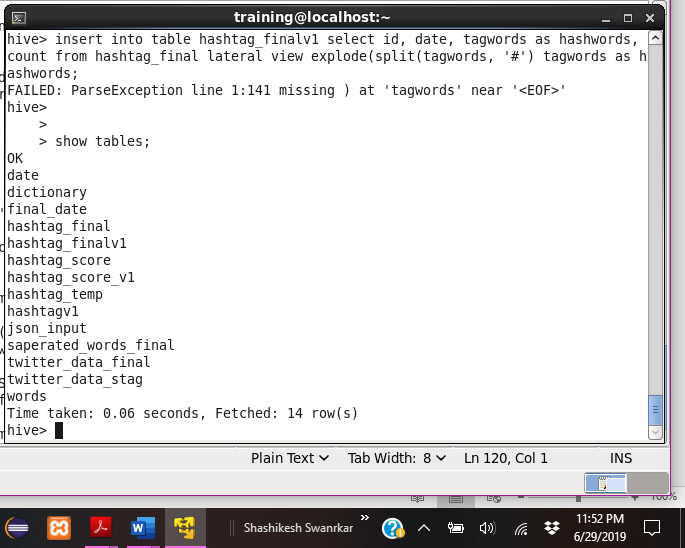
The purpose of below project is to understand the sentiments of the tweets. The data is collected from twitter and is in .json format. In order to carry out the analysis process used tools like Sqoop and Hive. The data is exported to Hadoop filesystem and then converting it into hive tables.

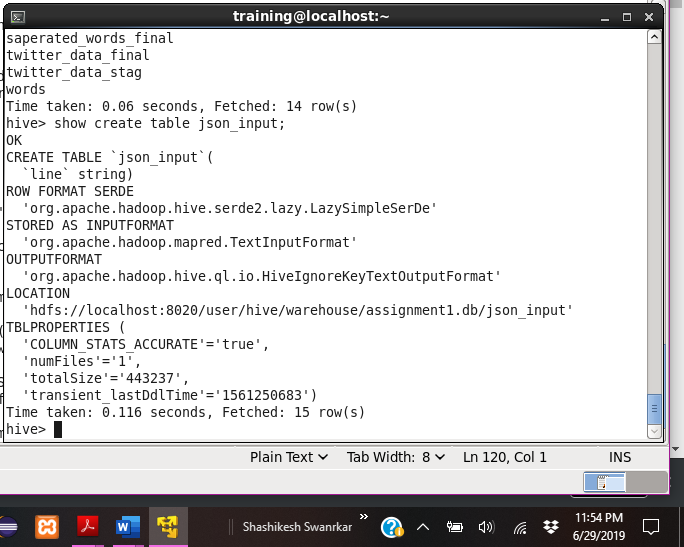
**Assignment 1**

**Solutions:**

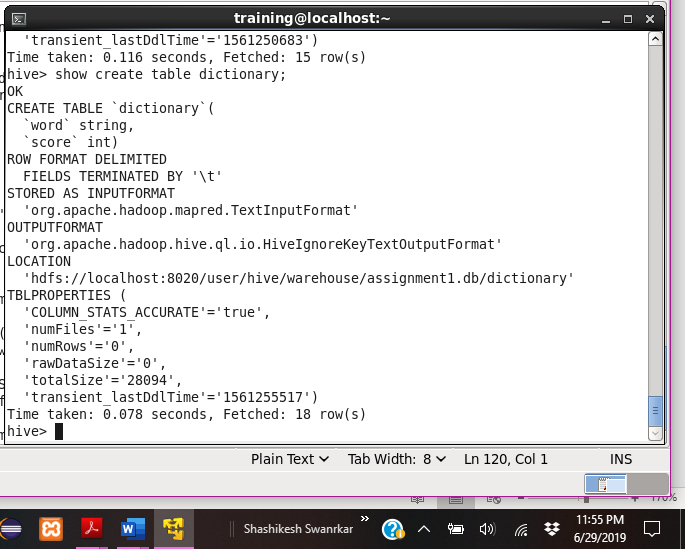
In order to carry out the analysis created several tables to break and process the data.



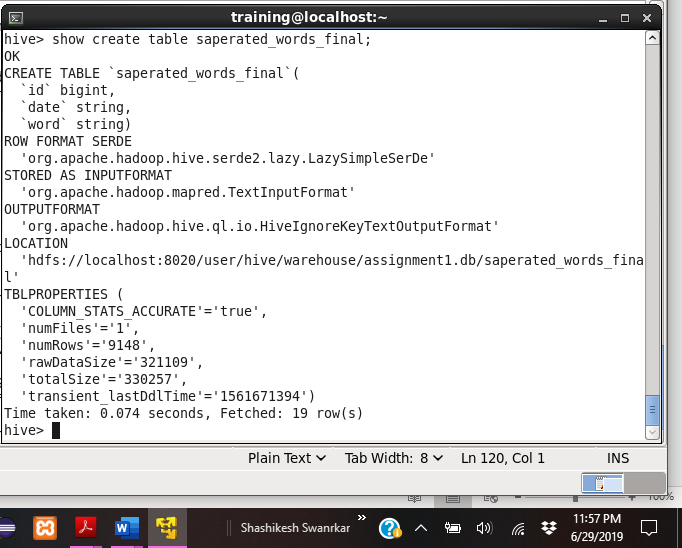
Json\_input – Below table is used to keep the json file into table.



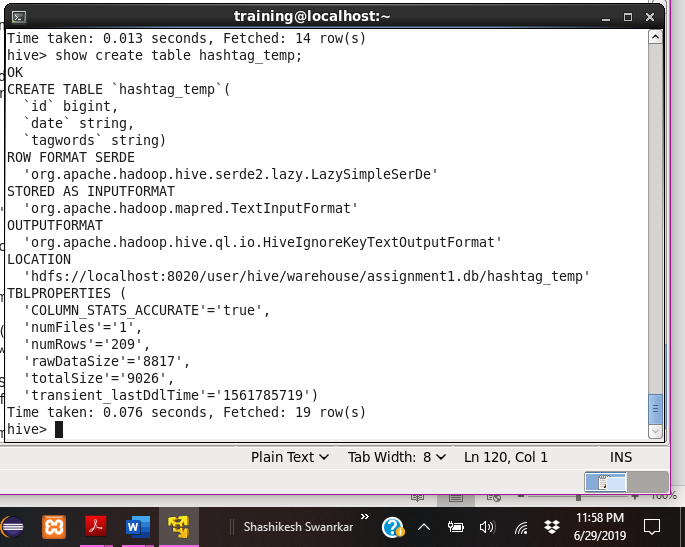
Dictionary – Below table in which the words scores are stored.



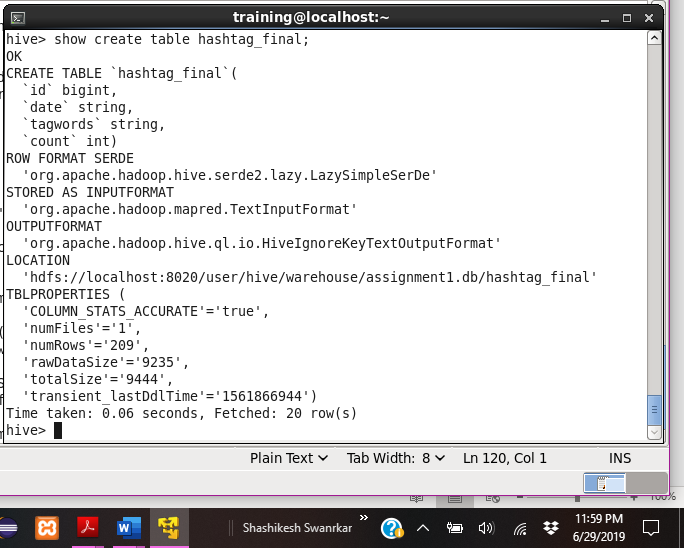
Saperated\_words\_final – Below table contains the words of each tweets separated.



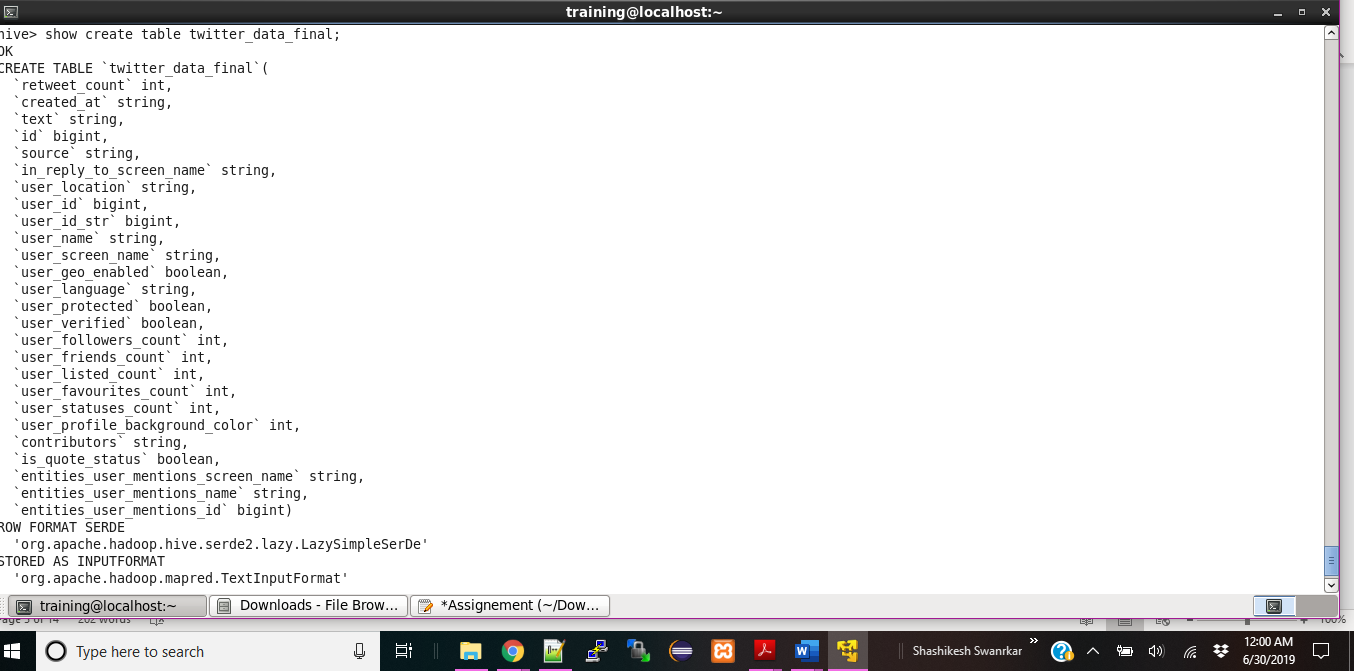
Hashtag\_temp – Below table contains partially cleaned hashtags.



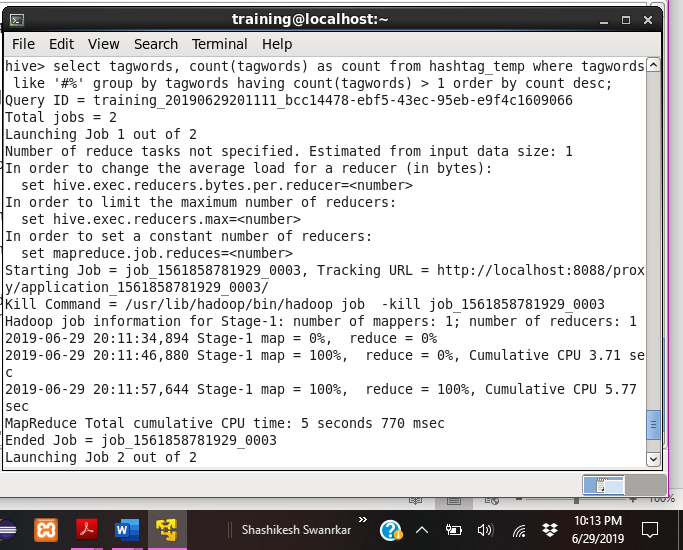
Hashtag\_final – Below table has completely cleaned hashtags.

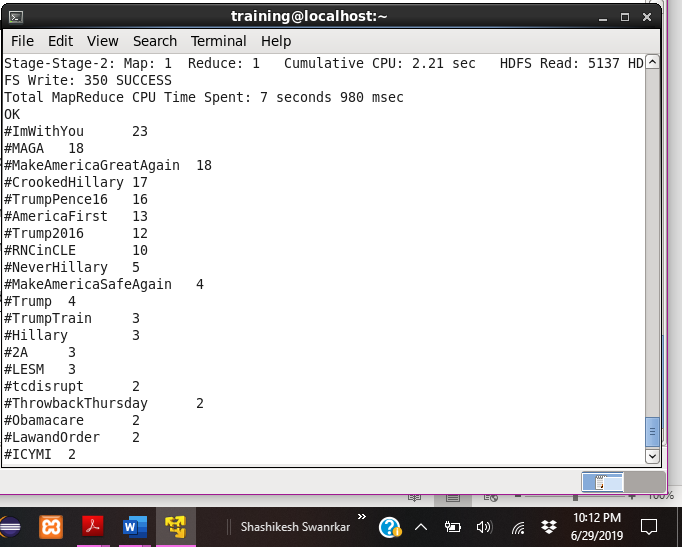


Twitter\_data\_final – Below is the query to break the the json data into columns.

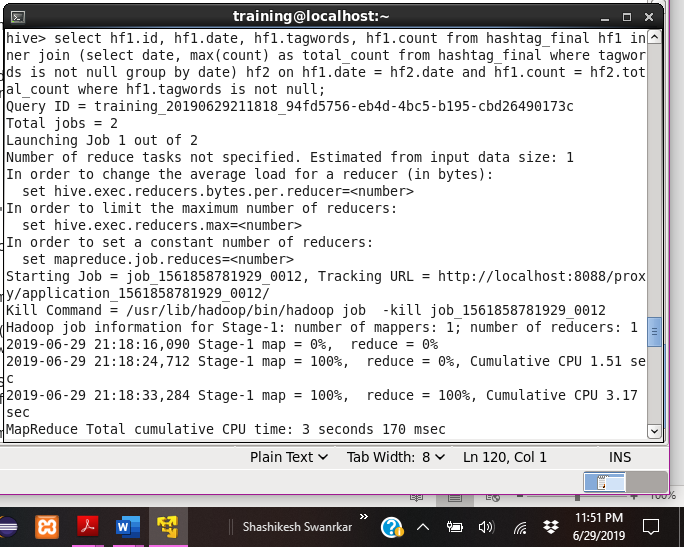


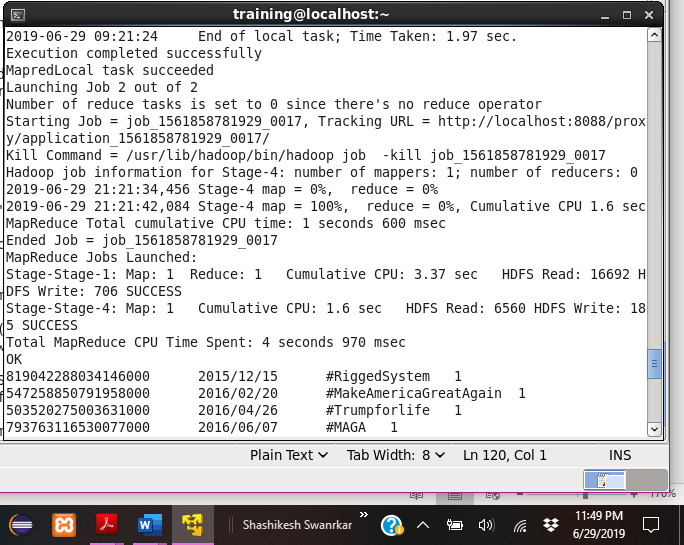
* Below is the screenshot of the query and output for hashtags used and how many times each are used.



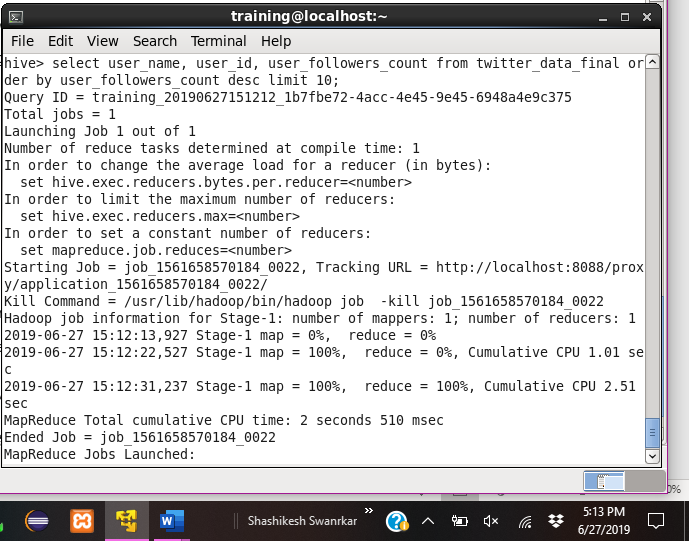


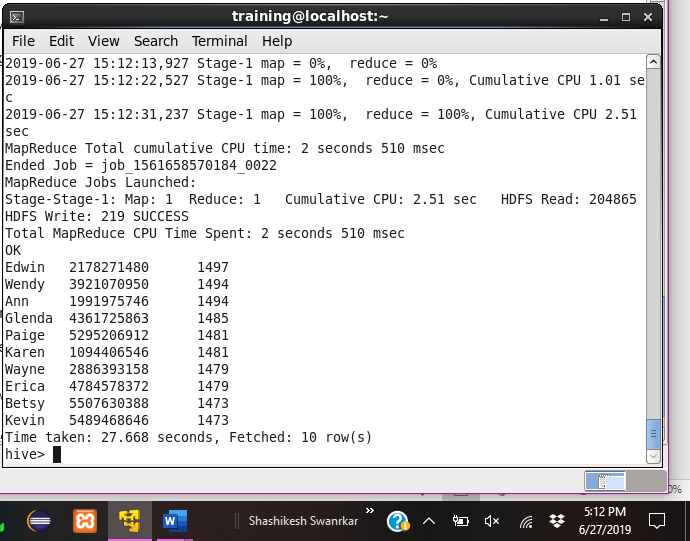
* Below is the screenshot of the query and output for the most trending hashtag in a day and how many times is it tweeted





* Below is the screenshot of the query and output the user’s followers count, who are the top ten users who have tweeted





Below is the screenshot of the query and output for the score for each tweet that was posted and Does the tweet have a positive or negative sentiment.

