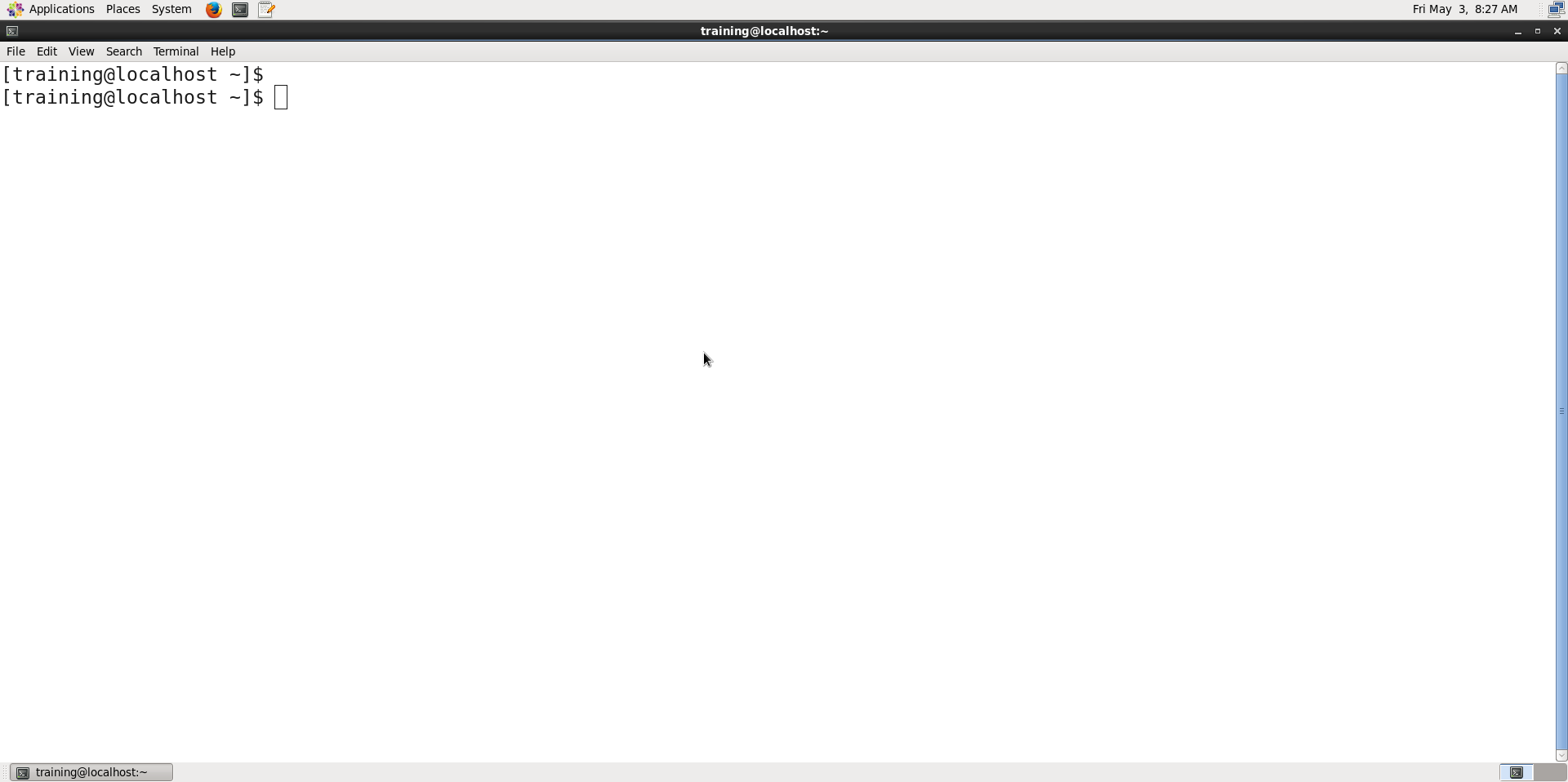
**Nisha Iyer: Big Data – Case Study**

Sentiment analysis on tweets regarding Demonetization in India.

Sentiment Analysis: In this section we will perform a sentiment analysis on tweets regarding Demonetization in India.

**Part 1:**

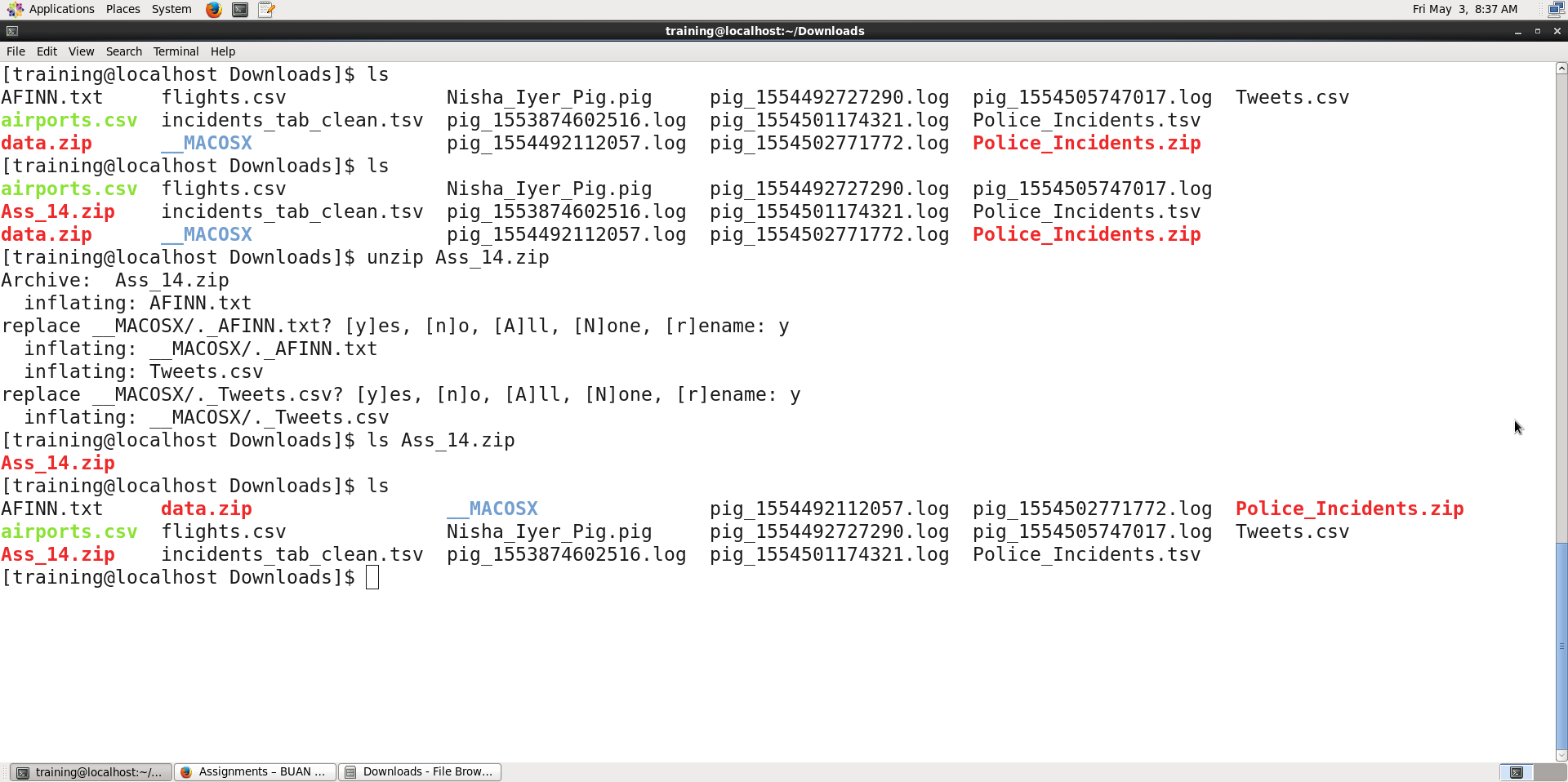
**Step 1:**



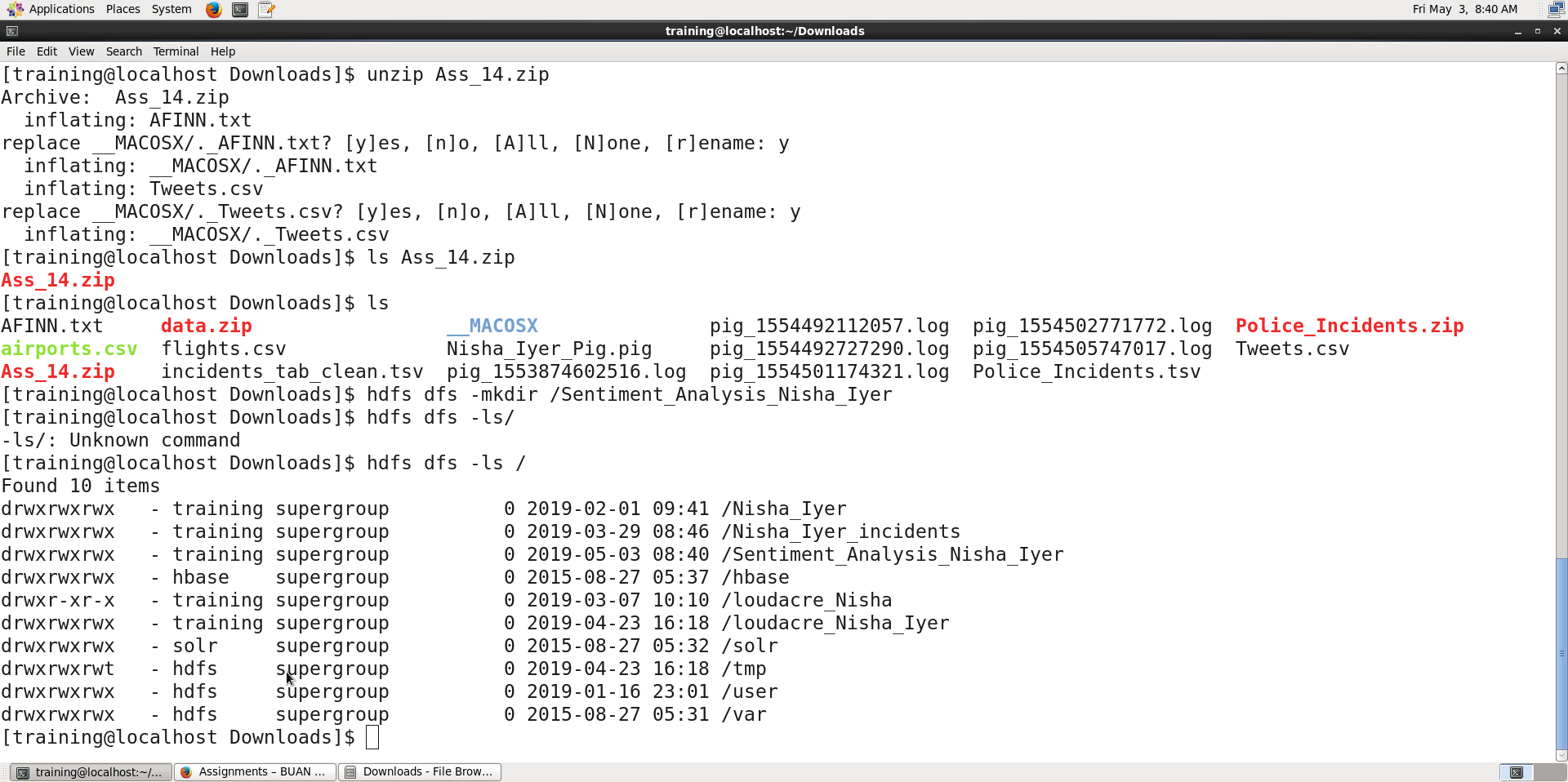
**Step 3:**



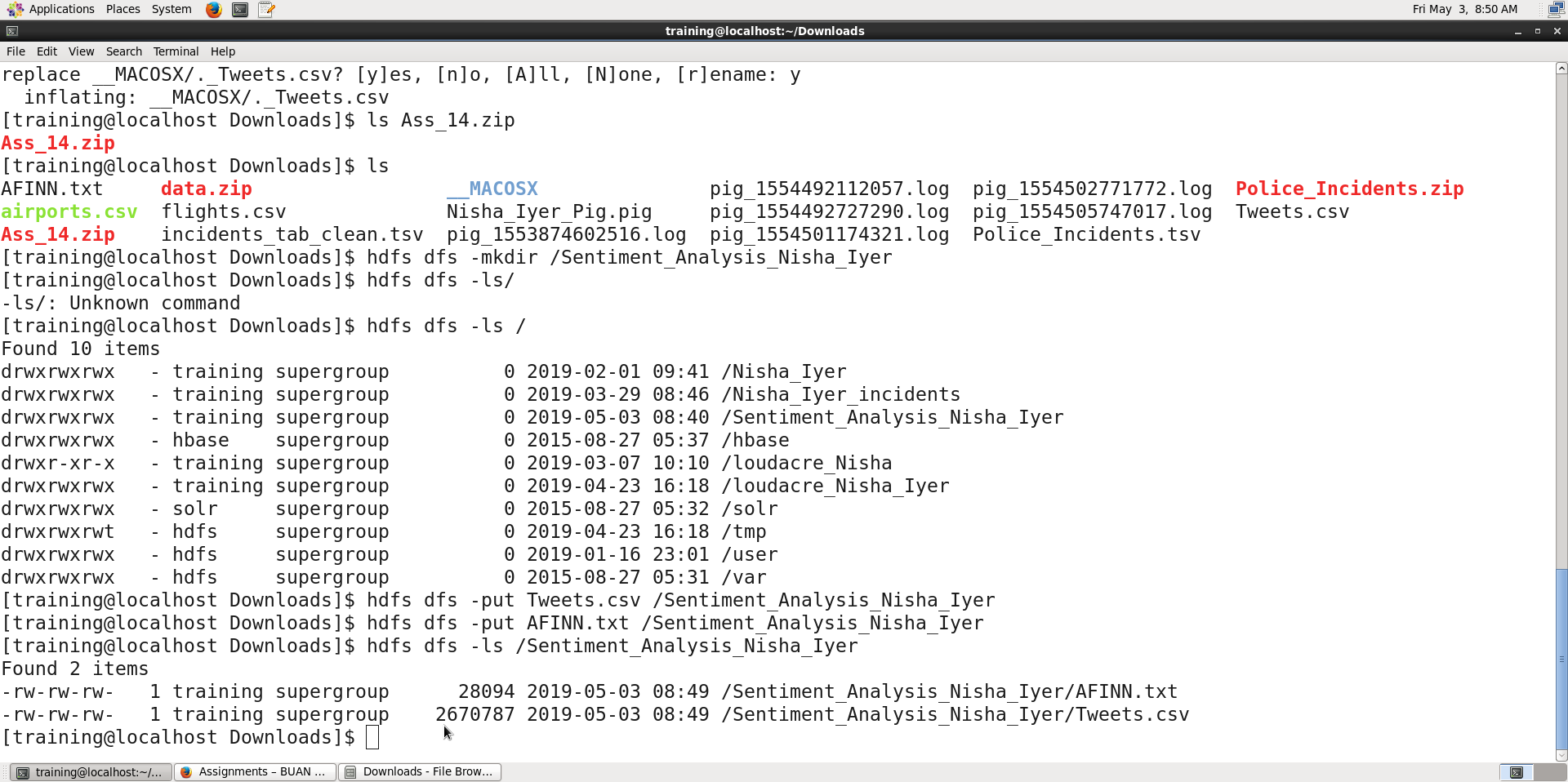
**Step 4:**



**Step 5:**

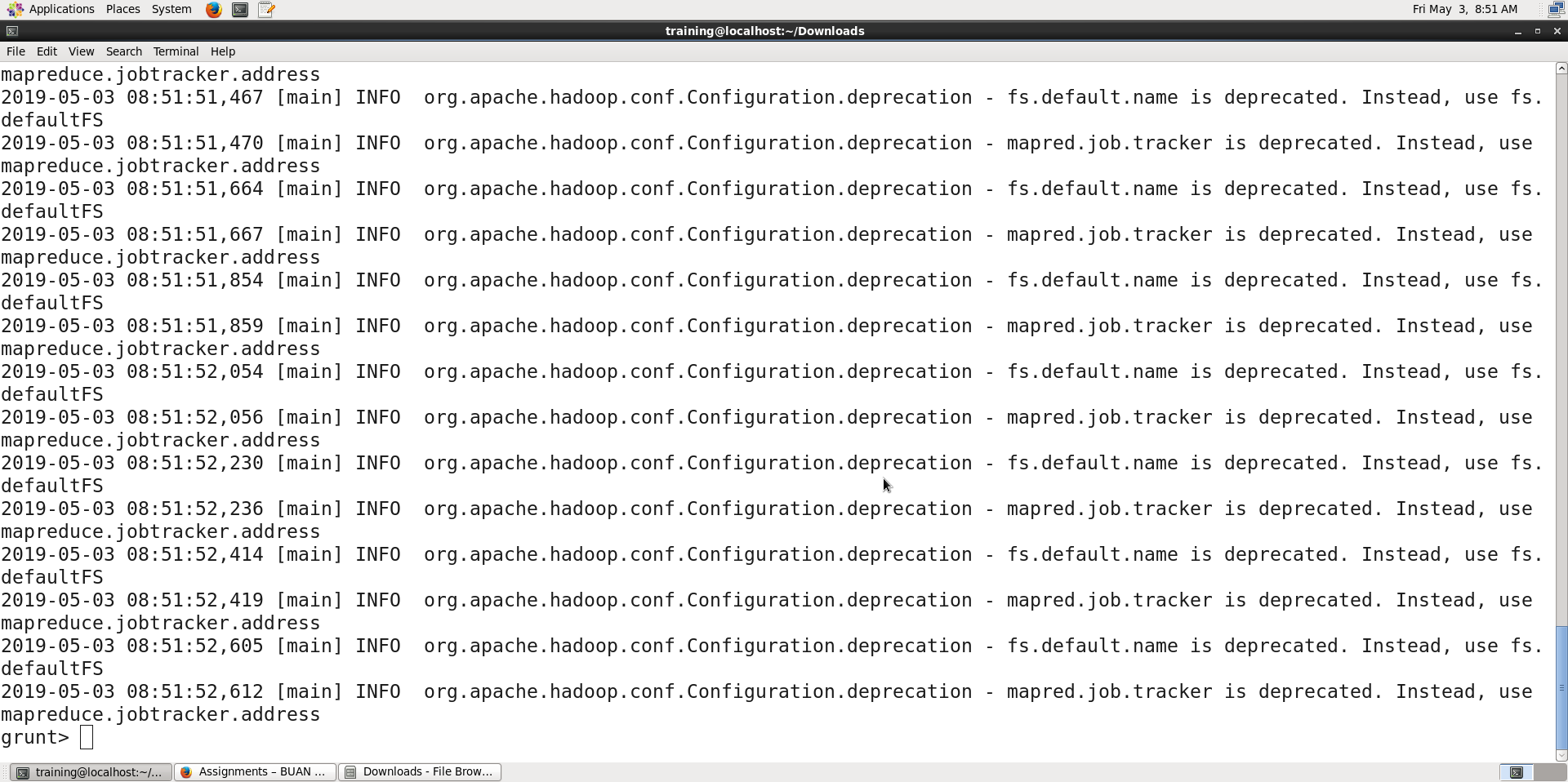


**Step 6:**



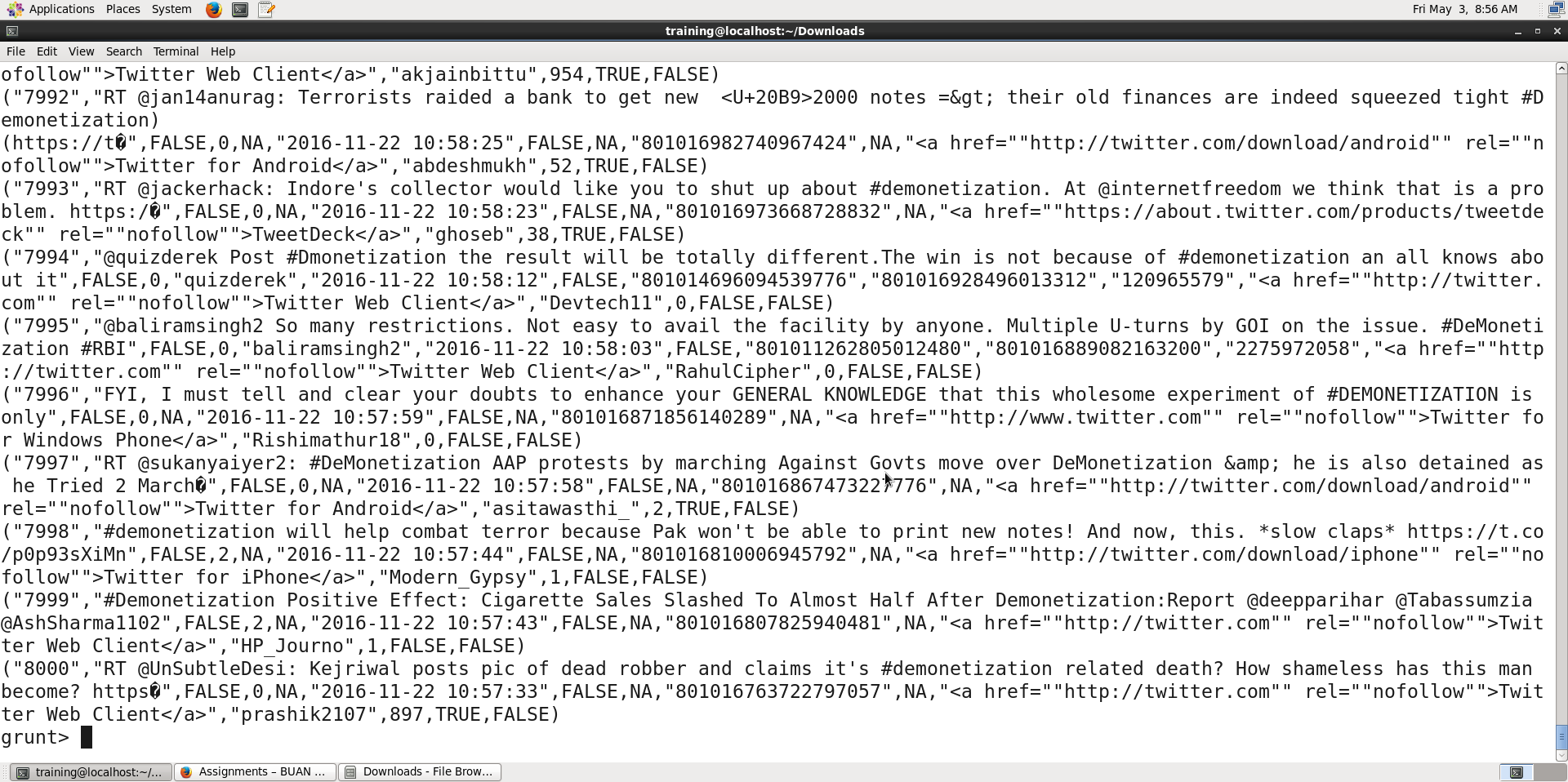
**Part 2:**

**Step 7:**



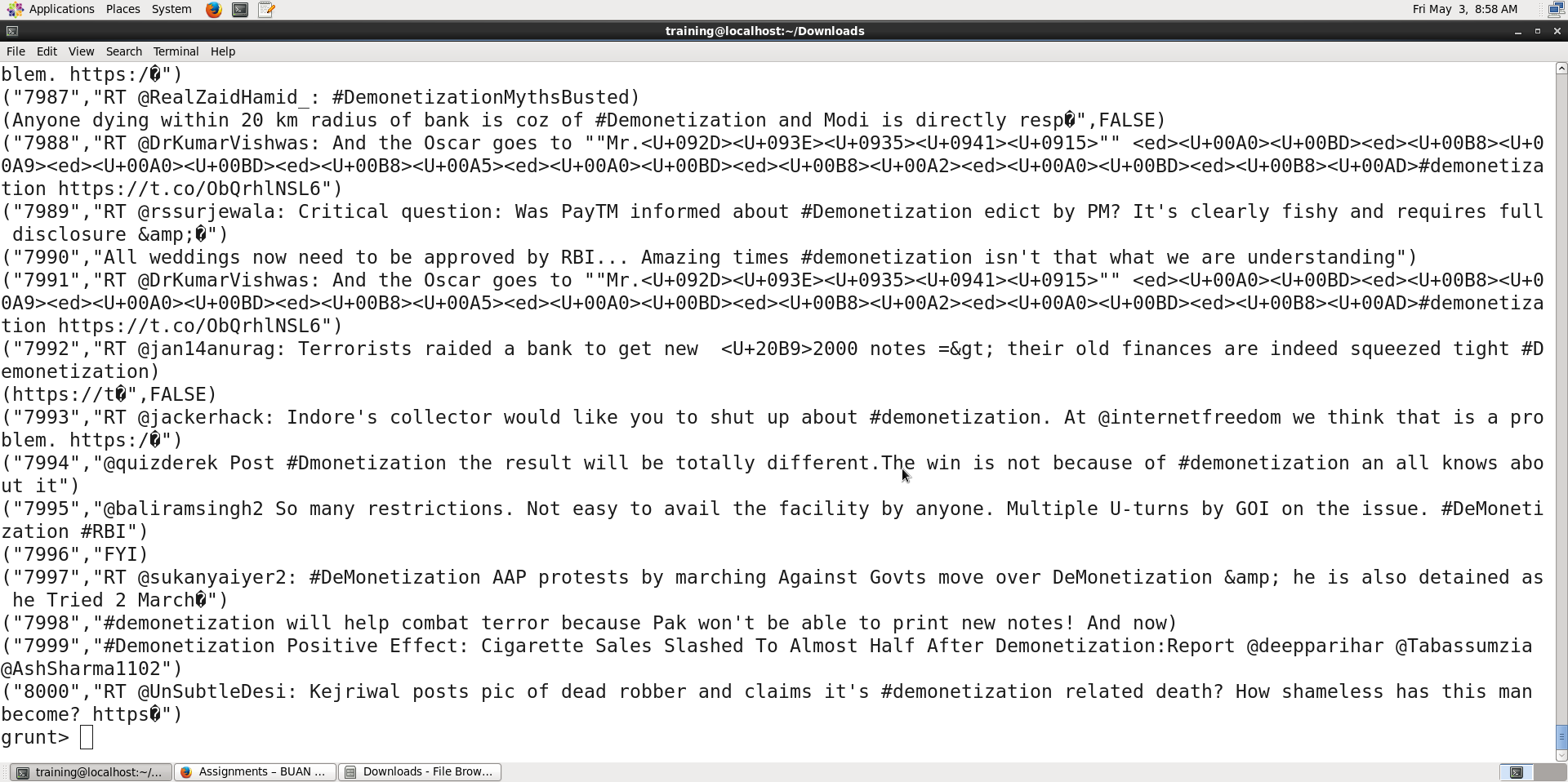
Explanation: The above command invokes the grunt shell. The grunt shell is an interactive shell to execute pig commands & scripts. The command above can be followed by various additional configurations but for our use case the stand grunt shell will suffice.

**Step 8:**



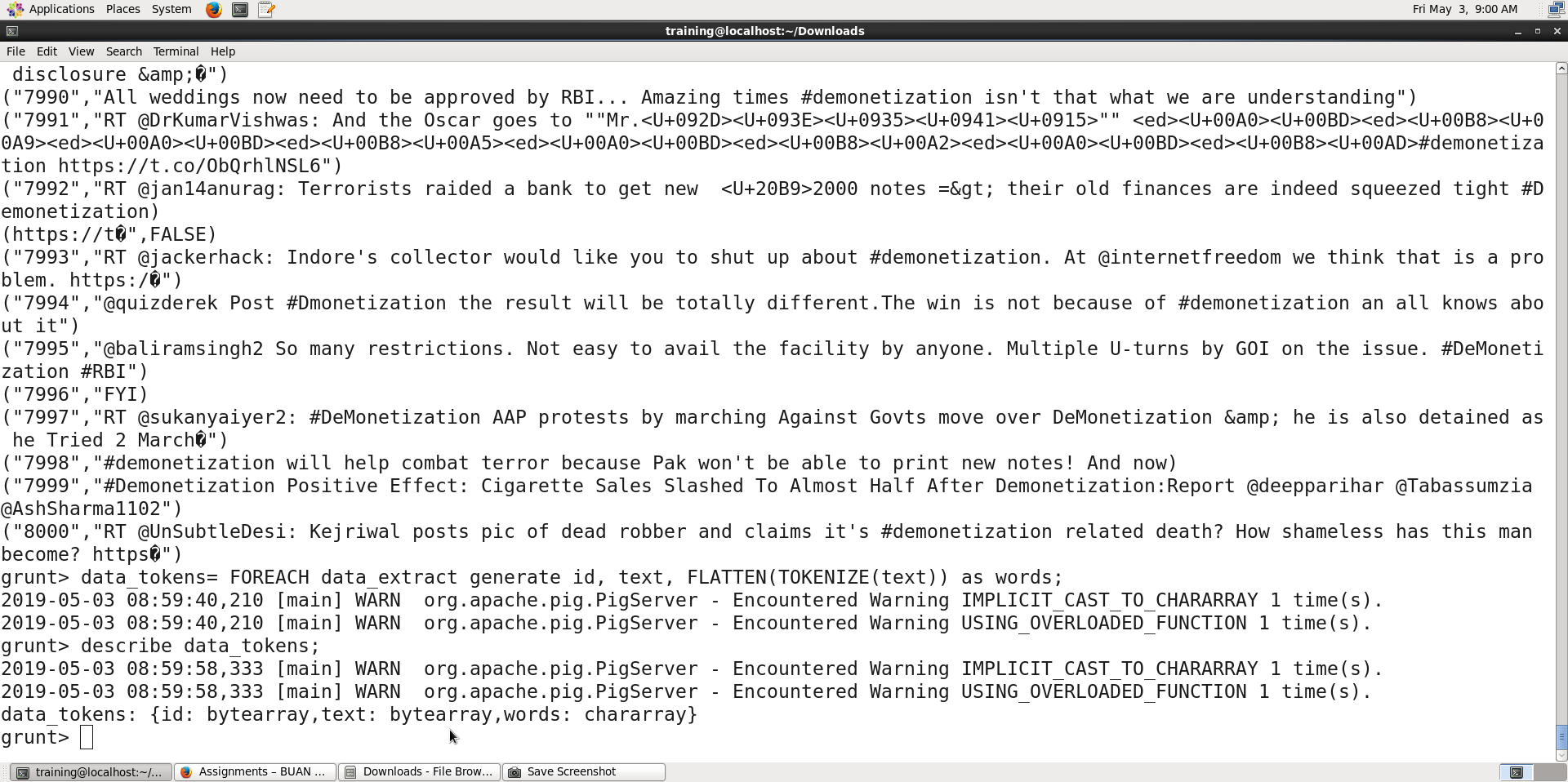
Explanation: The dump command displays the data for the variable specified in the grunt shell.

**Step 9:**



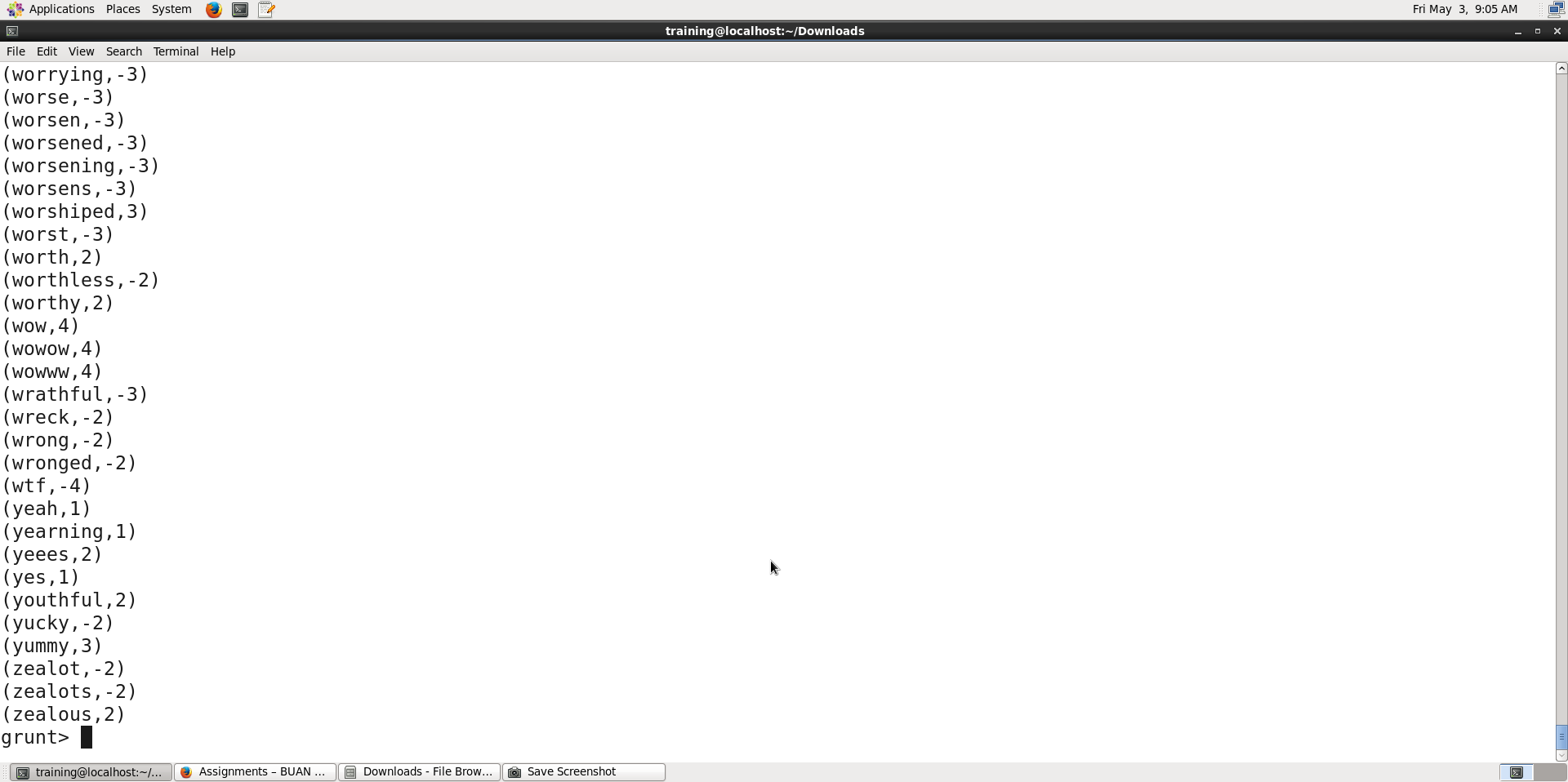
Explanation: We are defining and collecting the first two data elements in the csv file and dumping the data at the grunt shell.

**Step 10:**



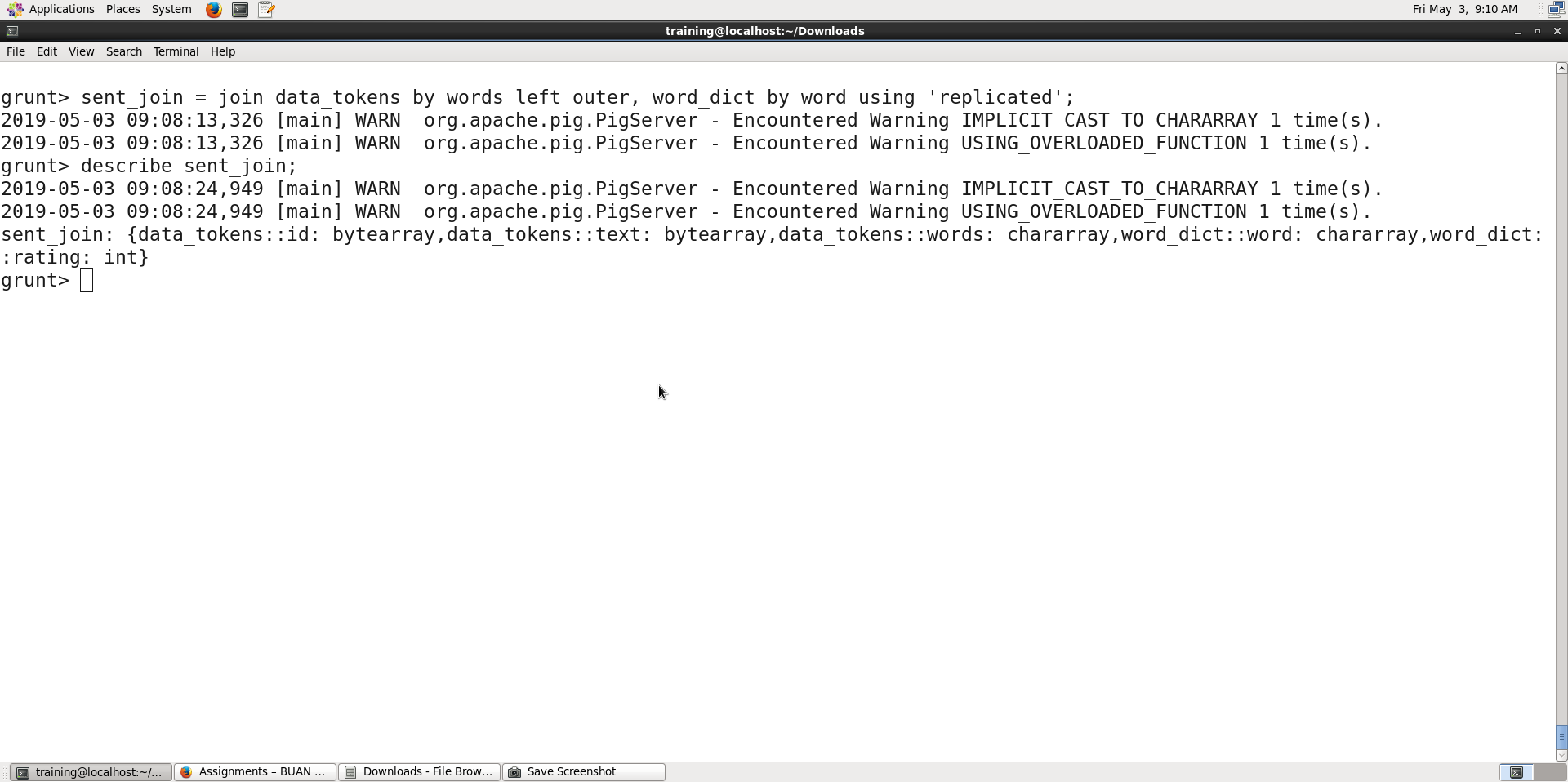
Explanation: We are defining a third element by breaking down the text information in the variable ‘text’ into words. The describe operation shows us the variable names and types of the data in the variable specified.

**Step 11:**



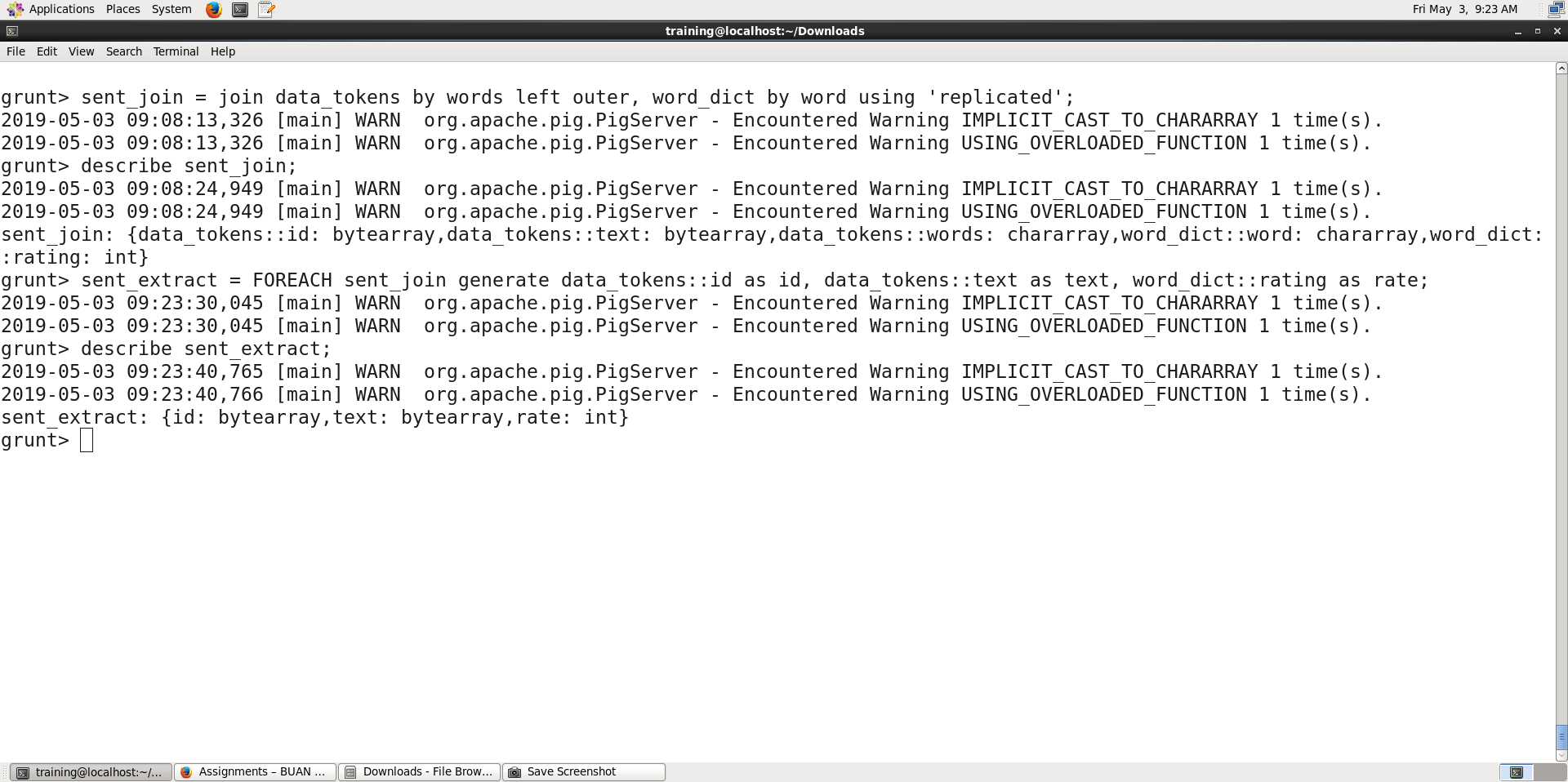
Explanation: AFINN dictionary is a dictionary consisting of 2500 words that are rated on a scale from -5 to +5 depending on the sentiment the word conveys. We will be using this dictionary to perform the sentiment analysis on our tweet data.

**Step 12:**



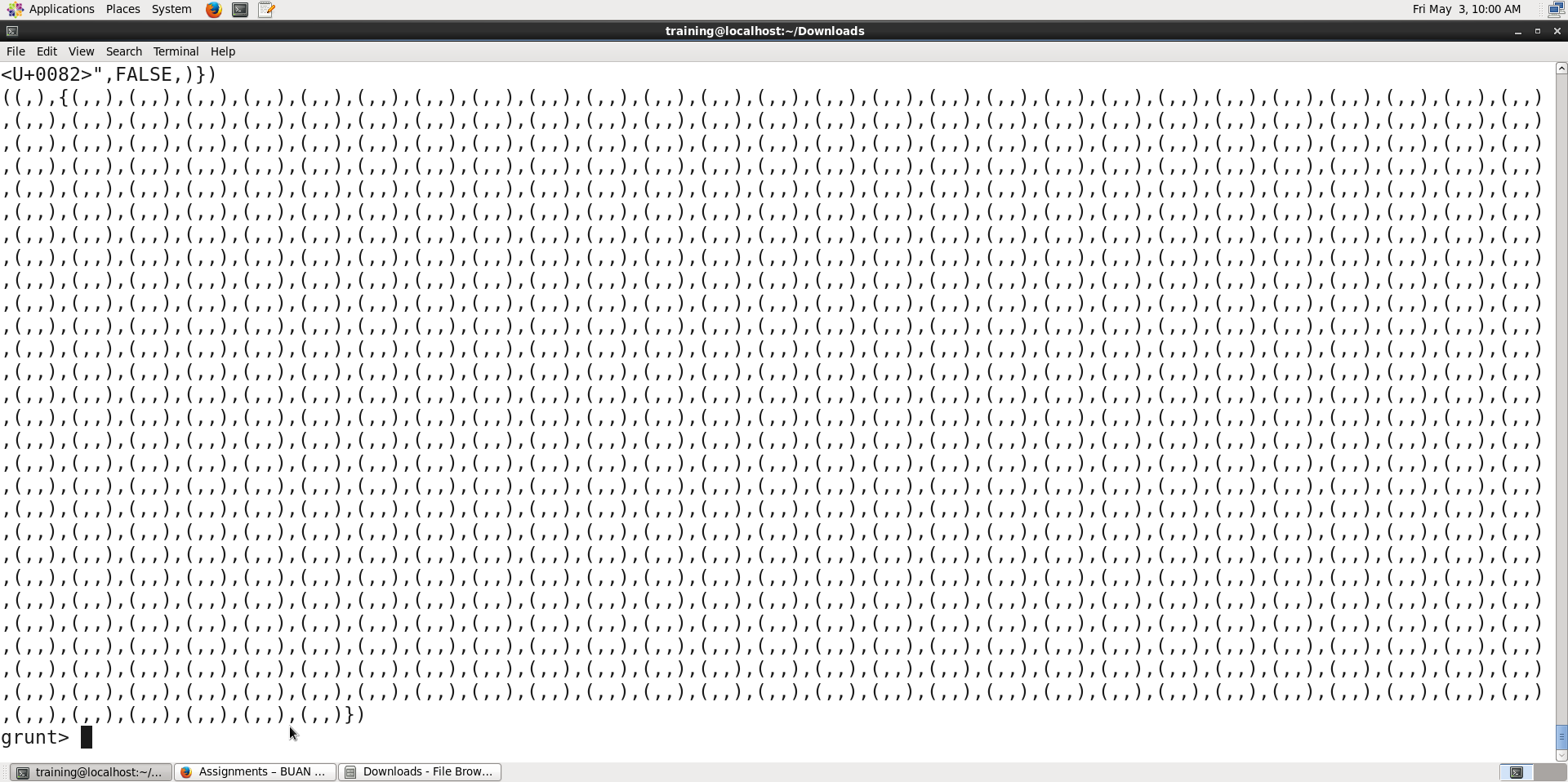
Explanation: We are joining the sentiments values from the AFINN dictionary with the words in the text from each tweet.

**Step 13:**



Explanation: We are now extracting the data from the variable in step 12 that we need to complete our sentiment analysis.

**Step 14:**



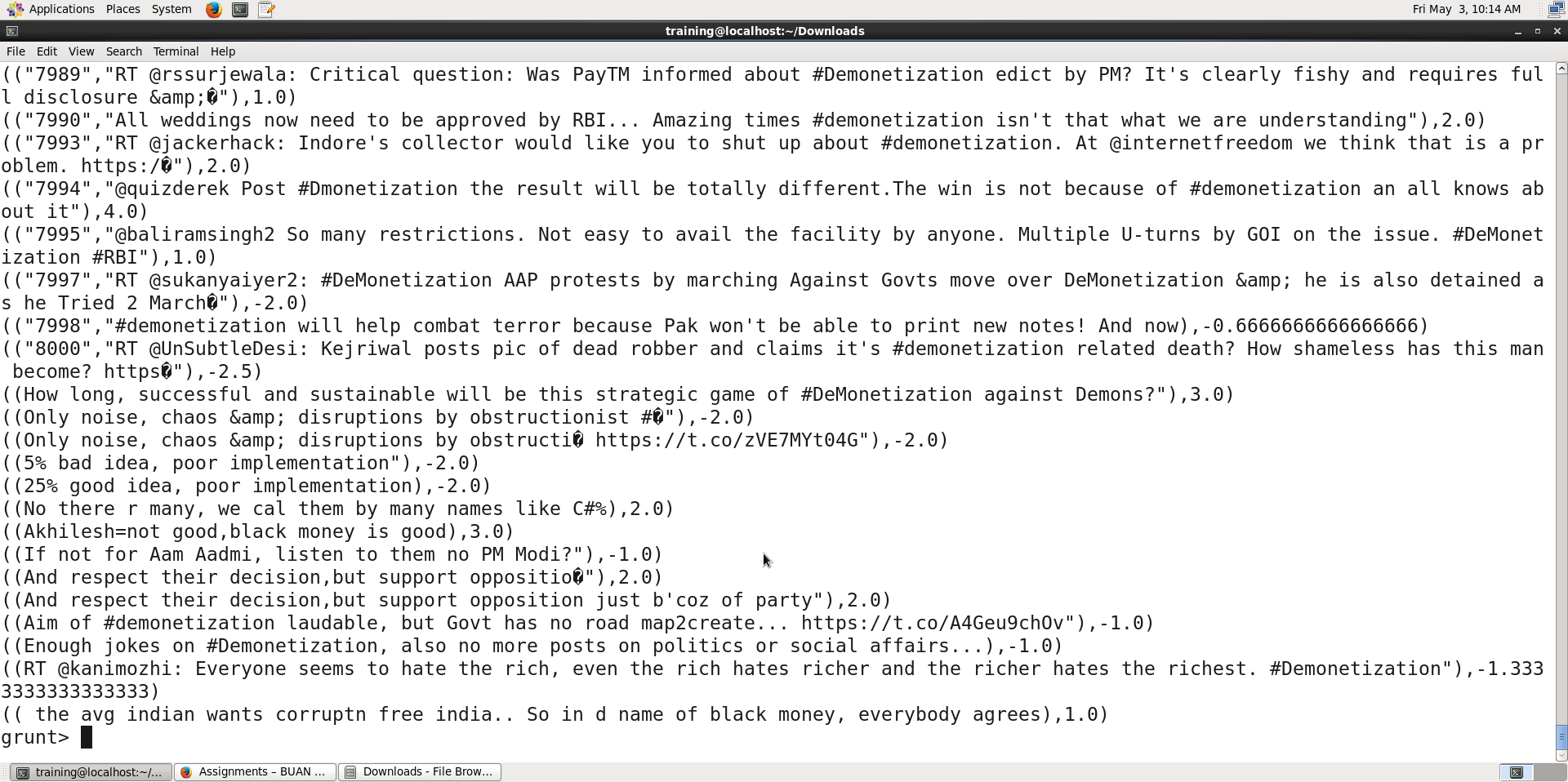
Explanation: We are grouping each record by its id and text.

**Step 15:**



Explanation: We are averaging the rating for all the words in each tweet and storing the result in average\_rating.

**Step 16:**



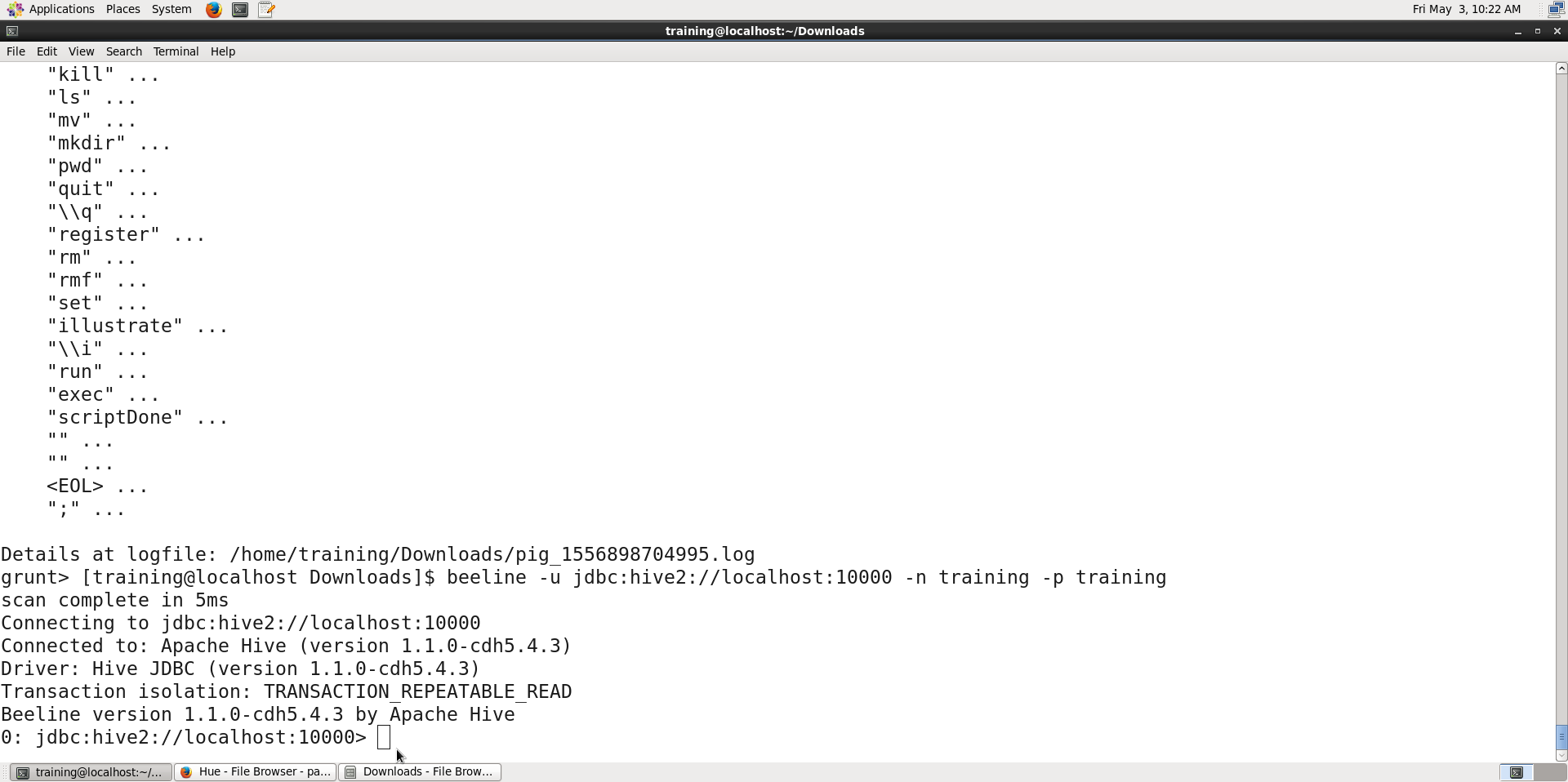
Explanation: We are filtering all the tweets for those tweets that have an average rating >= -5.

**Step 17:**



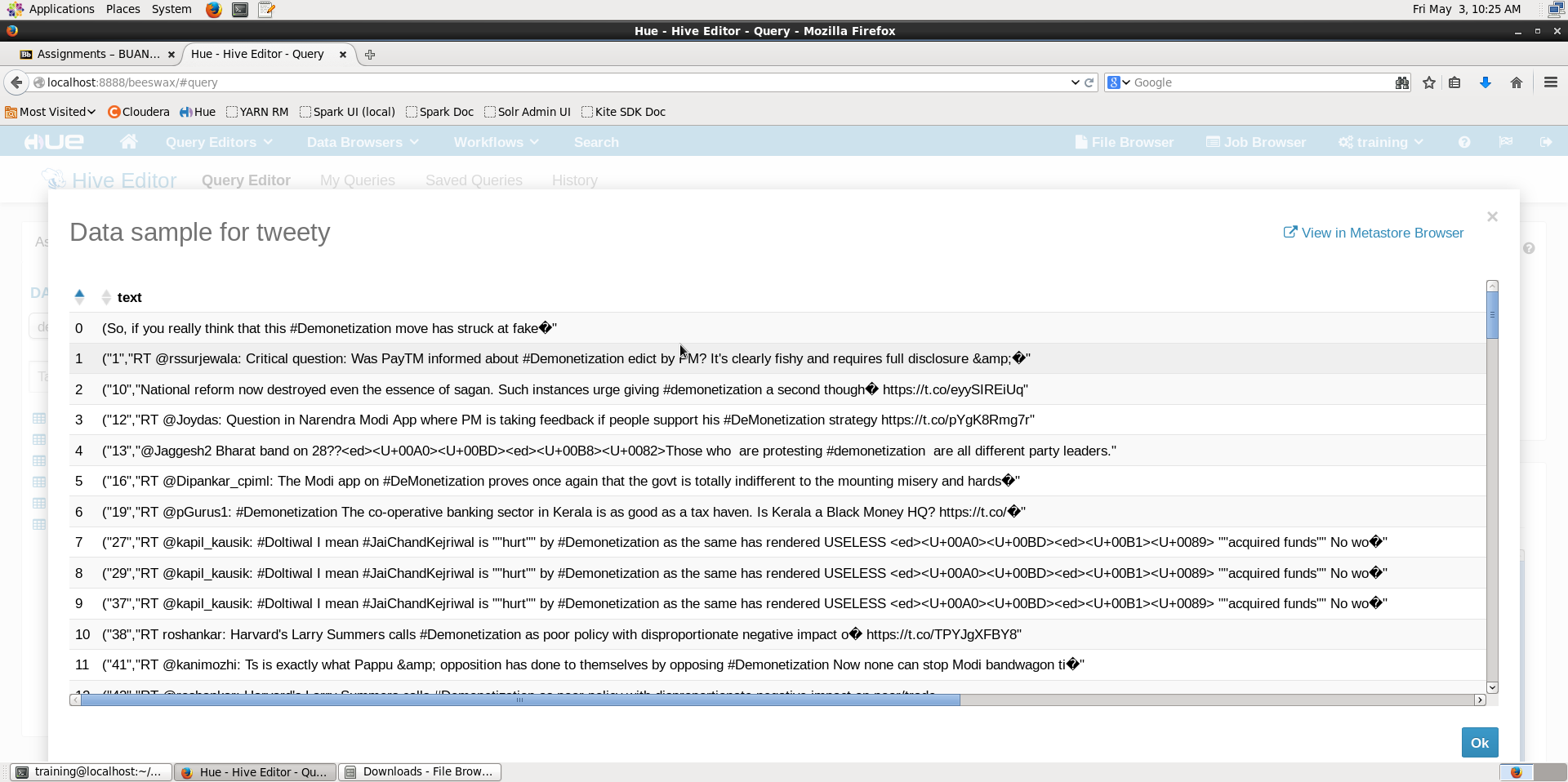
Explanation: We are viewing the output of the Analyzed\_Tweets file in HUE.

**Step 18:**



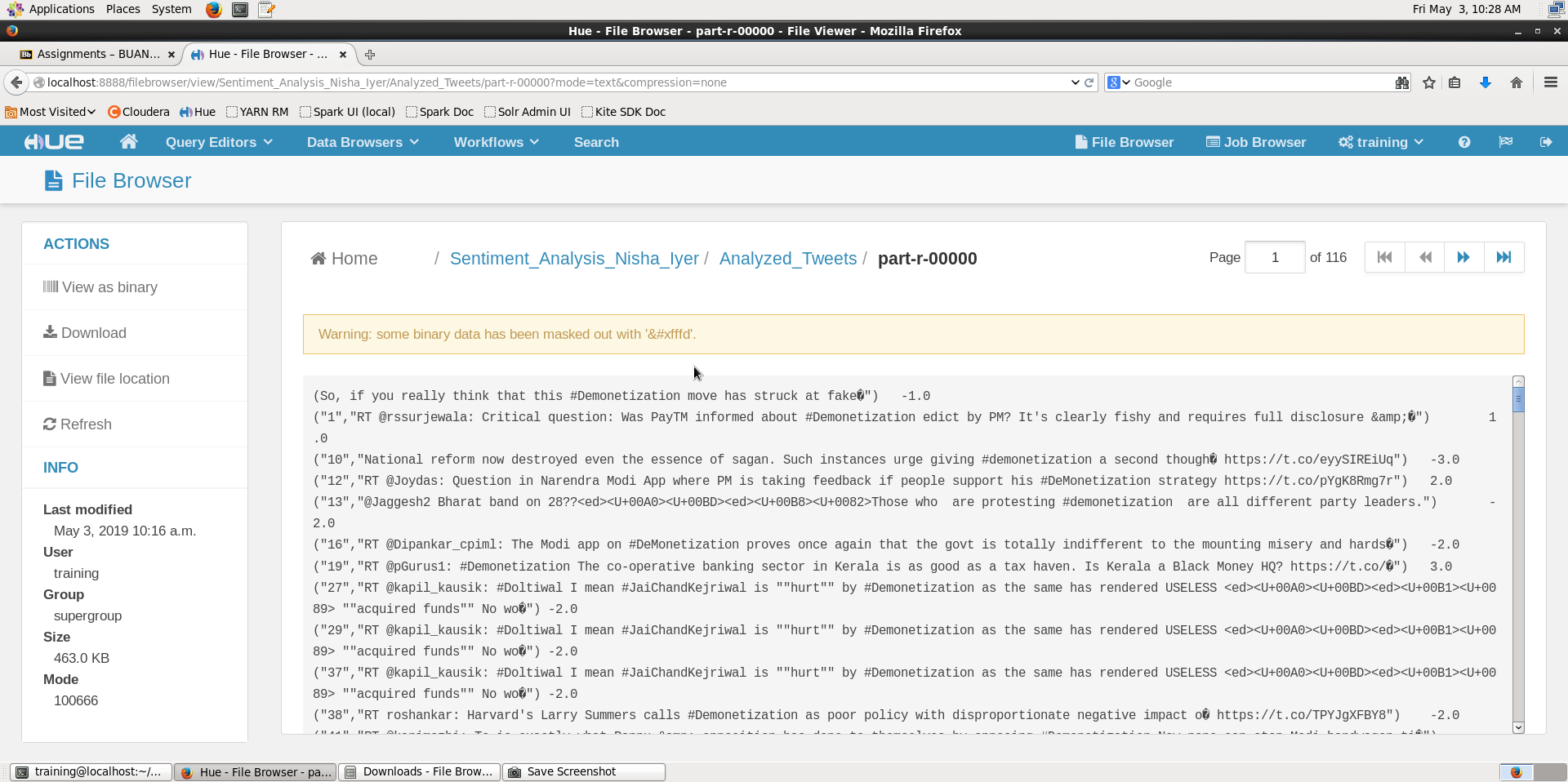
Explanation: We are starting up beeline, to execute the hive commands. Beeline shell is an interactive shell based on the SQLLine utility.

**Step 19:**



Explanation: We are creating a hive table to structure our unstructured data.

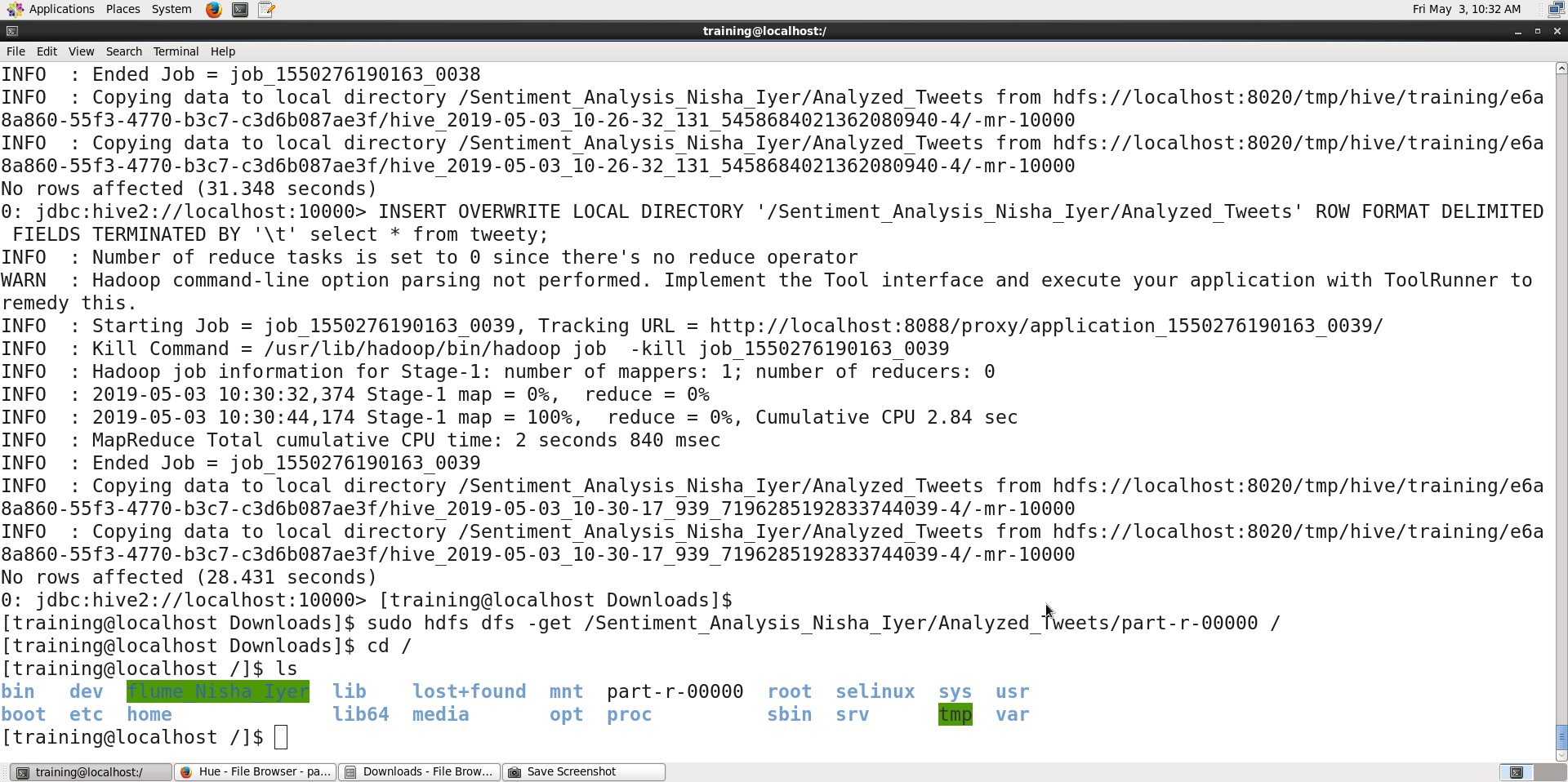
**Step 20:**



Explanation: We are now creating a tsv file from the table that we created in Step-20.

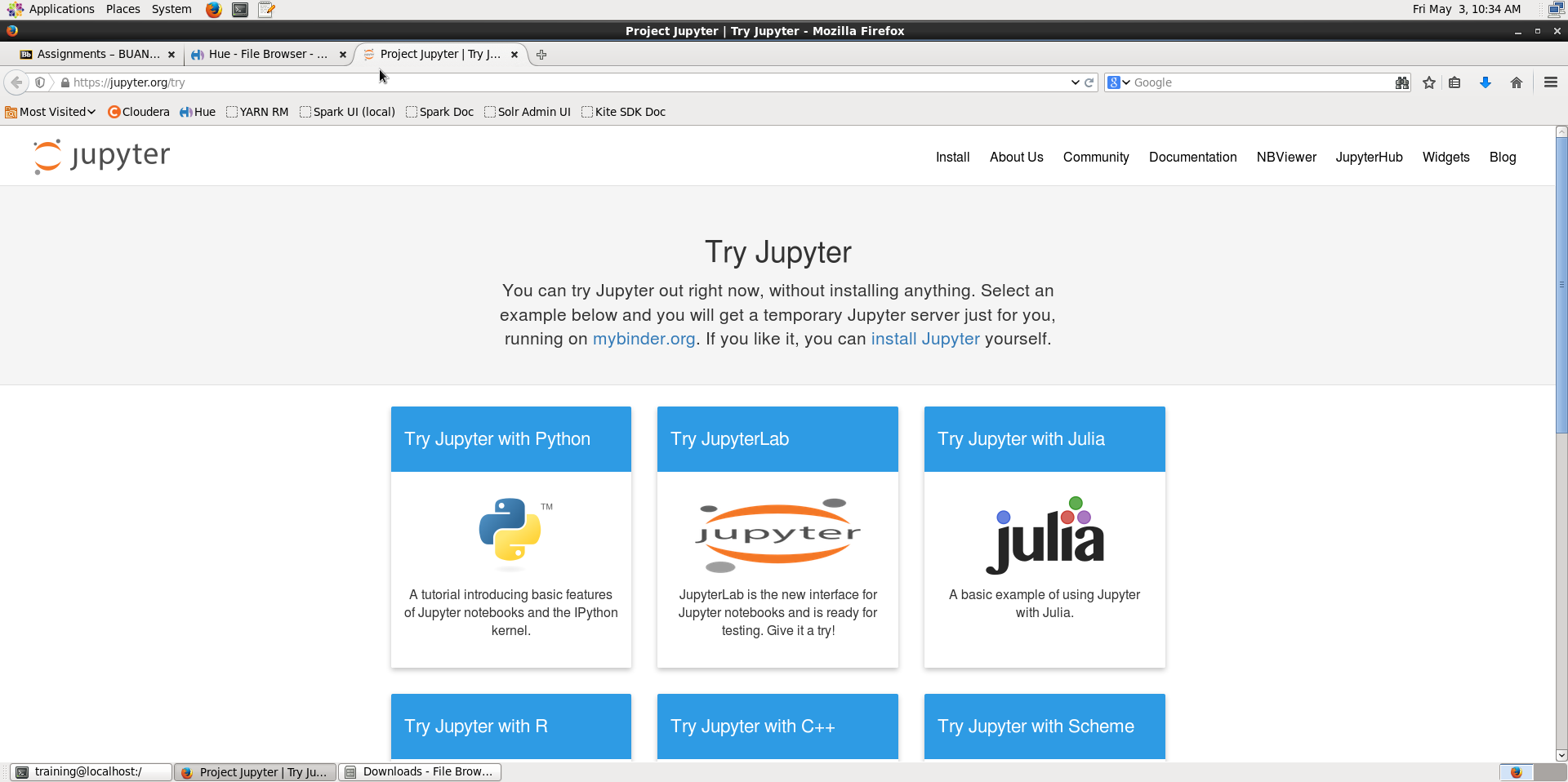
**Part 3:**

**Step 21:**

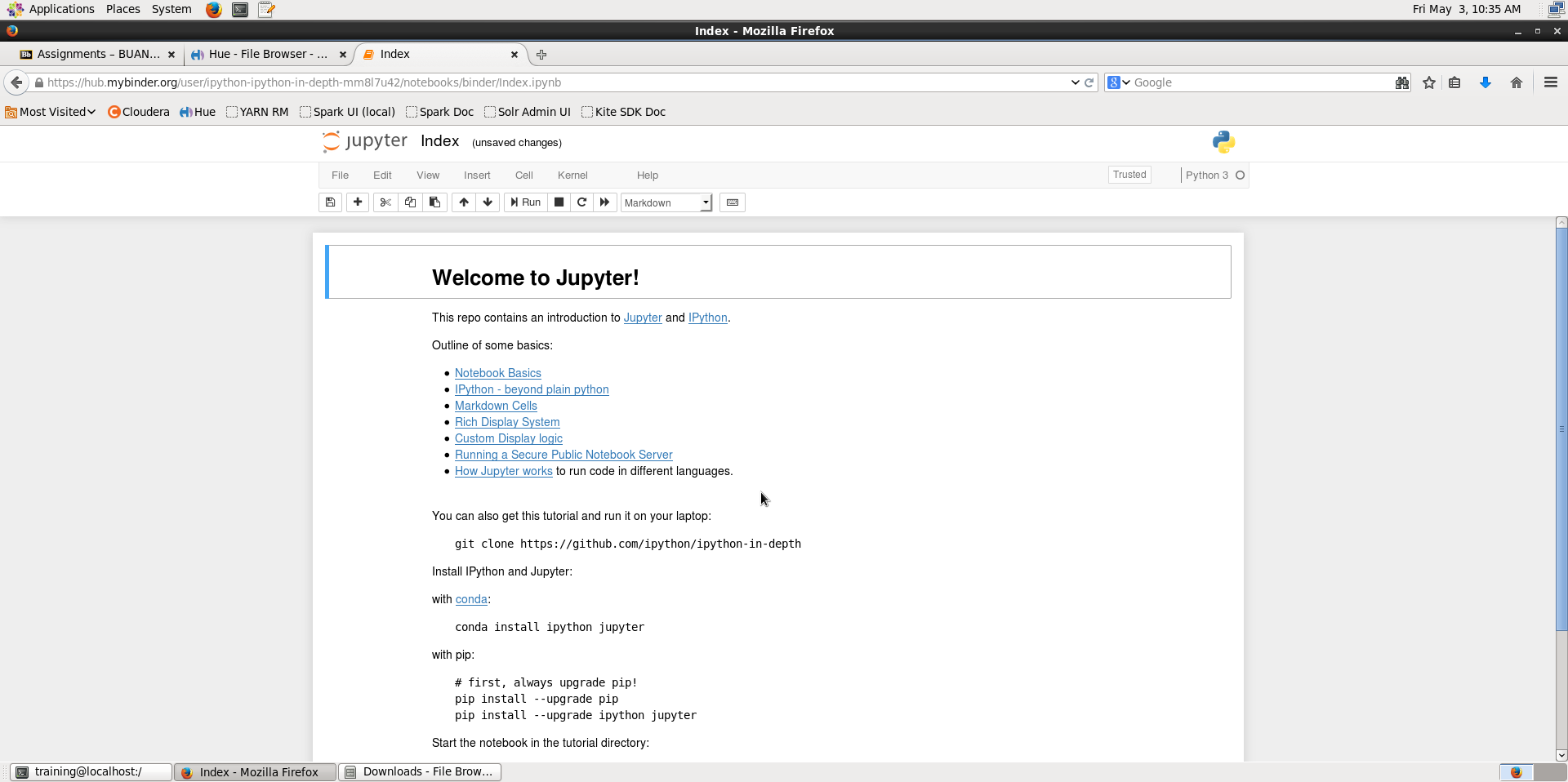


Explanation: We are copying the results of our sentiment analysis back on the local file system to visualize.

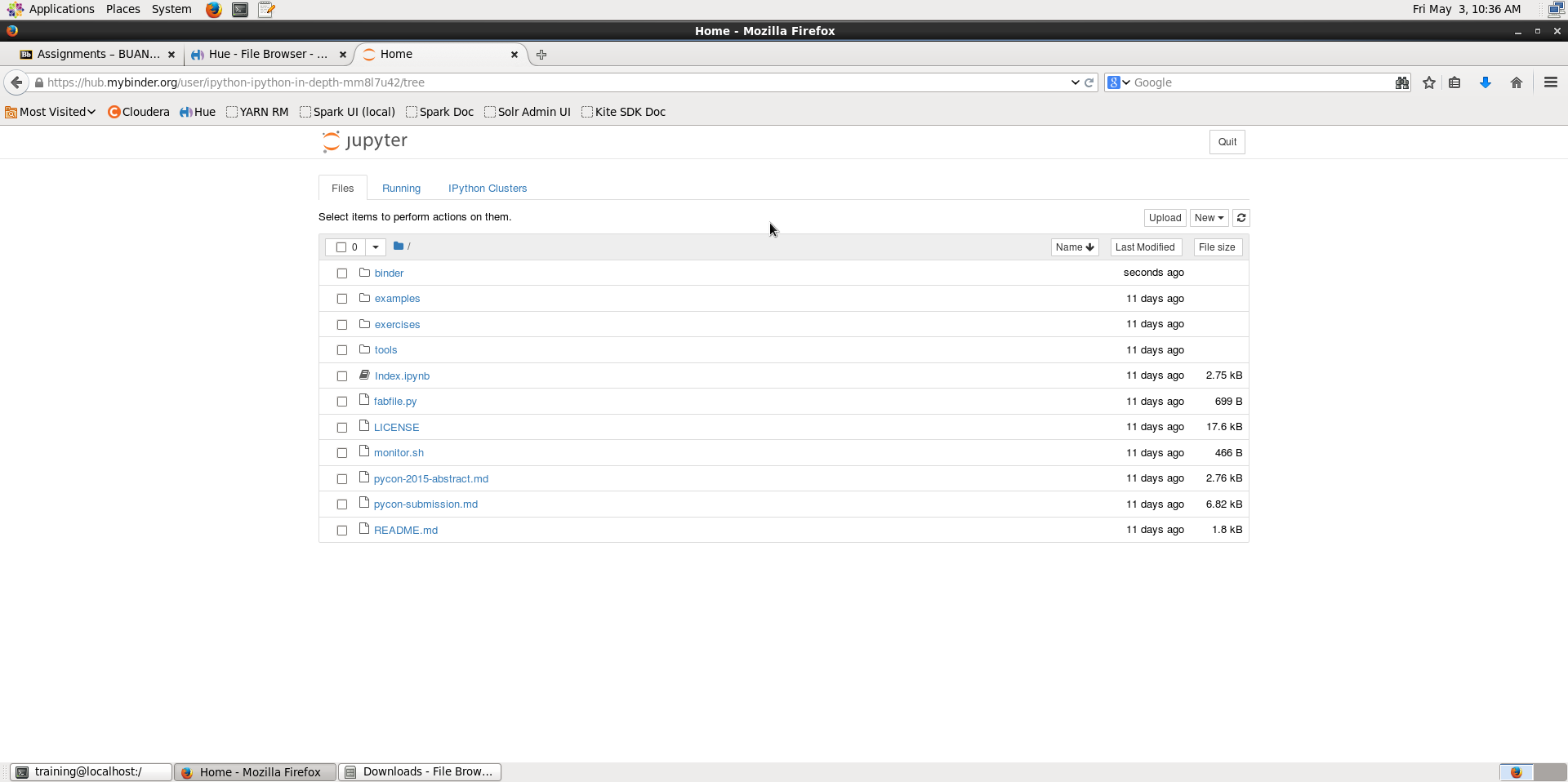
**Step 22:**



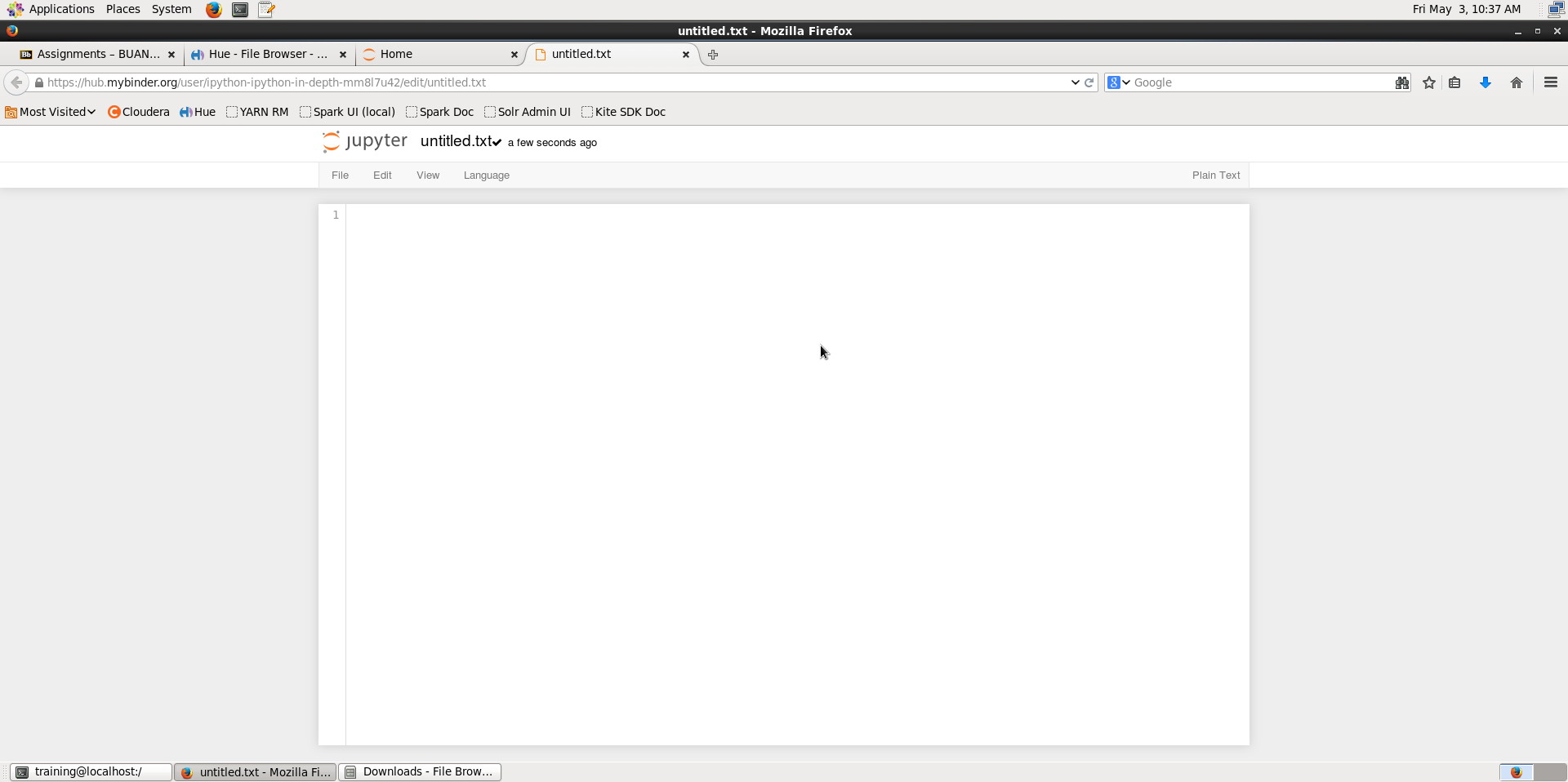
**Step 23:**



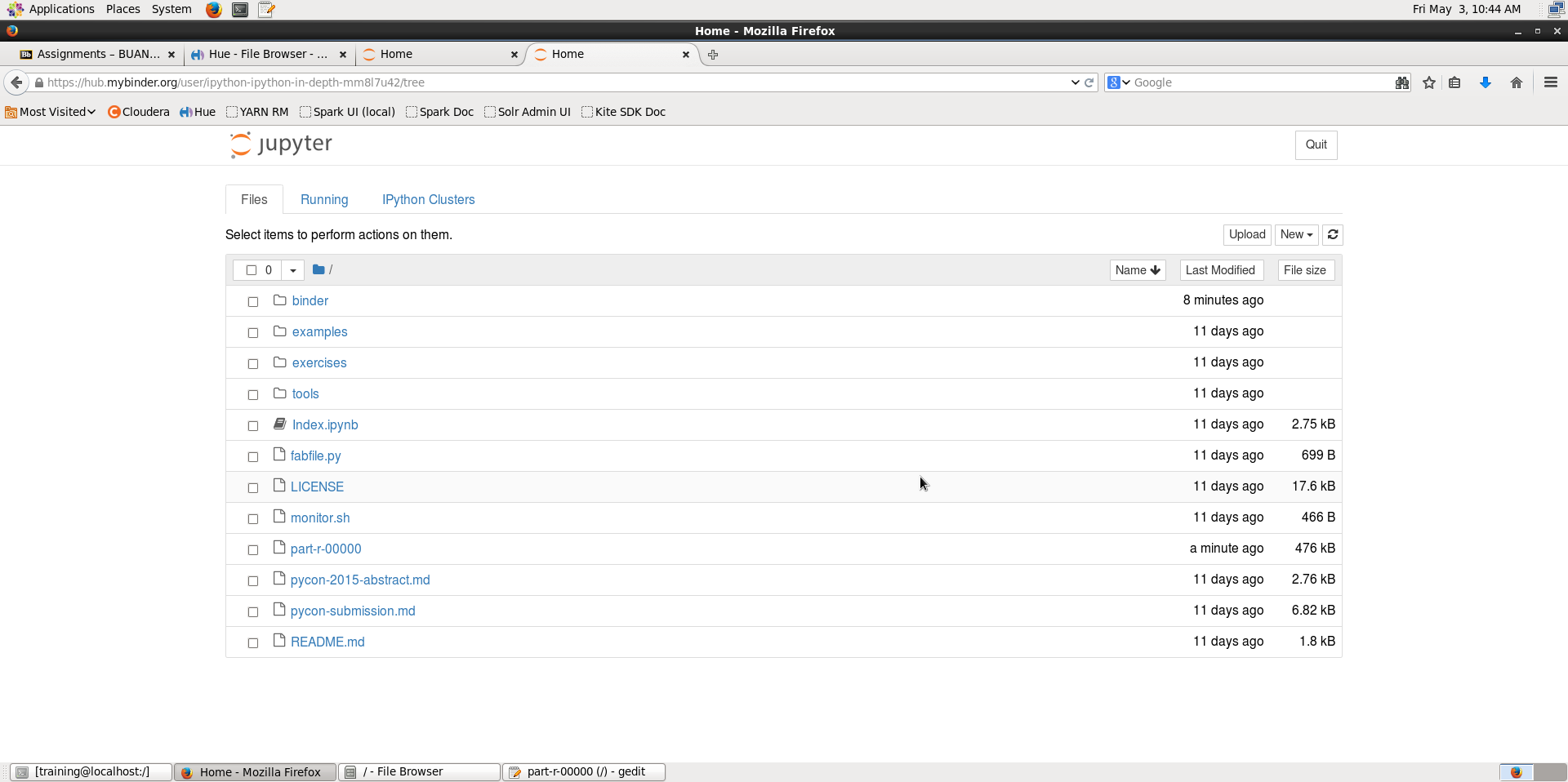
**Step 24:**



**Step 25:**

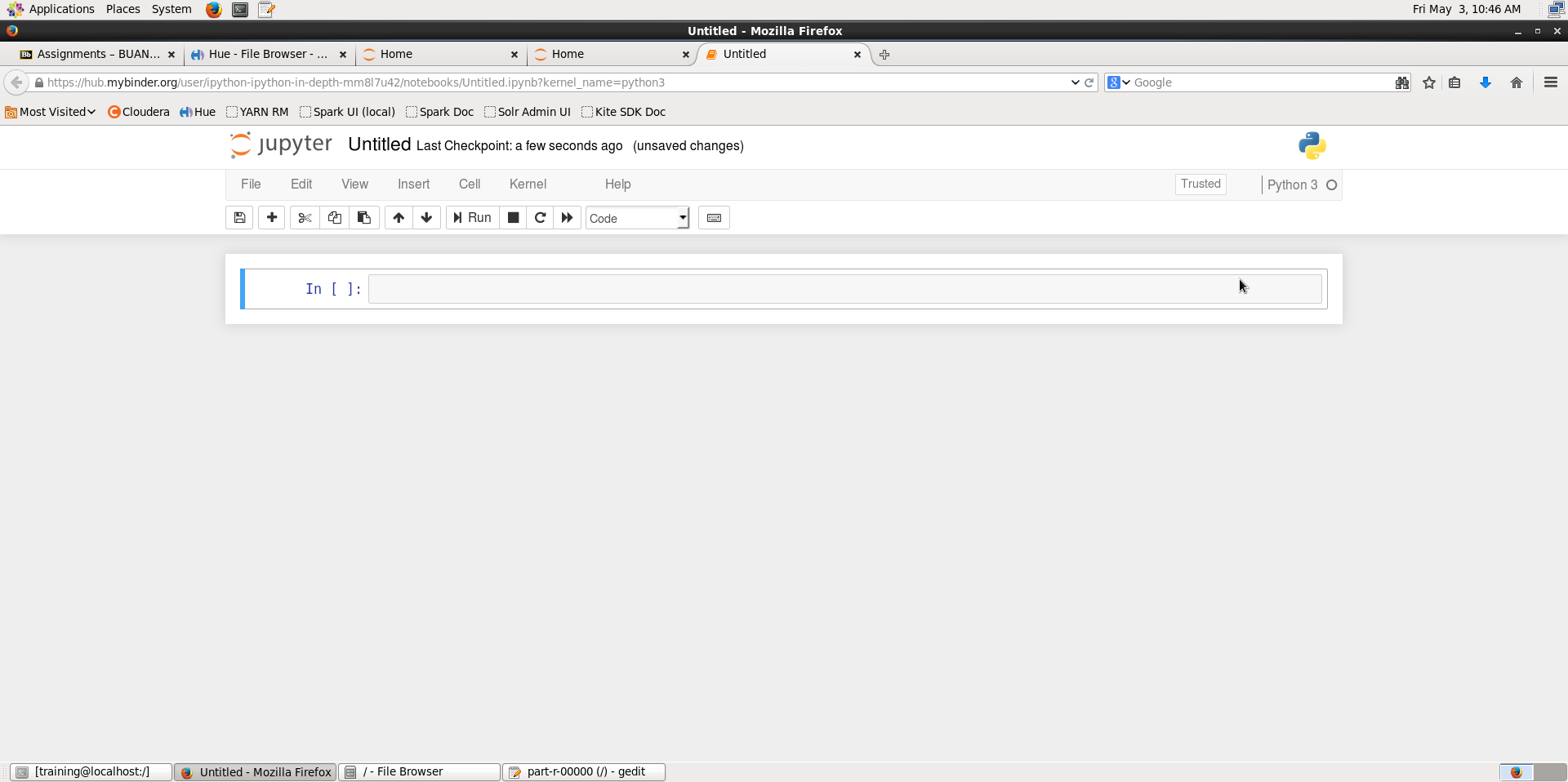


**Step 26:**

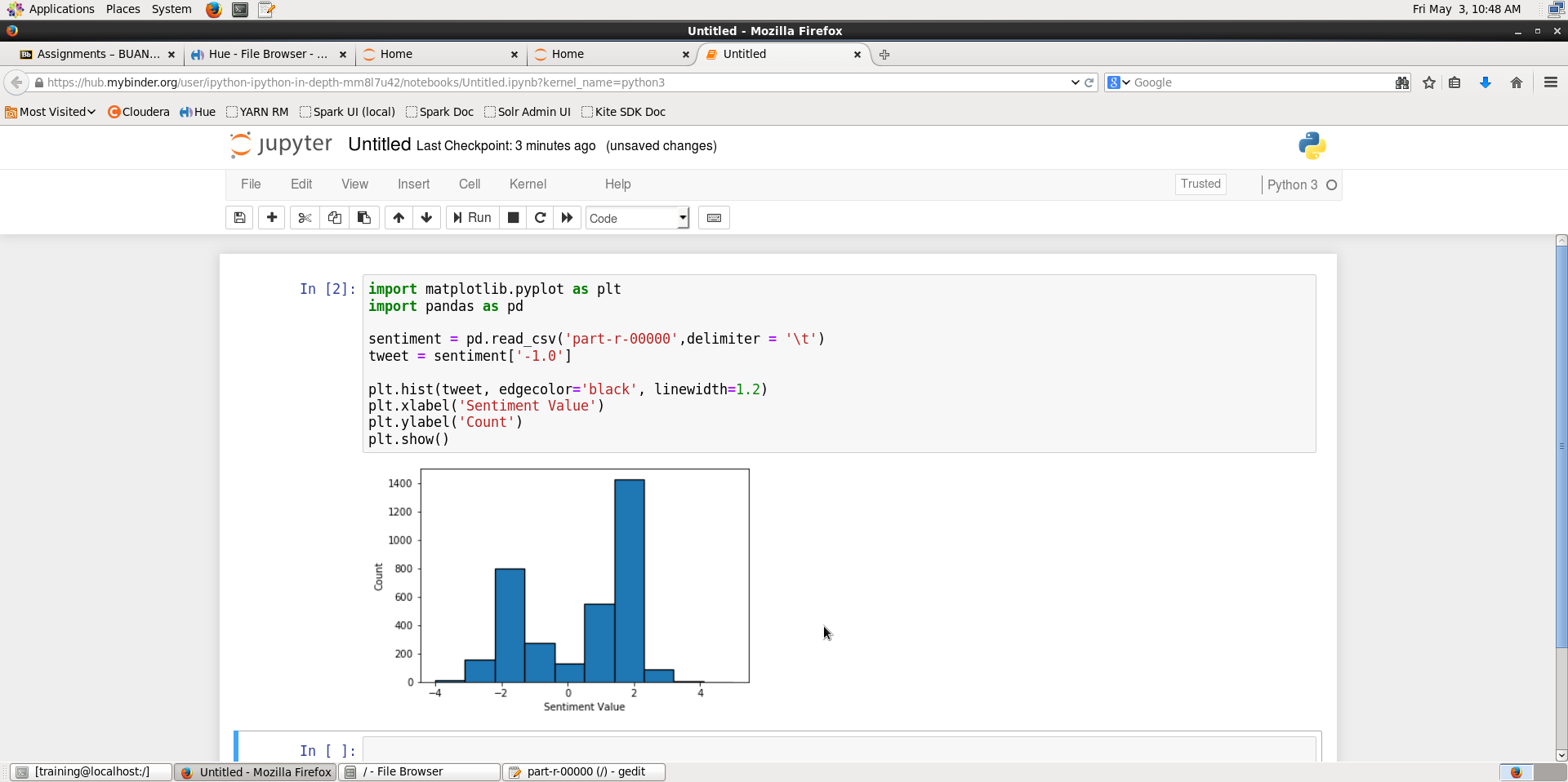


Open the part-r-00000 and copy the contents and paste it in the browser running Jupyter we opened in Step-25. Save the file as “part-r-00000”. Force save the file.

**Step 27:**



**Step 28:**



**Answer:** The inference is that percentage of the sentiment of people being positive is more than that of being negative. The highest rating of the sentiment of people is 2.