

Sentiment analysis on demonetization

Let us find out the views of different people on the demonetization by analysing the tweets from twitter. Here is the dataset where twitter tweets are gathered in CSV format. You can download the dataset from the below link

<https://drive.google.com/open?id=0ByJLBTmJojizNkRsZWJiY1VGc28>

```
val tweets =
sc.textFile("file:///home/kiran/Documents/datasets/demonetization-tweets.csv").map(x
=> x.split(",")).filter(x=>x.length>=2).map(x => (x(0).replaceAll("\\\"", ""),x(1).replaceAll("\\\"", "").toLowerCase)).map(x => (x._1,x._2.split("
"))).toDF("id","words")
tweets.registerTempTable("tweets")
```

```
val explode = spark.sql("select id as id,explode(words) as word from
tweets").registerTempTable("tweet_word")
```

```
val afinn = sc.textFile("file:///home/kiran/Documents/datasets/AFINN.txt").map(x =>
x.split("\t")).map(x => (x(0),x(1))).toDF("word","rating").registerTempTable("afinn")
```

```
val join = spark.sql("select t.id,AVG(a.rating) as rating from tweet_word t join afinn a on
t.word=a.word group by t.id order by rating desc").show
```

```
scala> val join = spark.sql("select t.id,AVG(a.rating) as rating from tweet_word t join afinn a on t.word=a.word group by t.id order by rating d
esc").show
+-----+
| id|rating|
+-----+
|7281| 4.0|
|6610| 4.0|
|6546| 4.0|
|7994| 4.0|
|3822| 4.0|
|4185| 4.0|
|5733| 4.0|
|7025| 4.0|
| 308| 3.5|
|1500| 3.0|
|2654| 3.0|
|4144| 3.0|
|4484| 3.0|
|4862| 3.0|
|3494| 3.0|
|5497| 3.0|
|1497| 3.0|
|5829| 3.0|
|5473| 3.0|
| 938| 3.0|
+-----+
only showing top 20 rows
```