**Session 7**

**Assignment 7.2**

Student Name: Karthik K

Course: Big Data Hadoop & Spark Training

Problem Statement

Implement the below blog at your end and send the complete documentation. <https://docs.google.com/document/d/14YUd_wi-KJTBEqqtyoNtPMFeORjFU-_yId4hJ11q90o/edit>

Sentiment analysis on demonetization

Let us find out the views of different people on the demonetization by analyzing 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=0ByJLBTmJojjzNkRsZWJiY1VGc28>

Source codes

***val tweets = sc.textFile("/home/acadgild/hadoop/demonetization-tweets.csv").map(x*** ***=>***

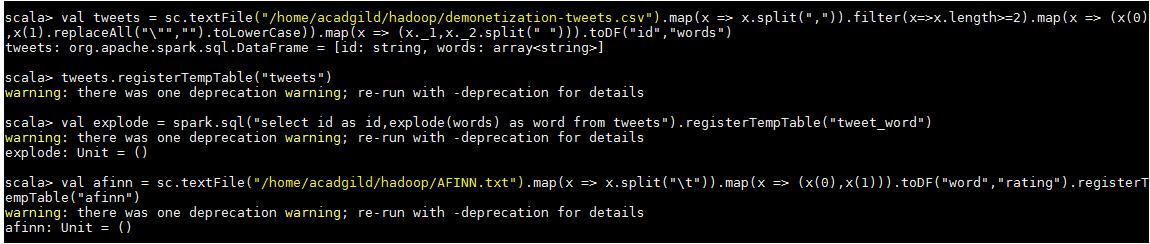
***x.split(",")).filter(x=>x.length>=2).map(x*** ***=>***

***(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("/home/acadgild/hadoop/AFINN.txt").map(x => x.split("\t")).map(x => (x(0),x(1))).toDF("word","rating").registerTempTable("afinn")***

******

Desired result

***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***

Expected Output

