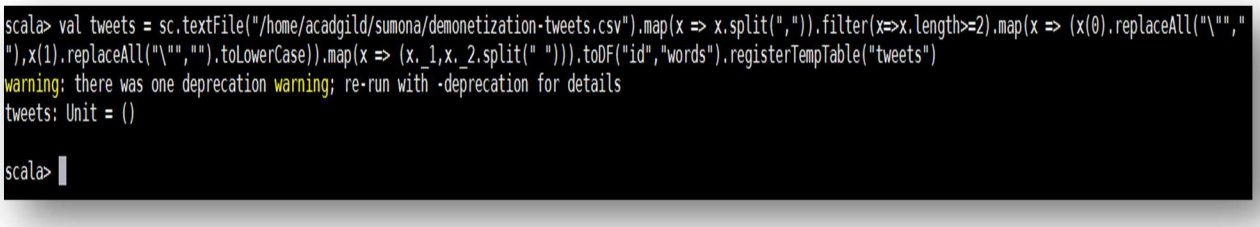
Assignment 22.2

Downloaded the dataset demonetization-tweets.csv and AFFIN.txt and loaded in the local file system **/home/acadgild/Sumona**

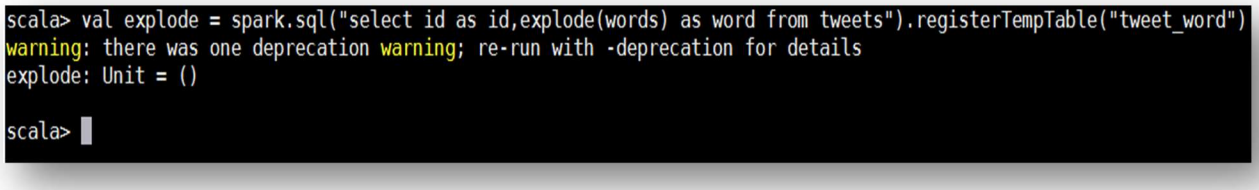
1. First we will read the csv file and then split the columns to get and create a DataFrame. Also, we will create a temporary table named **tweets**

val tweets = sc.textFile("/home/acadgild/sumona/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").registerTempTable("tweets")



2. Now from the above temporary table we will select the ID, words and form another temporary table **tweet_word**

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



3. Here we will read the **AFFIN** file and create a temporary table **affin**

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

Then we will join both the tables **tweet_word** and **affin** and get the views of different people on demonetization

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