Assignment 20.2

Step1: Download the demonetization\_tweets.csv from the Assignment Link given and copy to /home/acadgild/assignment\_20.2

Step2: Download AFINN.txt from the URL below

<https://raw.githubusercontent.com/wendykan/twitter-sentiment-analysis/master/AFINN-111.txt>

Step3: Process the tweets from demonetization-tweets.csv and store to temporary table tweets

val tweets\_rdd\_with\_header = sc.textFile("/home/acadgild/hadoop/demonetization-tweets.csv")

val header = tweets\_rdd\_with\_header.first()

val tweets\_rdd = tweets\_rdd\_with\_header.filter(row => row != header)

val tweets\_filtered\_rdd = tweets\_rdd.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(" ")))

val tweets\_df = tweets\_filtered\_rdd.toDF("id","words")

tweets\_df.registerTempTable("tweets")

sqlContext.sql("select id as id,explode(words) as word from tweets").registerTempTable("tweet\_word")

Screenshot is as below:



Step4:

val afinn\_rdd = sc.textFile("/home/acadgild/hadoop/assignment\_20.2/AFINN.txt")

val afinn\_df = afinn\_rdd.map(x => x.split("\t")).map(x => (x(0),x(1))).toDF("word","rating")

afinn\_df.registerTempTable("afinn")

sqlContext.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(100)

Step5: Find the the rating for tweet

sqlContext.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(100)



