**SESSION 20 ASSIGNMENT 2**

* Download and copy the demonetization\_tweets.csv from the Assignment Link to the local directory.
* Download AFINN.txt from the URL below

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

* Creating and filtering the RDD for twitter data and creating a temp table for the same

scala> val tweets\_rdd\_with\_header = sc.textFile("/home/acadgild/assignment\_20.2/demonetization-tweets.csv")

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

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

scala> 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(" ")))

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

scala> tweets\_df.registerTempTable("tweets")

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



* Creating RDD for AFINN data and storing it in a temp table

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

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

scala> afinn\_df.registerTempTable("afinn")

scala> 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)

* Joining the two tables on word and finding the rating of the tweets

scala> 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)



