CS 6350

ASSIGNMENT 2

Names of students in your group: 2

Sriprasath Gujuluva Parthasarathy(sxg180154)

Shariq Ali(sxa190016)

Number of free late days used: 0  
Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

1. Data Intensive Text Processing using MapReduce
2. https://www.transtats.bts.gov/DL\_SelectFields.asp?Table\_ID=236
3. <https://www.kaggle.com/crowdflower/twitter-airline-sentiment>
4. https://spark.apache.org/docs/2.2.0/api/scala/index.html#org.apache.spark. ml.tuning.CrossValidator
5. https://spark.apache.org/docs/2.2.0/ mllib-evaluation-metrics.html
6. <https://github.com/databricks/spark-deep-learning>
7. https://www.kaggle.com/datasets?search=image

Part 1

Dataset: 407161809\_T\_ONTIME\_REPORTING.csv

Code: sparkwordcount\_2.11-0.1.jar

Part 2

Dataset: tweets.csv

Code: sparkwordcount\_2.11-0.1.jar

Part 3

Dataset: s3://sg-cs6350/Assignment2/part3

Code: https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/32072069969770/2766025117377373/478845738118579/latest.html