**#Assignment**

**SIRSS1027**

**Aanchal Saxena**

**Q1 Get all the airport with latitude value greater than 40 ( make sure to pass the latitude value as int for float)  
  
Q2 Find out the working of mapValues function, reduceByKey, countByKey**

Ans.1.

import sys

sys.path.insert(0, '.')

from pyspark import SparkContext, SparkConf

from commons.Utils import Utils

def splitComma(line: str):

splits = Utils.COMMA\_DELIMITER.split(line)

return "{}, {}".format(splits[1], splits[6])

if \_\_name\_\_ == "\_\_main\_\_":

conf = SparkConf().setAppName("airports").setMaster("local[\*]")

sc = SparkContext(conf = conf)

airports = sc.textFile("in/airports.text")

airportsInUSA = airports.filter(lambda line: float(Utils.COMMA\_DELIMITER.split(line)[6]) > 40)

airportsNameAndCityNames = airportsInUSA.map(splitComma)

airportsNameAndCityNames.saveAsTextFile("out/airports\_by\_latitude.text")

Ans.2.

* **mapValues(func)**

Even without changing the key, mapValues operation applies a function to each value of a paired RDD of spark.  
rdd.mapValues(x => x+1)

* **countByKey()**

Through countByKey operation, we can count the number of elements for each key.  
rdd.countByKey()

* **reduceByKey(fun)**

Here, the reduceByKey operation generally combines values with the same key.  
add.reduceByKey( (x, y) => x + y)