Working with Sensor Data

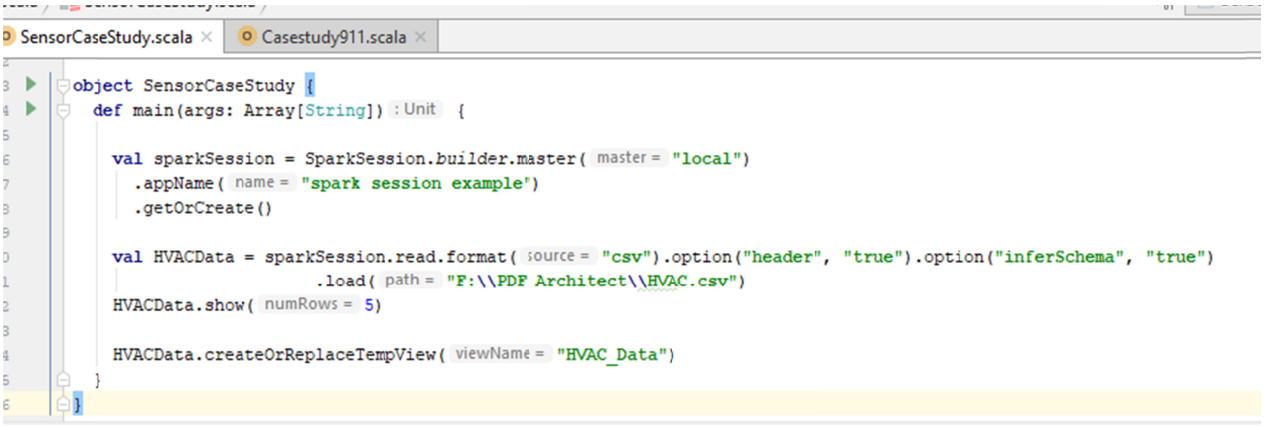
Columns those are present in the datasets:

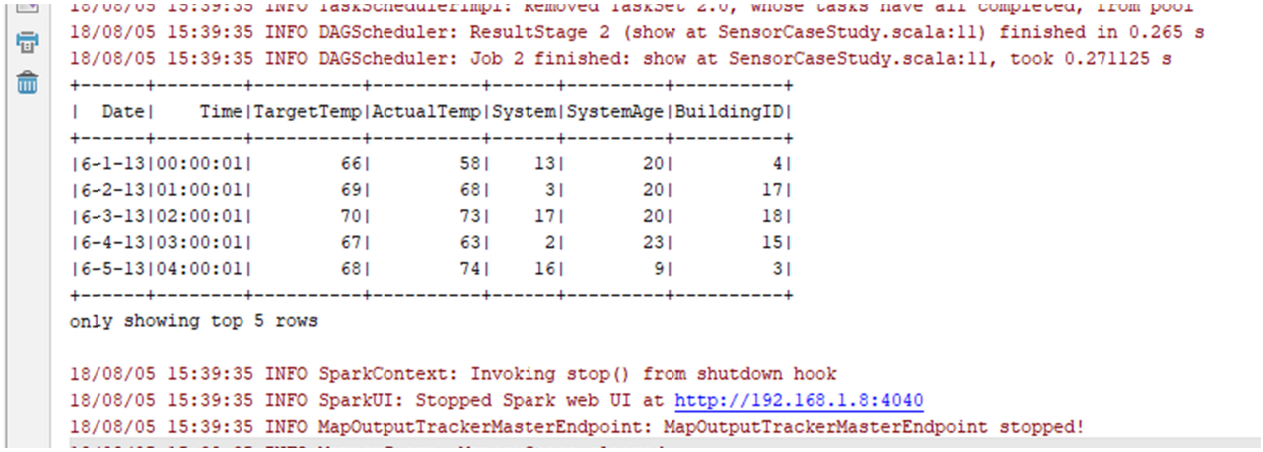
Building.csv – BuildingID, BuildingMgr, BuildingAge, HVACproduct,Country

HVAC.csv – Date, Time, TargetTemp, ActualTemp, System, SystemAge, BuildingID

Objective 1

- Load HVAC.csv file into temporary table





- Add a new column, tempchange - set to 1, if there is a change of greater than +/-5 between actual and target temperature

```
val HVACData = sparkSession.read.format( source = "csv").option("header", "true").option("inferSchema", "true")
                  .load( path = "F:\\PDF Architect\\HVAC.csv")
HVACData.show( numRows = 5)
import sparkSession.implicits._

HVACData.createOrReplaceTempView( viewName = "HVAC_Data")
val newhvac = HVACData.select( cols = $"Date", $"Time", $"TargetTemp".cast( to = "Int"), $"ActualTemp".
                cast( to = "Int"), $"System".cast( to = "Int"), $"SystemAge".cast( to = "Int"), $"BuildingID")
val newcolhvac = sparkSession.sql( sqlText = "select *,IF((targettemp - actualtemp) > 5, '1', IF" +
                  "((targettemp - actualtemp) < -5, '1', 0)) AS tempchange from HVAC_Data")
newcolhvac.show()
```

```
18/08/05 17:29:56 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
18/08/05 17:29:56 INFO DAGScheduler: ResultStage 3 (show at SensorCaseStudy.scala:20) finished in 0.211 s
18/08/05 17:29:56 INFO DAGScheduler: Job 3 finished: show at SensorCaseStudy.scala:20, took 0.219789 s
+------+--------+----------+----------+------+---------+----------+----------+
|  Date|    Time|TargetTemp|ActualTemp|System|SystemAge|BuildingID|tempchange|
+------+--------+----------+----------+------+---------+----------+----------+
| 6-1-13|00:00:01|        66|        58|    13|       20|         4|         1|
| 6-2-13|01:00:01|        69|        68|     3|       20|        17|         0|
| 6-3-13|02:00:01|        70|        73|    17|       20|        18|         0|
| 6-4-13|03:00:01|        67|        63|     2|       23|        15|         0|
| 6-5-13|04:00:01|        68|        74|    16|        9|         3|         1|
| 6-6-13|05:00:01|        67|        56|    13|       28|         4|         1|
| 6-7-13|06:00:01|        70|        58|    12|       24|         2|         1|
| 6-8-13|07:00:01|        70|        73|    20|       26|        16|         0|
| 6-9-13|08:00:01|        66|        69|    16|        9|         9|         0|
|6-10-13|09:00:01|        65|        57|     6|        5|        12|         1|
|6-11-13|10:00:01|        67|        70|    10|       17|        15|         0|
|6-12-13|11:00:01|        69|        62|     2|       11|         7|         1|
|6-13-13|12:00:01|        69|        73|    14|        2|        15|         0|
|6-14-13|13:00:01|        65|        61|     3|        2|         6|         0|
|6-15-13|14:00:01|        67|        59|    19|       22|        20|         1|
|6-16-13|15:00:01|        65|        56|    19|       11|         8|         1|
|6-17-13|16:00:01|        67|        57|    15|        7|         6|         1|
|6-18-13|17:00:01|        66|        57|    12|        5|        13|         1|
|6-19-13|18:00:01|        69|        58|     8|       22|         4|         1|
|6-20-13|19:00:01|        67|        55|    17|        5|         7|         1|
+------+--------+----------+----------+------+---------+----------+----------+
only showing top 20 rows

18/08/05 17:29:56 INFO SparkContext: Invoking stop() from shutdown hook
18/08/05 17:29:56 INFO SparkUI: Stopped Spark web UI at http://192.168.1.8:4040
```

Objective 2

Load building.csv file into temporary table

```
val BuildingData = sparkSession.read.format( source = "csv").option("header", "true").option("inferSchema", "true")
  .load( path = "F:\\PDF Architect\\building.csv")
BuildingData.createOrReplaceTempView( viewName = "Building_Data")
BuildingData.show( numRows = 5)

}
```

```
18/08/05 17:33:06 INFO TaskSetManager: Finished task 0.0 in stage 5.0 (TID 5) in 217 ms on localhost (executo
18/08/05 17:33:06 INFO TaskSchedulerImpl: Removed TaskSet 5.0, whose tasks have all completed, from pool
18/08/05 17:33:06 INFO DAGScheduler: ResultStage 5 (show at SensorCaseStudy.scala:24) finished in 0.244 s
18/08/05 17:33:06 INFO DAGScheduler: Job 5 finished: show at SensorCaseStudy.scala:24, took 0.263478 s
+----------+-----------+-----------+-----------+---------+
|BuildingID|BuildingMgr|BuildingAge|HVACproduct|  Country|
+----------+-----------+-----------+-----------+---------+
|         1|         M1|         25|     AC1000|      USA|
|         2|         M2|         27|     FN39TG|   France|
|         3|         M3|         28|     JDNS77|   Brazil|
|         4|         M4|         17|     GG1919|  Finland|
|         5|         M5|          3|    ACMAX22|Hong Kong|
+----------+-----------+-----------+-----------+---------+
only showing top 5 rows

18/08/05 17:33:06 INFO SparkContext: Invoking stop() from shutdown hook
18/08/05 17:33:06 INFO SparkUI: Stopped Spark web UI at http://192.168.1.8:4040
```

Objective 3

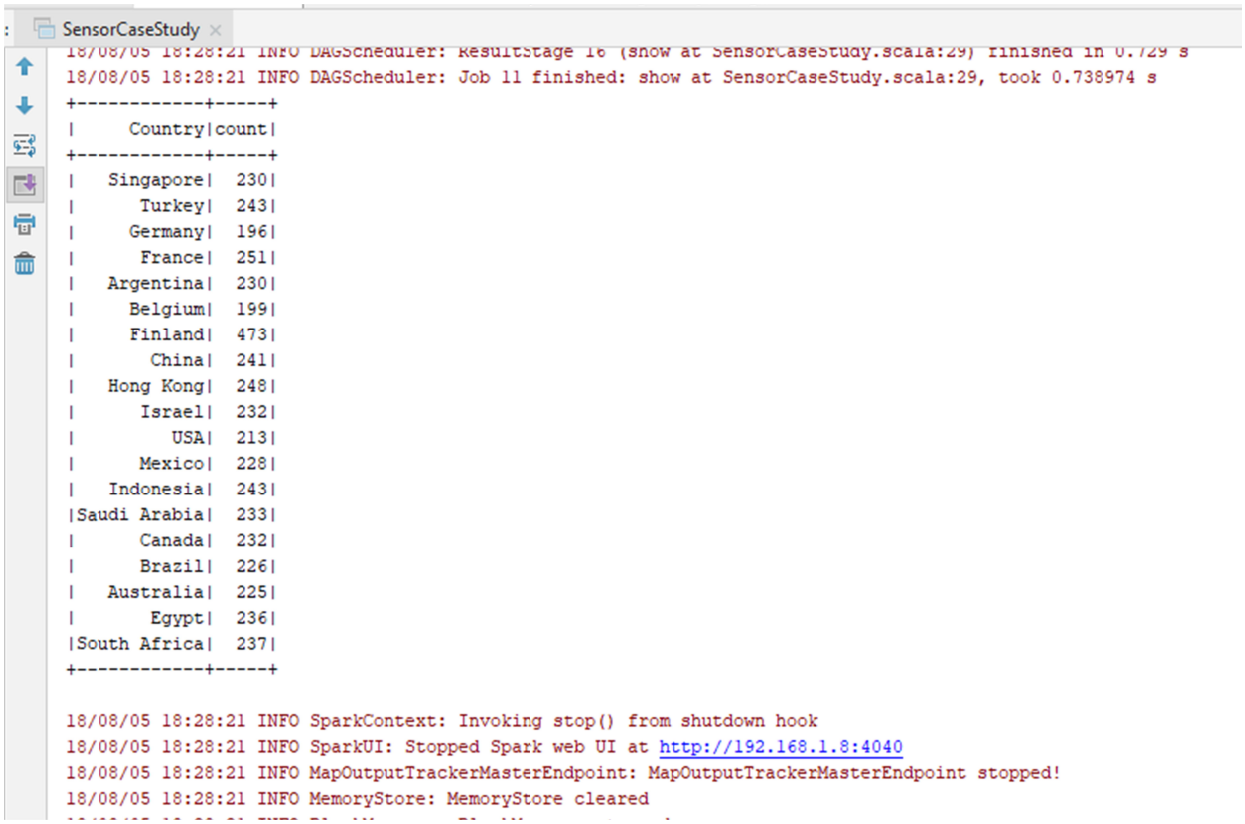
Figure out the number of times, temperature has changed by 5 degrees or more for each country:

- Join both the tables.
- Select tempchange and country column
- Filter the rows where tempchange is 1 and count the number of occurrence for each country

```
val newcolhvac = sparkSession.sql( sqlText = "select *,IF((targettemp - actualtemp) > 5, '1', IF" +
                    "((targettemp - actualtemp) < -5, '1', 0)) AS tempchange from HVAC_Data").toDF()
newcolhvac.createOrReplaceTempView( viewName = "newColHvac")
newcolhvac.printSchema()

val buildingData = sparkSession.read.format( source = "csv").option("header", "true").option("inferSchema", "true")
  .load( path = "F:\\PDF Architect\\building.csv").toDF()
buildingData.createOrReplaceTempView( viewName = "Building_Data")
buildingData.show( numRows = 5)

val joinedDF = newcolhvac.as( alias = "HD").join(buildingData.as( alias = "BD"),
        joinExprs = $"BD.BuildingID" === $"HD.BuildingID").filter( condition = $"tempchange" === 1).groupBy( col1 = "Country").count().show()
}
```


SensorCaseStudy
18/08/05 18:28:21 INFO DAGScheduler: ResultStage 16 (show at SensorCaseStudy.scala:29) finished in 0.729 s
18/08/05 18:28:21 INFO DAGScheduler: Job 11 finished: show at SensorCaseStudy.scala:29, took 0.738974 s
+------------+-----+
| Country|count|
+------------+-----+
Singapore	230
Turkey	243
Germany	196
France	251
Argentina	230
Belgium	199
Finland	473
China	241
Hong Kong	248
Israel	232
USA	213
Mexico	228
Indonesia	243
Saudi Arabia	233
Canada	232
Brazil	226
Australia	225
Egypt	236
South Africa	237
+------------+-----+
18/08/05 18:28:21 INFO SparkContext: Invoking stop() from shutdown hook
18/08/05 18:28:21 INFO SparkUI: Stopped Spark web UI at http://192.168.1.8:4040
18/08/05 18:28:21 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/08/05 18:28:21 INFO MemoryStore: MemoryStore cleared