# Case Study3

```
import org.apache.spark.sql.SparkSession
object CaseStudy3 {
  case class
hvac_cls(Date:String,Time:String,TargetTemp:Int,ActualTemp:Int,System:Int,SystemAge:Int,
BuildingId:Int)
  case class
building(buildid:Int,buildmgr:String,buildAge:Int,hvacproduct:String,Country:String)

  def main(args: Array[String]): Unit = {

    println("hey scala")
    val spark = SparkSession
      .builder()
      .master("local")
      .appName("Spark SQL basic example")
      .config("spark.some.config.option", "some-value")
      .getOrCreate()

    println("Spark Session Object created")

    val data = spark.sparkContext.textFile("E:\\Avani\\Acadgild\\Case Study 3\\HVAC.csv");
    val header=data.first()
    val data1 = data.filter(row => row != header)  //removing of header
    println("HVAC Data->>"+data1.count())
    println("removed header")

    //For implicit conversions like converting RDDs and sequences  to DataFrames
    import spark.implicits._

    //creation of HVAC temporary table
    val hvac = data1.map(x=>x.split(",")).map(x =>
hvac_cls(x(0),x(1),x(2).toInt,x(3).toInt,x(4).toInt,x(5).toInt,x(6).toInt)).toDF()
    //val hvac = data.map(x=>x.split(",")).map(x =>
hvac_cls(x(0),x(1),x(2).toInt,x(3).toInt,x(4).toInt,x(5).toInt,x(6).toInt)).toDF()
    hvac.show()
    println("HVAC Dataframe created !")
    hvac.registerTempTable("HVAC")
    println("HVAC Data loaded into temporary table!")

 val data2 = spark.sparkContext.textFile("E:\\Avani\\Acadgild\\Case Study 3\\building.csv");
    val header2=data2.first()
    val data3 = data2.filter(row => row != header2)
    println("building Data->>"+data3.count())
    println("removed header")

//creation of building temporary table
 val build = data3.map(x=> x.split(",")).map(x =>
building(x(0).toInt,x(1),x(2).toInt,x(3),x(4))).toDF
    build.show()
    println("Building Dataframe created !")
    build.registerTempTable("building")
    println("Building Data loaded into temporary table !")
```

**Objective 1**

- Load HVAC.csv file into temporary table

```
  hvac.registerTempTable("HVAC")
  println("HVAC Data loaded 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 hvac1 = spark.sql("select *,IF((targettemp - actualtemp) > 5, '1', IF((targettemp -
actualtemp) < -5, '1', 0)) AS tempchange from HVAC")

  hvac1.show()

  hvac1.registerTempTable("HVAC1")

  println("Tempchange created in HVAC table !")

  println("Objective 1 completed")
```

**Objective2**

- Load building.csv file into temporary table

```
  build.registerTempTable("building")
  println("Building Data loaded into temporary table !")
```

## Objective3

Figure out the number of times, temperature has changed by 5 degrees

or more for each country:

- Join both the tables

```
val build1 = spark.sql("select h.*, b.country, b.hvacproduct from building b join hvac1 h on
b.buildid = h.buildingid")
build1.show()
```

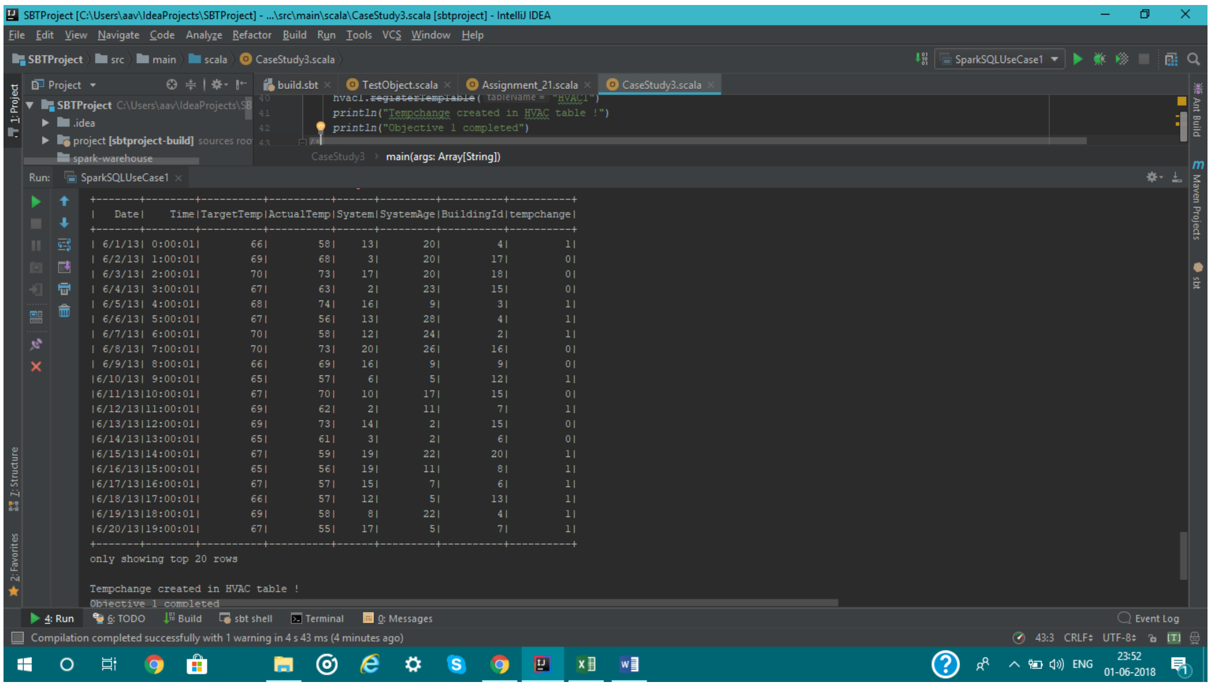
- Select tempchange and country column

```
val tempCountry = build1.map(x => (new Integer(x(7).toString),x(8).toString))
tempCountry.show()
```

- Filter the rows where tempchange is 1 and count the number of occurrence for each country

```
val tempCountryOnes = tempCountry.filter(x=> {if(x._1==1) true else false})
tempCountryOnes.show()
//counting the occurence of column_2
tempCountryOnes.groupBy("_2").count.show
println("Objective 3 completed!")
```

## Screenshot


SBTProject [C:\Users\aav\IdeaProjects\SBTProject] - ...\src\main\scala\CaseStudy3.scala [sbtproject] - IntelliJ IDEA
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
SBTProject src main scala CaseStudy3.scala
SparkSQLUseCase1
build.sbt TestObject.scala Assignment_21.scala CaseStudy3.scala
CaseStudy3 main(args: Array[String])
Run: SparkSQLUseCase1
4: Run 6: TODO Build sbt shell Terminal 0: Messages Event Log
43:3 CRLF UTF-8
23:52 01-06-2018

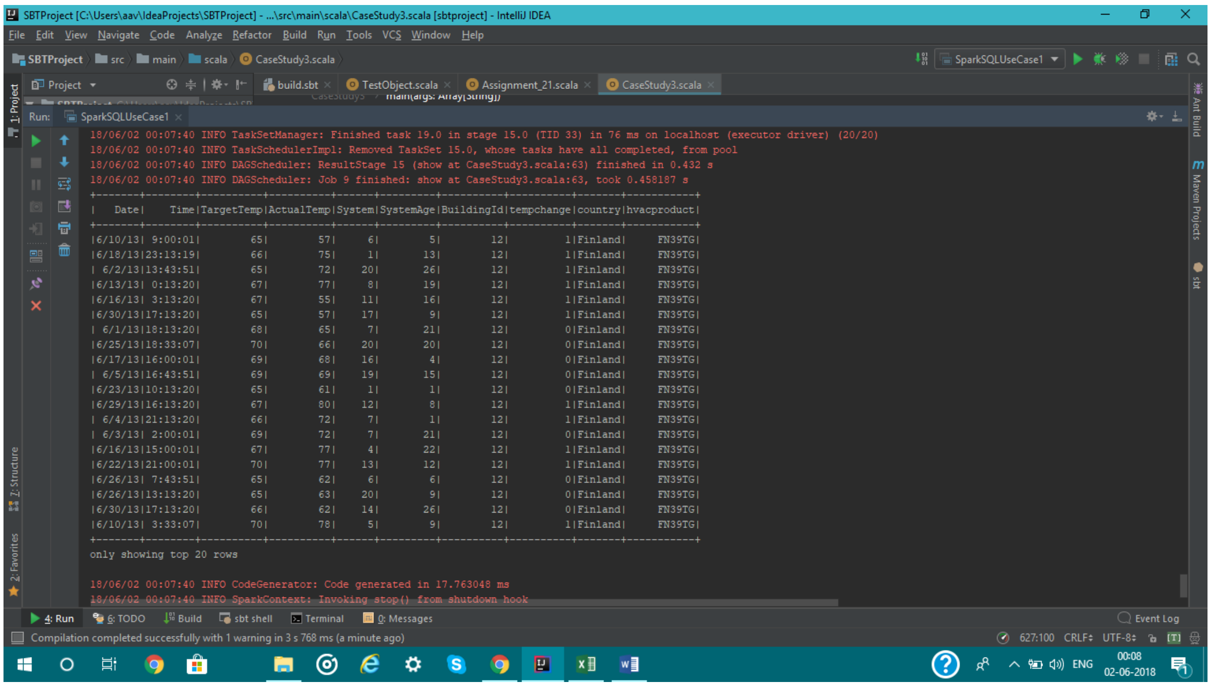
```
hvac1.registerTempTable( tableName = "HVAC1")
println("Tempchange created in HVAC table !")
println("Objective 1 completed")
```

```
+-------+--------+----------+----------+------+---------+----------+----------+
|   Date|    Time|TargetTemp|ActualTemp|System|SystemAge|BuildingId|tempchange|
+-------+--------+----------+----------+------+---------+----------+----------+
| 6/1/13| 0:00:01|        66|        58|    13|       20|         4|         1|
| 6/2/13| 1:00:01|        69|        68|     3|       20|        17|         0|
| 6/3/13| 2:00:01|        70|        73|    17|       20|        18|         0|
| 6/4/13| 3:00:01|        67|        63|     2|       23|        15|         0|
| 6/5/13| 4:00:01|        68|        74|    16|        9|         3|         1|
| 6/6/13| 5:00:01|        67|        56|    13|       28|         4|         1|
| 6/7/13| 6:00:01|        70|        58|    12|       24|         2|         1|
| 6/8/13| 7:00:01|        70|        73|    20|       26|        16|         0|
| 6/9/13| 8:00:01|        66|        69|    16|        9|         9|         0|
|6/10/13| 9:00:01|        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

Tempchange created in HVAC table !
Objective 1 completed
```

Compilation completed successfully with 1 warning in 4 s 43 ms (4 minutes ago)

**a.**

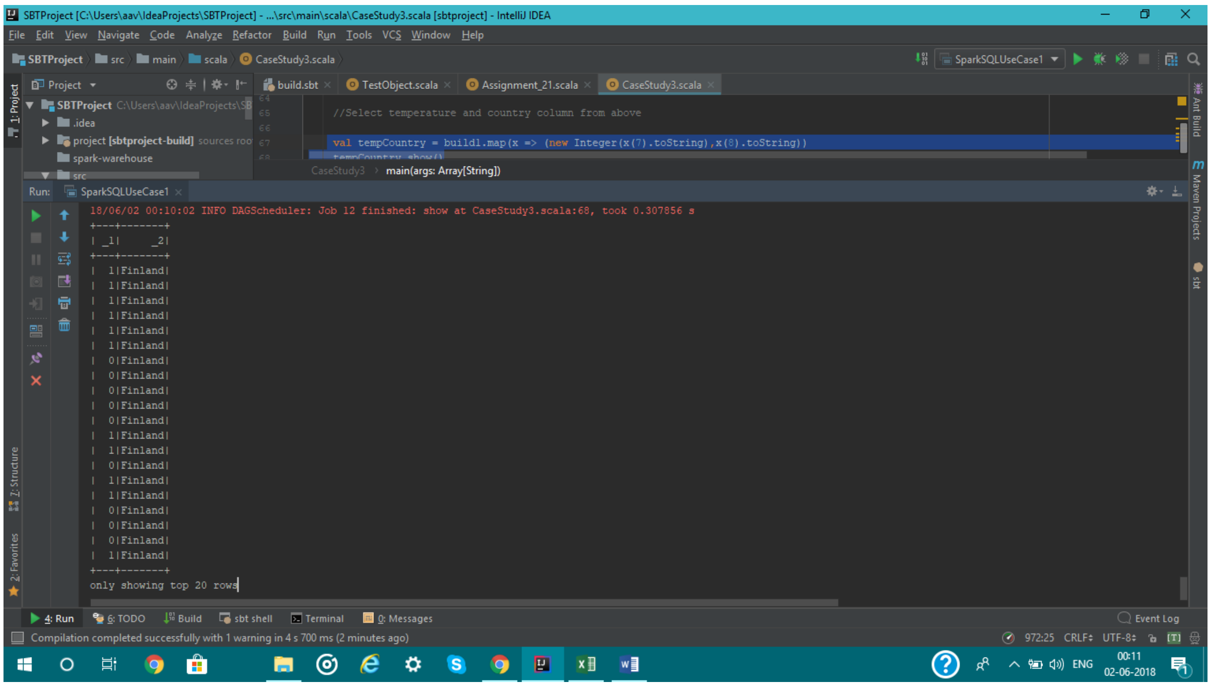

SBTProject [C:\Users\aav\IdeaProjects\SBTProject] - ...\src\main\scala\CaseStudy3.scala [sbtproject] - IntelliJ IDEA
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
SBTProject src main scala CaseStudy3.scala
SparkSQLUseCase1
build.sbt TestObject.scala Assignment_21.scala CaseStudy3.scala
Run: SparkSQLUseCase1
4: Run 6: TODO Build sbt shell Terminal 0: Messages Event Log
627:100 CRLF UTF-8
00:08 02-06-2018

```
18/06/02 00:07:40 INFO TaskSetManager: Finished task 19.0 in stage 15.0 (TID 33) in 76 ms on localhost (executor driver) (20/20)
18/06/02 00:07:40 INFO TaskSchedulerImpl: Removed TaskSet 15.0, whose tasks have all completed, from pool
18/06/02 00:07:40 INFO DAGScheduler: ResultStage 15 (show at CaseStudy3.scala:63) finished in 0.432 s
18/06/02 00:07:40 INFO DAGScheduler: Job 9 finished: show at CaseStudy3.scala:63, took 0.458187 s
+-------+--------+----------+----------+------+---------+----------+----------+-------+----------+
|   Date|    Time|TargetTemp|ActualTemp|System|SystemAge|BuildingId|tempchange|country|hvacproduct|
+-------+--------+----------+----------+------+---------+----------+----------+-------+----------+
|6/10/13| 9:00:01|        65|        57|     6|        5|        12|         1|Finland|    FN39TG|
|6/18/13|23:13:19|        66|        75|     1|       13|        12|         1|Finland|    FN39TG|
| 6/2/13|13:43:51|        65|        72|    20|       26|        12|         1|Finland|    FN39TG|
|6/13/13| 0:13:20|        67|        77|     8|       19|        12|         1|Finland|    FN39TG|
|6/16/13| 3:13:20|        67|        55|    11|       16|        12|         1|Finland|    FN39TG|
|6/30/13|17:13:20|        65|        57|    17|        9|        12|         1|Finland|    FN39TG|
| 6/1/13|18:13:20|        68|        65|     7|       21|        12|         0|Finland|    FN39TG|
|6/25/13|18:33:07|        70|        66|    20|       20|        12|         0|Finland|    FN39TG|
|6/17/13|16:00:01|        69|        68|    16|        4|        12|         0|Finland|    FN39TG|
| 6/5/13|16:43:51|        69|        69|    19|       15|        12|         0|Finland|    FN39TG|
|6/23/13|10:13:20|        65|        61|     1|        1|        12|         0|Finland|    FN39TG|
|6/29/13|16:13:20|        67|        80|    12|        8|        12|         1|Finland|    FN39TG|
| 6/4/13|21:13:20|        66|        72|     7|        1|        12|         1|Finland|    FN39TG|
| 6/3/13| 2:00:01|        69|        72|     7|       21|        12|         0|Finland|    FN39TG|
|6/16/13|15:00:01|        67|        77|     4|       22|        12|         1|Finland|    FN39TG|
|6/22/13|21:00:01|        70|        77|    13|       12|        12|         1|Finland|    FN39TG|
|6/26/13| 7:43:51|        65|        62|     6|        6|        12|         0|Finland|    FN39TG|
|6/26/13|13:13:20|        65|        63|    20|        9|        12|         0|Finland|    FN39TG|
|6/30/13|17:13:20|        66|        62|    14|       26|        12|         0|Finland|    FN39TG|
|6/10/13| 3:33:07|        70|        78|     5|        9|        12|         1|Finland|    FN39TG|
+-------+--------+----------+----------+------+---------+----------+----------+-------+----------+
only showing top 20 rows

18/06/02 00:07:40 INFO CodeGenerator: Code generated in 17.763048 ms
18/06/02 00:07:40 INFO SparkContext: Invoking stop() from shutdown hook
```

Compilation completed successfully with 1 warning in 3 s 768 ms (a minute ago)

**3.1**



SBTProject [C:\Users\aav\IdeaProjects\SBTProject] - ...\src\main\scala\CaseStudy3.scala [sbtproject] - IntelliJ IDEA

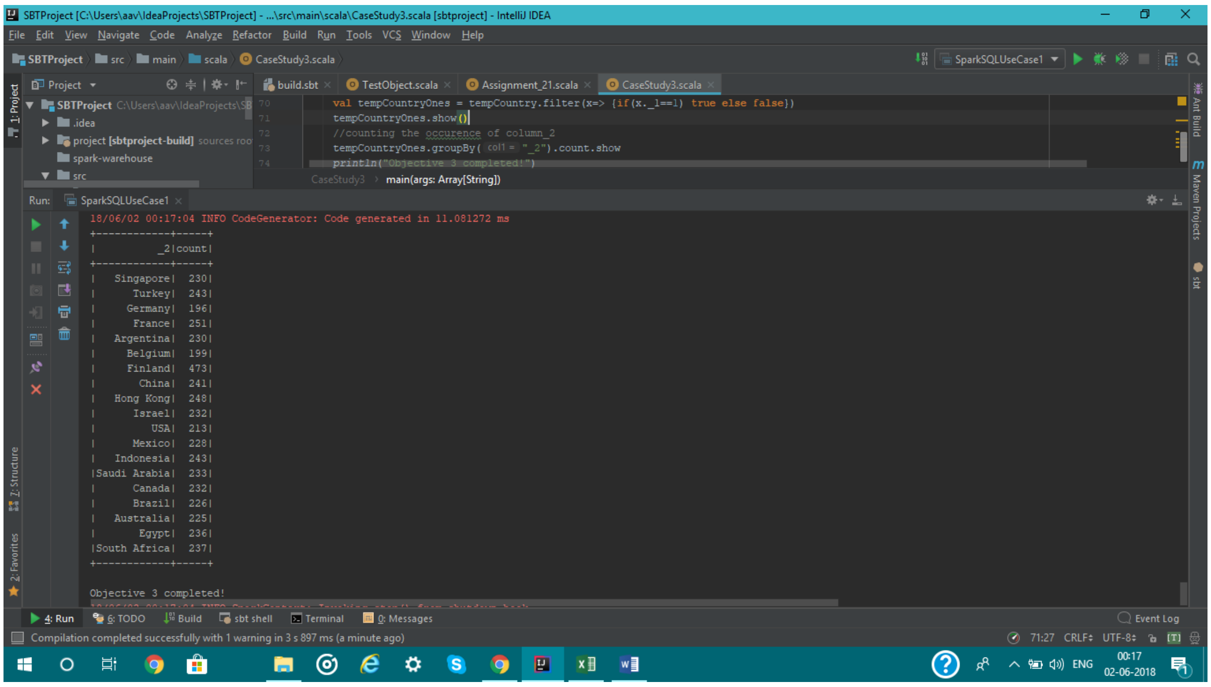
```
//Select temperature and country column from above

val tempCountry = buildl.map(x => (new Integer(x(7).toString),x(8).toString))
tempCountry.show()
```

```
18/06/02 00:10:02 INFO DAGScheduler: Job 12 finished: show at CaseStudy3.scala:68, took 0.307856 s
+---+-------+
| _1|     _2|
+---+-------+
|  1|Finland|
|  1|Finland|
|  1|Finland|
|  1|Finland|
|  1|Finland|
|  1|Finland|
|  0|Finland|
|  0|Finland|
|  0|Finland|
|  0|Finland|
|  0|Finland|
|  1|Finland|
|  1|Finland|
|  0|Finland|
|  1|Finland|
|  1|Finland|
|  0|Finland|
|  0|Finland|
|  0|Finland|
|  1|Finland|
+---+-------+
only showing top 20 rows
```

Compilation completed successfully with 1 warning in 4 s 700 ms (2 minutes ago)

3.2



SBTProject [C:\Users\aav\IdeaProjects\SBTProject] - ...\src\main\scala\CaseStudy3.scala [sbtproject] - IntelliJ IDEA

```
val tempCountryOnes = tempCountry.filter(x=> {if(x._1==1) true else false})
tempCountryOnes.show()
//counting the occurence of column_2
tempCountryOnes.groupBy( col1 = "_2").count.show
println("Objective 3 completed!")
```

```
18/06/02 00:17:04 INFO CodeGenerator: Code generated in 11.081272 ms
+------------+-----+
|          _2|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|
+------------+-----+

Objective 3 completed!
```

Compilation completed successfully with 1 warning in 3 s 897 ms (a minute ago)

3.3

## Complete code

```
import org.apache.spark.sql.SparkSession
object CaseStudy3 {
  case class
hvac_cls(Date:String,Time:String,TargetTemp:Int,ActualTemp:Int,System:Int,SystemAge:Int,
BuildingId:Int)
  case class
building(buildid:Int,buildmgr:String,buildAge:Int,hvacproduct:String,Country:String)

  def main(args: Array[String]): Unit = {

    println("hey scala")
    val spark = SparkSession
      .builder()
      .master("local")
      .appName("Spark SQL basic example")
      .config("spark.some.config.option", "some-value")
      .getOrCreate()

    println("Spark Session Object created")

    val data = spark.sparkContext.textFile("E:\\Avani\\Acadgild\\Case Study 3\\HVAC.csv");
    val header=data.first()
    val data1 = data.filter(row => row != header)  //removing of header
    println("HVAC Data->>"+data1.count())
    println("removed header")

    //For implicit conversions like converting RDDs and sequences  to DataFrames
    import spark.implicits._

    val hvac = data1.map(x=>x.split(",")).map(x =>
hvac_cls(x(0),x(1),x(2).toInt,x(3).toInt,x(4).toInt,x(5).toInt,x(6).toInt)).toDF()
    //val hvac = data.map(x=>x.split(",")).map(x =>
hvac_cls(x(0),x(1),x(2).toInt,x(3).toInt,x(4).toInt,x(5).toInt,x(6).toInt)).toDF()
    hvac.show()
    println("HVAC Dataframe created !")

    val data2 = spark.sparkContext.textFile("E:\\Avani\\Acadgild\\Case Study
3\\building.csv");
    val header2=data2.first()
    val data3 = data2.filter(row => row != header2)
    println("building Data->>"+data3.count())
    println("removed header")
    val build = data3.map(x=> x.split(",")).map(x =>
building(x(0).toInt,x(1),x(2).toInt,x(3),x(4))).toDF
    build.show()
    println("Building Dataframe created !")

    //creation of HVAC temporary table
    hvac.registerTempTable("HVAC")
    println("HVAC Data loaded into temporary table!")

    //creation of building temporary table
    build.registerTempTable("building")
    println("Building Data loaded into temporary table for objective 2!")

    //creation of a new column based on conditions
    val hvac1 = spark.sql("select *,IF((targettemp - actualtemp) > 5, '1', IF((targettemp -
actualtemp) < -5, '1', 0)) AS tempchange from HVAC")
    hvac1.show()
    hvac1.registerTempTable("HVAC1")
    println("Tempchange created in HVAC table !")
    println("Objective 1 completed")

    //Now join the two tables
    val build1 = spark.sql("select h.*, b.country, b.hvacproduct from building b join hvac1 h on
b.buildid = h.buildingid")
    build1.show()

    //Select temperature and country column from above

    val tempCountry = build1.map(x => (new Integer(x(7).toString),x(8).toString))
    tempCountry.show()

    val tempCountryOnes = tempCountry.filter(x=> {if(x._1==1) true else false})
    tempCountryOnes.show()
    //counting the occurence of column_2
    tempCountryOnes.groupBy("_2").count.show
    println("Objective 3 completed!")

  }
}
```