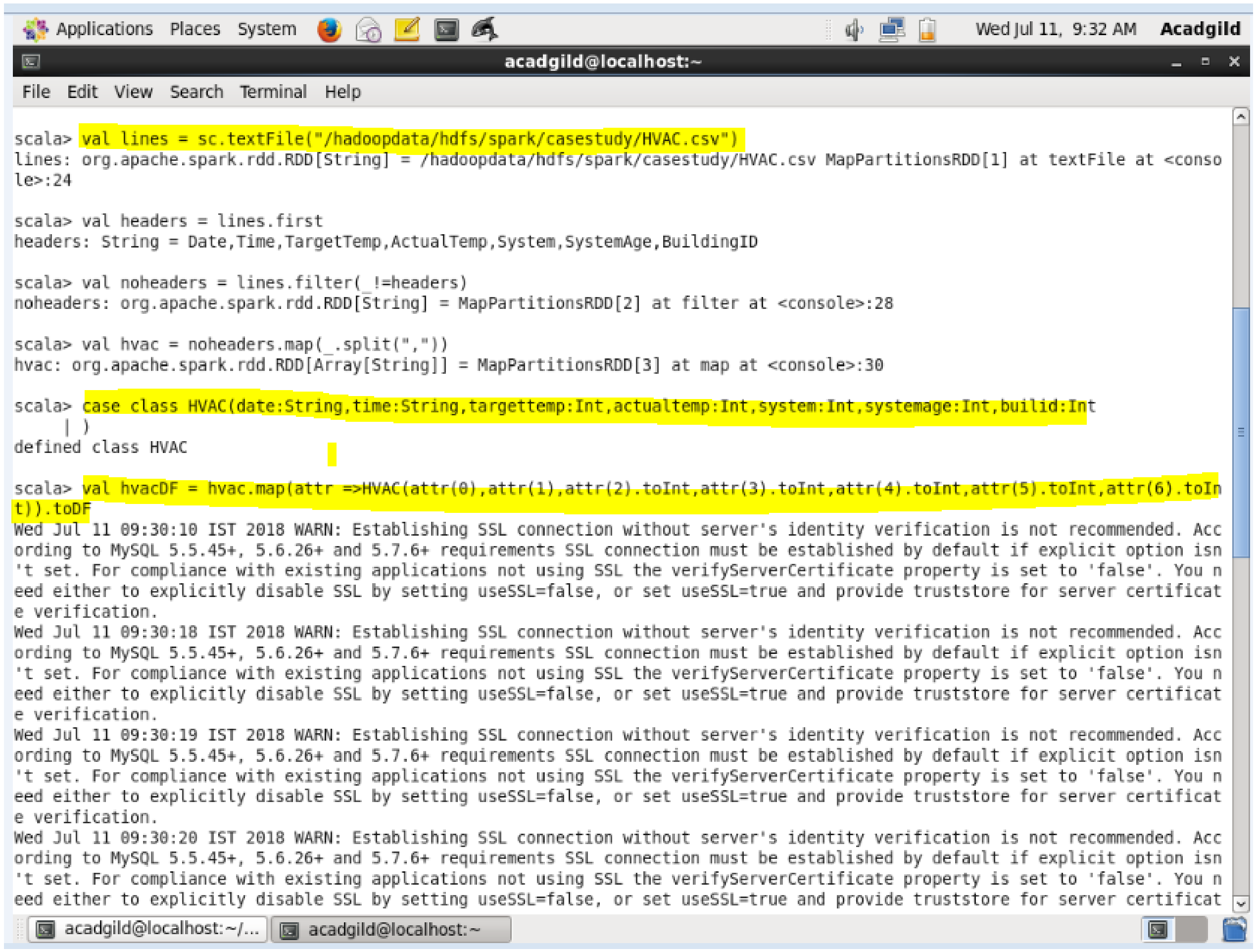
# CASE STUDY 3-Working with Sensor Data

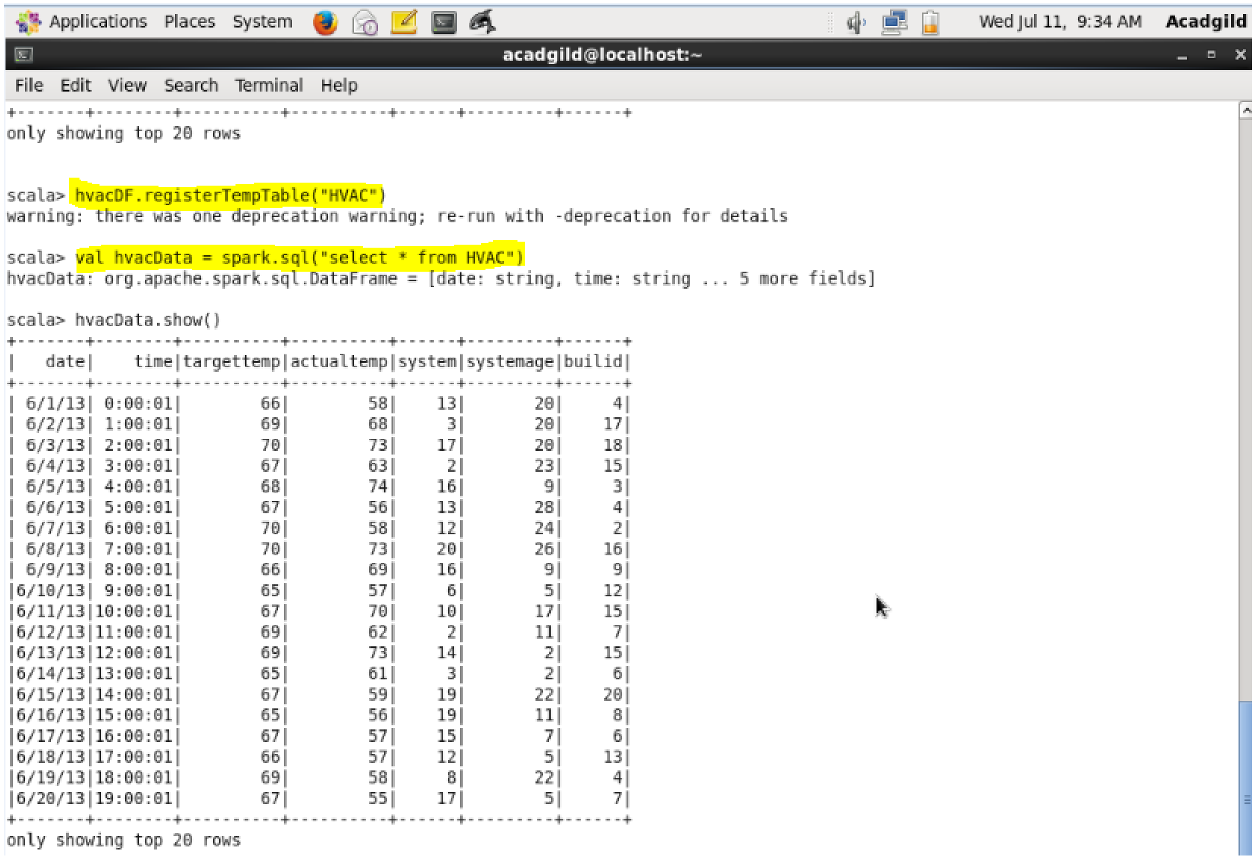
## Objective 1: Load HVAC.csv file into temporary table


Applications Places System
Wed Jul 11, 9:32 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
acadgild@localhost:~/...
acadgild@localhost:~

```
scala> val lines = sc.textFile("/hadoopdata/hdfs/spark/casestudy/HVAC.csv")
lines: org.apache.spark.rdd.RDD[String] = /hadoopdata/hdfs/spark/casestudy/HVAC.csv MapPartitionsRDD[1] at textFile at <conso
le>:24

scala> val headers = lines.first
headers: String = Date,Time,TargetTemp,ActualTemp,System,SystemAge,BuildingID

scala> val noheaders = lines.filter(_!=headers)
noheaders: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at filter at <console>:28

scala> val hvac = noheaders.map(_.split(","))
hvac: org.apache.spark.rdd.RDD[Array[String]] = MapPartitionsRDD[3] at map at <console>:30

scala> case class HVAC(date:String,time:String,targettemp:Int,actualtemp:Int,system:Int,systemage:Int,builid:Int
     | )
defined class HVAC

scala> val hvacDF = hvac.map(attr =>HVAC(attr(0),attr(1),attr(2).toInt,attr(3).toInt,attr(4).toInt,attr(5).toInt,attr(6).toIn
t)).toDF
Wed Jul 11 09:30:10 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. Acc
ording to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn
't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You n
eed either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificat
e verification.
Wed Jul 11 09:30:18 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. Acc
ording to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn
't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You n
eed either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificat
e verification.
Wed Jul 11 09:30:19 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. Acc
ording to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn
't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You n
eed either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificat
e verification.
Wed Jul 11 09:30:20 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. Acc
ording to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn
't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You n
eed either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificat
```

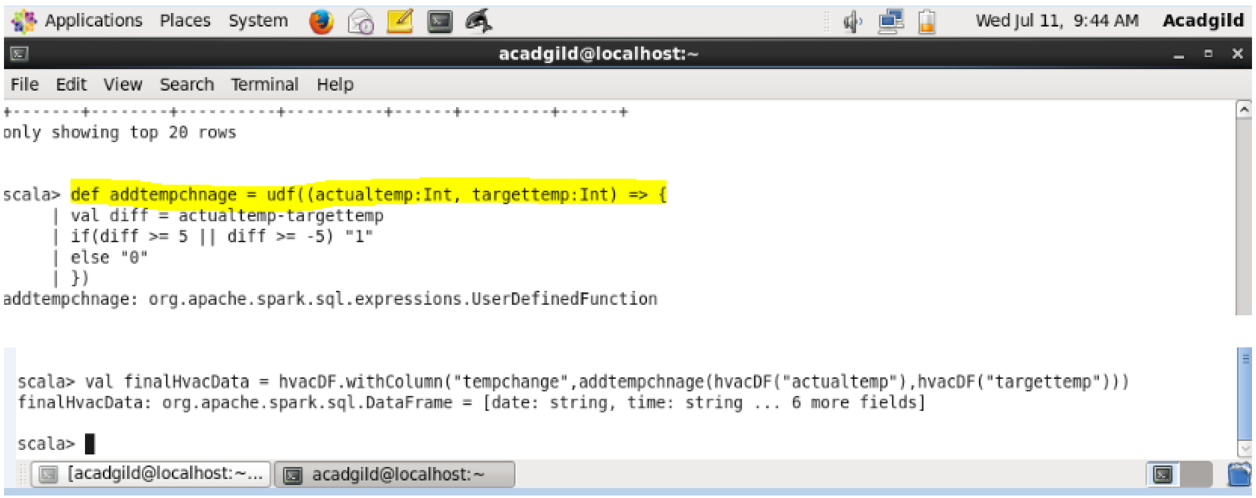
# CASE STUDY 3-Working with Sensor Data

Loaded the data in to temporary table called HVAC


Applications Places System
Wed Jul 11, 9:34 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help

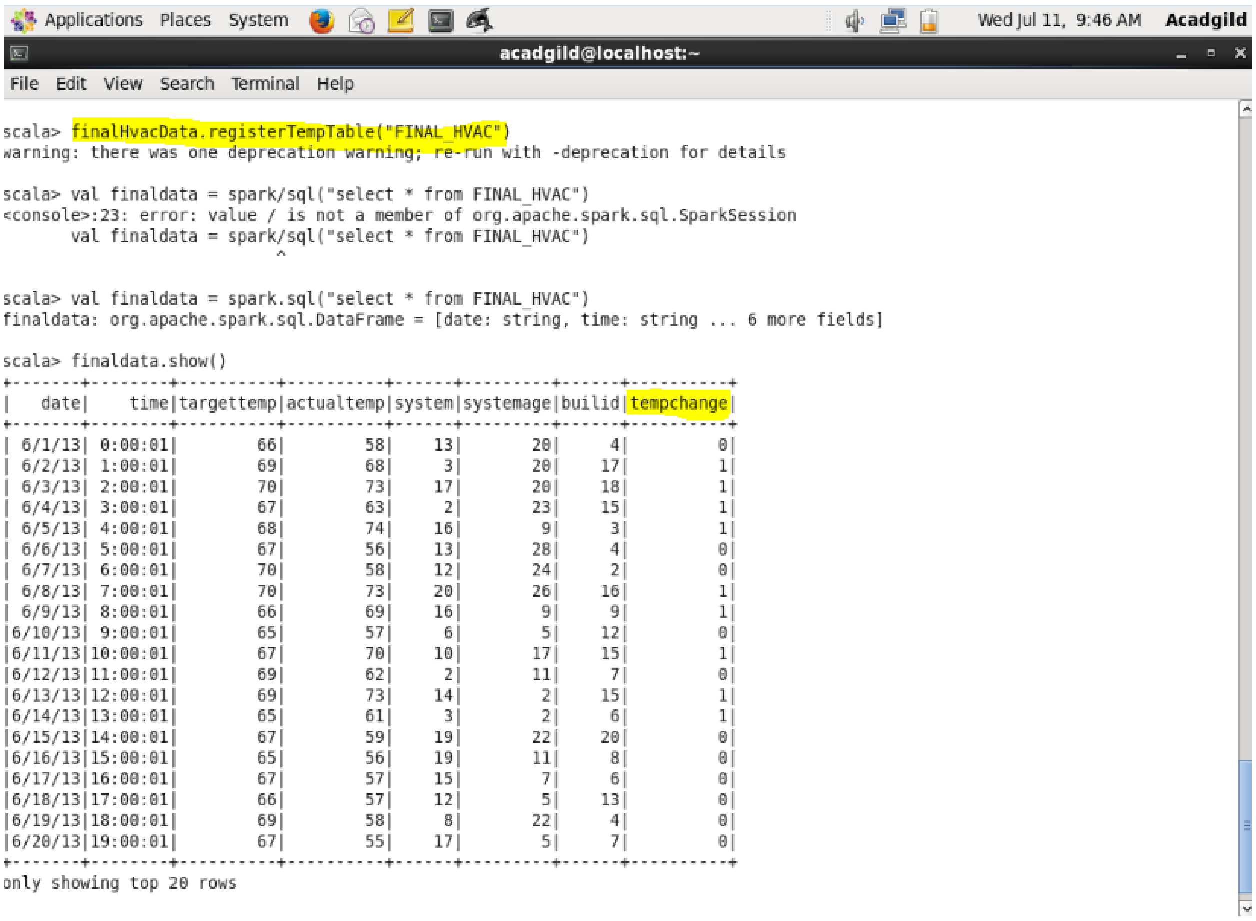
```
+-------+--------+----------+----------+------+---------+------+
only showing top 20 rows


scala> hvacDF.registerTempTable("HVAC")
warning: there was one deprecation warning; re-run with -deprecation for details

scala> val hvacData = spark.sql("select * from HVAC")
hvacData: org.apache.spark.sql.DataFrame = [date: string, time: string ... 5 more fields]

scala> hvacData.show()
+-------+--------+----------+----------+------+---------+------+
|   date|    time|targettemp|actualtemp|system|systemage|buildid|
+-------+--------+----------+----------+------+---------+------+
| 6/1/13| 0:00:01|        66|        58|    13|       20|     4|
| 6/2/13| 1:00:01|        69|        68|     3|       20|    17|
| 6/3/13| 2:00:01|        70|        73|    17|       20|    18|
| 6/4/13| 3:00:01|        67|        63|     2|       23|    15|
| 6/5/13| 4:00:01|        68|        74|    16|        9|     3|
| 6/6/13| 5:00:01|        67|        56|    13|       28|     4|
| 6/7/13| 6:00:01|        70|        58|    12|       24|     2|
| 6/8/13| 7:00:01|        70|        73|    20|       26|    16|
| 6/9/13| 8:00:01|        66|        69|    16|        9|     9|
|6/10/13| 9:00:01|        65|        57|     6|        5|    12|
|6/11/13|10:00:01|        67|        70|    10|       17|    15|
|6/12/13|11:00:01|        69|        62|     2|       11|     7|
|6/13/13|12:00:01|        69|        73|    14|        2|    15|
|6/14/13|13:00:01|        65|        61|     3|        2|     6|
|6/15/13|14:00:01|        67|        59|    19|       22|    20|
|6/16/13|15:00:01|        65|        56|    19|       11|     8|
|6/17/13|16:00:01|        67|        57|    15|        7|     6|
|6/18/13|17:00:01|        66|        57|    12|        5|    13|
|6/19/13|18:00:01|        69|        58|     8|       22|     4|
|6/20/13|19:00:01|        67|        55|    17|        5|     7|
+-------+--------+----------+----------+------+---------+------+
only showing top 20 rows
```

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


Applications Places System
Wed Jul 11, 9:44 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost:~...
acadgild@localhost:~

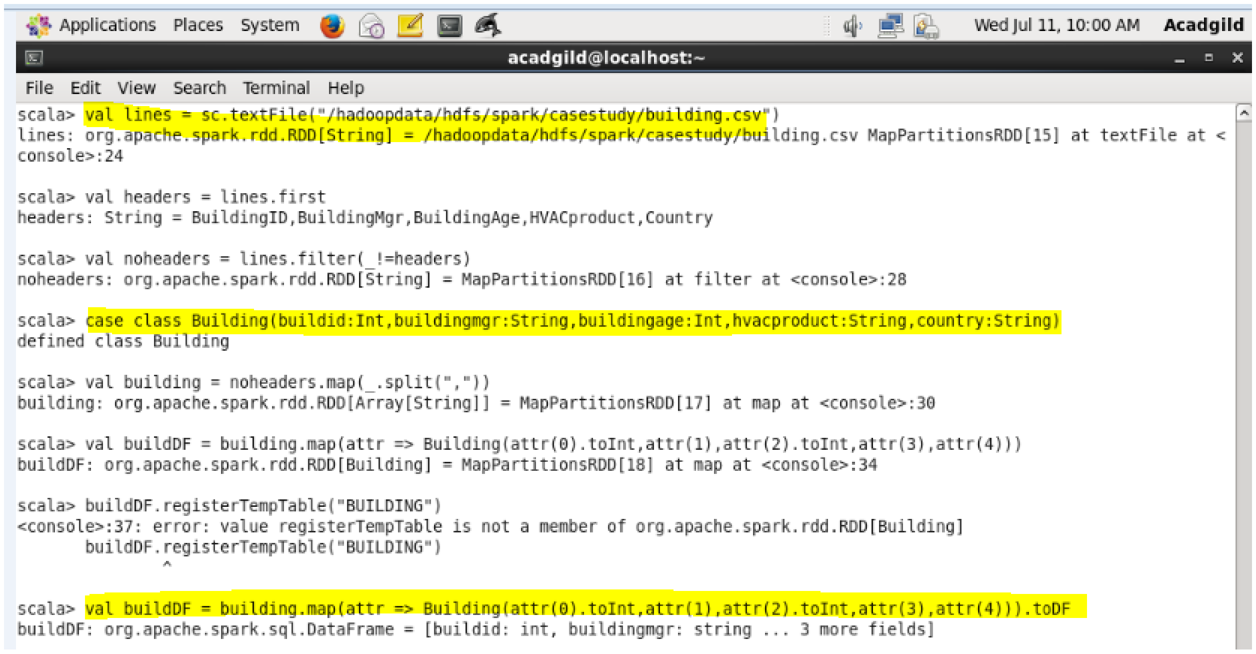
```
+-------+--------+----------+----------+------+---------+------+
only showing top 20 rows


scala> def addtempchnage = udf((actualtemp:Int, targettemp:Int) => {
     | val diff = actualtemp-targettemp
     | if(diff >= 5 || diff >= -5) "1"
     | else "0"
     | })
addtempchnage: org.apache.spark.sql.expressions.UserDefinedFunction


scala> val finalHvacData = hvacDF.withColumn("tempchange",addtempchnage(hvacDF("actualtemp"),hvacDF("targettemp")))
finalHvacData: org.apache.spark.sql.DataFrame = [date: string, time: string ... 6 more fields]

scala> 
```

# CASE STUDY 3-Working with Sensor Data


Applications Places System
Wed Jul 11, 9:46 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help

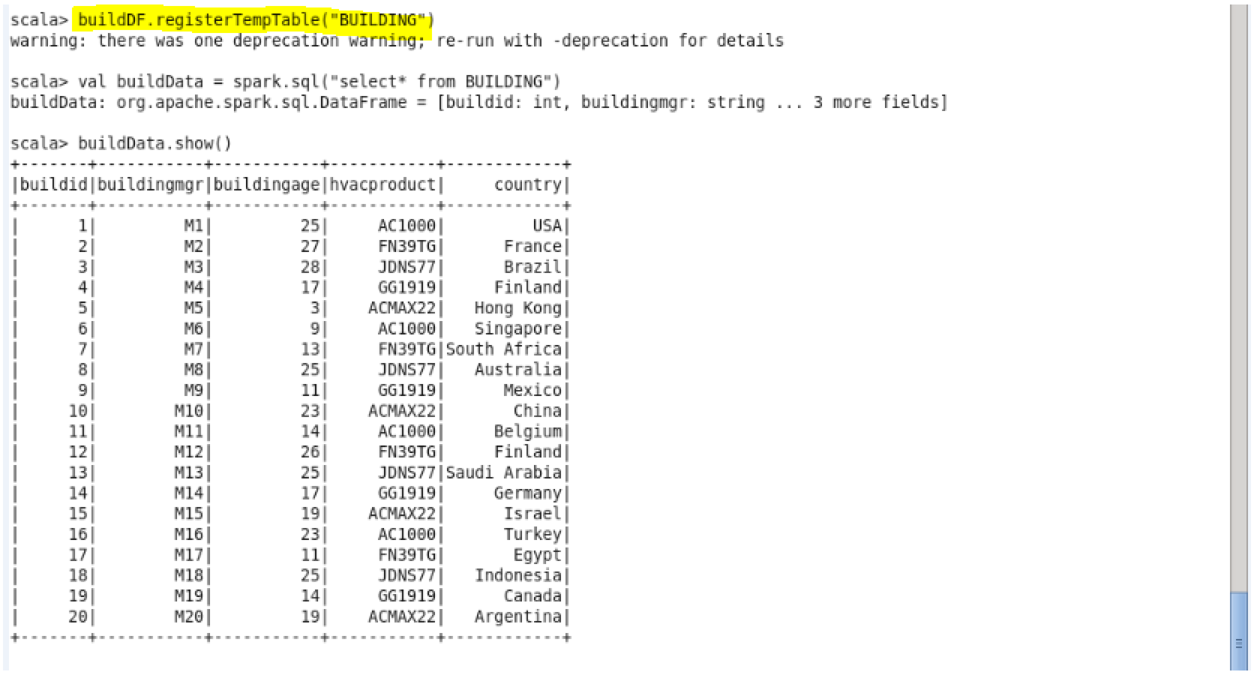
```
scala> finalHvacData.registerTempTable("FINAL_HVAC")
warning: there was one deprecation warning; re-run with -deprecation for details

scala> val finaldata = spark/sql("select * from FINAL_HVAC")
<console>:23: error: value / is not a member of org.apache.spark.sql.SparkSession
       val finaldata = spark/sql("select * from FINAL_HVAC")
                            ^

scala> val finaldata = spark.sql("select * from FINAL_HVAC")
finaldata: org.apache.spark.sql.DataFrame = [date: string, time: string ... 6 more fields]

scala> finaldata.show()
+-------+--------+----------+----------+------+---------+------+----------+
|   date|    time|targettemp|actualtemp|system|systemage|builid|tempchange|
+-------+--------+----------+----------+------+---------+------+----------+
| 6/1/13| 0:00:01|        66|        58|    13|       20|     4|         0|
| 6/2/13| 1:00:01|        69|        68|     3|       20|    17|         1|
| 6/3/13| 2:00:01|        70|        73|    17|       20|    18|         1|
| 6/4/13| 3:00:01|        67|        63|     2|       23|    15|         1|
| 6/5/13| 4:00:01|        68|        74|    16|        9|     3|         1|
| 6/6/13| 5:00:01|        67|        56|    13|       28|     4|         0|
| 6/7/13| 6:00:01|        70|        58|    12|       24|     2|         0|
| 6/8/13| 7:00:01|        70|        73|    20|       26|    16|         1|
| 6/9/13| 8:00:01|        66|        69|    16|        9|     9|         1|
|6/10/13| 9:00:01|        65|        57|     6|        5|    12|         0|
|6/11/13|10:00:01|        67|        70|    10|       17|    15|         1|
|6/12/13|11:00:01|        69|        62|     2|       11|     7|         0|
|6/13/13|12:00:01|        69|        73|    14|        2|    15|         1|
|6/14/13|13:00:01|        65|        61|     3|        2|     6|         1|
|6/15/13|14:00:01|        67|        59|    19|       22|    20|         0|
|6/16/13|15:00:01|        65|        56|    19|       11|     8|         0|
|6/17/13|16:00:01|        67|        57|    15|        7|     6|         0|
|6/18/13|17:00:01|        66|        57|    12|        5|    13|         0|
|6/19/13|18:00:01|        69|        58|     8|       22|     4|         0|
|6/20/13|19:00:01|        67|        55|    17|        5|     7|         0|
+-------+--------+----------+----------+------+---------+------+----------+
only showing top 20 rows
```

**<u>Ojective 2:</u> Load building.csv file into temporary table**


Applications Places System
Wed Jul 11, 10:00 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help

```
scala> val lines = sc.textFile("/hadoopdata/hdfs/spark/casestudy/building.csv")
lines: org.apache.spark.rdd.RDD[String] = /hadoopdata/hdfs/spark/casestudy/building.csv MapPartitionsRDD[15] at textFile at <
console>:24

scala> val headers = lines.first
headers: String = BuildingID,BuildingMgr,BuildingAge,HVACproduct,Country

scala> val noheaders = lines.filter(_!=headers)
noheaders: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[16] at filter at <console>:28

scala> case class Building(buildid:Int,buildingmgr:String,buildingage:Int,hvacproduct:String,country:String)
defined class Building

scala> val building = noheaders.map(_.split(","))
building: org.apache.spark.rdd.RDD[Array[String]] = MapPartitionsRDD[17] at map at <console>:30

scala> val buildDF = building.map(attr => Building(attr(0).toInt,attr(1),attr(2).toInt,attr(3),attr(4)))
buildDF: org.apache.spark.rdd.RDD[Building] = MapPartitionsRDD[18] at map at <console>:34

scala> buildDF.registerTempTable("BUILDING")
<console>:37: error: value registerTempTable is not a member of org.apache.spark.rdd.RDD[Building]
       buildDF.registerTempTable("BUILDING")
               ^

scala> val buildDF = building.map(attr => Building(attr(0).toInt,attr(1),attr(2).toInt,attr(3),attr(4))).toDF
buildDF: org.apache.spark.sql.DataFrame = [buildid: int, buildingmgr: string ... 3 more fields]
```

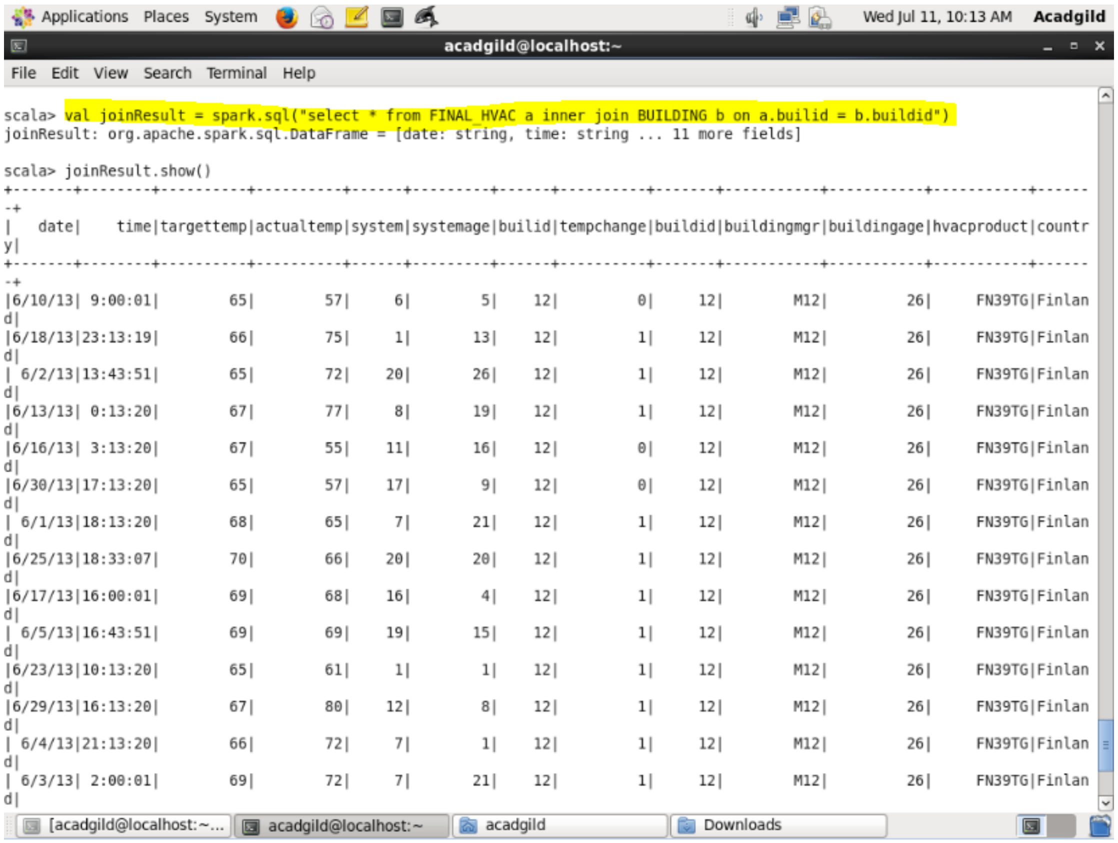
# CASE STUDY 3-Working with Sensor Data



```
scala> buildDF.registerTempTable("BUILDING")
warning: there was one deprecation warning; re-run with -deprecation for details

scala> val buildData = spark.sql("select* from BUILDING")
buildData: org.apache.spark.sql.DataFrame = [buildid: int, buildingmgr: string ... 3 more fields]

scala> buildData.show()
+-------+-----------+-----------+-----------+------------+
|buildid|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|
|      6|         M6|          9|     AC1000|   Singapore|
|      7|         M7|         13|     FN39TG|South Africa|
|      8|         M8|         25|     JDNS77|   Australia|
|      9|         M9|         11|     GG1919|      Mexico|
|     10|        M10|         23|    ACMAX22|       China|
|     11|        M11|         14|     AC1000|     Belgium|
|     12|        M12|         26|     FN39TG|     Finland|
|     13|        M13|         25|     JDNS77|Saudi Arabia|
|     14|        M14|         17|     GG1919|     Germany|
|     15|        M15|         19|    ACMAX22|      Israel|
|     16|        M16|         23|     AC1000|      Turkey|
|     17|        M17|         11|     FN39TG|       Egypt|
|     18|        M18|         25|     JDNS77|   Indonesia|
|     19|        M19|         14|     GG1919|      Canada|
|     20|        M20|         19|    ACMAX22|   Argentina|
+-------+-----------+-----------+-----------+------------+
```

**<u>Objective3</u>: Figure out the number of times, temperature has changed by 5 degrees or more for each country:**

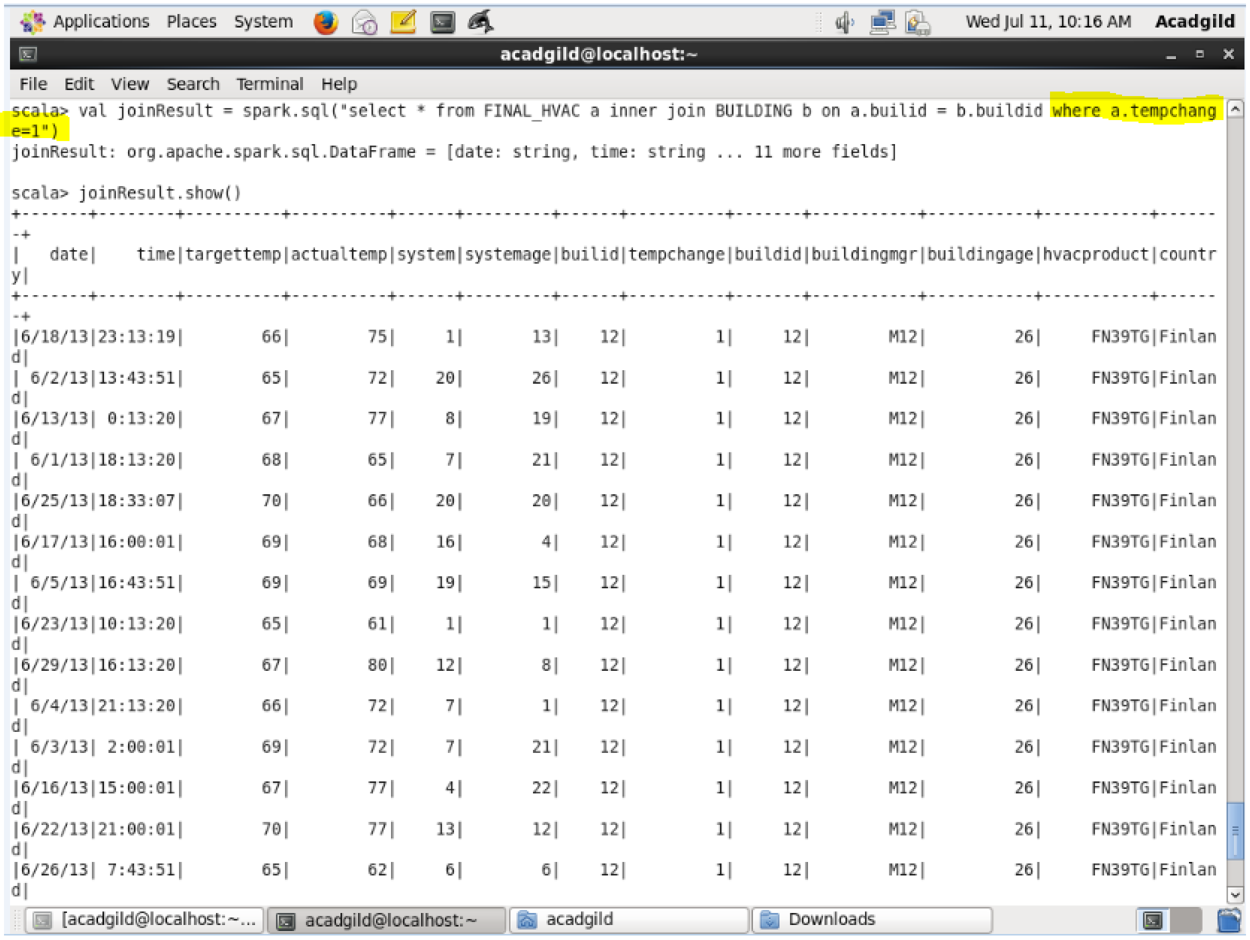
1)Join both the tables.



```
scala> val joinResult = spark.sql("select * from FINAL_HVAC a inner join BUILDING b on a.builid = b.buildid")
joinResult: org.apache.spark.sql.DataFrame = [date: string, time: string ... 11 more fields]

scala> joinResult.show()
+-------+--------+----------+----------+------+---------+------+----------+-------+-----------+-----------+-----------+------
-+
|   date|    time|targettemp|actualtemp|system|systemage|builid|tempchange|buildid|buildingmgr|buildingage|hvacproduct|countr
y|
+-------+--------+----------+----------+------+---------+------+----------+-------+-----------+-----------+-----------+------
-+
|6/10/13| 9:00:01|        65|        57|     6|        5|    12|         0|     12|        M12|         26|     FN39TG|Finlan
d|
|6/18/13|23:13:19|        66|        75|     1|       13|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
| 6/2/13|13:43:51|        65|        72|    20|       26|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
|6/13/13| 0:13:20|        67|        77|     8|       19|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
|6/16/13| 3:13:20|        67|        55|    11|       16|    12|         0|     12|        M12|         26|     FN39TG|Finlan
d|
|6/30/13|17:13:20|        65|        57|    17|        9|    12|         0|     12|        M12|         26|     FN39TG|Finlan
d|
| 6/1/13|18:13:20|        68|        65|     7|       21|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
|6/25/13|18:33:07|        70|        66|    20|       20|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
|6/17/13|16:00:01|        69|        68|    16|        4|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
| 6/5/13|16:43:51|        69|        69|    19|       15|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
|6/23/13|10:13:20|        65|        61|     1|        1|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
|6/29/13|16:13:20|        67|        80|    12|        8|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
| 6/4/13|21:13:20|        66|        72|     7|        1|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
| 6/3/13| 2:00:01|        69|        72|     7|       21|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
```

CASE STUDY 3-Working with Sensor Data

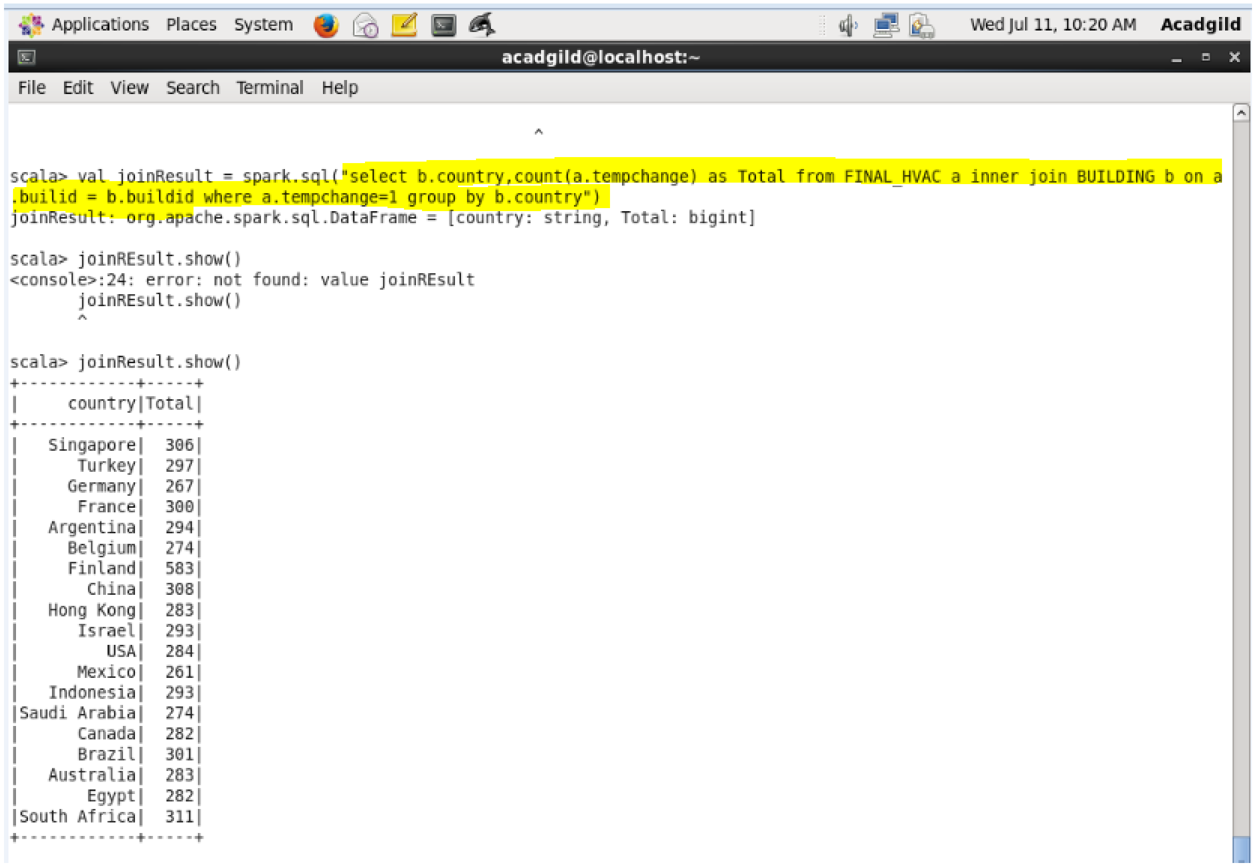
2)Select tempchange and country column



```
scala> val joinResult = spark.sql("select * from FINAL_HVAC a inner join BUILDING b on a.builid = b.buildid where a.tempchang
e=1")
joinResult: org.apache.spark.sql.DataFrame = [date: string, time: string ... 11 more fields]

scala> joinResult.show()
+-------+--------+----------+----------+------+---------+------+----------+-------+-----------+-----------+-----------+------
-+
|   date|    time|targettemp|actualtemp|system|systemage|builid|tempchange|buildid|buildingmgr|buildingage|hvacproduct|countr
y|
+-------+--------+----------+----------+------+---------+------+----------+-------+-----------+-----------+-----------+------
-+
|6/18/13|23:13:19|        66|        75|     1|       13|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
| 6/2/13|13:43:51|        65|        72|    20|       26|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
|6/13/13| 0:13:20|        67|        77|     8|       19|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
| 6/1/13|18:13:20|        68|        65|     7|       21|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
|6/25/13|18:33:07|        70|        66|    20|       20|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
|6/17/13|16:00:01|        69|        68|    16|        4|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
| 6/5/13|16:43:51|        69|        69|    19|       15|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
|6/23/13|10:13:20|        65|        61|     1|        1|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
|6/29/13|16:13:20|        67|        80|    12|        8|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
| 6/4/13|21:13:20|        66|        72|     7|        1|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
| 6/3/13| 2:00:01|        69|        72|     7|       21|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
|6/16/13|15:00:01|        67|        77|     4|       22|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
|6/22/13|21:00:01|        70|        77|    13|       12|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
|6/26/13| 7:43:51|        65|        62|     6|        6|    12|         1|     12|        M12|         26|     FN39TG|Finlan
d|
```

# CASE STUDY 3-Working with Sensor Data

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


Applications Places System
Wed Jul 11, 10:20 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help

```
scala> val joinResult = spark.sql("select b.country,count(a.tempchange) as Total from FINAL_HVAC a inner join BUILDING b on a
.builid = b.buildid where a.tempchange=1 group by b.country")
joinResult: org.apache.spark.sql.DataFrame = [country: string, Total: bigint]

scala> joinREsult.show()
<console>:24: error: not found: value joinREsult
       joinREsult.show()
       ^

scala> joinResult.show()
+------------+-----+
|     country|Total|
+------------+-----+
|   Singapore|  306|
|      Turkey|  297|
|     Germany|  267|
|      France|  300|
|   Argentina|  294|
|     Belgium|  274|
|     Finland|  583|
|       China|  308|
|   Hong Kong|  283|
|      Israel|  293|
|         USA|  284|
|      Mexico|  261|
|   Indonesia|  293|
|Saudi Arabia|  274|
|      Canada|  282|
|      Brazil|  301|
|   Australia|  283|
|       Egypt|  282|
|South Africa|  311|
+------------+-----+
```