

- Load HVAC.csv file into temporary table

```
scala> val hvacDF = spark.read.csv("/user/spark/HVAC.csv").toDF("date","time","targettemp","actualtemp","system","systemage","buildid")
hvacDF: org.apache.spark.sql.DataFrame = [date: string, time: string ... 5 more fields]

scala> hvacDF.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
```

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

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

scala> finalHVACData.show
+-----+-----+-----+-----+-----+-----+-----+
| date|   time|targettemp|actualtemp|system|systemage|buildid|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
```



```
scala> spark.sql("select * from Result where tempchange = 1").show
```

buildid	date	time	targettemp	actualtemp	system	systemage	tempchange	buildingmgr	buildingage	hvacproduct	country
17	5/2/13	1:00:01	69	68	3	20	1	M17	11	FN39TG	Egypt
18	5/3/13	2:00:01	70	73	17	20	1	M18	25	JDNS77	Indonesia
15	5/4/13	3:00:01	67	63	2	23	1	M15	19	ACMAX22	Israel
3	5/5/13	4:00:01	68	74	16	9	1	M3	28	JDNS77	Brazil
16	5/8/13	7:00:01	70	73	20	26	1	M16	23	AC1090	Turkey
9	5/9/13	8:00:01	66	69	16	9	1	M9	11	GG19L9	Mexico
15	6/11/13	10:00:01	67	70	10	17	1	M15	19	ACMAX22	Israel
15	6/13/13	12:00:01	69	73	14	2	1	M15	19	ACMAX22	Israel
6	6/14/13	13:00:01	65	61	3	2	1	M6	9	AC1090	Singapore
17	6/21/13	20:00:01	69	72	7	5	1	M17	11	FN39TG	Egypt
9	6/22/13	21:00:01	66	69	6	29	1	M9	11	GG19L9	Mexico
20	6/23/13	22:00:01	67	65	6	18	1	M20	19	ACMAX22	Argentina
18	6/25/13	0:13:19	70	71	19	14	1	M18	25	JDNS77	Indonesia
8	6/26/13	1:13:19	67	71	1	6	1	M8	25	JDNS77	Australia
14	6/27/13	2:13:19	69	68	16	14	1	M14	17	GG19L9	Germany
13	6/29/13	4:13:19	66	77	7	6	1	M13	25	JDNS77	Saudi Arabia
3	5/1/13	6:13:19	68	66	11	15	1	M3	28	JDNS77	Brazil
11	5/3/13	8:13:19	66	65	14	18	1	M11	14	AC1090	Belgium
9	5/4/13	9:13:19	67	69	1	5	1	M9	11	GG19L9	Mexico
10	5/5/13	10:13:19	70	78	5	5	1	M10	23	ACMAX22	China

only showing top 20 rows

```
scala> spark.sql("select country,count(country) from Result group by country").show
```

country	count(country)
Singapore	415
Turkey	414
Germany	363
France	400
Argentina	400
Belgium	376
Finland	814
China	425
Hong Kong	406
Israel	399
USA	393
Mexico	383
Indonesia	421
Saudi Arabia	396
Canada	380
Brazil	406
Australia	380
Egypt	406
South Africa	415