OBJECTIVE 1:-

(i)Load HVAC.csv file into temporarytable

(ii)Add a new column, tempchange -set to 1, if there is a change of greater than +/-5 betweenactual
and target temperature

(Here I load the SparkSQLUseCase1.scala and related .csv files.Here I use the IntelliJ IDEA Community Edition 2018.1.2 x64)

COMMAND:-

(i)
val data = spark.sparkContext.textFile("D:/hadoopstudy/interviewsessionII/session21/HVAC.csv");

(ii)

```
case class
hvac_cls(Date:String,Time:String,TargetTemp:Int,ActualTemp:Int,System:Int,SystemAge:Int,BuildingId:Int)
```

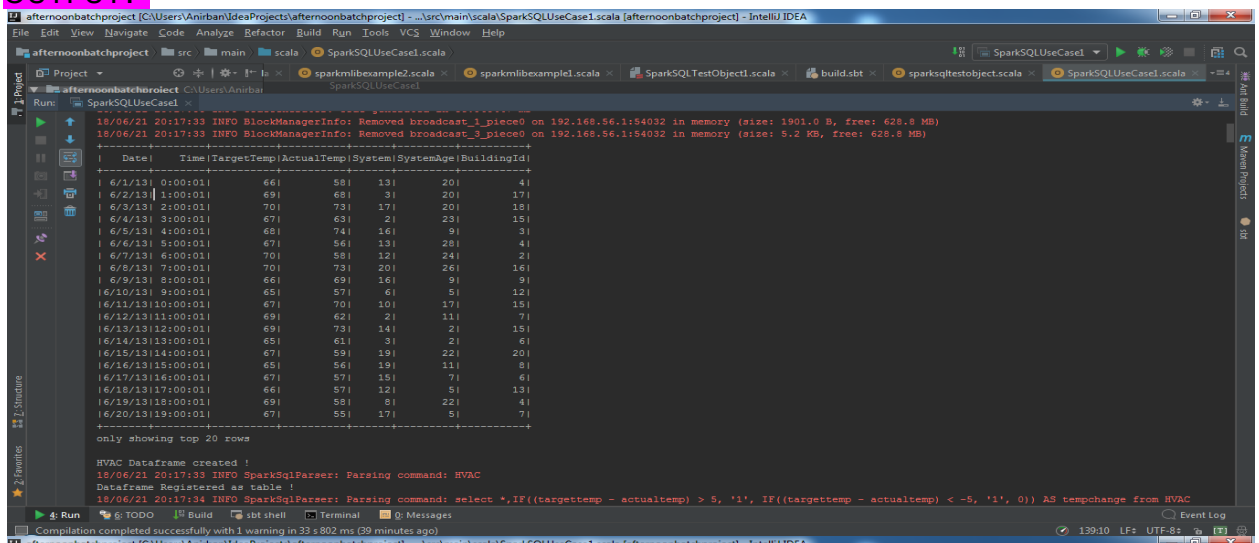
```
val header = data.first()

val data1 = data.filter(row => row != header)

println("Header removed from the data !")


//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()

hvac.show()

println("HVAC Dataframe created !")

hvac.registerTempTable("HVAC")


println("Dataframe Registered as table !")

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")
```

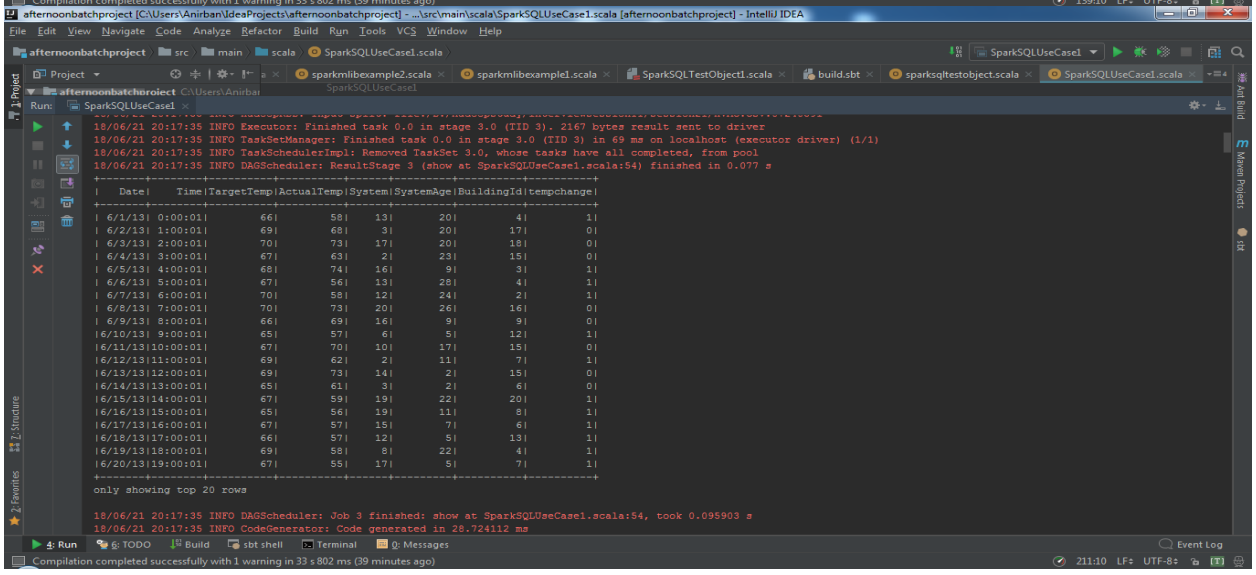
```
println("Data Frame Registered as HVAC1 table !")
```

EXPLANATION:-

(i)Here I load the HVAC.csv into the temporaytable.
(ii) Here I add a new column, tempchange -set to 1, if there is a change of greater than +/-5 betweenactual
and target temperature

OUTPUT:-





OBJECTIVE 2:-

(i)Load building.csv file into temporarytable

COMMAND:-

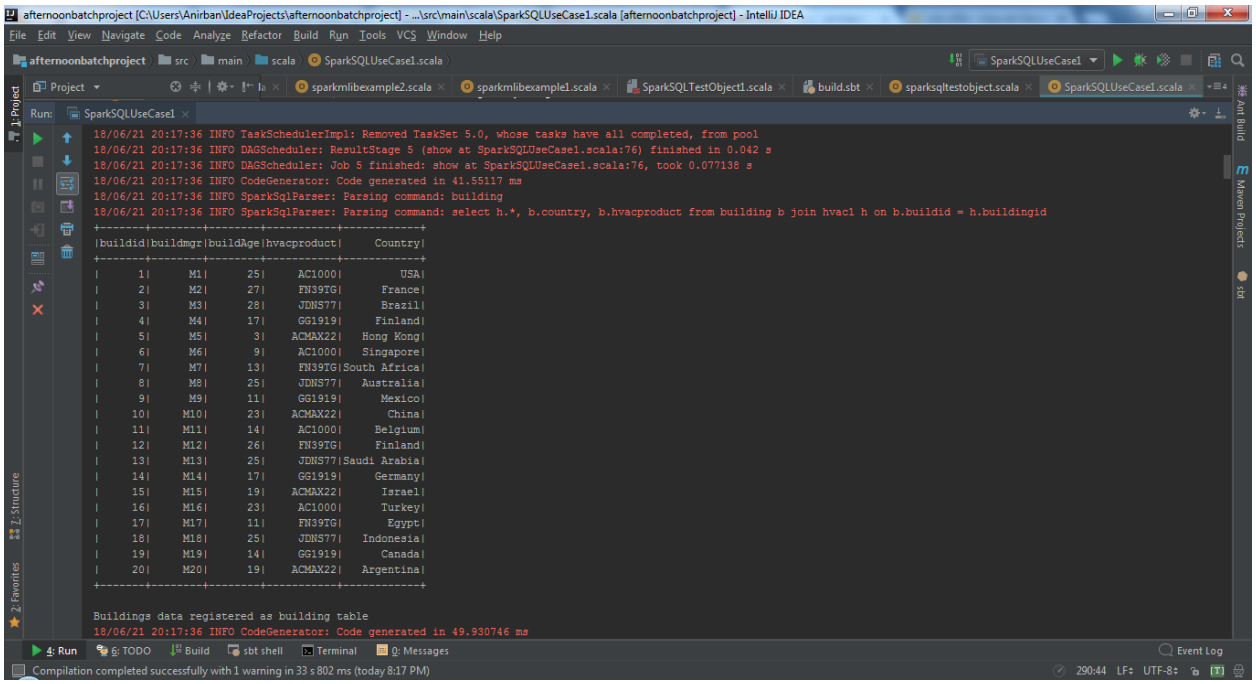
```
case class
building(buildid:Int,buildmgr:String,buildAge:Int,hvacproduct:String,Country:String)
```

```
val data2 =
spark.sparkContext.textFile("D:/hadoopstudy/interviewsessionII/session21/building.csv"
);

   val header1 = data2.first()

   val data3 = data2.filter(row => row != header1)


   println("Header removed from the building data")

   //Now let us create the building dataframe

   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()

build.registerTempTable("building")

   println("Buildings data registered as building table")
```

EXPLANATION:-

Here I load building.csv file into temporarytable.

OUTPUT:-



OBJECT 3:-
Figure out the number of times, temperature has changed by 5 degrees or more for eachcountry:
(1)Join both the tables.

(2)Select tempchange and countrycolumn

(3)Filter the rows where tempchange is 1 and count the numberof occurrence for eachcountry

COMMAND:-

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

build1.show()
```

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

tempCountry.show()
```

```
(3)val tempCountryOnes = tempCountry.filter(x=> {if(x._1==1) true else false})

tempCountryOnes.show()

tempCountryOnes.groupBy("_2").count.show
```

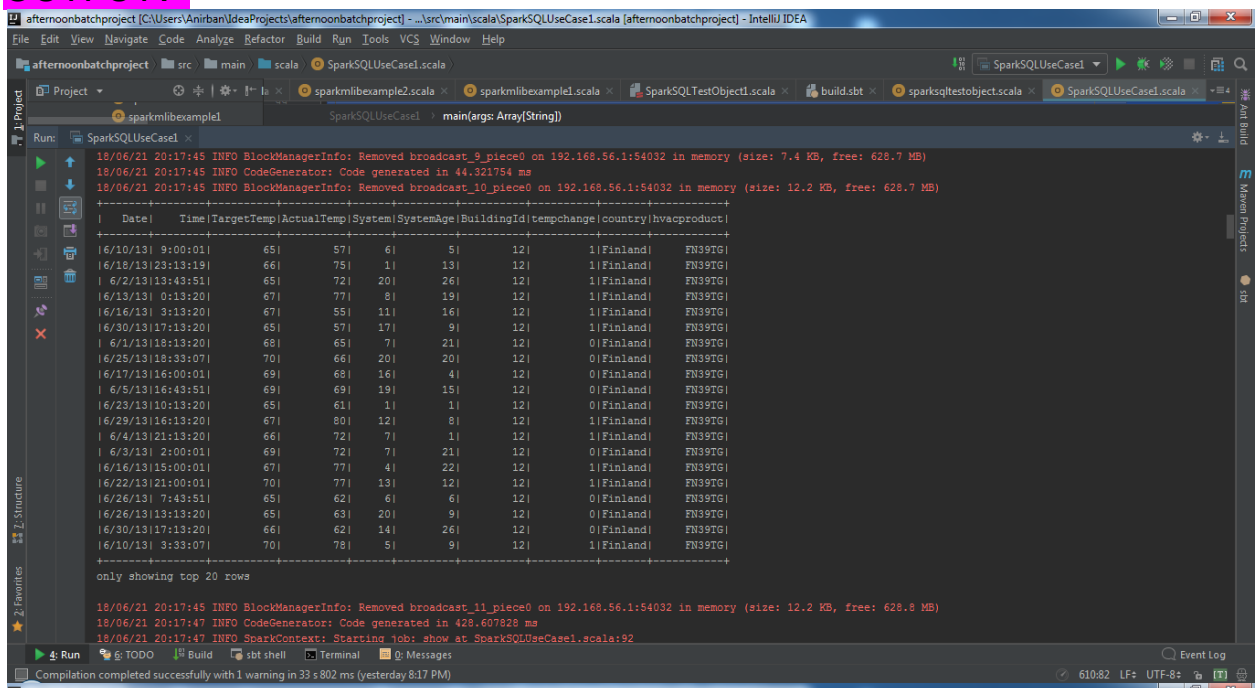
EXPLANATION:-

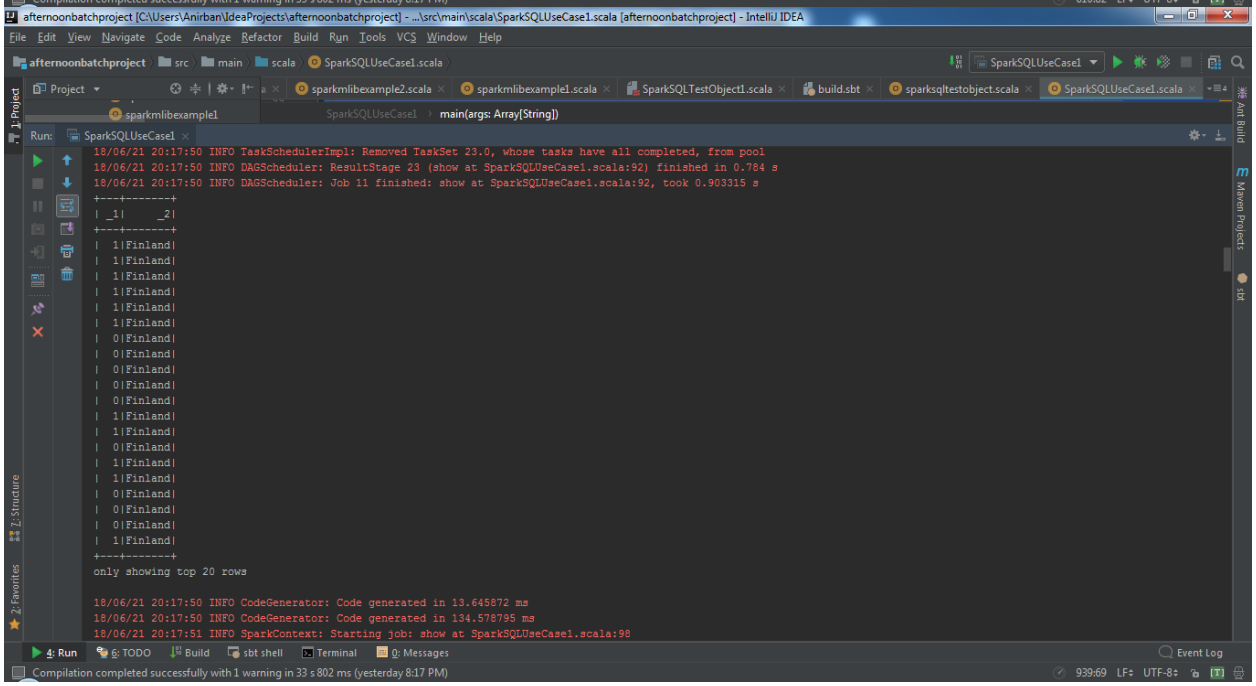
(1)Here I Join both the tables.

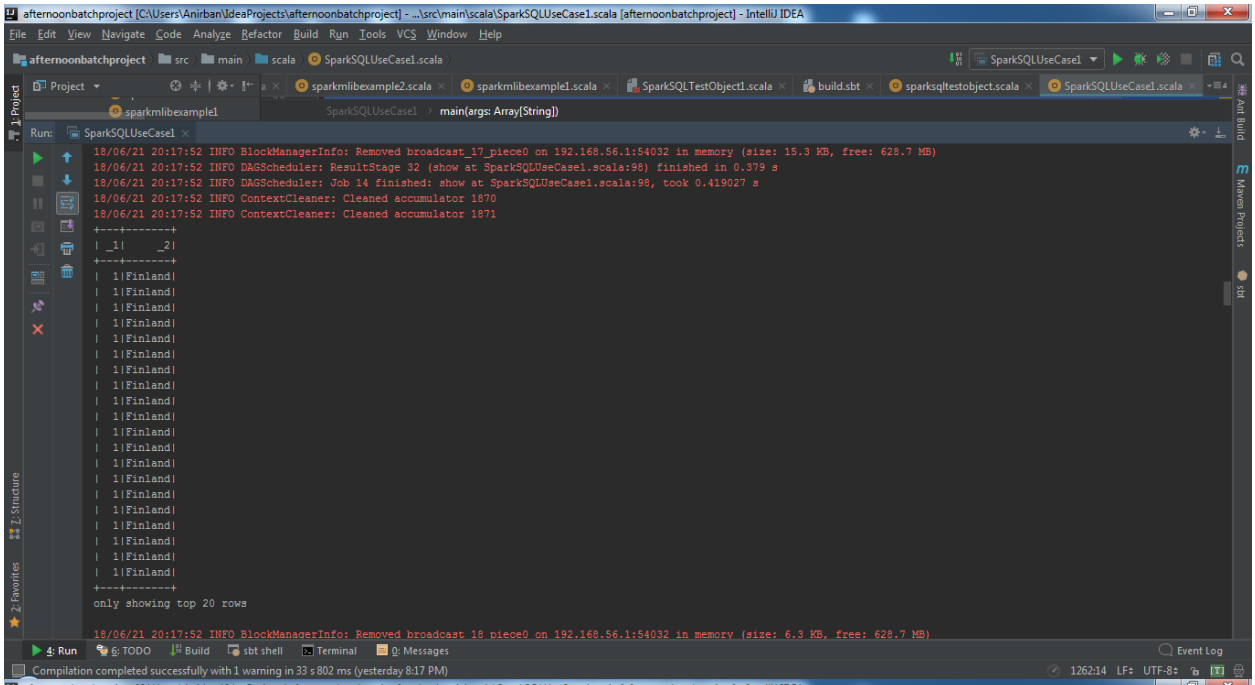
(2)Here I Select tempchange and countrycolumn

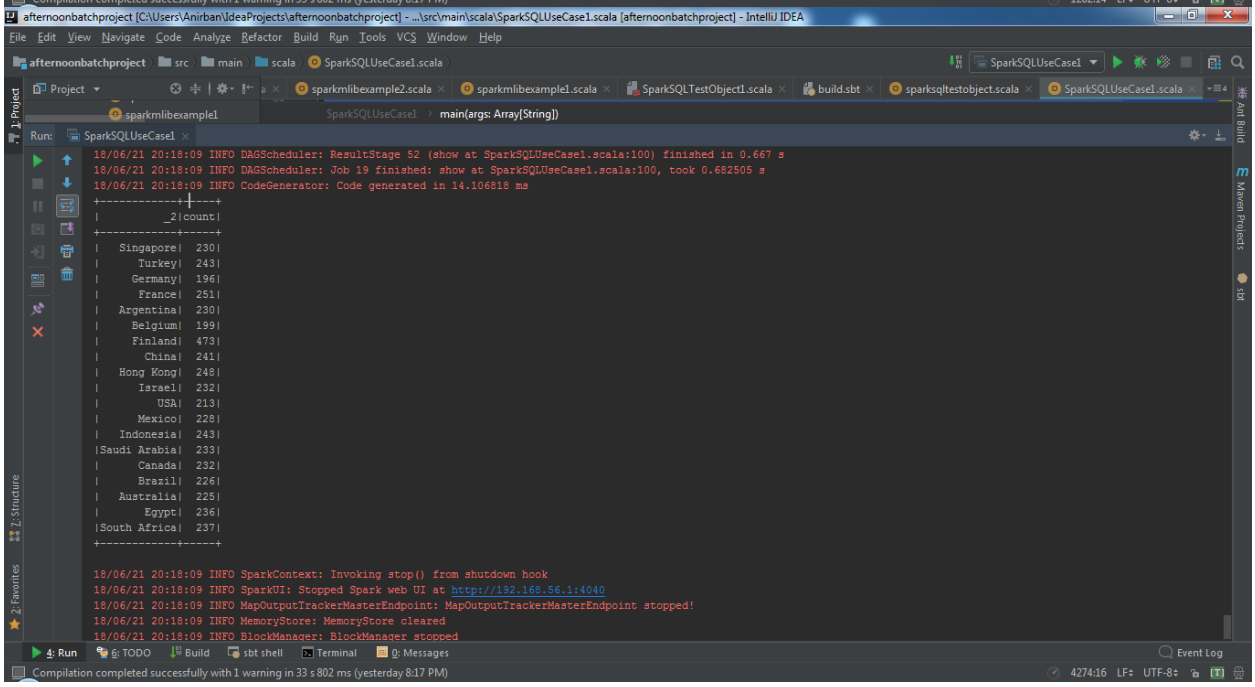
(3)Here I Filter the rows where tempchange is 1 and count the numberof occurrence for eachcountry

OUTPUT:-


afternoonbatchproject [C:\Users\Anirban\IdeaProjects\afternoonbatchproject] - ...\src\main\scala\SparkSQLUseCase1.scala [afternoonbatchproject] - IntelliJ IDEA
Run: SparkSQLUseCase1
18/06/21 20:17:45 INFO BlockManagerInfo: Removed broadcast_9_piece0 on 192.168.56.1:54032 in memory (size: 7.4 KB, free: 628.7 MB)
18/06/21 20:17:45 INFO CodeGenerator: Code generated in 44.321754 ms
18/06/21 20:17:45 INFO BlockManagerInfo: Removed broadcast_10_piece0 on 192.168.56.1:54032 in memory (size: 12.2 KB, free: 628.7 MB)
| 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
18/06/21 20:17:45 INFO BlockManagerInfo: Removed broadcast_11_piece0 on 192.168.56.1:54032 in memory (size: 12.2 KB, free: 628.8 MB)
18/06/21 20:17:47 INFO CodeGenerator: Code generated in 428.607828 ms
18/06/21 20:17:47 INFO SparkContext: Starting job: show at SparkSQLUseCase1.scala:92
Compilation completed successfully with 1 warning in 33 s 802 ms (yesterday 8:17 PM)


afternoonbatchproject [C:\Users\Anirban\IdeaProjects\afternoonbatchproject] - ...\src\main\scala\SparkSQLUseCase1.scala [afternoonbatchproject] - IntelliJ IDEA
Run: SparkSQLUseCase1
18/06/21 20:17:50 INFO TaskSchedulerImpl: Removed TaskSet 23.0, whose tasks have all completed, from pool
18/06/21 20:17:50 INFO DAGScheduler: ResultStage 23 (show at SparkSQLUseCase1.scala:92) finished in 0.784 s
18/06/21 20:17:50 INFO DAGScheduler: Job 11 finished: show at SparkSQLUseCase1.scala:92, took 0.903315 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
18/06/21 20:17:50 INFO CodeGenerator: Code generated in 13.645872 ms
18/06/21 20:17:50 INFO CodeGenerator: Code generated in 134.578795 ms
18/06/21 20:17:51 INFO SparkContext: Starting job: show at SparkSQLUseCase1.scala:98
Compilation completed successfully with 1 warning in 33 s 802 ms (yesterday 8:17 PM)


afternoonbatchproject [C:\Users\Anirban\IdeaProjects\afternoonbatchproject] - ...\src\main\scala\SparkSQLUseCase1.scala [afternoonbatchproject] - IntelliJ IDEA
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
SparkSQLUseCase1
sparkmlibexample2.scala
sparkmlibexample1.scala
SparkSQLTestObject1.scala
build.sbt
sparksqltestobject.scala
SparkSQLUseCase1.scala
main(args: Array[String])
Run: SparkSQLUseCase1
18/06/21 20:17:52 INFO BlockManagerInfo: Removed broadcast_17_piece0 on 192.168.56.1:54032 in memory (size: 15.3 KB, free: 628.7 MB)
18/06/21 20:17:52 INFO DAGScheduler: ResultStage 32 (show at SparkSQLUseCase1.scala:98) finished in 0.379 s
18/06/21 20:17:52 INFO DAGScheduler: Job 14 finished: show at SparkSQLUseCase1.scala:98, took 0.419027 s
18/06/21 20:17:52 INFO ContextCleaner: Cleaned accumulator 1870
18/06/21 20:17:52 INFO ContextCleaner: Cleaned accumulator 1871
+---+-------+
| _1| _2|
+---+-------+
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
1	Finland
+---+-------+
only showing top 20 rows
18/06/21 20:17:52 INFO BlockManagerInfo: Removed broadcast_18_piece0 on 192.168.56.1:54032 in memory (size: 6.3 KB, free: 628.7 MB)
4: Run
6: TODO
Build
sbt shell
Terminal
0: Messages
Event Log
Compilation completed successfully with 1 warning in 33 s 802 ms (yesterday 8:17 PM)
1262:14 LF: UTF-8:


afternoonbatchproject [C:\Users\Anirban\IdeaProjects\afternoonbatchproject] - ...\src\main\scala\SparkSQLUseCase1.scala [afternoonbatchproject] - IntelliJ IDEA
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
SparkSQLUseCase1
sparkmlibexample2.scala
sparkmlibexample1.scala
SparkSQLTestObject1.scala
build.sbt
sparksqltestobject.scala
SparkSQLUseCase1.scala
main(args: Array[String])
Run: SparkSQLUseCase1
18/06/21 20:18:09 INFO DAGScheduler: ResultStage 52 (show at SparkSQLUseCase1.scala:100) finished in 0.667 s
18/06/21 20:18:09 INFO DAGScheduler: Job 19 finished: show at SparkSQLUseCase1.scala:100, took 0.682505 s
18/06/21 20:18:09 INFO CodeGenerator: Code generated in 14.106818 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
+------------+-----+
18/06/21 20:18:09 INFO SparkContext: Invoking stop() from shutdown hook
18/06/21 20:18:09 INFO SparkUI: Stopped Spark web UI at http://192.168.56.1:4040
18/06/21 20:18:09 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/06/21 20:18:09 INFO MemoryStore: MemoryStore cleared
18/06/21 20:18:09 INFO BlockManager: BlockManager stopped
4: Run
6: TODO
Build
sbt shell
Terminal
0: Messages
Event Log
Compilation completed successfully with 1 warning in 33 s 802 ms (yesterday 8:17 PM)
4274:16 LF: UTF-8: