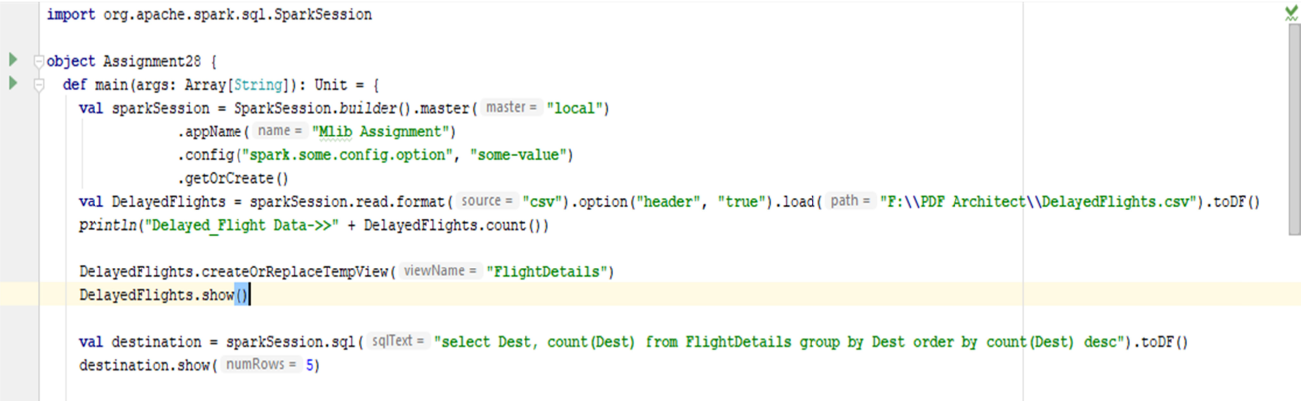
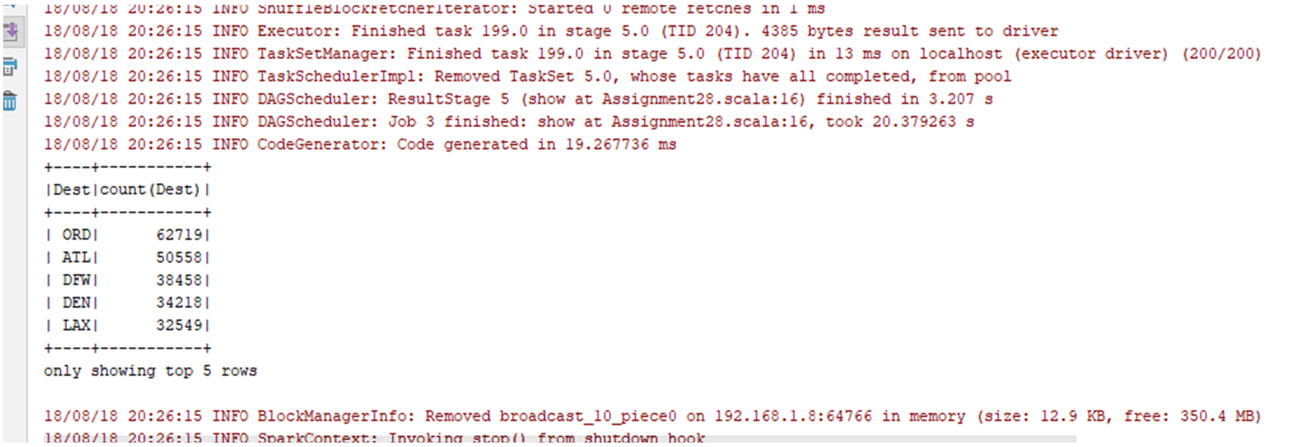
Problem Statement 1

Find out the top 5 most visited destinations.



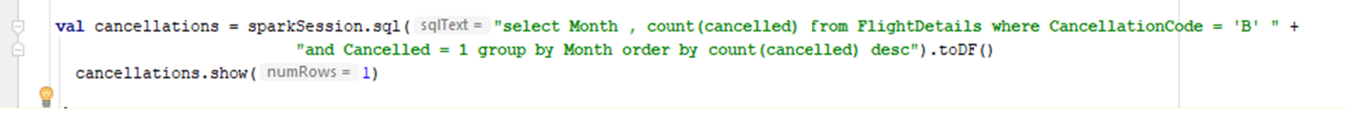
```
import org.apache.spark.sql.SparkSession

object Assignment28 {
  def main(args: Array[String]): Unit = {
    val sparkSession = SparkSession.builder().master( master = "local")
            .appName( name = "Mlib Assignment")
            .config("spark.some.config.option", "some-value")
            .getOrCreate()
    val DelayedFlights = sparkSession.read.format( source = "csv").option("header", "true").load( path = "F:\\PDF Architect\\DelayedFlights.csv").toDF()
    println("Delayed_Flight Data->>" + DelayedFlights.count())

    DelayedFlights.createOrReplaceTempView( viewName = "FlightDetails")
    DelayedFlights.show()

    val destination = sparkSession.sql( sqlText = "select Dest, count(Dest) from FlightDetails group by Dest order by count(Dest) desc").toDF()
    destination.show( numRows = 5)
```



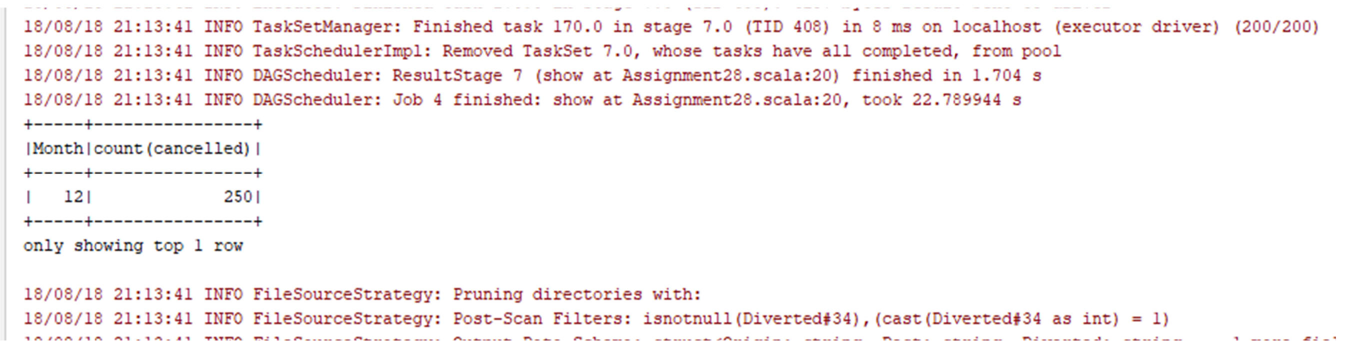
```
18/08/18 20:26:15 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
18/08/18 20:26:15 INFO Executor: Finished task 199.0 in stage 5.0 (TID 204). 4385 bytes result sent to driver
18/08/18 20:26:15 INFO TaskSetManager: Finished task 199.0 in stage 5.0 (TID 204) in 13 ms on localhost (executor driver) (200/200)
18/08/18 20:26:15 INFO TaskSchedulerImpl: Removed TaskSet 5.0, whose tasks have all completed, from pool
18/08/18 20:26:15 INFO DAGScheduler: ResultStage 5 (show at Assignment28.scala:16) finished in 3.207 s
18/08/18 20:26:15 INFO DAGScheduler: Job 3 finished: show at Assignment28.scala:16, took 20.379263 s
18/08/18 20:26:15 INFO CodeGenerator: Code generated in 19.267736 ms
+----+-----------+
|Dest|count(Dest)|
+----+-----------+
| ORD|      62719|
| ATL|      50558|
| DFW|      38458|
| DEN|      34218|
| LAX|      32549|
+----+-----------+
only showing top 5 rows

18/08/18 20:26:15 INFO BlockManagerInfo: Removed broadcast_10_piece0 on 192.168.1.8:64766 in memory (size: 12.9 KB, free: 350.4 MB)
18/08/18 20:26:15 INFO SparkContext: Invoking stop() from shutdown hook
```

Problem Statement 2

Which month has seen the most number of cancellations due to bad weather?



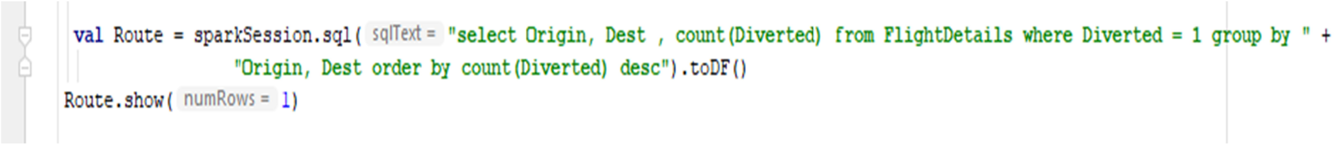
```
 val cancellations = sparkSession.sql( sqlText = "select Month , count(cancelled) from FlightDetails where CancellationCode = 'B' " +
                              "and Cancelled = 1 group by Month order by count(cancelled) desc").toDF()
  cancellations.show( numRows = 1)
```



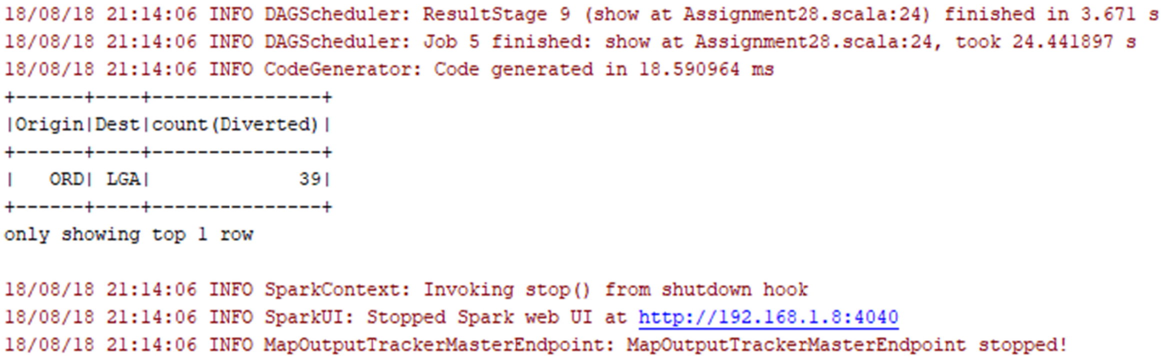
```
18/08/18 21:13:41 INFO TaskSetManager: Finished task 170.0 in stage 7.0 (TID 408) in 8 ms on localhost (executor driver) (200/200)
18/08/18 21:13:41 INFO TaskSchedulerImpl: Removed TaskSet 7.0, whose tasks have all completed, from pool
18/08/18 21:13:41 INFO DAGScheduler: ResultStage 7 (show at Assignment28.scala:20) finished in 1.704 s
18/08/18 21:13:41 INFO DAGScheduler: Job 4 finished: show at Assignment28.scala:20, took 22.789944 s
+-----+----------------+
|Month|count(cancelled)|
+-----+----------------+
|   12|             250|
+-----+----------------+
only showing top 1 row

18/08/18 21:13:41 INFO FileSourceStrategy: Pruning directories with:
18/08/18 21:13:41 INFO FileSourceStrategy: Post-Scan Filters: isnotnull(Diverted#34),(cast(Diverted#34 as int) = 1)
18/08/18 21:13:41 INFO FileSourceStrategy: Output Data Schema: struct<Origin: string, Dest: string, Diverted: string ... 1 more fiel
```

Problem Statement 3

Which route (origin & destination) has seen the maximum diversion?



```
val Route = sparkSession.sql( sqlText = "select Origin, Dest , count(Diverted) from FlightDetails where Diverted = 1 group by " +
                "Origin, Dest order by count(Diverted) desc").toDF()
Route.show( numRows = 1)
```



```
18/08/18 21:14:06 INFO DAGScheduler: ResultStage 9 (show at Assignment28.scala:24) finished in 3.671 s
18/08/18 21:14:06 INFO DAGScheduler: Job 5 finished: show at Assignment28.scala:24, took 24.441897 s
18/08/18 21:14:06 INFO CodeGenerator: Code generated in 18.590964 ms
+------+----+---------------+
|Origin|Dest|count(Diverted)|
+------+----+---------------+
|   ORD| LGA|             39|
+------+----+---------------+
only showing top 1 row

18/08/18 21:14:06 INFO SparkContext: Invoking stop() from shutdown hook
18/08/18 21:14:06 INFO SparkUI: Stopped Spark web UI at http://192.168.1.8:4040
18/08/18 21:14:06 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
```