

Hospital CASE STUDY

Task 1: Load file into spark

Code:

```
package SQL

import org.apache.spark.sql._

object HospitalCaseStudy {

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

    println("hey scala")

    val spark = SparkSession
      .builder()
      .master(master = "local")
      .appName(name = "Hospital Case Study")
      .config("spark.some.config.option", "some-value")
      .getOrCreate()

    println("Spark Session Object Created")

    spark.sparkContext.setLogLevel("WARN")

    val df1 = spark.sqlContext.read.option("header", "true").option("inferSchema",
      "true").csv("C:/Users/Gaurav/Desktop/inpatientCharges.csv")
    df1.show()
    df1.createOrReplaceTempView("hospital_charges")

  }
}
```

Output:

```

ScalaSparkProject [C:\Users\Gaurav\IdeaProjects\ScalaSparkProject] - ...src\main\scala\SQL\HospitalCaseStudy.scala [ScalaSparkProject] - IntelliJ IDEA
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
ScalaSparkProject src main scala SQL HospitalCaseStudy.scala
Run: HospitalCaseStudy x
"C:\Program Files\Java\jdk1.8.0_181\bin\java.exe" ...
hey scala
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
19/01/08 22:42:53 INFO SparkContext: Running Spark version 2.1.0
19/01/08 22:42:54 INFO SecurityManager: Changing view acls to: Gaurav
19/01/08 22:42:54 INFO SecurityManager: Changing modify acls to: Gaurav
19/01/08 22:42:54 INFO SecurityManager: Changing view acls groups to:
19/01/08 22:42:54 INFO SecurityManager: Changing modify acls groups to:
19/01/08 22:42:54 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(Gaurav); groups with view permissions: Set();
19/01/08 22:42:56 INFO Utils: Successfully started service 'sparkDriver' on port 53137.
19/01/08 22:42:56 INFO SparkEnv: Registering MapOutputTracker
19/01/08 22:42:56 INFO SparkEnv: Registering BlockManagerMaster
19/01/08 22:42:56 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
19/01/08 22:42:56 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
19/01/08 22:42:56 INFO DiskBlockManager: Created local directory at C:\Users\Gaurav\AppData\Local\Temp\blockmgr-9b5ab775-74a5-4381-8524-235b453f5749
19/01/08 22:42:56 INFO MemoryStore: MemoryStore started with capacity 1170.6 MB
19/01/08 22:42:56 INFO SparkEnv: Registering OutputCommitCoordinator
19/01/08 22:42:56 INFO Utils: Successfully started service 'SparkUI' on port 4040.
19/01/08 22:42:56 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://192.168.213.1:4040
19/01/08 22:42:57 INFO Executor: Starting executor ID driver on host localhost
19/01/08 22:42:57 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 53146.
19/01/08 22:42:57 INFO NettyBlockTransferService: Server created on 192.168.213.1:53146
19/01/08 22:42:57 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
19/01/08 22:42:57 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 192.168.213.1, 53146, None)
19/01/08 22:42:57 INFO BlockManagerMasterEndpoint: Registering block manager 192.168.213.1:53146 with 1170.6 MB RAM, BlockManagerId(driver, 192.168.213.1, 53146, None)
19/01/08 22:42:57 INFO BlockManager: Registered BlockManager BlockManagerId(driver, 192.168.213.1, 53146, None)
19/01/08 22:42:57 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, 192.168.213.1, 53146, None)
19/01/08 22:42:57 INFO SharedState: Warehouse path is 'file:/C:/Users/Gaurav/IdeaProjects/ScalaSparkProject/spark-warehouse/'.
Spark Session Object Created
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| DRGDefinition|ProviderId| ProviderName|ProviderStreetAddress|ProviderCity|ProviderState|ProviderZipCode|HospitalReferralRegionDescription|TotalDischarges|AverageCov|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 039 - EXTRACRANIA...| 10001|SOUTHEAST ALABAMA...| 1108 ROSS CLARK C...| DOTHAN| AL| 36301| AL - Dothan| 91|
| 039 - EXTRACRANIA...| 10005|MARSHALL MEDICAL...| 2505 U S HIGHWAY ...| BOAZ| AL| 35957| AL - Birmingham| 14|
| 039 - EXTRACRANIA...| 10006|ELIZA COFFEE MEMO...| 205 MARENGO STREET| FLORENCE| AL| 35631| AL - Birmingham| 24|
| 039 - EXTRACRANIA...| 10011| ST VINCENT'S EAST| 50 MEDICAL PARK E...| BIRMINGHAM| AL| 35235| AL - Birmingham| 25|
| 039 - EXTRACRANIA...| 10016|SHELLEY BAPTIST ME...| 1000 FIRST STREET...| ALABASTER| AL| 35007| AL - Birmingham| 18|
| 039 - EXTRACRANIA...| 10023|BAPTIST MEDICAL C...| 2105 EAST SOUTH B...| MONTGOMERY| AL| 36116| AL - Montgomery| 67|
| 039 - EXTRACRANIA...| 10029|EAST ALABAMA MEDI...| 2000 PEPPERELL PA...| OPELIKA| AL| 36801| AL - Birmingham| 51|
| 039 - EXTRACRANIA...| 10033|UNIVERSITY OF ALA...| 619 SOUTH 19TH ST...| BIRMINGHAM| AL| 35233| AL - Birmingham| 32|
| 039 - EXTRACRANIA...| 10039| HUNTSVILLE HOSPITAL| 163 CIVIL RD| HUNTSVILLE| AL| 35891| AL - Huntsville| 135|
| 039 - EXTRACRANIA...| 10040|GADSDEN REGIONAL ...| 1007 GOODYEAR AVENUE| GADSDEN| AL| 35903| AL - Birmingham| 34|
| 039 - EXTRACRANIA...| 10046|RIVERVIEW REGIONA...| 600 SOUTH THIRD S...| GADSDEN| AL| 35901| AL - Birmingham| 14|
| 039 - EXTRACRANIA...| 10055| FLOWERS HOSPITAL| 4370 WEST MAIN ST...| DOTHAN| AL| 36305| AL - Dothan| 45|
| 039 - EXTRACRANIA...| 10056|ST VINCENT'S BIRM...| 810 ST VINCENT'S ...| BIRMINGHAM| AL| 35205| AL - Birmingham| 43|
| 039 - EXTRACRANIA...| 10078|NORTHEAST ALABAMA...| 400 EAST 10TH STREET| ANNISTON| AL| 36207| AL - Birmingham| 21|
| 039 - EXTRACRANIA...| 10083|SOUTH BALDWIN REG...| 1613 NORTH MCKENZ...| FOLEY| AL| 36535| AL - Mobile| 15|
| 039 - EXTRACRANIA...| 10085|DECATUR GENERAL H...| 1201 7TH STREET SE| DECATUR| AL| 35609| AL - Huntsville| 27|
| 039 - EXTRACRANIA...| 10090| PROVIDENCE HOSPITAL| 6801 AIRPORT BOUL...| MOBILE| AL| 36608| AL - Mobile| 27|
| 039 - EXTRACRANIA...| 10092|D C H REGIONAL ME...| 809 UNIVERSITY BO...| TUSCALOOSA| AL| 35401| AL - Tuscaloosa| 31|
| 039 - EXTRACRANIA...| 10100| THOMAS HOSPITAL| 750 MORPHY AVENUE| FAIRHOPE| AL| 36532| AL - Mobile| 18|
| 039 - EXTRACRANIA...| 10103|BAPTIST MEDICAL C...| 701 PRINCETON AVE...| BIRMINGHAM| AL| 35211| AL - Birmingham| 33|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows
Process finished with exit code 0
Compilation completed successfully in 6 s 442 ms (2 minutes ago)
653 chars, 27 line breaks 1:1 CRLF + UTF-8 +
22:44 08-01-2019

```

```

ScalaSparkProject [C:\Users\Gaurav\IdeaProjects\ScalaSparkProject] - ...src\main\scala\SQL\HospitalCaseStudy.scala [ScalaSparkProject] - IntelliJ IDEA
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
ScalaSparkProject src main scala SQL HospitalCaseStudy.scala
Run: HospitalCaseStudy x
"C:\Program Files\Java\jdk1.8.0_181\bin\java.exe" ...
hey scala
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
19/01/08 22:42:53 INFO SparkContext: Running Spark version 2.1.0
19/01/08 22:42:54 INFO SecurityManager: Changing view acls to: Gaurav
19/01/08 22:42:54 INFO SecurityManager: Changing modify acls to: Gaurav
19/01/08 22:42:54 INFO SecurityManager: Changing view acls groups to:
19/01/08 22:42:54 INFO SecurityManager: Changing modify acls groups to:
19/01/08 22:42:54 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(Gaurav); groups with view permissions: Set();
19/01/08 22:42:56 INFO Utils: Successfully started service 'sparkDriver' on port 53137.
19/01/08 22:42:56 INFO SparkEnv: Registering MapOutputTracker
19/01/08 22:42:56 INFO SparkEnv: Registering BlockManagerMaster
19/01/08 22:42:56 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
19/01/08 22:42:56 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
19/01/08 22:42:56 INFO DiskBlockManager: Created local directory at C:\Users\Gaurav\AppData\Local\Temp\blockmgr-9b5ab775-74a5-4381-8524-235b453f5749
19/01/08 22:42:56 INFO MemoryStore: MemoryStore started with capacity 1170.6 MB
19/01/08 22:42:56 INFO SparkEnv: Registering OutputCommitCoordinator
19/01/08 22:42:56 INFO Utils: Successfully started service 'SparkUI' on port 4040.
19/01/08 22:42:56 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://192.168.213.1:4040
19/01/08 22:42:57 INFO Executor: Starting executor ID driver on host localhost
19/01/08 22:42:57 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 53146.
19/01/08 22:42:57 INFO NettyBlockTransferService: Server created on 192.168.213.1:53146
19/01/08 22:42:57 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
19/01/08 22:42:57 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 192.168.213.1, 53146, None)
19/01/08 22:42:57 INFO BlockManagerMasterEndpoint: Registering block manager 192.168.213.1:53146 with 1170.6 MB RAM, BlockManagerId(driver, 192.168.213.1, 53146, None)
19/01/08 22:42:57 INFO BlockManager: Registered BlockManager BlockManagerId(driver, 192.168.213.1, 53146, None)
19/01/08 22:42:57 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, 192.168.213.1, 53146, None)
19/01/08 22:42:57 INFO SharedState: Warehouse path is 'file:/C:/Users/Gaurav/IdeaProjects/ScalaSparkProject/spark-warehouse/'.
Spark Session Object Created
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| DRGDefinition|ProviderId| ProviderName|ProviderStreetAddress|ProviderCity|ProviderState|ProviderZipCode|HospitalReferralRegionDescription|TotalDischarges|AverageCov|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 039 - EXTRACRANIA...| 10001|SOUTHEAST ALABAMA...| 1108 ROSS CLARK C...| DOTHAN| AL| 36301| AL - Dothan| 91|
| 039 - EXTRACRANIA...| 10005|MARSHALL MEDICAL...| 2505 U S HIGHWAY ...| BOAZ| AL| 35957| AL - Birmingham| 14|
| 039 - EXTRACRANIA...| 10006|ELIZA COFFEE MEMO...| 205 MARENGO STREET| FLORENCE| AL| 35631| AL - Birmingham| 24|
| 039 - EXTRACRANIA...| 10011| ST VINCENT'S EAST| 50 MEDICAL PARK E...| BIRMINGHAM| AL| 35235| AL - Birmingham| 25|
| 039 - EXTRACRANIA...| 10016|SHELLEY BAPTIST ME...| 1000 FIRST STREET...| ALABASTER| AL| 35007| AL - Birmingham| 18|
| 039 - EXTRACRANIA...| 10023|BAPTIST MEDICAL C...| 2105 EAST SOUTH B...| MONTGOMERY| AL| 36116| AL - Montgomery| 67|
| 039 - EXTRACRANIA...| 10029|EAST ALABAMA MEDI...| 2000 PEPPERELL PA...| OPELIKA| AL| 36801| AL - Birmingham| 51|
| 039 - EXTRACRANIA...| 10033|UNIVERSITY OF ALA...| 619 SOUTH 19TH ST...| BIRMINGHAM| AL| 35233| AL - Birmingham| 32|
| 039 - EXTRACRANIA...| 10039| HUNTSVILLE HOSPITAL| 163 CIVIL RD| HUNTSVILLE| AL| 35891| AL - Huntsville| 135|
| 039 - EXTRACRANIA...| 10040|GADSDEN REGIONAL ...| 1007 GOODYEAR AVENUE| GADSDEN| AL| 35903| AL - Birmingham| 34|
| 039 - EXTRACRANIA...| 10046|RIVERVIEW REGIONA...| 600 SOUTH THIRD S...| GADSDEN| AL| 35901| AL - Birmingham| 14|
| 039 - EXTRACRANIA...| 10055| FLOWERS HOSPITAL| 4370 WEST MAIN ST...| DOTHAN| AL| 36305| AL - Dothan| 45|
| 039 - EXTRACRANIA...| 10056|ST VINCENT'S BIRM...| 810 ST VINCENT'S ...| BIRMINGHAM| AL| 35205| AL - Birmingham| 43|
| 039 - EXTRACRANIA...| 10078|NORTHEAST ALABAMA...| 400 EAST 10TH STREET| ANNISTON| AL| 36207| AL - Birmingham| 21|
| 039 - EXTRACRANIA...| 10083|SOUTH BALDWIN REG...| 1613 NORTH MCKENZ...| FOLEY| AL| 36535| AL - Mobile| 15|
| 039 - EXTRACRANIA...| 10085|DECATUR GENERAL H...| 1201 7TH STREET SE| DECATUR| AL| 35609| AL - Huntsville| 27|
| 039 - EXTRACRANIA...| 10090| PROVIDENCE HOSPITAL| 6801 AIRPORT BOUL...| MOBILE| AL| 36608| AL - Mobile| 27|
| 039 - EXTRACRANIA...| 10092|D C H REGIONAL ME...| 809 UNIVERSITY BO...| TUSCALOOSA| AL| 35401| AL - Tuscaloosa| 31|
| 039 - EXTRACRANIA...| 10100| THOMAS HOSPITAL| 750 MORPHY AVENUE| FAIRHOPE| AL| 36532| AL - Mobile| 18|
| 039 - EXTRACRANIA...| 10103|BAPTIST MEDICAL C...| 701 PRINCETON AVE...| BIRMINGHAM| AL| 35211| AL - Birmingham| 33|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows
Process finished with exit code 0
Compilation completed successfully in 6 s 442 ms (2 minutes ago)
653 chars, 27 line breaks 1:1 CRLF + UTF-8 +
22:44 08-01-2019

```

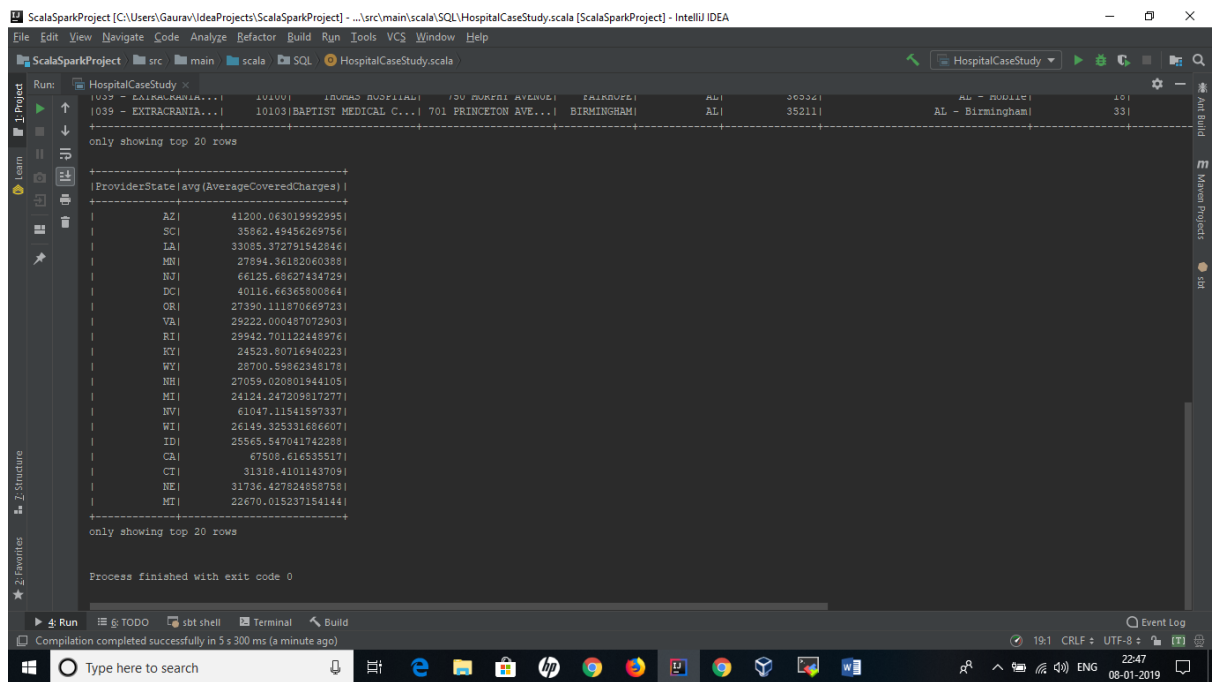
Task 2:

1. What is the average amount of AverageCoveredCharges per state.

Code:

```
df1.groupBy("ProviderState").avg("AverageCoveredCharges").show
```

Output:

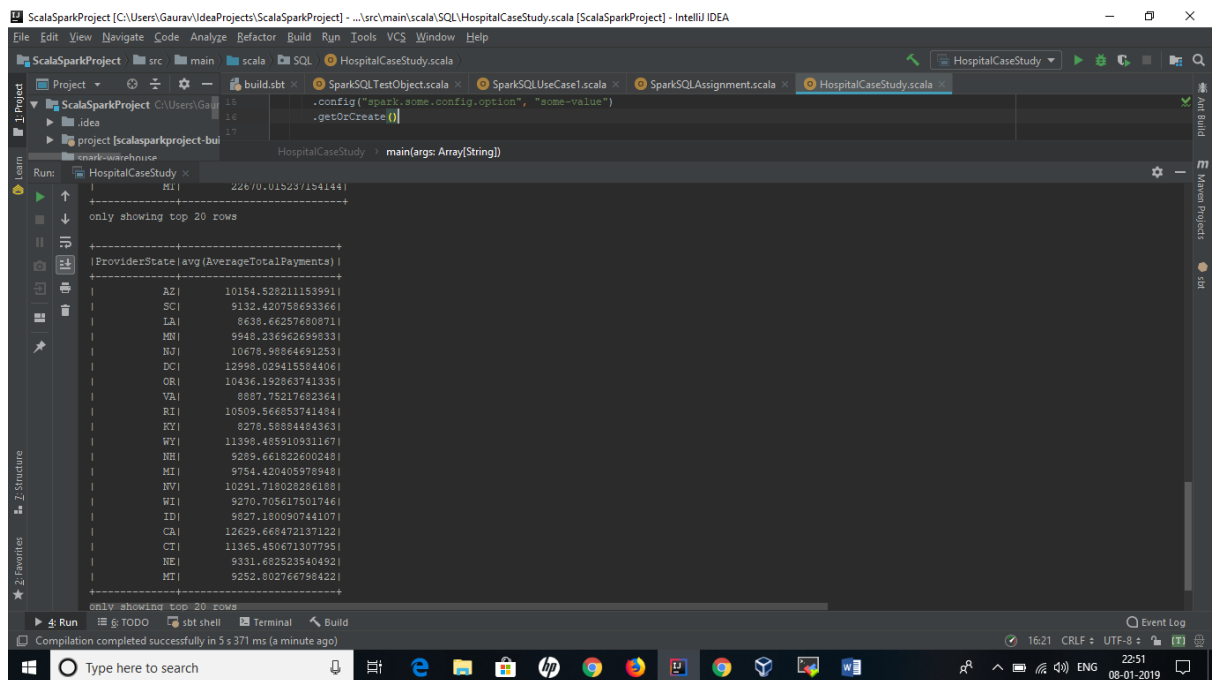


2. find out the AverageTotalPayments charges per state.

Code:

```
df1.groupBy("ProviderState").avg("AverageTotalPayments").show
```

Output:

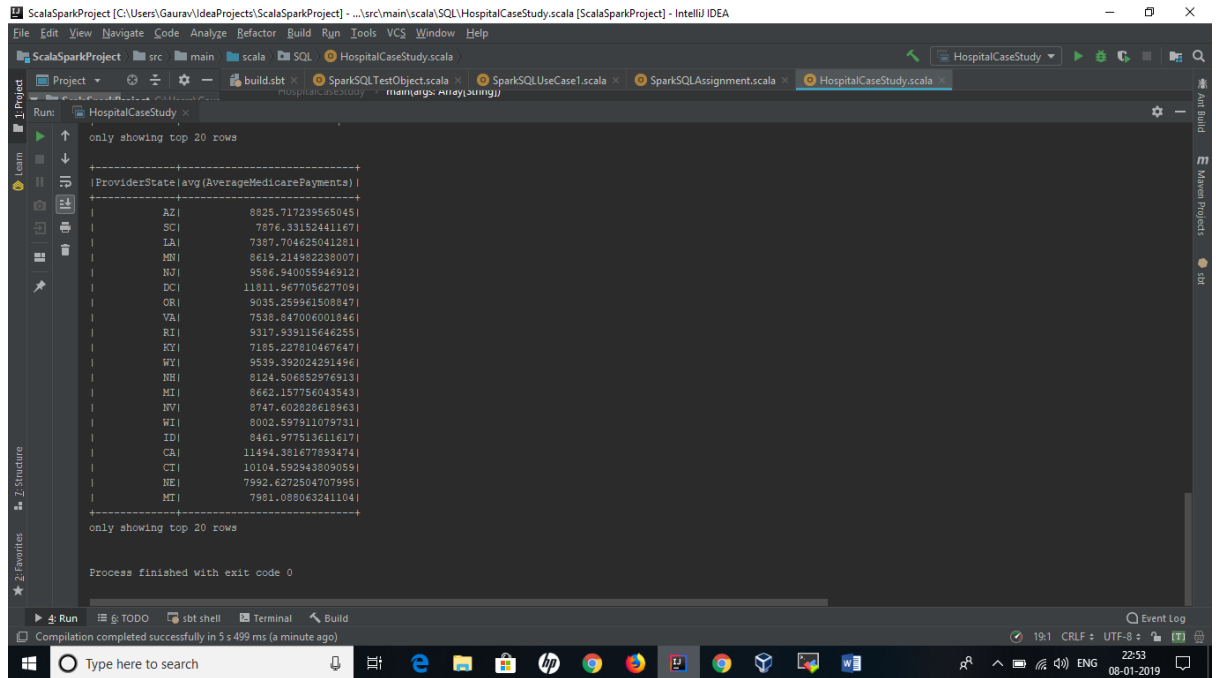


3. find out the AverageMedicarePayments charges per state.

Code:

```
df1.groupBy("ProviderState").avg("AverageMedicarePayments").show
```

Output:



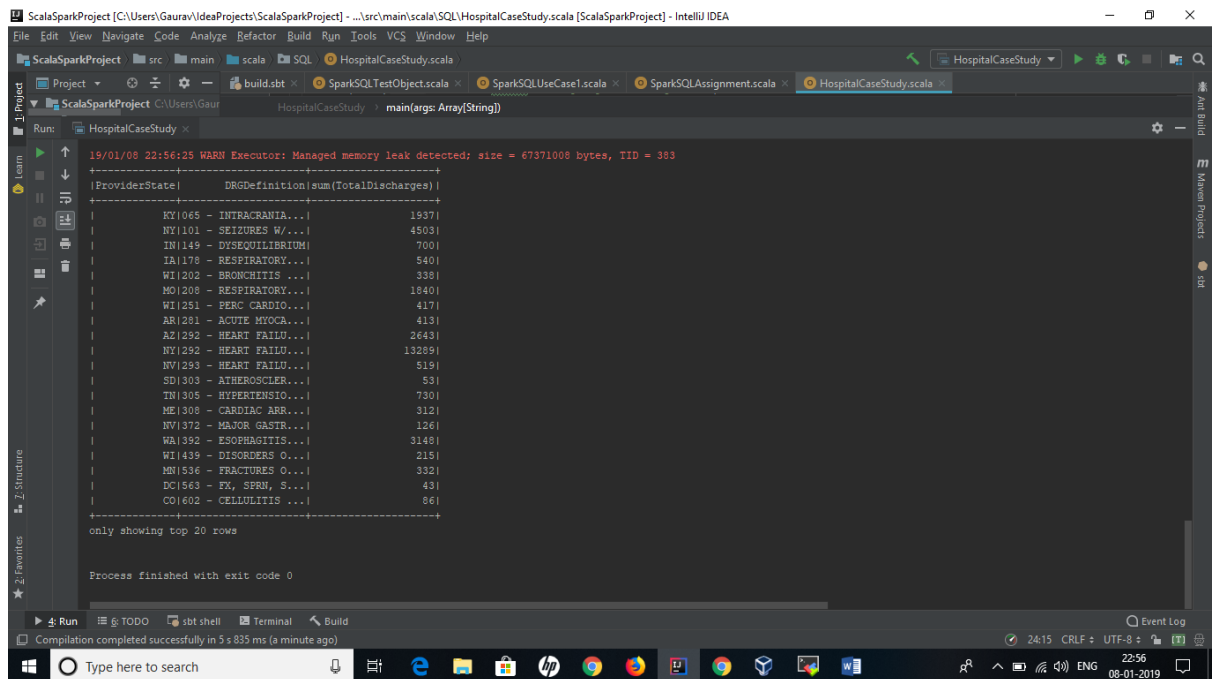
Task 3:

- Find out the total number of Discharges per state and for each disease.

Code:

```
df1.groupBy("ProviderState","DRGDefinition").sum("TotalDischarges").show
```

Output:



- Sort the output in descending order of totalDischarges.

Code:

```
df1.groupBy("ProviderState","DRGDefinition").sum("TotalDischarges").orderBy(desc  
(sum("TotalDischarges").toString)).show
```