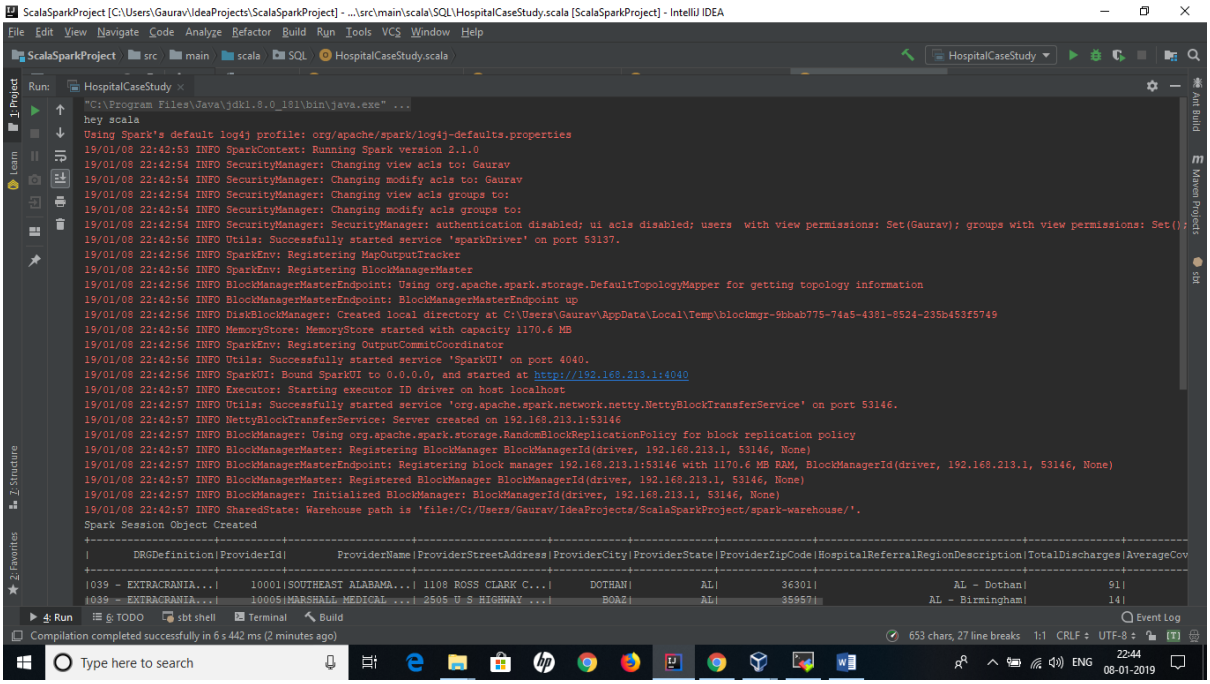
Hospital CASE STUDY

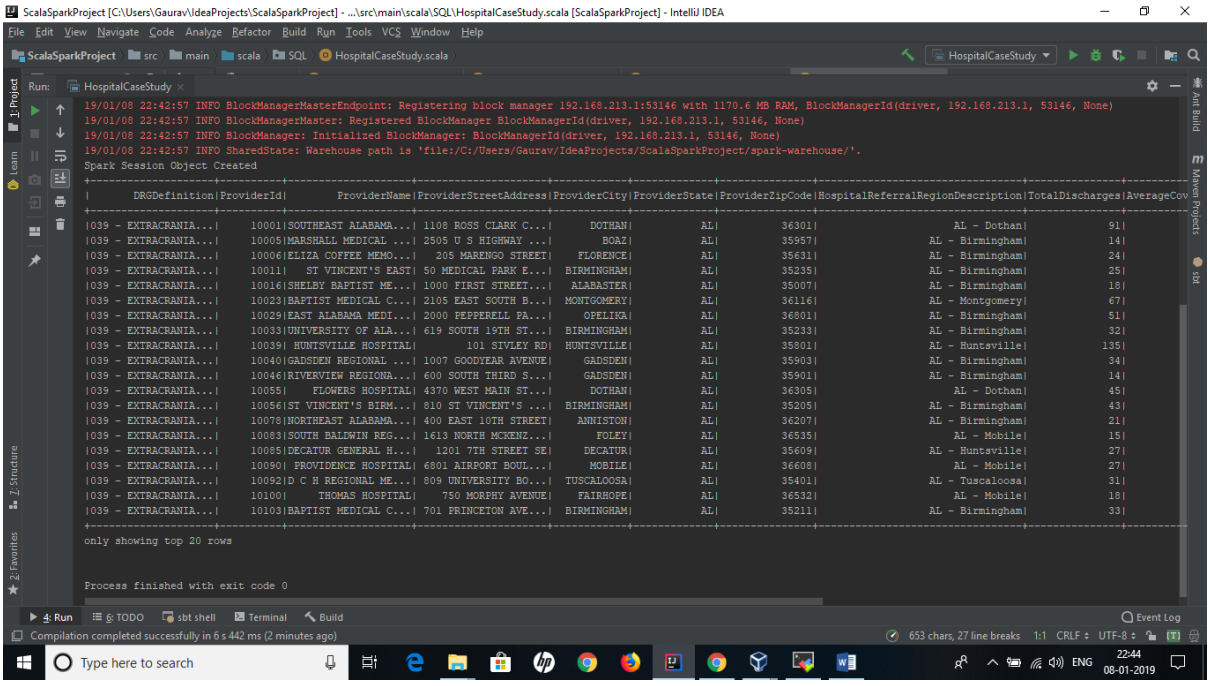
Task 1: Load file into spark

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:





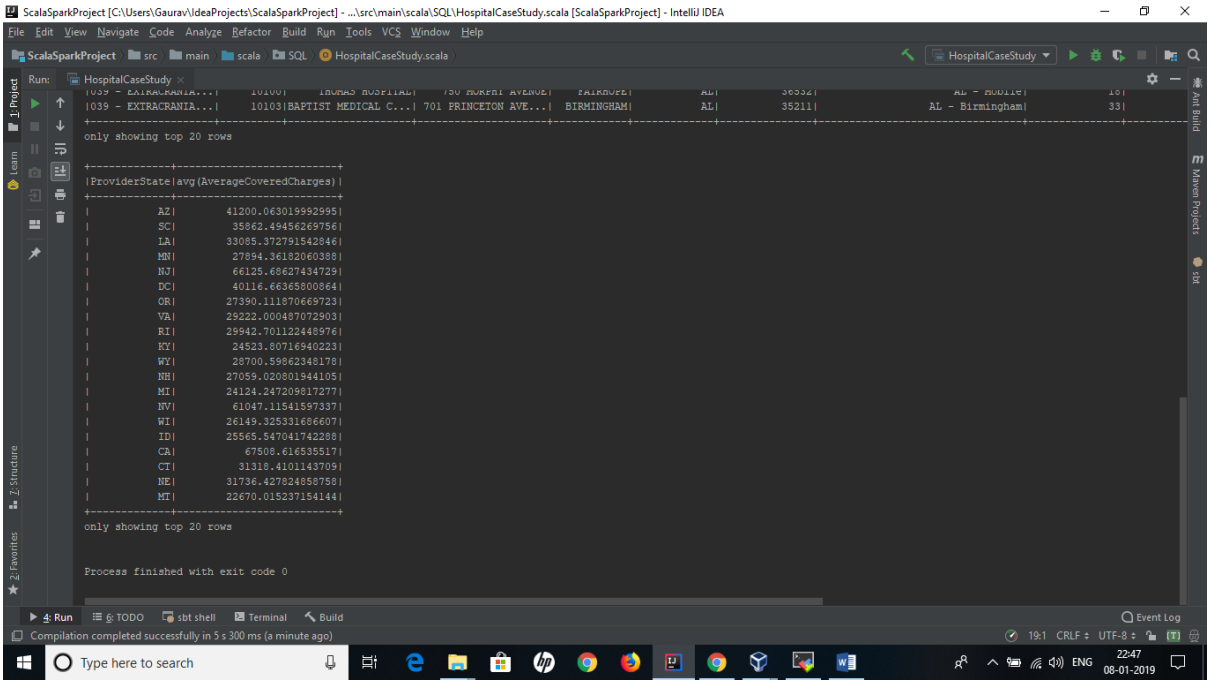
Task 2:

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

Code:

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

Output:

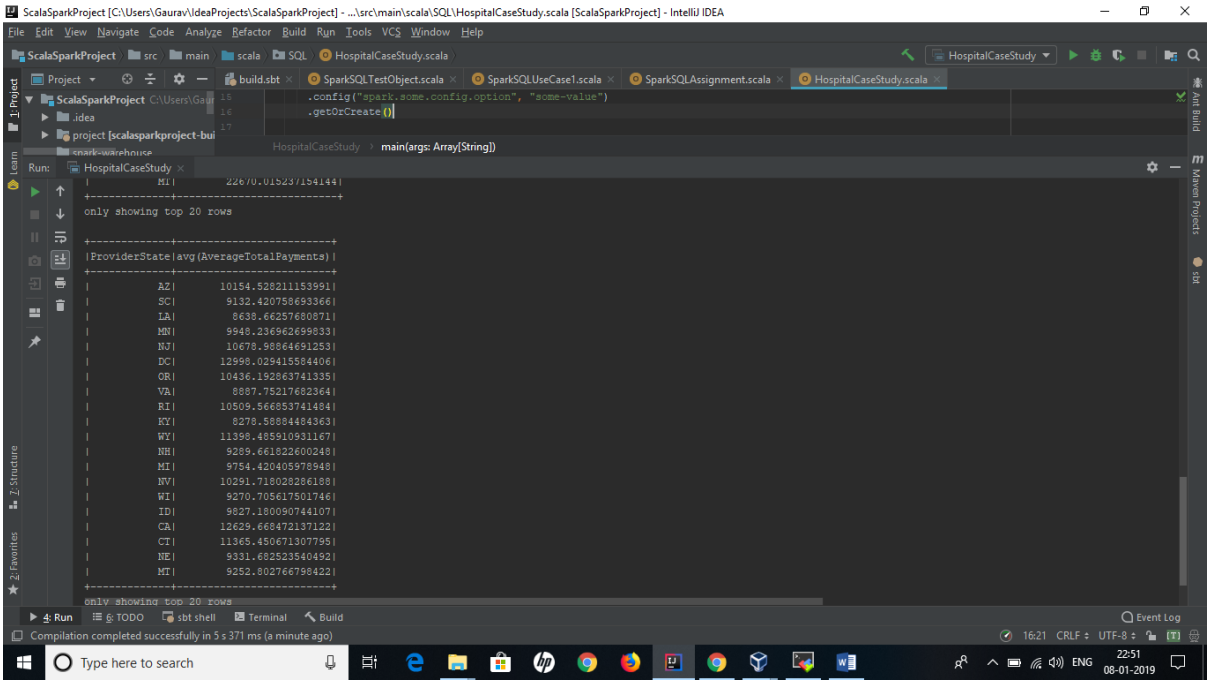


2. find out the AverageTotalPayments charges per state.

Code:

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

Output:

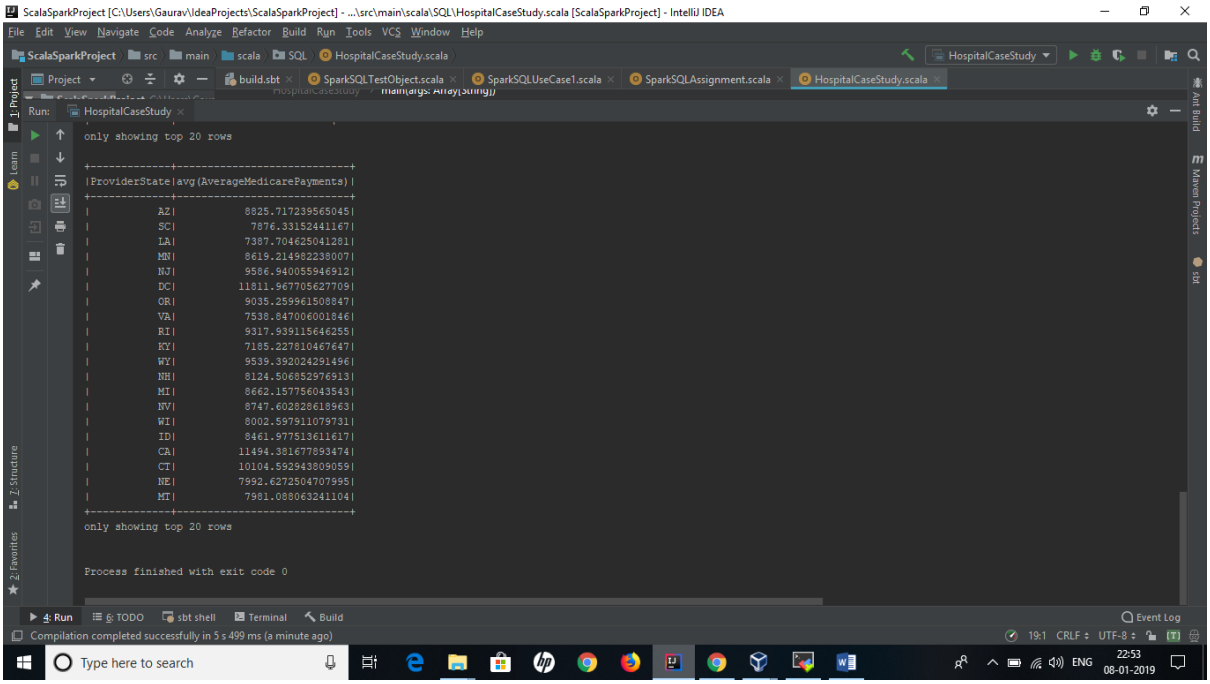


3. find out the AverageMedicarePayments charges per state.

Code:

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

Output:



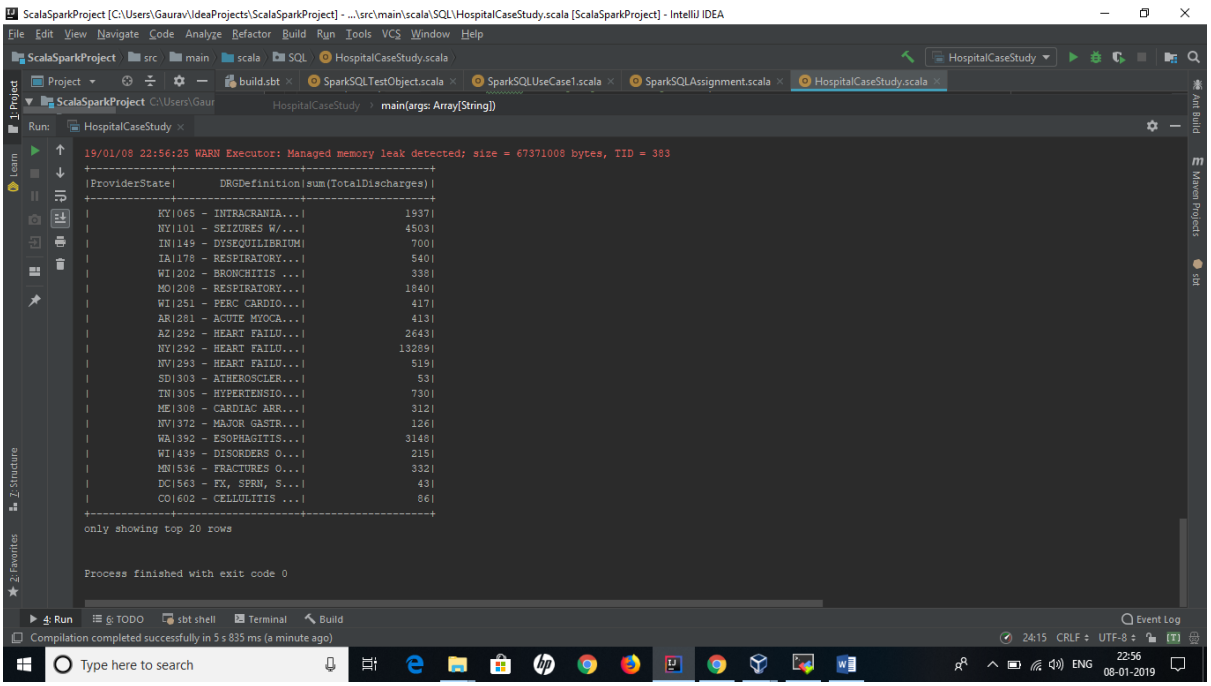
Task 3:

- 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
```