

Hospital Charges Analysis

CASE STUDY – 4


Date: 11-Mar-2019

Input File: inpatientCharges.csv

Objective – 1

Load file in to Spark

Program:

```
object HospitalCaseStudy {  
  def main(args : Array[String]) : Unit = {  
    Logger.getLogger("org").setLevel(Level.ERROR)  
  
    // Creating a spark context object to run in local machine  
    val spark = SparkSession.builder.appName("Hospital Charges Analysis").master("local").getOrCreate()  
    import spark.implicits._  
  
    // Create a dataframe and load the building csv file  
    val dfInpatientCharges = spark.sqlContext.read.format("csv")  
      .option("header", "true")  
      .option("inferSchema", "true")  
      .load("E:\\Acadgild\\Data\\inpatientCharges.csv")  
  
    dfInpatientCharges.printSchema()   
    //dfInpatientCharges.registerTempTable("hospital_data")  
  }  
}
```

Input csv file is loaded in to spark context and we have printed the schema of data frame using printSchema method.

Execution:

```
at com.spark.hospital.HospitalCaseStudy.main(HospitalCaseStudy.scala)  
root  
 |-- DRGDefinition: string (nullable = true)  
 |-- ProviderId: integer (nullable = true)  
 |-- ProviderName: string (nullable = true)  
 |-- ProviderStreetAddress: string (nullable = true)  
 |-- ProviderCity: string (nullable = true)  
 |-- ProviderState: string (nullable = true)  
 |-- ProviderZipCode: integer (nullable = true)  
 |-- HospitalReferralRegionDescription: string (nullable = true)  
 |-- TotalDischarges: integer (nullable = true)  
 |-- AverageCoveredCharges: double (nullable = true)  
 |-- AverageTotalPayments: double (nullable = true)  
 |-- AverageMedicarePayments: double (nullable = true)
```

Objective – 2

- What is the average amount of **AverageCoveredCharges** per state
- find out the AverageTotalPayments charges per state
- find out the AverageMedicarePayments charges per state.

Objective – 3

- Find out the total number of Discharges per state and for each disease
- Sort the output in descending order of totalDischarges

Program:

```
Console | TemperatureCaseStudy.scala | HospitalCaseStudy.scala |
import org.apache.spark.sql.Row
import org.apache.spark.sql.functions._

object HospitalCaseStudy {
  def main(args: Array[String]): Unit = {
    Logger.getLogger("org").setLevel(Level.ERROR)

    // Creating a spark context object to run in local machine
    val spark = SparkSession.builder.appName("Hospital Charges Analysis").master("local").getOrCreate()
    import spark.implicits._

    // Create a dataframe and load the building csv file
    val dfInpatientCharges = spark.sqlContext.read.format("csv")
      .option("header", "true")
      .option("inferSchema", "true")
      .load("E:\\Acadgild\\Data\\inpatientCharges.csv")

    //dfInpatientCharges.printSchema()
    dfInpatientCharges.registerTempTable("hospital_data")
    val avgCoveredChargesPerState = spark.sql("SELECT ProviderState, AVG(AverageCoveredCharges) FROM hospital_data GROUP BY ProviderState")
    val avgTotalPaymentsPerState = spark.sql("SELECT ProviderState, AVG(AverageTotalPayments) FROM hospital_data GROUP BY ProviderState")
    val avgMedicarePaymentsPerState = spark.sql("SELECT ProviderState, AVG(AverageMedicarePayments) FROM hospital_data GROUP BY ProviderState")
    val totalDischargesPerStatePerDisease = spark.sql("SELECT DRGDefinition, ProviderState, SUM(TotalDischarges) as TotalDischarges FROM hospital_data GROUP BY DRGDefinition, ProviderState ORDER BY TotalDischarges DESC")

    println("***** AVERAGE COVERED CHARGES PER STATE *****")
    avgCoveredChargesPerState.show()
    println("***** AVERAGE TOTAL PAYMENTS PER STATE *****")
    avgTotalPaymentsPerState.show()
    println("***** AVERAGE MEDICARE PAYMENTS PER STATE *****")
    avgMedicarePaymentsPerState.show()
    println("***** TOTAL DISCHARGES PER DISEASE PER STATE BY DESCENDING ORDER *****")
    totalDischargesPerStatePerDisease.show()
  }
}
```

Execution:

***** AVERAGE COVERED CHARGES PER STATE *****		
+-----+		
ProviderState avg(AverageCoveredCharges)		
+-----+		
	AZ	41200.063019992995
	SC	35862.49456269756
	LA	33085.372791542846
	MN	27894.36182060388
	NJ	66125.68627434729
	DC	40116.66365800864
	OR	27390.111870669723
	VA	29222.000487072903
	RI	29942.701122448976
	KY	24523.80716940223
	WY	28700.59862348178
	NH	27059.020801944105
	MI	24124.247209817277
	NV	61047.11541597337
	WI	26149.325331686607
	ID	25565.547041742288
	CA	67508.616535517
	CT	31318.4101143709
	NE	31736.427824858758
	MT	22670.015237154144
+-----+		
only showing top 20 rows		

***** AVERAGE TOTAL PAYMENTS PER STATE *****		
+-----+		
ProviderState avg(AverageTotalPayments)		
+-----+		
	AZ	10154.528211153991
	SC	9132.420758693366
	LA	8638.66257680871
	MN	9948.236962699833
	NJ	10678.98864691253
	DC	12998.029415584406
	OR	10436.192863741335
	VA	8887.75217682364
	RI	10509.566853741484
	KY	8278.58884484363
	WY	11398.485910931167
	NH	9289.661822600248
	MI	9754.420405978948
	NV	10291.718028286188
	WI	9270.705617501746
	ID	9827.180090744107
	CA	12629.668472137122
	CT	11365.450671307795
	NE	9331.682523540492
	MT	9252.802766798422
+-----+		
only showing top 20 rows		

***** AVERAGE MEDICARE PAYMENTS PER STATE *****		
+-----+		
ProviderState avg(AverageMedicarePayments)		
+-----+		
	AZ	8825.717239565045
	SC	7876.33152441167
	LA	7387.704625041281
	MN	8619.214982238007
	NJ	9586.940055946912
	DC	11811.967705627709
	OR	9035.259961508847
	VA	7538.847006001846
	RI	9317.939115646255
	KY	7185.227810467647
	WY	9539.392024291496
	NH	8124.506852976913
	MI	8662.157756043543
	NV	8747.602828618963
	WI	8002.597911079731
	ID	8461.977513611617
	CA	11494.381677893474
	CT	10104.592943809059
	NE	7992.6272504707995
	MT	7981.088063241104
+-----+		
only showing top 20 rows		

***** TOTAL DISCHARGES PER DISEASE PER STATE BY DESCENDING ORDER **			
+-----+			
DRGDefinition ProviderState TotalDischarges			
+-----+			
	[871 - SEPTICEMIA ...]	CA	34284
	[470 - MAJOR JOINT...]	TX	30095
	[470 - MAJOR JOINT...]	FL	29985
	[470 - MAJOR JOINT...]	CA	29731
	[871 - SEPTICEMIA ...]	TX	23144
	[871 - SEPTICEMIA ...]	NY	21970
	[392 - ESOPHAGITIS...]	FL	21298
	[470 - MAJOR JOINT...]	IL	20095
	[470 - MAJOR JOINT...]	NY	19371
	[871 - SEPTICEMIA ...]	FL	18660
	[690 - KIDNEY & UR...]	TX	17384
	[392 - ESOPHAGITIS...]	NY	17337
	[470 - MAJOR JOINT...]	MI	16847
	[470 - MAJOR JOINT...]	PA	16712
	[292 - HEART FAILU...]	FL	16639
	[690 - KIDNEY & UR...]	FL	16405
	[470 - MAJOR JOINT...]	OH	16062
	[470 - MAJOR JOINT...]	NC	15820
	[871 - SEPTICEMIA ...]	IL	15610
	[871 - SEPTICEMIA ...]	MI	15548
+-----+			
only showing top 20 rows			