

Assignment Submission- Case Study 4: Hospital Data

- Dataset Description
- DRG Definition: The code and description identifying the MS-DRG. MS-DRGs are a classification system that groups similar clinical conditions (diagnoses) and procedures furnished by the hospital during their stay.
- Provider Id: The CMS Certification Number (CCN) assigned to the Medicare-certified hospital facility.
- Provider Name: The name of the provider.
- Provider Street Address: The provider's street address.
- Provider City: The city where the provider is located.
- Provider State: The state where the provider is located.
- Provider Zip Code: The provider's zip code.
- Provider HRR: The Hospital Referral Region (HRR) where the provider is located.
- Total Discharges: The number of discharges billed by the provider for inpatient hospital services.
- Hospital Charges Data Analysis in the United States

SOLUTION:

Comments are added in the code for each step performed.

```
package SQL

import org.apache.spark.sql.SparkSession

object Hospital_Charge_Analysis {

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

    //Let us create a spark session object
    val spark = SparkSession
      .builder()
      .master("local")
      .appName("Hospital charge analysis")
      .config("Some-spark.config.option", "some-value")
      .getOrCreate()

    println("Spark Session Object created")

    //Set the log level as warning
    spark.sparkContext.setLogLevel("WARN")

    val data = spark.sqlContext.read.option("header", "true").option("inferSchema",
```

```

"true").csv("E:\\ACADGILD\\SCALA_SPARK\\CaseStudy_Hospital\\inpatientCharges.csv")

data.show()

//calculate average amount of "Average Covered Change" per state.
val avgCoveredChargePerState =
data.groupBy("ProviderState").avg("AverageCoveredCharges")

avgCoveredChargePerState.show()

//calculate average amount of "Average Total Payment" per state.

val avgTotalPayPerState =
data.groupBy("ProviderState").avg("AverageTotalPayments")

avgTotalPayPerState.show()

//calculate average amount of "Average Medicare Payment" per state.

val avgMedicarePerState =
data.groupBy("ProviderState").avg("AverageMedicarePayments")

avgMedicarePerState.show()

// total number of discharges per state for each disease
val numDisPerStatePerDisease = data.groupBy("\\uFEFFDRGDefinition",
"ProviderState").sum("TotalDischarges")

numDisPerStatePerDisease.show()

}

}

```

OUTPUT:

```

+-----+-----+
| 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 |
+-----+-----+

```

```

+-----+-----+
|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|
+-----+-----+

```

```

+-----+-----+
|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|
+-----+-----+

```

DRGDefinition	ProviderState	sum(TotalDischarges)
1064 - INTRACRANIA...	AR	659
1069 - TRANSIENT I...	OH	3382
189 - PULMONARY E...	OK	1102
193 - SIMPLE PNEU...	KS	1213
208 - RESPIRATORY...	ME	230
208 - RESPIRATORY...	SD	186
254 - OTHER VASCU...	TX	1999
280 - ACUTE MYOCA...	MA	1956
300 - PERIPHERAL ...	WV	511
309 - CARDIAC ARR...	OK	1205
378 - G.I. HEMORR...	VT	268
392 - ESOPHAGITIS...	KY	5822
394 - OTHER DIGES...	HI	16
418 - LAPAROSCOPI...	WA	309
481 - HIP & FEMUR...	MT	294
641 - MISC DISORD...	WV	1271
684 - RENAL FAILU...	IN	541
689 - KIDNEY & UR...	ME	92
812 - RED BLOOD C...	NM	161
897 - ALCOHOL/DRU...	RI	314

only showing top 20 rows