Hospital CASE STUDY 4

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.

Average Covered Charges: The provider's average charge for services covered by Medicare for all discharges in the
MS-DRG. These will vary from hospital to hospital because of the differences in hospital charge structures.

Average Total Payments: The average total payments to all providers for the MS-DRG including the MSDRG amount,

teaching, disproportionate share, capital, and outlier payments for all cases. Also included in the average total payments are co-payment and deductible amounts that the patient is responsible for and any additional payments by third parties for coordination of benefits.

Average Medicare Payments: The average amount that Medicare pays to the provider for Medicare's share of the MS-DRG. Average Medicare payment amounts include the MS-DRG amount, teaching, disproportionate share, capital, and outlier payments for all cases. Medicare payments DO NOT include beneficiary co-payments and deductible amounts nor any additional payments from third parties for coordination of benefits.

Load file into spark

Data Set:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	DRGDefin	Providerlc	ProviderN	ProviderS	ProviderC	ProviderS	ProviderZ	HospitalR	TotalDisch	AverageCc	AverageTc	AverageMedicarePayments						
	039 - EXTR	10001	SOUTHEAS	1108 ROSS	DOTHAN	AL	36301	AL - Dotha	91	32963.07	5777.24	4763.73						
	039 - EXTR	10005	MARSHAL	2505 U S H	BOAZ	AL	35957	AL - Birmii	14	15131.85	5787.57	4976.71						
	039 - EXTR	10006	ELIZA COF	205 MARE	FLORENCE	AL	35631	AL - Birmii	24	37560.37	5434.95	4453.79						
	039 - EXTR	10011	ST VINCEN	50 MEDIC/	BIRMINGI-	AL	35235	AL - Birmii	25	13998.28	5417.56	4129.16						
	039 - EXTR	10016	SHELBY BA	1000 FIRST	ALABASTE	AL	35007	AL - Birmii	18	31633.27	5658.33	4851.44						
	039 - EXTR	10023	BAPTIST N	2105 EAST	MONTGOI	AL	36116	AL - Mont|	67	16920.79	6653.8	5374.14						
	039 - EXTR	10029	EAST ALAI	2000 PEPP	OPELIKA	AL	36801	AL - Birmii	51	11977.13	5834.74	4761.41						
	039 - EXTR	10033	UNIVERSI	619 SOUTI-	BIRMINGI-	AL	35233	AL - Birmii	32	35841.09	8031.12	5858.5						
	039 - EXTR	10039	HUNTSVIL	101 SIVLEY	HUNTSVIL	AL	35801	AL - Hunts	135	28523.39	6113.38	5228.4						
	039 - EXTR	10040	GADSDEN	1007 GOO	GADSDEN	AL	35903	AL - Birmii	34	75233.38	5541.05	4386.94						
	039 - EXTR	10046	RIVERVIEV	600 SOUTI-	GADSDEN	AL	35901	AL - Birmii	14	67327.92	5461.57	4493.57						
	039 - EXTR	10055	FLOWERS	4370 WES	DOTHAN	AL	36305	AL - Dotha	45	39607.28	5356.28	4408.2						
	039 - EXTR	10056	ST VINCEN	810 ST VIN	BIRMINGI-	AL	35205	AL - Birmii	43	22862.23	5374.65	4186.02						
	039 - EXTR	10078	NORTHEA	400 EAST 1	ANNISTOI	AL	36207	AL - Birmii	21	31110.85	5366.23	4376.23						
	039 - EXTR	10083	SOUTH BA	1613 NOR	FOLEY	AL	36535	AL - Mobil	15	25411.33	5282.93	4383.73						
	039 - EXTR	10085	DECATUR	1201 7TH S	DECATUR	AL	35609	AL - Hunts	27	9234.51	5676.55	4509.11						
	039 - EXTR	10090	PROVIDEN	6801 AIRP	MOBILE	AL	36608	AL - Mobil	27	15895.85	5930.11	3972.85						
	039 - EXTR	10092	D C H REGI	809 UNIVE	TUSCALOC	AL	35401	AL - Tusca	31	19721.16	6192.54	5179.38						
	039 - EXTR	10100	THOMAS I-	750 MORP	FAIRHOPE	AL	36532	AL - Mobil	18	10710.88	4968	3898.88						
	039 - EXTR	10103	BAPTIST N	701 PRINC	BIRMINGI-	AL	35211	AL - Birmii	33	51343.75	5996	4962.45						
	039 - EXTR	10104	TRINITY M	800 MONT	BIRMINGI-	AL	35213	AL - Birmii	29	55219.31	5710.31	4471.68						
	039 - EXTR	10113	MOBILE IN	5 MOBILE	MOBILE	AL	36652	AL - Mobil	66	14948.15	5550.9	4219.9						

SOLUTION:

Code :

```scala
package Hospital_UseCase

import org.apache.spark.sql.SparkSession
import org.apache.spark.sql.functions._
object UseCase {

  def main(args: Array[String]): Unit = {
    println("hey scala") //Let us create a spark session object

    val spark = SparkSession
      .builder()
      .master("local")
      .appName("Spark SQL Use Case 1 ")
      .config("spark.some.config.option", "some-value")
      .getOrCreate()
    println("Hospital Usecase Study")

    val session =
org.apache.spark.sql.SparkSession.builder.master("local").appName("Spark CSV
Reader").getOrCreate

    val df = session.read.format("com.databricks.spark.csv").option("header",
"true").option("inferSchema", "true").load("C:/Users/Admin/Desktop/Assignment_to_be
submitted/Hospital Case Study 5/inpatientCharges.csv")
    df.show()
    df.registerTempTable("hospital_charges")

    //What is the average amount of AverageCoveredCharges per state
    df.groupBy("ProviderState").avg("AverageCoveredCharges").show
    println("Average amount of AverageCoveredCharges per state")

    //find out the AverageTotalPayments charges per state
    df.groupBy("ProviderState").avg("AverageTotalPayments").show
    println("AverageTotalPayments charges per state")

    //find out the AverageMedicarePayments charges per state.
  df.groupBy("ProviderState").avg("AverageMedicarePayments").show
   println("AverageMedicarePayments charges per state")

    //Find out the total number of Discharges per state and for each disease
    df.groupBy(("ProviderState"),("DRGDefinition")).sum("TotalDischarges").show
   println("Total number of Discharges per state and for each disease")

    //Sort the output in descending order of totalDischarges

df.groupBy(("ProviderState"),("DRGDefinition")).sum("TotalDischarges").sort(desc(su
m("TotalDischarges").toString)).show

df.groupBy(("ProviderState"),("DRGDefinition")).sum("TotalDischarges").orderBy(desc
(sum("TotalDischarges").toString)).show
    println("Sort the output in descending order of totalDischarges")
  }

}
```

Output :

What is the average amount of **AverageCoveredCharges** per state

18/09/22 16:45:38 INFO CodeGenerator: Code generated in 67.959723 ms
+------------+------------------------+
|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 amount of AverageCoveredCharges per state
18/09/22 16:45:38 INFO SparkContext: Invoking stop() from shutdown hook

Find out the **AverageTotalPayments** charges per state

18/09/22 16:47:37 INFO CodeGenerator: Code generated in 30.807365 ms
+------------+------------------------+
|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

AverageTotalPayments charges per state
18/09/22 16:47:37 INFO SparkContext: Invoking stop() from shutdown hook
18/09/22 16:47:37 INFO SparkUI: Stopped Spark web UI at http://192.168.100.4:4040

Find out the **AverageMedicarePayments** charges per state

```
18/09/22 16:49:36 INFO DAGScheduler: Job 7 finished: show at UseCase.scala:33, took 2.275711 s
+------------+--------------------------+
|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

AverageMedicarePayments charges per state
18/09/22 16:49:36 INFO CodeGenerator: Code generated in 19.150122 ms
```

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

```
18/09/22 16:51:38 INFO CodeGenerator: Code generated in 12.393372 ms
+------------+--------------------+--------------------+
|ProviderState|       DRGDefinition|sum(TotalDischarges)|
+------------+--------------------+--------------------+
|          KY|065 - INTRACRANIA...|                1937|
|          NY|101 - SEIZURES W/...|                4503|
|          IN|149 - DYSEQUILIBRIUM|                 700|
|          IA|178 - RESPIRATORY...|                 540|
|          WI|202 - BRONCHITIS ...|                 338|
|          MO|208 - RESPIRATORY...|                1840|
|          WI|251 - PERC CARDIO...|                 417|
|          AR|281 - ACUTE MYOCA...|                 413|
|          AZ|292 - HEART FAILU...|                2643|
|          NY|292 - HEART FAILU...|               13289|
|          NV|293 - HEART FAILU...|                 519|
|          SD|303 - ATHEROSCLER...|                  53|
|          TN|305 - HYPERTENSIO...|                 730|
|          ME|308 - CARDIAC ARR...|                 312|
|          NV|372 - MAJOR GASTR...|                 126|
|          WA|392 - ESOPHAGITIS...|                3148|
|          WI|439 - DISORDERS O...|                 215|
|          MN|536 - FRACTURES O...|                 332|
|          DC|563 - FX, SPRN, S...|                  43|
|          CO|602 - CELLULITIS ...|                  86|
+------------+--------------------+--------------------+
only showing top 20 rows

Total number of Discharges per state and for each disease
18/09/22 16:51:38 INFO SparkContext: Invoking stop() from shutdown hook
```

Sort the output in descending order of **totalDischarges**

```
18/09/22 16:53:17 INFO DAGScheduler: Job 4 finished: show at UseCase.scala:41, took 6.631393 s
18/09/22 16:53:17 INFO CodeGenerator: Code generated in 17.413127 ms
+-------------+--------------------+--------------------+
|ProviderState|       DRGDefinition|sum(TotalDischarges)|
+-------------+--------------------+--------------------+
|           CA|871 - SEPTICEMIA ...|               34284|
|           TX|470 - MAJOR JOINT...|               30095|
|           FL|470 - MAJOR JOINT...|               29985|
|           CA|470 - MAJOR JOINT...|               29731|
|           TX|871 - SEPTICEMIA ...|               23144|
|           NY|871 - SEPTICEMIA ...|               21970|
|           FL|392 - ESOPHAGITIS...|               21298|
|           IL|470 - MAJOR JOINT...|               20095|
|           NY|470 - MAJOR JOINT...|               19371|
|           FL|871 - SEPTICEMIA ...|               18660|
|           TX|690 - KIDNEY & UR...|               17384|
|           NY|392 - ESOPHAGITIS...|               17337|
|           MI|470 - MAJOR JOINT...|               16847|
|           PA|470 - MAJOR JOINT...|               16712|
|           FL|292 - HEART FAILU...|               16639|
|           FL|690 - KIDNEY & UR...|               16405|
|           OH|470 - MAJOR JOINT...|               16062|
|           NC|470 - MAJOR JOINT...|               15820|
|           IL|871 - SEPTICEMIA ...|               15610|
|           MI|871 - SEPTICEMIA ...|               15548|
+-------------+--------------------+--------------------+
only showing top 20 rows

18/09/22 16:53:17 INFO FileSourceStrategy: Pruning directories with:
18/09/22 16:53:17 INFO FileSourceStrategy: Post-Scan Filters:
```

18/09/22 16:53:21 INFO TaskSchedulerImpl: Removed TaskSet 7.0, whose tasks have all completed, from pool
18/09/22 16:53:21 INFO DAGScheduler: ResultStage 7 (show at UseCase.scala:42) finished in 1.839 s
18/09/22 16:53:21 INFO DAGScheduler: Job 5 finished: show at UseCase.scala:42, took 3.403778 s
```
+-------------+--------------------+--------------------+
|ProviderState|       DRGDefinition|sum(TotalDischarges)|
+-------------+--------------------+--------------------+
|           CA|871 - SEPTICEMIA ...|               34284|
|           TX|470 - MAJOR JOINT...|               30095|
|           FL|470 - MAJOR JOINT...|               29985|
|           CA|470 - MAJOR JOINT...|               29731|
|           TX|871 - SEPTICEMIA ...|               23144|
|           NY|871 - SEPTICEMIA ...|               21970|
|           FL|392 - ESOPHAGITIS...|               21298|
|           IL|470 - MAJOR JOINT...|               20095|
|           NY|470 - MAJOR JOINT...|               19371|
|           FL|871 - SEPTICEMIA ...|               18660|
|           TX|690 - KIDNEY & UR...|               17384|
|           NY|392 - ESOPHAGITIS...|               17337|
|           MI|470 - MAJOR JOINT...|               16847|
|           PA|470 - MAJOR JOINT...|               16712|
|           FL|292 - HEART FAILU...|               16639|
|           FL|690 - KIDNEY & UR...|               16405|
|           OH|470 - MAJOR JOINT...|               16062|
|           NC|470 - MAJOR JOINT...|               15820|
|           IL|871 - SEPTICEMIA ...|               15610|
|           MI|871 - SEPTICEMIA ...|               15548|
+-------------+--------------------+--------------------+
only showing top 20 rows
```

Sort the output in descending order of totalDischarges

18/09/22 16:53:21 INFO SparkContext: Invoking stop() from shutdown hook
18/09/22 16:53:21 INFO SparkUI: Stopped Spark web UI at http://192.168.100.4:4040