

## CaseStudy4:

-----

### 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 MS-DRG 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

## Objective1:

**Note: Source code is uploaded as separate file.**

## Load file into spark

### Query used:

```
val Hospital_data = spark.read.format("csv")
    .option("header", "true")
    .schema (Manual_schema)
    .load("D:\\Lavanya\\inpatientCharges.csv") . toDF ()
```

### Output:

DRGDefinition	ProviderId	ProviderName	ProviderStreetAddress	ProviderCity	ProviderState	ProviderZipCode	HospitalReferralRegionDescription	TotalDischarges	AverageCover
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	SHELBY 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	36901	AL - Birmingham	51	
039 - EXTRACRANIA...	10033	UNIVERSITY OF ALA...	619 SOUTH 19TH ST...	BIRMINGHAM	AL	35233	AL - Birmingham	32	
039 - EXTRACRANIA...	10039	HUNTSVILLE HOSPITAL	101 SIVLEY RD	HUNTSVILLE	AL	35801	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

## Objective2:

➤ What is the average amount of AverageCoveredCharges per state

Query used:

-----

```
val data1 = spark.sql("select ProviderState, avg(AverageCoveredCharges) from  
IPatientCharges group by ProviderState ")  
data1.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|  
+-----+-----+  
only showing top 20 rows
```

➤ find out the AverageTotalPayments charges per state

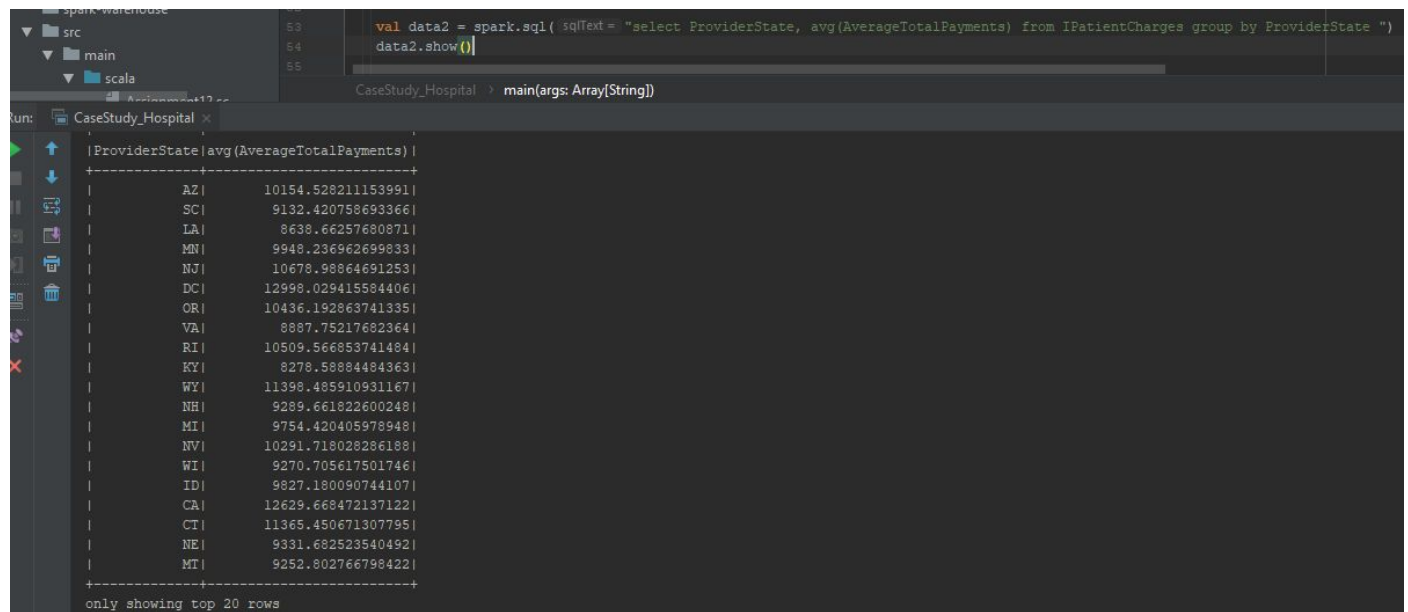
## Query used:

-----

```
val data2 = spark.sql("select ProviderState, avg(AverageTotalPayments) from IPatientCharges group by ProviderState ")
data2.show()
```

## Output:

-----



The screenshot shows a Scala IDE with a project named 'CaseStudy\_Hospital'. The code in the editor defines a Spark SQL query to calculate the average total payments by provider state. The output window displays the results of this query, showing the top 20 rows. The output is a table with two columns: 'ProviderState' and 'avg(AverageTotalPayments)'.

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

➤ find out the AverageMedicarePayments charges per state.

## Query used:

-----

```
val data3 = spark.sql("select ProviderState, avg(AverageMedicarePayments) from IPatientCharges group by ProviderState ")
data3.show()
```

## Output:

-----

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

## Objective3:

-----

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

## Query used:

-----

```
val data4 = spark.sql("select ProviderState, DRGDefinition, sum(TotalDischarges) as  
TotalDischarge from IPatientCharges group by ProviderState, DRGDefinition order by  
TotalDischarge desc")
```

```
data4.show()
```

## Output:

```
-----  
+-----+-----+-----+  
|ProviderState|      DRGDefinition|TotalDischarge|  
+-----+-----+-----+  
|      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
```