

Case-Study Day 25 (Hospital Data)

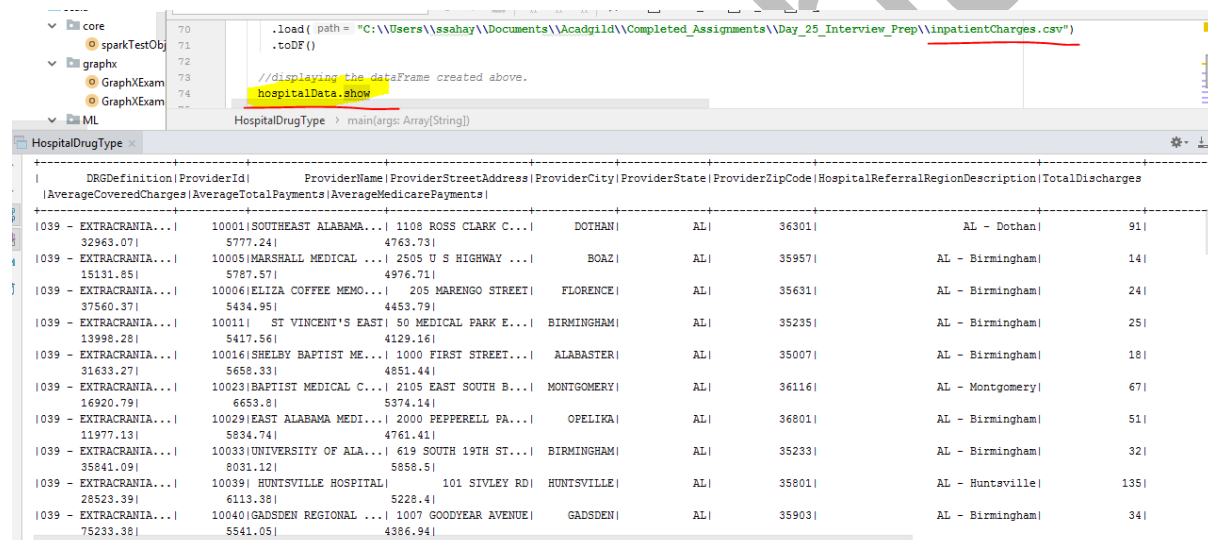
Objective 1:

Load file into spark

Ans:

Note: Program files are properly documented for a detailed description of each instruction used within the program.

ScreenShot:



```
70 .load(path = "C:\\Users\\ssahay\\Documents\\Acadgild\\Completed_Assignments\\Day_25_Interview_Prep\\inpatientCharges.csv")
71 .toDF()
72
73 //displaying the dataframe created above.
74 hospitalData.show
```

DRGDefinition	ProviderId	ProviderName	ProviderStreetAddress	ProviderCity	ProviderState	ProviderZipCode	HospitalReferralRegionDescription	TotalDischarges
039 - EXTRACRANIA...	10001	SOUTHEAST ALABAMA...	1108 ROSS CLARK C...	DOETHAN	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	36801	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

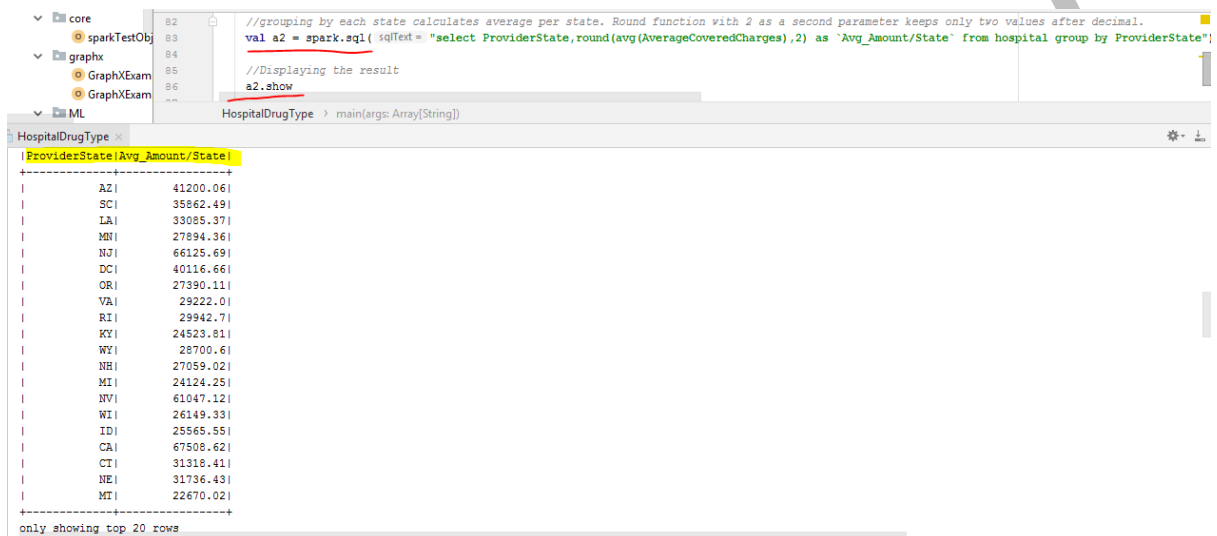
Objective 2:

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

Ans:

Note: Program files are properly documented for a detailed description of each instruction used within the program.

ScreenShot:

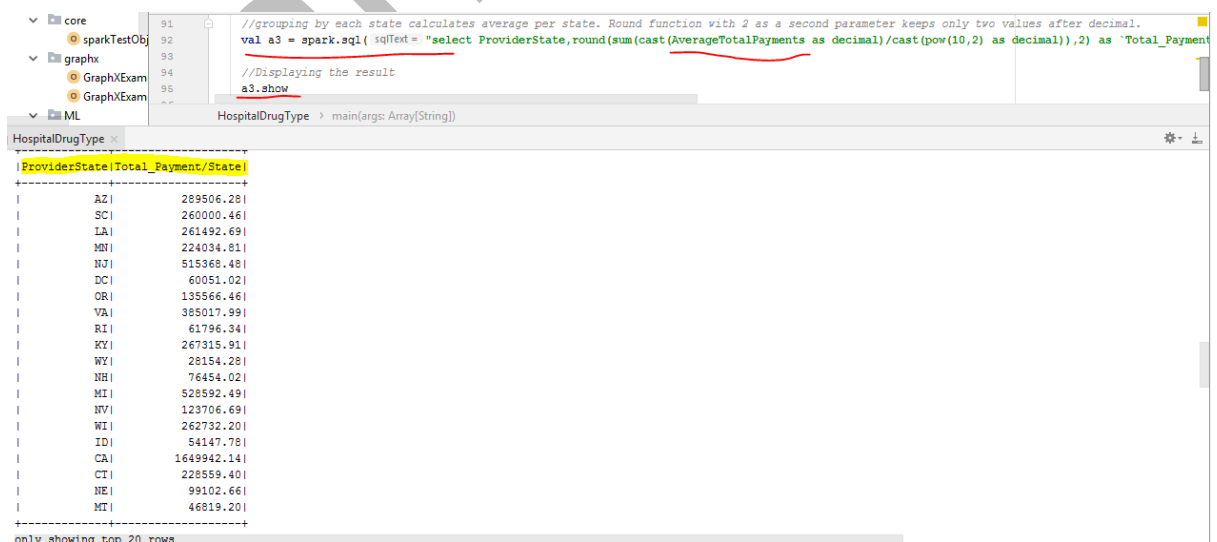


```
//grouping by each state calculates average per state. Round function with 2 as a second parameter keeps only two values after decimal.
val a2 = spark.sql("select ProviderState, round(avg(AverageCoveredCharges), 2) as 'Avg_Amount/State' from hospital group by ProviderState")
//Displaying the result
a2.show
```

ProviderState	Avg_Amount/State
AZ	41200.06
SC	35862.49
LA	33085.37
MN	27894.36
NJ	66125.69
DC	40116.66
OR	27390.11
VA	29222.0
RI	29942.7
KY	24523.81
WY	28700.6
NH	27059.02
MI	24124.25
NV	61047.12
WI	26149.33
ID	25565.55
CA	67508.62
CT	31318.41
NE	31736.43
MT	22670.02

only showing top 20 rows

Find out the **AverageTotalPayments** charges per state



```
//grouping by each state calculates average per state. Round function with 2 as a second parameter keeps only two values after decimal.
val a3 = spark.sql("select ProviderState, round(sum(cast(AverageTotalPayments as decimal)/cast(pow(10,2) as decimal)), 2) as 'Total_Payment/State' from hospital group by ProviderState")
//Displaying the result
a3.show
```

ProviderState	Total_Payment/State
AZ	289506.28
SC	260000.46
LA	261492.69
MN	224034.81
NJ	515368.48
DC	60051.02
OR	135566.46
VA	385017.99
RI	61796.34
KY	267315.91
WY	28154.28
NH	76454.02
MI	528592.49
NV	123706.69
WI	262732.20
ID	54147.78
CA	1649942.14
CT	228559.40
NE	99102.66
MT	46819.20

only showing top 20 rows

Find out the **AverageMedicarePayments** charges per state

```
100 //grouping by each state calculates average per state. Round function with 2 as a second parameter keeps only two values after decimal.
101 val a4 = spark.sql( sqlText = "select ProviderState,round(sum(cast(AverageMedicarePayments as decimal)/cast(pow(10,2) as decimal)),2) as `Total_Medicare`"
102
103 //Displaying the result
104 a4.show
```

HospitalDrugType > main(args: Array[String])

ProviderState	Total_Medicare_Payment/State
AZ	251621.67
SC	224239.61
LA	223626.26
MI	194104.94
NJ	462666.61
DC	54571.33
OR	117368.35
VA	326583.58
RI	54789.50
KY	232011.54
WY	23562.37
NH	68864.97
MI	469403.22
NV	105146.54
WI	226794.06
ID	46625.52
CA	1501628.09
CT	203203.55
NE	84881.92
MT	40384.33

only showing top 20 rows

Objective 3:

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

Ans:

Note: Program files are properly documented for a detailed description of each instruction used within the program.

ScreenShot:

```
109 //grouping the result based on DRGDefinition,ProviderState
110 val a5 = spark.sql( sqlText = "select DRGDefinition,ProviderState, sum(TotalDischarges) as `#Discharges/State/Disease` from hospital group by DRGDefinition,ProviderState"
111
112 //Displaying the result
113 a5.show
```

HospitalDrugType > main(args: Array[String])

DRGDefinition	ProviderState	#Discharges/State/Disease
1064 - INTRACRANIA...	AR	659
1069 - TRANSIENT I...	OH	3382
1189 - PULMONARY E...	OK	1102
1193 - SIMPLE PNEU...	KS	1213
1208 - RESPIRATORY...	ME	230
1208 - RESPIRATORY...	SD	186
1254 - OTHER VASCU...	TX	1999
1280 - ACUTE MYOCA...	MA	1956
1300 - PERIPHERAL ...	WV	511
1309 - CARDIAC ARR...	OK	1205
1378 - G.I. HEMORR...	VT	268
1392 - ESOPHAGITIS...	KY	5822
1394 - OTHER DIGES...	HI	16
1418 - LAPAROSCOPI...	WA	309
1481 - HIP & FEMUR...	MT	294
1641 - MISC DISORD...	WV	1271
1684 - RENAL FAILU...	IN	541
1689 - KIDNEY & UR...	ME	92
1812 - RED BLOOD C...	NM	161
1897 - ALCOHOL/DRU...	RI	314

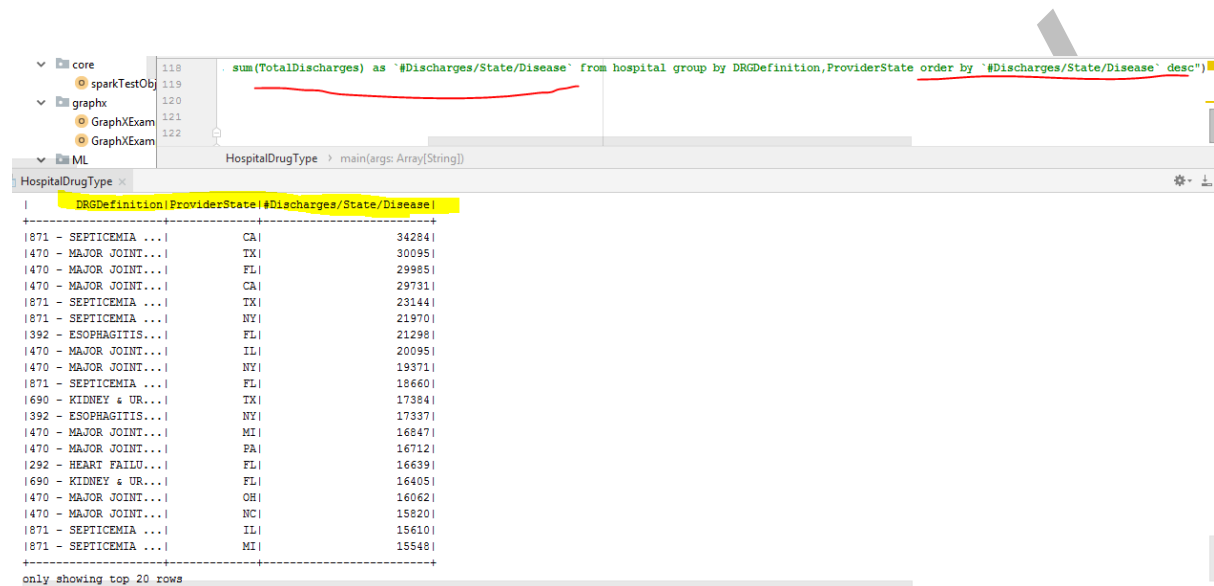
only showing top 20 rows

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

Ans:

Note: Program files are properly documented for a detailed description of each instruction used within the program.

ScreenShot:



```
118 .sum(TotalDischarges) as '#Discharges/State/Disease' from hospital group by DRGDefinition,ProviderState order by '#Discharges/State/Disease' desc"
```

DRGDefinition	ProviderState	#Discharges/State/Disease
1871 - SEPTICEMIA ...	CA	34284
1470 - MAJOR JOINT...	TX	30095
1470 - MAJOR JOINT...	FL	29985
1470 - MAJOR JOINT...	CA	29731
1871 - SEPTICEMIA ...	TX	23144
1871 - SEPTICEMIA ...	NY	21970
1392 - ESOPHAGITIS...	FL	21298
1470 - MAJOR JOINT...	IL	20095
1470 - MAJOR JOINT...	NY	19371
1871 - SEPTICEMIA ...	FL	18660
1690 - KIDNEY & UR...	TX	17384
1392 - ESOPHAGITIS...	NY	17337
1470 - MAJOR JOINT...	MI	16847
1470 - MAJOR JOINT...	PA	16712
1292 - HEART FAILU...	FL	16639
1690 - KIDNEY & UR...	FL	16405
1470 - MAJOR JOINT...	OH	16062
1470 - MAJOR JOINT...	NC	15820
1871 - SEPTICEMIA ...	IL	15610
1871 - SEPTICEMIA ...	MI	15548

only showing top 20 rows

End
