Case-Study Day (Hospital Data)

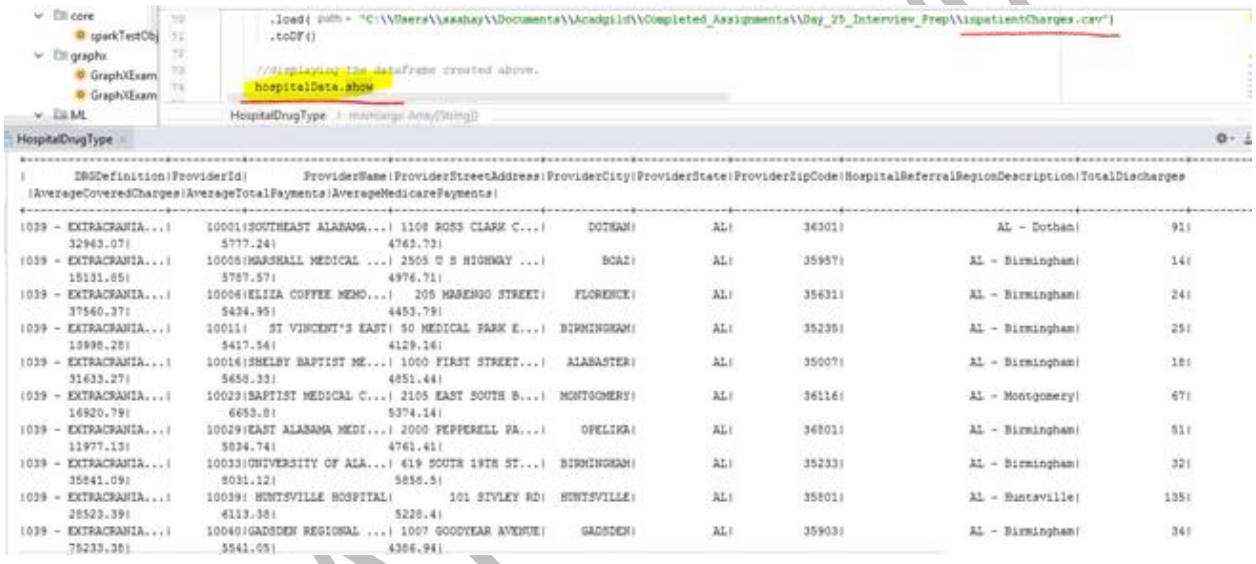
Objective 1:

Load file into spark

Ans:-

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

ScreenShot:-



```
.load( path = "C:\\Users\\asahay\\Documents\\Acadgild\\Completed_Assignments\\Day_25_Interview_Prep\\inpatientCharges.csv")
.toDF()

//displaying the dataframe created above.
hospitalData.show
```

HospitalDrugType

DRGDefinition	ProviderId	ProviderName	ProviderStreetAddress	ProviderCity	ProviderState	ProviderZipCode	HospitalReferralRegionDescription	TotalDischarges	AverageCoveredCharges	AverageTotalPayments	AverageMedicarePayments
039 - EXTRACRANIA...	10001	SOUTHEAST ALABAMA...	1108 ROSS CLARK C...	DOTHAN	AL	36301	AL - Dothan	91	32963.07	5777.24	4763.73
039 - EXTRACRANIA...	10005	MARSHALL MEDICAL ...	2505 U S HIGHWAY ...	BOAZ	AL	35957	AL - Birmingham	14	15131.85	5787.57	4976.71
039 - EXTRACRANIA...	10006	ELIZA COFFEE MEMO...	205 MARENGO STREET	FLORENCE	AL	35631	AL - Birmingham	24	37560.37	5434.95	4453.79
039 - EXTRACRANIA...	10011	ST VINCENT'S EAST	50 MEDICAL PARK E...	BIRMINGHAM	AL	35235	AL - Birmingham	25	13998.28	5417.56	4129.16
039 - EXTRACRANIA...	10016	SHELBY BAPTIST ME...	1000 FIRST STREET...	ALABASTER	AL	35007	AL - Birmingham	18	31633.27	5658.33	4851.44
039 - EXTRACRANIA...	10023	BAPTIST MEDICAL C...	2105 EAST SOUTH B...	MONTGOMERY	AL	36116	AL - Montgomery	67	16920.79	6653.8	5374.14
039 - EXTRACRANIA...	10029	EAST ALABAMA MEDI...	2000 PEPPERELL PA...	OPELIKA	AL	36801	AL - Birmingham	51	11977.13	5834.74	4761.41
039 - EXTRACRANIA...	10033	UNIVERSITY OF ALA...	619 SOUTH 19TH ST...	BIRMINGHAM	AL	35233	AL - Birmingham	32	35841.09	8031.12	5856.5
039 - EXTRACRANIA...	10039	HUNTSVILLE HOSPITAL	101 SIVLEY RD	HUNTSVILLE	AL	35801	AL - Huntsville	135	28523.39	6113.38	5228.4
039 - EXTRACRANIA...	10040	GADSDEN REGIONAL ...	1007 GOODYEAR AVENUE	GADSDEN	AL	35903	AL - Birmingham	34	75233.38	5541.05	4386.94

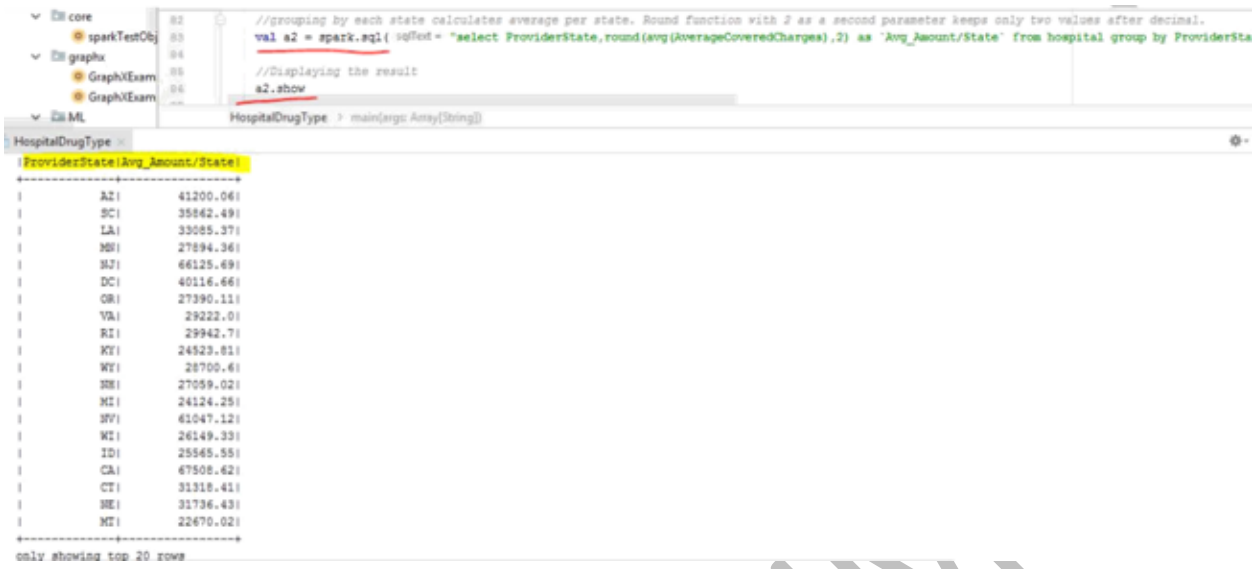
Objective 2:

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

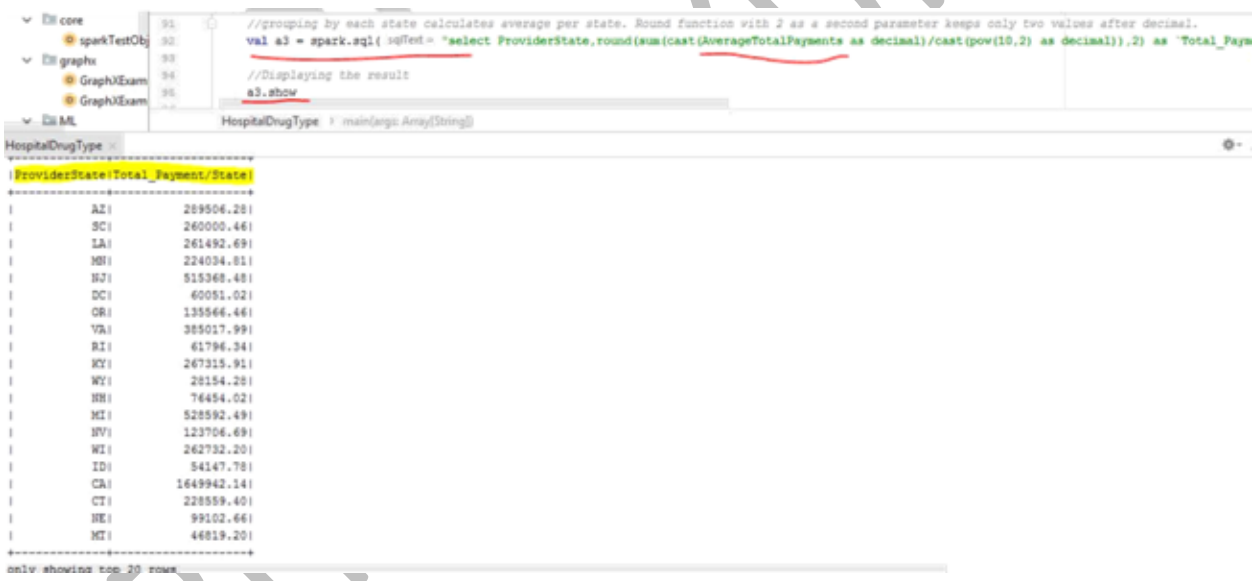
Ans:

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

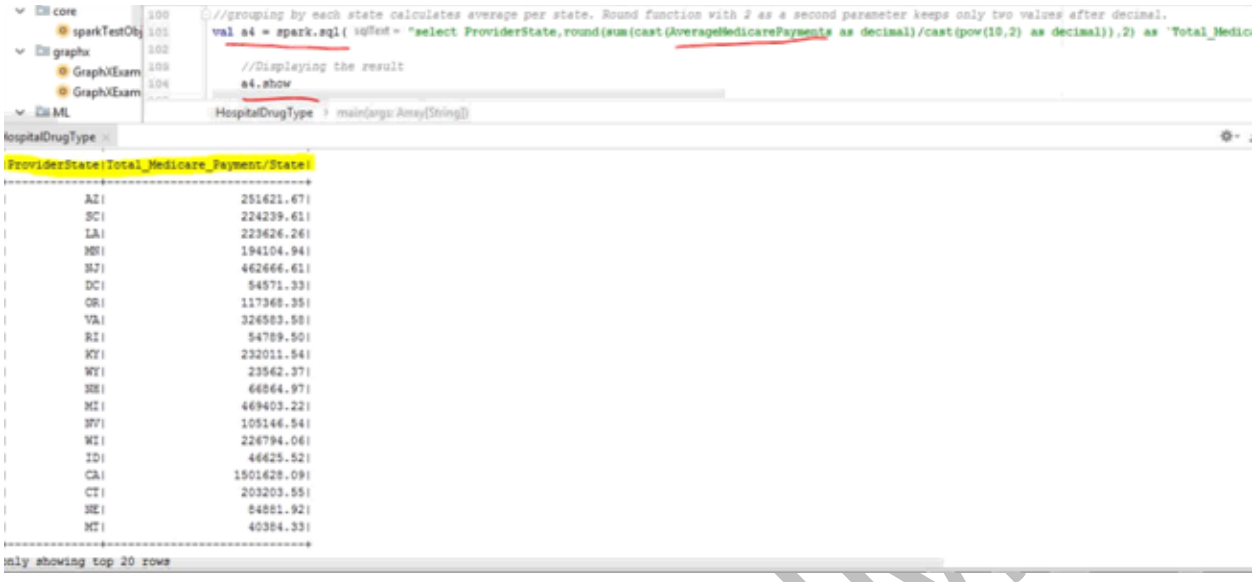
ScreenShot:-



Find out the **AverageTotalPayments** charges per state



Find out the **AverageMedicarePayments** 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 a4 = spark.sql( sqlText = "select ProviderState,round(sum(cast(AverageMedicarePayments as decimal)/cast(pow(10,2) as decimal)),2) as `Total_Medic
//Displaying the result
a4.show
```

```
+-------------+-------------------------------+
|ProviderState|Total_Medicare_Payment/State|
+-------------+-------------------------------+
|           AZ|                      251621.67|
|           SC|                      224239.61|
|           LA|                      223626.26|
|           MS|                      194104.94|
|           NJ|                      462666.61|
|           DC|                       54571.33|
|           OR|                      117368.35|
|           VA|                      326583.58|
|           RI|                       54789.50|
|           KY|                      232011.54|
|           WY|                       23562.37|
|           NH|                       66864.97|
|           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
```

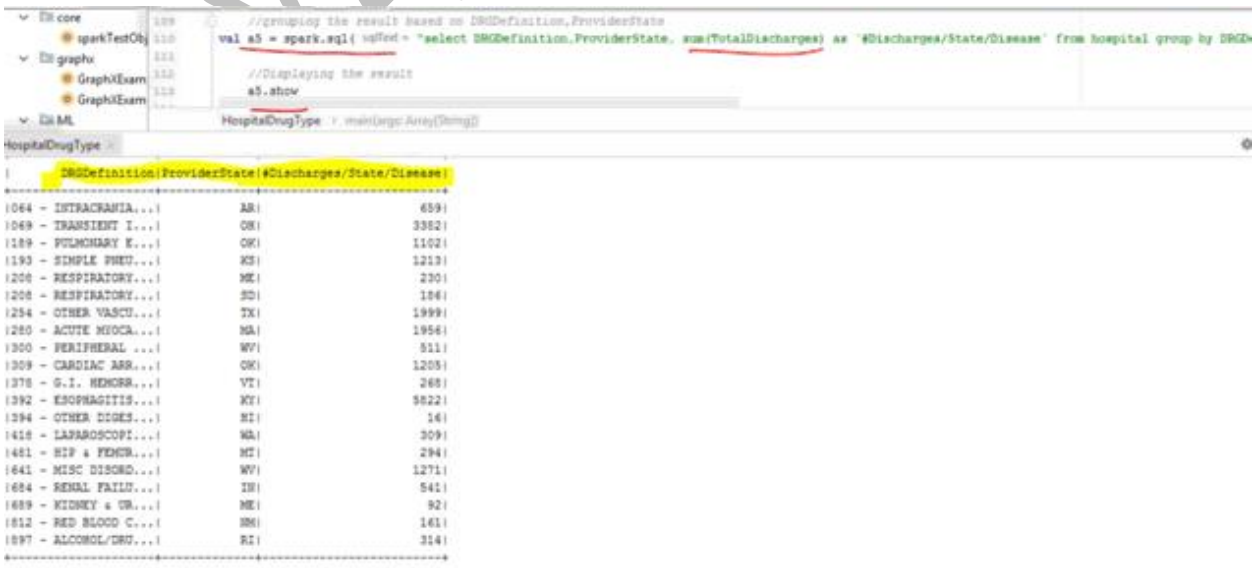
Objective 3:

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:-



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

```
+--------------------+-------------+-------------------------+
|      DRGDefinition|ProviderState|#Discharges/State/Disease|
+--------------------+-------------+-------------------------+
|064 - INTRACRANIA...|           AR|                      659|
|069 - TRANSIENT I...|           OK|                     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 MIOCA...|           NA|                     1956|
|300 - PERIPHERAL ...|           WV|                      511|
|309 - CARDIAC ARR...|           OK|                     1205|
|378 - G.I. HEMORR...|           VI|                      268|
|392 - ESOPHAGITIS...|           KY|                     5822|
|394 - OTHER DIGES...|           NI|                       14|
|418 - LAPAROSCOPI...|           WA|                      309|
|481 - HIP & FEMUR...|           MI|                      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|
+--------------------+-------------+-------------------------+
```

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:-