

Case_Study_5

Objective-1:

Load file into spark

Objective-2:

- What is the average amount of AverageCoveredCharges per state
- find out the AverageTotalPayments charges per state
- find out the AverageMedicarePayments charges per state.

Objective-3:

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

Scala code written for all the above objective is shown in the below screenshot.

```
package SQL
import ...

object SparkSQLUseCase3 {

  def main(args: Array[String]):Unit = {
    println("hey scala")
    val spark = SparkSession
      .builder()
      .master("local")
      .appName("Spark SQL Use Case 1 ")
      .config("spark.some.config.option", "some-value")
      .getOrCreate()
    println("Spark Session Object created")
    //Set the log level as warning
    spark.sparkContext.setLogLevel("WARN")
    //file loaded into Spark
    val df1 = spark.read.option("header", "true").option("inferSchema", "true").csv( path = "C:\\Users\\shiva\\Downloads\\inpatientCharges.csv")
    println("File loaded")
    df1.show()
    df1.createOrReplaceTempView( viewName = "hospital_charges")
    println("Header removed from the data !")

    //average amount of AverageCoveredCharges per state
    println("average amount of AverageCoveredCharges per state")
    df1.groupBy( col = "ProviderState").avg( colNames = "AverageCoveredCharges").show
    //the AverageTotalPayments charges per state
    println("the AverageTotalPayments charges per state")
    df1.groupBy( col = "ProviderState").avg( colNames = "AverageTotalPayments").show
    //the AverageMedicarePayments charges per state.
    println("the AverageMedicarePayments charges per state.")
    df1.groupBy( col = "ProviderState").avg( colNames = "AverageMedicarePayments").show
    //the total number of Discharges per state and for each disease
    println("the total number of Discharges per state and for each disease")
    df1.groupBy( col = "ProviderState", cols = "DRGDefinition").sum( colNames = "TotalDischarges").show
    //the output in descending order of totalDischarges
    println("the output in descending order of totalDischarges")
    df1.groupBy( col = "ProviderState", cols = "DRGDefinition").sum( colNames = "TotalDischarges").sort(desc(sum( columnName = "TotalDischarges").toString)).show
    df1.groupBy( col = "ProviderState", cols = "DRGDefinition").sum( colNames = "TotalDischarges").orderBy(desc(sum( columnName = "TotalDischarges").toString)).show
  }
}
```

Data loaded in to the spark. Output is shown in the below screenshot.

File loaded

	DRGDefinition	ProviderID	ProviderName	ProviderStreetAddress	ProviderCity	ProviderState	ProviderZipCode	HospitalReferralRegionDescription	TotalDischarges	AverageCoveredCharges	Average
(039 - EXTRACRANIA...)	10001	SOUTHEAST ALABAMA...	1100 ROSS CLARK C...	DOthan	AL	36201		AL - Dothan	91	32963.07	
(039 - EXTRACRANIA...)	10005	MARSHALL MEDICAL ...	2505 U S HIGHWAY ...	BOAd	AL	35957		AL - Birmingham	14	15131.85	
(039 - EXTRACRANIA...)	10006	ELISA COFFE MEMO...	205 MARENGO STREET	FLORENCE	AL	35631		AL - Birmingham	24	37560.37	
(039 - EXTRACRANIA...)	10011	ST VINCENT'S EAST	50 MEDICAL PARK E...	BIRMINGHAM	AL	35235		AL - Birmingham	25	13998.28	
(039 - EXTRACRANIA...)	10016	SHELBY BAPTIST ME...	1000 FIRST STREET...	ALABASTER	AL	35007		AL - Birmingham	18	31633.27	
(039 - EXTRACRANIA...)	10023	BAPTIST MEDICAL C...	2105 EAST SOUTH B...	MONTGOMERY	AL	36116		AL - Montgomery	67	16920.79	
(039 - EXTRACRANIA...)	10029	EAST ALABAMA MEDI...	2000 PEPPERELL PA...	OPELIKA	AL	36801		AL - Birmingham	51	11977.13	
(039 - EXTRACRANIA...)	10033	UNIVERSITY OF ALA...	619 SOUTH 19TH ST...	BIRMINGHAM	AL	35233		AL - Birmingham	32	35841.09	
(039 - EXTRACRANIA...)	10039	HUNTSVILLE HOSPITAL	101 BIVLEY RD	HUNTSVILLE	AL	35804		AL - Huntsville	139	28523.39	
(039 - EXTRACRANIA...)	10040	GADSDEN REGIONAL ...	1007 GOODMAN AVENUE	GADSDEN	AL	35903		AL - Birmingham	34	75223.38	
(039 - EXTRACRANIA...)	10046	RIVERVIEW REGIONA...	600 SOUTH THIRD S...	GADSDEN	AL	35901		AL - Birmingham	14	67327.52	
(039 - EXTRACRANIA...)	10055	FLOWERS HOSPITAL	4370 WEST MAIN ST...	DOthan	AL	36205		AL - Dothan	45	39607.28	
(039 - EXTRACRANIA...)	10056	ST VINCENT'S BIRM...	810 ST VINCENT'S ...	BIRMINGHAM	AL	35205		AL - Birmingham	43	22862.23	
(039 - EXTRACRANIA...)	10078	NORTHEAST ALABAMA...	400 EAST 10TH STREET	ANNISTON	AL	36207		AL - Birmingham	21	31110.85	
(039 - EXTRACRANIA...)	10083	SOUTH BALDWIN REG...	1613 NORTH MCKENZ...	FOLEY	AL	36535		AL - Mobile	15	25411.33	
(039 - EXTRACRANIA...)	10085	DECATUR GENERAL H...	1201 7TH STREET SE	DECATUR	AL	35609		AL - Huntsville	27	9234.51	
(039 - EXTRACRANIA...)	10090	PROVIDENCE HOSPITAL	6901 AIRPORT BOUL...	MOBILE	AL	36608		AL - Mobile	27	15895.85	
(039 - EXTRACRANIA...)	10092	D C H REGIONAL ME...	809 UNIVERSITY BO...	TUSCALOOSA	AL	35401		AL - Tuscaloosa	31	19721.16	
(039 - EXTRACRANIA...)	10100	THOMAS HOSPITAL	750 MORRIS AVENUE	FAIRHOPE	AL	36532		AL - Mobile	18	10710.88	
(039 - EXTRACRANIA...)	10103	BAPTIST MEDICAL C...	701 PRINCETON AVE...	BIRMINGHAM	AL	35211		AL - Birmingham	33	51349.75	

only showing top 20 rows

The output of average amount of AverageCoveredCharges per state output is shown in the below screenshot.

average amount of AverageCoveredCharges per state

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

The output of AverageTotalPayments charges per state is shown in the below screenshot.

```
the AverageTotalPayments charges per state
+-----+-----+
| 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
```

The output of AverageMedicarePayments charges occurred per state is shown in the below screenshot.

```
the AverageMedicarePayments charges per state.
+-----+-----+
|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
```

The output of total number of discharges per state and for each disease is shown in the below screenshot.

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

The output of total discharges in descending order is shown in the below screenshot.

```
the output in descending order of totalDischarges
+-----+-----+-----+
|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
```