Load file into spark

```
scala> val hospitalDataDF = spark.read.csv("/user/spark/inpatientCharges.csv").toDF("drgdefinition","providerdid","providerna
me","providerstreetaddr","providercity","providerstate","providerzipcode","region","totaldischarges","avgcoveredcharges","avg
totalpayment","avgmedicarepayment")
hospitalDataDF: org.apache.spark.sql.DataFrame = [drgdefinition: string, providerdid: string ... 10 more fields]

scala> hospitalDataDF.createOrReplaceTempView("hospitalData")

scala> spark.sql("select * from hospitalData").show
+--------------------+-----------+--------------------+--------------------+------------+-------------+--------------+------+
|       drgdefinition|providerdid|        providername|  providerstreeaddr|providercity|providerstate|providerzipcode|
  region|totaldischarges|avgcoveredcharges|avgtotalpayment|avgmedicarepayment|
+--------------------+-----------+--------------------+--------------------+------------+-------------+--------------+------+
|839 - EXTRACRANIA...|       10001|SOUTHEAST ALABAM...|1108 ROSS CLARK C...|      DOTHAN|           AL|         36301|   AL
  - Dothan|          91|       32963.07|       5777.24|        4763.73|
|839 - EXTRACRANIA...|       10005|MARSHALL MEDICAL ...|2505 U S HIGHWAY ...|        BOAZ|           AL|         35957|AL - B
  immingham|          14|       15131.85|       5787.57|        4976.71|
|839 - EXTRACRANIA...|       10006|ELIZA COFFEE MEMC...| 205 MARENGO STREET|    FLORENCE|           AL|         35631|AL - B
  immingham|          24|       37560.37|       5434.95|        4453.79|
|839 - EXTRACRANIA...|       10011|   ST VINCENT'S EAST|50 MEDICAL PARK E...|  BIRMINGHAM|           AL|         35235|AL - B
  immingham|          25|       13990.20|       5417.56|        4129.16|
|839 - EXTRACRANIA...|       10016|SHELBY BAPTIST ME...|1000 FIRST STREET...|   ALABASTER|           AL|         35007|AL - B
  immingham|          18|       31633.27|       5658.33|        4851.44|
|839 - EXTRACRANIA...|       10023|BAPTIST MEDICAL C...|2105 EAST SOUTH B...|  MONTGOMERY|           AL|         36116|AL - M
  ortgomery|          67|       16920.79|        6653.8|        5374.14|
|839 - EXTRACRANIA...|       10029|EAST ALABAMA MEDI...|2000 PEPPERELL PA...|     OPELIKA|           AL|         36801|AL - B
  immingham|          51|       11977.13|       5834.74|        4761.41|
|839 - EXTRACRANIA...|       10033|UNIVERSITY OF ALA...|619 SOUTH 19TH ST...|  BIRMINGHAM|           AL|         35233|AL - B
  immingham|          32|       35841.09|       8831.12|         5858.5|
|839 - EXTRACRANIA...|       10039| HUNTSVILLE HOSPITAL|  101 SIVLEY RD|   HUNTSVILLE|           AL|         35801|AL - H
  urtsville|         135|       28523.39|       6113.38|         5229.4|
|839 - EXTRACRANIA...|       10040|GADSDEN REGIONAL ...|1007 GOODYEAR AVENUE|     GADSDEN|           AL|         35903|AL - B
  immingham|          34|       75233.38|       5541.05|        4386.94|
|839 - EXTRACRANIA...|       10046|RIVERVIEW REGIONA...|600 SOUTH THIRD S...|     GADSDEN|           AL|         35901|AL - B
  immingham|          14|       67327.92|        546.57|        4493.57|
```

What is the average amount of AverageCoveredCharges per state

```
scala> spark.sql("select providerstate,avg(avgcoveredcharges) from hospitalData group by providerstate").show
+-------------+---------------------------------------+
|providerstate|avg(CAST(avgcoveredcharges AS DOUBLE))|
+-------------+---------------------------------------+
|           AZ|                      44349.5698722374|
|           SC|                      38562.94126548668|
|           LA|                      37289.3364466177|
|           MN|                      30153.315841013828|
|           NJ|                      70646.93355236136|
|           DC|                      43710.14887640451|
|           OR|                      28253.11391812865|
|           VA|                      31765.24006489671|
|           RI|                      33147.171034482744|
|           KY|                      27084.887072243317|
|           WY|                      30710.63651162791|
|           NH|                      28916.959882352927|
|           MI|                      25930.067257769628|
|           NV|                      68842.74955584478|
|           WI|                      28457.56578225695|
|           ID|                      26821.237962962958|
|           CA|                      72712.25290346794|
|           CT|                      33668.55779319045|
|           NE|                      34024.06050113896|
|           MT|                      23271.31260663985|
+-------------+---------------------------------------+
only showing top 20 rows
```

find out the AverageTotalPayments charges per state

```
scala> spark.sql("select providerstate,sum(avgtotalpayment) from hospitalData group by providerstate").show
+-------------+--------------------------------+
|providerstate|sum(CAST(avgtotalpayment AS DOUBLE))|
+-------------+--------------------------------+
|           AZ|               1.2467909240600006E7|
|           SC|               1.0853977930000005E7|
|           LA|               1.1424323609999983E7|
|           MN|                    9011841.169999996|
|           NJ|               2.1959887200000003E7|
|           DC|                   2446024.2000000007|
|           OR|                    5352073.699999996|
|           VA|               1.5860084840000002E7|
|           RI|                   2518819.2599999993|
|           KY|               1.1715002420000015E7|
|           WY|                          1040815.01|
|           NH|                   3243190.3460000003|
|           MI|               2.2284718339999966E7|
|           NV|                    5172079.550000002|
|           WI|               1.0745732200000005E7|
|           ID|                          2160222.41|
|           CA|                 6.864022839000001E7|
|           CT|                    9420934.359999994|
|           NE|                    4190321.7200000007|
|           MT|                   1950643.9060000008|
+-------------+--------------------------------+
only showing top 20 rows

scala>
```

find out the AverageMedicarePayments charges per state.

```
scala> spark.sql("select providerstate,sum(avgmedicarepayment) from hospitalData group by providerstate").show
+-------------+----------------------------------+
|providerstate|sum(CAST(avgmedicarepayment AS DOUBLE))|
+-------------+----------------------------------+
|           AZ|                 1.6843064419999998E7|
|           SC|                    9435461.236000094|
|           LA|                    9829321.819999997|
|           MN|                    7822389.949999995|
|           NJ|                 1.984741647600066E7|
|           DC|                    2238687.3600000003|
|           OR|                           4658864.83|
|           VA|                 1.3468689689999985E7|
|           RI|                   2233758.0699999994|
|           KY|                 1.021231428600001E7|
|           WY|                    866076.6699999998|
|           NH|                    2839560.120000001|
|           MI|                 1.987673922600002E7|
|           NV|                    4430086.249999997|
|           WI|                           9367783.57|
|           ID|                   1811042.0700000098|
|           CA|                 6.268313838999914E7|
|           CT|                    8414424.320000096|
|           NE|                   3586662.6500000013|
|           MT|                   1668555.3099999996|
+-------------+----------------------------------+
only showing top 20 rows
```

-

-

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

```
scala> spark.sql("select providerstate,sum(totaldischarges) from hospitalData group by providerstate").show
+-------------+--------------------------------+
|providerstate|sum(CAST(totaldischarges AS DOUBLE))|
+-------------+--------------------------------+
|           AZ|                         39743.9|
|           SC|                         48159.9|
|           LA|                         44996.9|
|           MN|                         33992.9|
|           NJ|                         99357.9|
|           DC|                          7891.9|
|           OR|                         14513.9|
|           VA|                         68995.9|
|           RI|                          8792.9|
|           KY|                         67887.9|
|           NY|                          7457.9|
|           NH|                         12931.9|
|           MI|                        117468.9|
|           NV|                         16142.9|
|           WI|                         33634.9|
|           ID|                          5313.9|
|           CA|                        173766.9|
|           CT|                         37285.9|
|           NE|                         15138.9|
|           MT|                          6136.9|
+-------------+--------------------------------+
only showing top 20 rows

scala>
```

Sort the output in descending order of totalDicharge

```
scala> spark.sql("select * from hospitalData order by totaldischarges desc").show
+-----------------+-----------+--------------------+--------------------+-------------+-------------+-------------+
|    drgdefinition|providerdid|        providername|providerstreetaddr| providercity|providerstate|providerzipcode|
|           region|totaldischarges|avgcoveredcharges|avgtotalpaymen|avgmedicarepayment|
+-----------------+-----------+--------------------+--------------------+-------------+-------------+-------------+
|069 - TRANSIENT I...|     310051|    OVERLOOK HOSPITAL|      99 BEAUVIOR AVE|      SUMMET|           NJ|         7902| N
- Morristown|         99|           27330.22|             5736.24|      4616.18|
|190 - CHRONIC OBS...|     490048|LEWISGALE MEDICAL...|  1900 ELECTRIC ROAD|       SALEM|           VA|        24153|
VA - Roanoke|         99|           23988.89|             6621.77|      4975.39|
|069 - TRANSIENT I...|     360020|SUMMA HEALTH SYST...|525 EAST MARKET S...|       AKRON|           OH|        44309|
OH - Akron|         99|           31521.83|             5816.77|      4806.82|
|039 - EXTRACRANIA...|     450851|BAYLOR HEART AND ...|621 NORTH HALL ST...|      DALLAS|           TX|        75226|
TX - Dallas|         99|           16224.38|             6828.77|       4764.6|
|101 - SEIZURES W/...|     340047|NORTH CAROLINA BA...|MEDICAL CENTER BO...|WINSTON-SALEM|           NC|    27157|NC -
Winston-Salem|         99|           24831.11|             7957.09|      6249.57|
|065 - INTRACRANIA...|     210054|SOUTHERN MARYLAND...|     7503 SURRATTS RD|     CLINTON|           MD|        20735| DC
- Washington|         99|            9231.38|             8677.51|      7848.94|
|177 - RESPIRATORY...|     140288|ADVOCATE GOOD SAM...|3815 HIGHLAND AVENUE|DOWNERS GROVE|           IL|        60515|
IL - Hinsdale|         99|           52288.24|            12649.27|     11075.81|
|065 - INTRACRANIA...|      50438|HUNTINGTON MEMORI...|100 W  CALIFORNIA...|    PASADENA|           CA|        91109| CA
- Los Angeles|         99|           54742.58|             8977.28|      7653.75|
|177 - RESPIRATORY...|     190015|NORTH OAKS MEDICA...|15790 PAUL VEGA M...|     HAMMOND|           LA|        70403| LA
- Baton Rouge|         99|          120135.51|            14531.45|     13066.74|
|065 - INTRACRANIA...|     210022|  SUBURBAN HOSPITAL|8600 OLD GEORGEIO...|    BETHESDA|           MD|        20814|  DC
- Washington|         99|            9509.47|             8990.43|      7675.67|
|177 - RESPIRATORY...|     420004| MUSC MEDICAL CENTER|     169 ASHLEY AVE|  CHARLESTON|           SC|        29425|  SC
- Charleston|         99|           39067.09|            20359.49|     19558.34|
|065 - INTRACRANIA...|     230020|OAKWOOD HOSPITAL ...|  18101 OAKWOOD BLVD|    DEARBORN|           MI|        48124|
MI - Dearborn|         99|           22958.29|             8259.64|      7366.58|
|189 - PULMONARY E...|     490050|VIRGINIA HOSPITAL...|1701 NORTH GEORGE...|   ARLINGTON|           VA|        22205|  V
A - Arlington|         99|           18840.75|             8479.69|      7748.86|
|065 - INTRACRANIA...|     260105|ST LOUIS UNIVERSI...|     3635 VISTA AVE|  SAINT LOUIS|           MO|        63110|  M
O - St. Louis|         99|           38156.67|            10715.07|      9675.52|
|190 - CHRONIC OBS...|     100191|MEDICAL CENTER OF...|         9330 SR 54|     TRINITY|           FL|        34655|
FL - Hudson|         99|           53682.82|             6597.35|      5461.56|
```