

# HOSPITAL CASESTUDY4- Hospital Data Analysis in the United States

## Objective 1 : Load file into spark

```
Applications Places System acadgild Wed Jul 11, 11:04 AM acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
scala> val hospitalDataDF= spark.read.csv("/hadoopdata/hdfs/spark/casestudy/inpatientCharges.csv").toDF("drgdefinition","prov
iderId","providerName","providerStreet","providerCity","providerState","providerZipCode","hospitalRefRegDesc","totalDischarge
s","AvgCovCharges","AvgTotPayments","AvgMedicarePaymets")
hospitalDataDF: org.apache.spark.sql.DataFrame = [drgdefinition: string, providerId: string ... 10 more fields]

scala> hospitalDataDF.createOrReplaceTempview("INPATIENT_CHARGES")
<console>:26: error: value createOrReplaceTempview is not a member of org.apache.spark.sql.DataFrame
    hospitalDataDF.createOrReplaceTempview("INPATIENT_CHARGES")
                    ^

scala> hospitalDataDF.createOrReplaceTempView("INPATIENT_CHARGES")
```

```
Applications Places System acadgild Wed Jul 11, 11:04 AM acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
... 48 elided

scala> spark.sql("select * from INPATIENT_CHARGES").show
+-----+-----+-----+-----+-----+-----+-----+-----+
|drgdefinition|providerId|providerName|providerStreet|providerCity|providerState|providerZipCode|hospi
talRefRegDesc|totalDischarges|AvgCovCharges|AvgTotPayments|AvgMedicarePaymets|
+-----+-----+-----+-----+-----+-----+-----+-----+
|DRGDefinition|ProviderId|ProviderName|ProviderStreetAdd...|ProviderCity|ProviderState|ProviderZipCode|Hospita
lReferralR...|TotalDischarges|AverageCoveredCha...|AverageTotalPayments|AverageMedicarePa...|
|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|5858.5|
|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|
|039 - EXTRACRANIA...|10046|RIVERVIEW REGIONA...|600 SOUTH THIRD S...|GADSDEN|AL|35901|AL
- Birmingham|14|67327.92|5461.57|4493.57|
|039 - EXTRACRANIA...|10055|FLOWERS HOSPITAL|4370 WEST MAIN ST...|DOTHAN|AL|36305|
AL - Dothan|45|39607.28|5356.28|4408.2|
|039 - EXTRACRANIA...|10056|ST VINCENT'S BIRM...|810 ST VINCENT'S ...|BIRMINGHAM|AL|35205|AL
- Birmingham|43|22862.23|5374.65|4186.02|
|039 - EXTRACRANIA...|10078|NORTHEAST ALABAMA...|400 EAST 10TH STREET|ANNISTON|AL|36207|AL
- Birmingham|21|31110.85|5366.23|4376.23|
```

## HOSPITAL CASESTUDY4- Hospital Data Analysis in the United States

- 1) What is the average amount of AverageCoveredCharges per state

```
scala> spark.sql("select providerState, avg(AvgCovCharges) as total_avg from INPATIENT_CHARGES group by providerState").show
```

providerState	total_avg
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	31736.427824058758
MT	22670.015237154144

only showing top 20 rows

```
scala>
```

- 2) find out the AverageTotalPayments charges per state

```
scala> spark.sql("select providerState, avg(AvgTotPayments) as total_avg from INPATIENT_CHARGES group by providerState").show
```

providerState	total_avg
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

```
scala>
```

## HOSPITAL CASESTUDY4- Hospital Data Analysis in the United States

- 3) find out the Average Medicare Payments charges per state.

```
scala> spark.sql("select providerState, avg(AvgMedicarePayments) as total_avg from INPATIENT_CHARGES group by providerState").show
```

providerState	total_avg
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

```
scala>
```

9 Items in Trash

[acadgild@loc... acadgild@local... [Downloads] [Gmail - Mozill... [acadgild] [inpatientChar...

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

```
scala> spark.sql("select providerState, drgdefinition, sum(totalDischarges) as total from INPATIENT_CHARGES group by providerState, drgdefinition order by sum(totalDischarges) desc").show
```

providerState	drgdefinition	total
CA	871 - SEPTICEMIA ...	34284.0
TX	470 - MAJOR JOINT...	30095.0
FL	470 - MAJOR JOINT...	29985.0
CA	470 - MAJOR JOINT...	29731.0
TX	871 - SEPTICEMIA ...	23144.0
NY	871 - SEPTICEMIA ...	21970.0
FL	392 - ESOPHAGITIS...	21298.0
IL	470 - MAJOR JOINT...	20095.0
NY	470 - MAJOR JOINT...	19371.0
FL	871 - SEPTICEMIA ...	18660.0
TX	690 - KIDNEY & UR...	17384.0
NY	392 - ESOPHAGITIS...	17337.0
MI	470 - MAJOR JOINT...	16847.0
PA	470 - MAJOR JOINT...	16712.0
FL	292 - HEART FAILU...	16639.0
FL	690 - KIDNEY & UR...	16405.0
OH	470 - MAJOR JOINT...	16062.0
NC	470 - MAJOR JOINT...	15820.0
IL	871 - SEPTICEMIA ...	15610.0
MI	871 - SEPTICEMIA ...	15548.0

only showing top 20 rows

```
scala>
```

[acadgild@loc... acadgild@local... [Downloads] [Gmail - Mozill... [acadgild] [inpatientChar...