

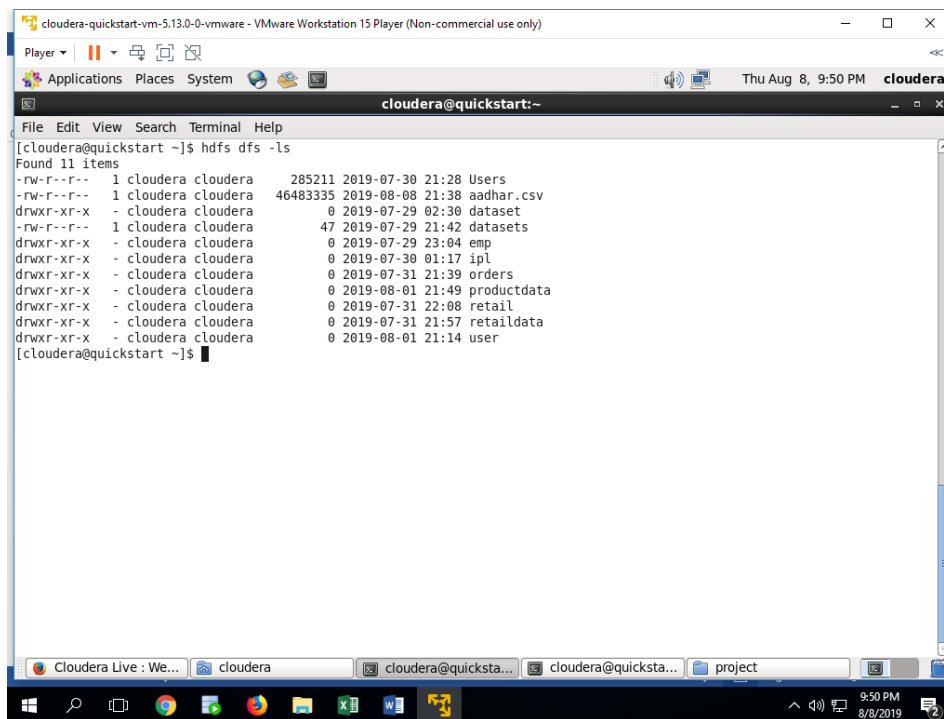
# Checkpoint 1

**1.//loading of adhaar.csv file in hdfs by first copying it to the home folder and then putting it in hdfs using**

hdfs dfs- put /home/cloudera/project/aadhar.csv

**//to check that the file exists in hdfs**

hdfs dfs-ls



```
cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 15 Player (Non-commercial use only)
Player
Applications Places System
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 11 items
-rw-r--r-- 1 cloudera cloudera 285211 2019-07-30 21:28 Users
-rw-r--r-- 1 cloudera cloudera 46483335 2019-08-08 21:38 aadhar.csv
drwxr-xr-x - cloudera cloudera 0 2019-07-29 02:30 dataset
-rw-r--r-- 1 cloudera cloudera 47 2019-07-29 21:42 datasets
drwxr-xr-x - cloudera cloudera 0 2019-07-29 23:04 emp
drwxr-xr-x - cloudera cloudera 0 2019-07-30 01:17 ipl
drwxr-xr-x - cloudera cloudera 0 2019-07-31 21:39 orders
drwxr-xr-x - cloudera cloudera 0 2019-08-01 21:49 productdata
drwxr-xr-x - cloudera cloudera 0 2019-07-31 22:08 retail
drwxr-xr-x - cloudera cloudera 0 2019-07-31 21:57 retaildata
drwxr-xr-x - cloudera cloudera 0 2019-08-01 21:14 user
[cloudera@quickstart ~]$
```

**//creating and using database in hive**

create database if not exists aadhar

use aadhar

**//creating managed table in hive and loading data into it**

create table if not exists aadharInfo(registrar string,enrolmentAgency string,state string,district string,subDistrict string,pinCode string,gender string, age int,aadharGenerated int,enrolmentRejected int,residentsProvidingEmail int, residentsProvidingMobileNumber int)

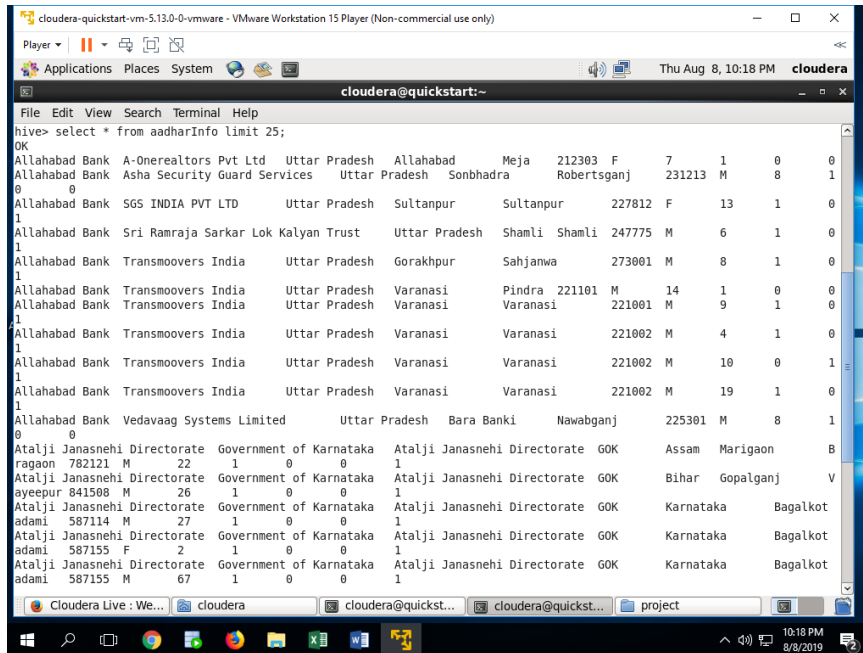
> row format delimited fields terminated by ','

```
> TBLPROPERTIES('skip.header.line.count'='1');
```

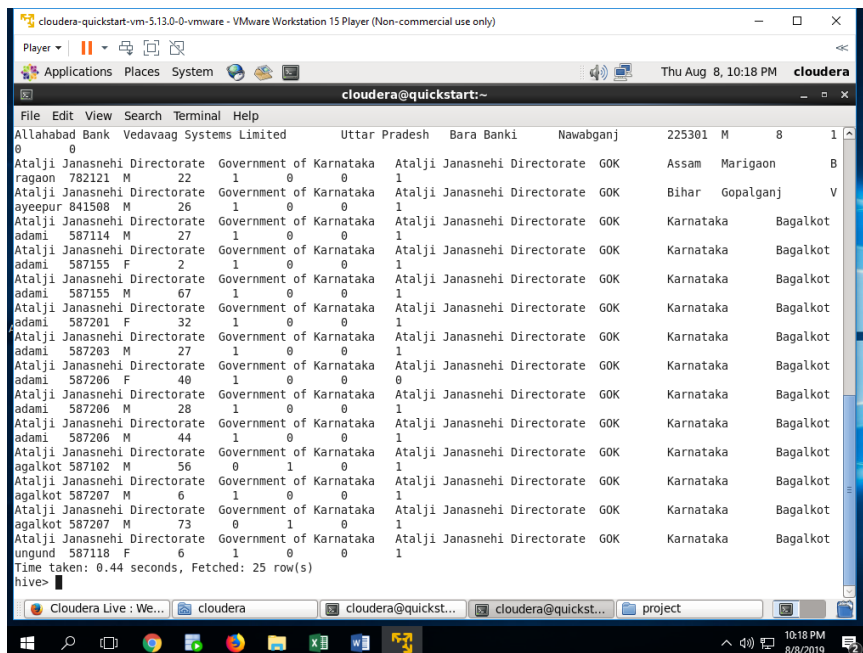
load data local inpath '/home/cloudera/project/aadhar.csv' into table aadharInfo;

//checking if data has been loaded properly

```
select * from aadharInfo limit 25;
```



```
hive> select * from aadharInfo limit 25;
OK
Allahabad Bank A-Onerealtors Pvt Ltd Uttar Pradesh Allahabad Meja 212303 F 7 1 0 0
Allahabad Bank Asha Security Guard Services Uttar Pradesh Sonbhadra 231213 M 8 1
0
Allahabad Bank SGS INDIA PVT LTD Uttar Pradesh Sultanpur Sultanpur 227812 F 13 1 0
1
Allahabad Bank Sri Ramraja Sarkar Lok Kalyan Trust Uttar Pradesh Shamli Shamli 247775 M 6 1 0
1
Allahabad Bank Transmoovers India Uttar Pradesh Gorakhpur Sahjanwa 273001 M 8 1 0
1
Allahabad Bank Transmoovers India Uttar Pradesh Varanasi Pindra 221101 M 14 1 0 0
Allahabad Bank Transmoovers India Uttar Pradesh Varanasi Varanasi 221001 M 9 1 0
1
Allahabad Bank Transmoovers India Uttar Pradesh Varanasi Varanasi 221002 M 4 1 0
1
Allahabad Bank Transmoovers India Uttar Pradesh Varanasi Varanasi 221002 M 10 0 1
1
Allahabad Bank Transmoovers India Uttar Pradesh Varanasi Varanasi 221002 M 19 1 0
1
Allahabad Bank Vedavaag Systems Limited Uttar Pradesh Bara Banki Nawabganj 225301 M 8 1
0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Assam Marigaon B
ragoon 782121 M 22 1 0 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Bihar Gopalganj V
ayeeapur 841508 M 26 1 0 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
adani 587114 M 27 1 0 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
adani 587155 F 2 1 0 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
adani 587155 M 67 1 0 0
```



```
hive> select * from aadharInfo limit 25;
OK
Allahabad Bank Vedavaag Systems Limited Uttar Pradesh Bara Banki Nawabganj 225301 M 8 1
0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Assam Marigaon B
ragoon 782121 M 22 1 0 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Bihar Gopalganj V
ayeeapur 841508 M 26 1 0 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
adani 587114 M 27 1 0 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
adani 587155 F 2 1 0 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
adani 587155 M 67 1 0 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
adani 587201 F 32 1 0 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
adani 587203 M 27 1 0 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
adani 587206 F 40 1 0 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
adani 587206 M 28 1 0 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
adani 587206 M 44 1 1 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
agalgot 587102 M 56 0 1 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
agalgot 587207 M 6 1 0 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
agalgot 587207 M 73 0 1 0
Atalji Janasnehi Directorate Government of Karnataka Atalji Janasnehi Directorate GOK Karnataka Bagalkot
ungund 587118 F 6 1 0 0
Time taken: 0.44 seconds, Fetched: 25 row(s)
hive>
```

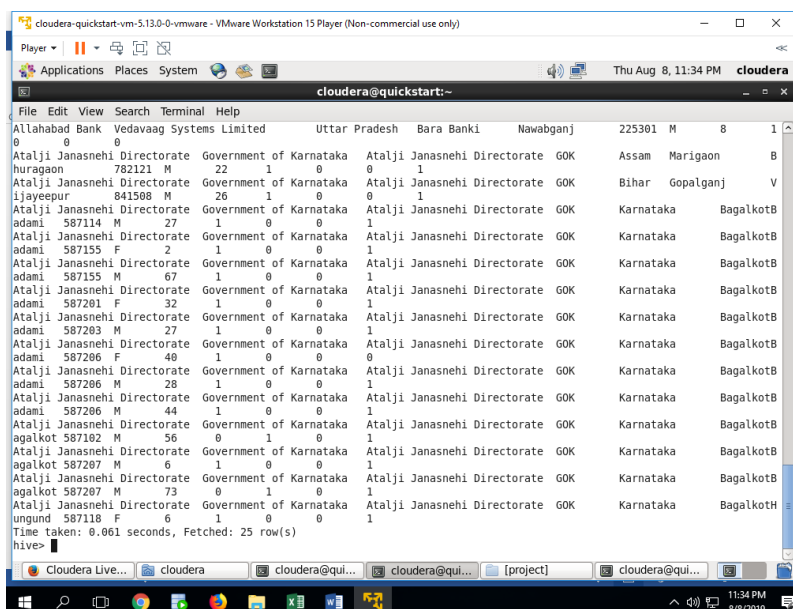
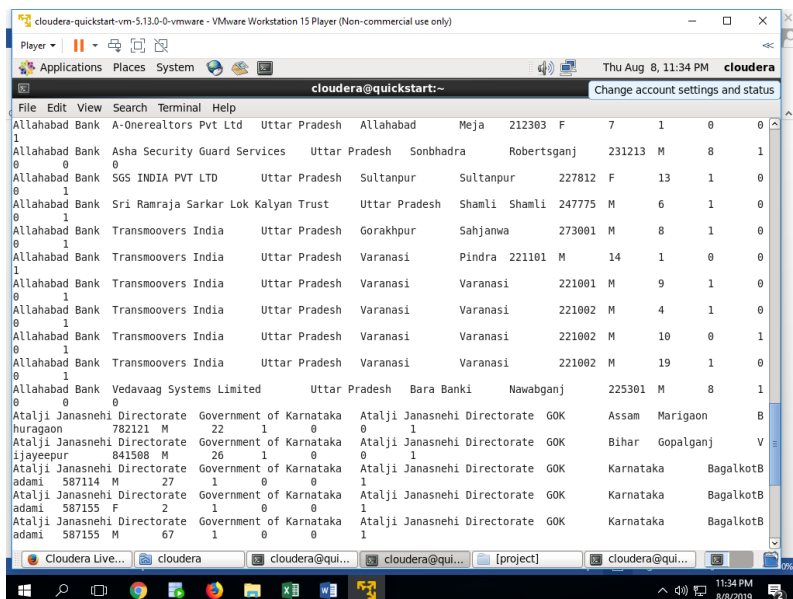
## //creating external table in hive and loading data into it

create external table if not exists aadharInfoExternal(registrar string,enrolmentAgency string,state string,district string, subDistrict string,pinCode string,gender string, age int,aadharGenerated int,enrolmentRejected int,residentsProvidingEmail int, residentsProvidingMobileNumber int)

- > row format delimited fields terminated by ','
- > location '/user/cloudera/project'
- > TBLPROPERTIES('skip.header.line.count'='1');

## //checking if data has been loaded properly

select \* from aadharInfoExternal limit 25;



## //creating dataframe in spark

```
val aadharRDD=sc.textFile("/user/cloudera/aadhar.csv")
```

## //removing header

```
val remHeader=aadharRDD.first()
```

remHeader: String = Registrar,Enrolment Agency,State,District,Sub District,Pin

Code,Gender,Age,Aadhaar generated,Enrolment Rejected,Residents providing email,Residents providing mobile number

```
val final_data=aadharRDD.filter(row=>row!=(remHeader))
```

## //creating dataframe

```
val
```

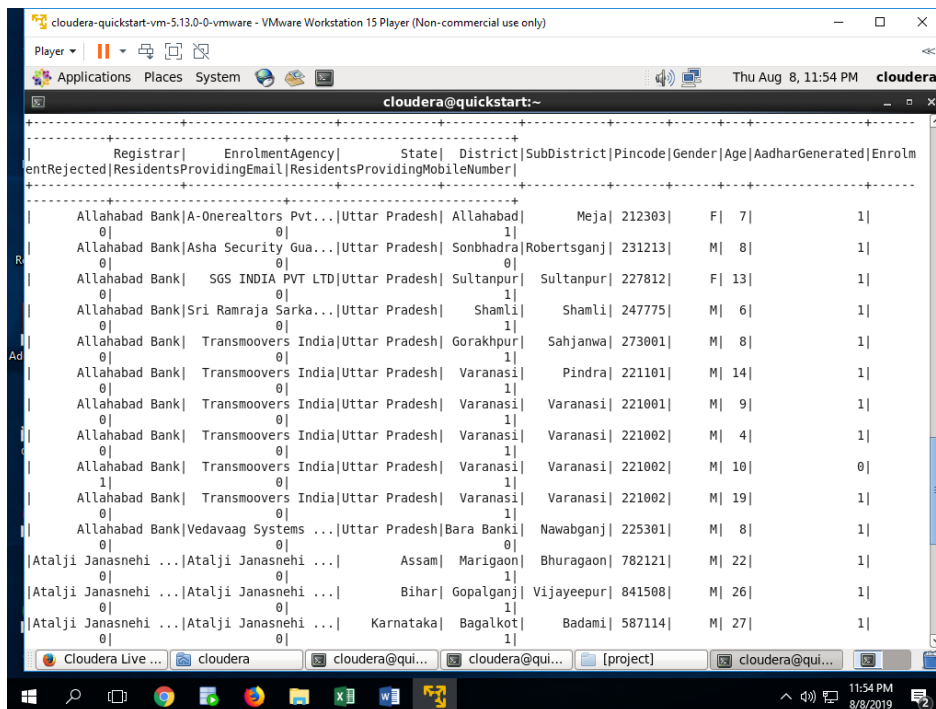
```
splitData=final_data.map(x=>(x.split(",")(0),x.split(",")(1),x.split(",")(2),x.split(",")(3),x.split(",")(4),x.split(",")(5),x.split(",")(6),x.split(",")(7)toInt,x.split(",")(8)toInt,x.split(",")(9)toInt,x.split(",")(10)toInt,x.split(",")(11)toInt));
```

```
val
```

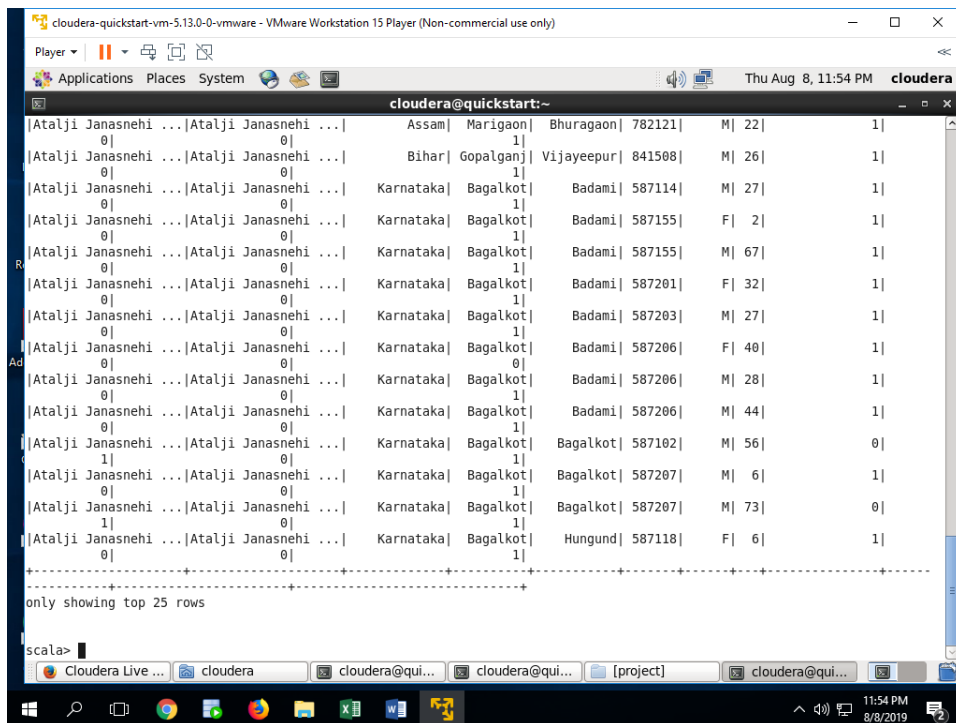
```
aadharDF=splitData.toDF("Registrar","EnrolmentAgency","State","District","SubDistrict","Pincode","Gender","Age","AadhaarGenerated","EnrolmentRejected","ResidentsProvidingEmail","ResidentsProvidingMobileNumber")
```

## //checking if properly loaded in Dataframe

```
aadharDF.show(25)
```



Registrar	EnrolmentAgency	State	District	SubDistrict	Pincode	Gender	Age	AadhaarGenerated	EnrolmentRejected	ResidentsProvidingEmail	ResidentsProvidingMobileNumber
Allahabad Bank	A-Onerealtors Pvt...	Uttar Pradesh	Allahabad	Meja	212303	F	7	1			
Allahabad Bank	Asha Security Gua...	Uttar Pradesh	Sonbhadra	Robertsganj	231213	M	8	1			
Allahabad Bank	SGS INDIA PVT LTD	Uttar Pradesh	Sultanpur	Sultanpur	227812	F	13	1			
Allahabad Bank	Sri Ramraja Sarka...	Uttar Pradesh	Shamli	Shamli	247775	M	6	1			
Allahabad Bank	Transmoovers India	Uttar Pradesh	Gorakhpur	Sahjanwa	273001	M	8	1			
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Pindra	221101	M	14	1			
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Varanasi	221001	M	9	1			
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Varanasi	221002	M	4	1			
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Varanasi	221002	M	10	0			
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Varanasi	221002	M	19	1			
Allahabad Bank	Vedavaag Systems ...	Uttar Pradesh	Bara Banki	Nawabganj	225301	M	8	1			
Atalji Janasnehi ...	Atalji Janasnehi ...	Assam	Marigaon	Bhuragaon	782121	M	22	1			
Atalji Janasnehi ...	Atalji Janasnehi ...	Bihar	Gopalganj	Vijayeeipur	841508	M	26	1			
Atalji Janasnehi ...	Atalji Janasnehi ...	Karnataka	Bagalkot	Badami	587114	M	27	1			



## Checkpoint 2

### 2. //Describing the schema

aadharDF.printSchema

#### OUTPUT

```
|-- Registrar: string (nullable = true)
|-- EnrolmentAgency: string (nullable = true)
|-- State: string (nullable = true)
|-- District: string (nullable = true)
|-- SubDistrict: string (nullable = true)
|-- Pincode: string (nullable = true)
```

```
|-- Gender: string (nullable = true)
|-- Age: integer (nullable = false)
|-- AadharGenerated: integer (nullable = false)
|-- EnrolmentRejected: integer (nullable = false)
|-- ResidentsProvidingEmail: integer (nullable = false)
|-- ResidentsProvidingMobileNumber: integer (nullable = false)
```

```
cloudera@quickstart:~$ scala
scala> aadharDF.printSchema
root
 |-- Registrar: string (nullable = true)
 |-- EnrolmentAgency: string (nullable = true)
 |-- State: string (nullable = true)
 |-- District: string (nullable = true)
 |-- SubDistrict: string (nullable = true)
 |-- Pincode: string (nullable = true)
 |-- Gender: string (nullable = true)
 |-- Age: integer (nullable = false)
 |-- AadharGenerated: integer (nullable = false)
 |-- EnrolmentRejected: integer (nullable = false)
 |-- ResidentsProvidingEmail: integer (nullable = false)
 |-- ResidentsProvidingMobileNumber: integer (nullable = false)

scala>
```

### 3.Count and names of registrars in the table.

```
select registrar , count(registrar) from aadharInfo group by registrar;
```

#### OUTPUT

...

Allahabad Bank 11

Atalji Janasnehi Directorate	Government of Karnataka	1458
Bank Of India		19791
Bank of Baroda		1412
CSC e-Governance Services India Limited		209771
Canara Bank		867
Commissioner Nagaland		25
DC Aalo		126
DC ITANAGAR CAPITAL COMPLEX		38
DC LOHIT		119
DC NAMSAI		154
DC PAPUMPARE		15
DC Siang		38
DENA BANK		33869
DIT Lakshadweep		1
Department of Information Technology Govt of Jharkhand		464
Dept of ITC Govt of Rajasthan		13565
Director General Health Services Health Deptt	Haryana	3
Director Health and Family Welfare	UT	165
Directorate of Public Health and Family Welfare	Govt of Andhra Pradesh	23
Directorate of Woman and Child Development	Government of Himachal Pradesh	33
FCR Govt of Haryana		1823
FCS Govt of Punjab		173
Govt of Goa		857
Govt of Gujarat		13894
Govt of Himachal Pradesh		583
Govt of Karnataka		5330
Govt of Kerala		11937
Govt of Maharashtra		241
Govt of Sikkim - Dept of Econo		32

Govt of UT of Chandigarh	95
Govt. of Mizoram	3220
Govt. of Uttarkhand	44
IDBI Bank Ltd	31
Information Technology & Communication Department	3958
Madhya Pradesh State Electronics Development Corporation Ltd.	17309
NSDL e-Governance Infrastructure Limited	54214
National Cooperative Consumers Federation Of India Limited	2590
Odisha Computer Application Center	1701
Punjab National Bank	1400
Punjab and Sind Bank	1543
RDD Govt of Tripura	606
Registrar General India BEL2	167
Registrar General India ECIL	757
Registrar General India Others	7
Registrar General of India ITI	55
Rural Development Department Bihar-1	640
Rural Development Dept Govt. of Bihar	4145
Secretery IT J&K	110
State Bank of India	3422
Tamil Nadu eGovernance Agency	15468
U P Electronics Corporation Limited	293
U.P. Development Systems Corporation Ltd	4139
UIDAI-Registrar	19
UT Govt. Of Dadra & Nagar Haveli	46
UT Of Daman and Diu	50
UT of Puducherry	1
UTI Infrastructure Technology & Services Limited	2395
Union Bank	5536



Time taken: 27.469 seconds, Fetched: 60 row(s)

#### 4.Number of states, districts in each state and sub-districts in each district.

select state,count(distinct district) from aadharInfo group by state;

#### OUTPUT

Andaman and Nicobar Islands 2

Andhra Pradesh13

Arunachal Pradesh 17

Assam 28

Bihar 38

Chandigarh 1

Chhattisgarh 30

Dadra and Nagar Haveli 1

Daman and Diu 2

Delhi 9

Goa 2

Gujarat 33

Haryana 21

Himachal Pradesh 11

Jammu and Kashmir 22

Jharkhand 24

Karnataka 30

Kerala 14

Lakshadweep 1

Madhya Pradesh 50

Maharashtra	36
Manipur	9
Meghalaya	8
Mizoram	8
Nagaland	11
Odisha	30
Others	1
Puducherry	2
Punjab	22
Rajasthan	33
Sikkim	4
Tamil Nadu	32
Telangana	10
Tripura	8
Uttar Pradesh	75
Uttarakhand	13
West Bengal	19

Time taken: 20.985 seconds, Fetched: 37 row(s)

```
select district,count(distinct subdistrict) from aadhar group by district;
```

## OUTPUT

...

Tawang	1
Tehri Garhwal	5
Thane	9
Thanjavur	8
The Dangs	4
The Nilgiris	6
Theni	5

Thiruvananthapuram 4

Thiruvavarur 7

Thoothukkudi 9

Thoubal 3

Thrissur5

Tikamgarh 5

Tinsukia 4

Tirap 1

Tiruchirappalli 9

Tirunelveli 13

Tiruppur 7

Tiruvallur 10

Tiruvannamalai 8

Tonk 7

Tuensang 2

Tumakuru 11

Udaipur 10

Udalguri 7

Udham Singh Nagar 5

Udhampur 4

Udupi 4

Ujjain 7

Ukhrul 5

Umaria 1

Una 9

Unakoti3

Unnao 5

Upper Siang 3

Upper Subansiri10

Uttar Bastar Kanker	7
Uttar Dinajpur	9
Uttara Kannada	13
Uttarkashi	4
Vadodara	8
Vaishali	16
Valsad	6
Varanasi	2
Vellore	9
Vidisha	7
Vijayapura	5
Villupuram	9
Virudhunagar	8
Visakhapatnam	41
Vizianagaram	34
Warangal	49
Wardha	8
Washim	6
Wayanad	3
West Champaran	18
West Delhi	3
West Garo Hills	8
West Godavari	46
West Kameng	3
West Khasi Hills	1
West Siang	14
West Sikkim	2
West Singhbhum	17
West Tripura	7

Wokha 3

Yadgir 3

Yamuna Nagar 2

Yavatmal 16

Zunheboto 5

Time taken: 21.806 seconds, Fetched: 664 row(s)

## 6.Names of private agencies for each state.

Select distinct(enrolmentAgency), state from aadharInfo;

### OUTPUT

...

West Bengal NPS Technologies Pvt. Ltd

West Bengal National Cooperative Consumers Federation of India Limited

West Bengal Nekton IT India Pvt Ltd.

West Bengal Netlink software Pvt Ltd

West Bengal Network for Information & Computer

West Bengal Nevaeh Technology Pvt. Ltd.

West Bengal Nielsen India Private Limited

West Bengal OSWAL COMPUTERS & CONSULTANTS

West Bengal Ojus Healthcare Private Limited

West Bengal Om Softwares

West Bengal Omnitech Infosolutions Ltd

West Bengal Organisation for Development Integrated Social & Health Action ODISHA

West Bengal Osiris Infotech Pvt. Ltd.

West Bengal PROTEX COMPUTER PVT LTD

West Bengal	Prakash Computer Services
West Bengal	Promind Solutions P Limited
West Bengal	Quick Data IT Services Pvt Ltd
West Bengal	RBS multisolutions private limited
West Bengal	RELIGARE SECURITIES LTD
West Bengal	Radiant Haroti Industries India Ltd
West Bengal	Rajcomp Info Services Ltd
West Bengal	SGS INDIA PVT LTD
West Bengal	SILVER JUBILEE MOTORS LTD.
West Bengal	SREI INFRASTRUCTURE FINANCES L
West Bengal	SRM Education And Social Welfare Society
West Bengal	SRR Infotech
West Bengal	SVG Express Services Pvt Ltd
West Bengal	Saket Advertising Pvt. Ltd
West Bengal	Sant Naval Institute of Information Technology
West Bengal	Sarvalabh Global Foundation
West Bengal	Seva Society Collector Kutch
West Bengal	SoftAge Information Technology Limited
West Bengal	Squaria Global India Private Limited
West Bengal	Sri Ramraja Sarkar Lok Kalyan Trust
West Bengal	Steel City Securities Limited
West Bengal	Super Printers
West Bengal	Synapses Solutions Private Limited
West Bengal	TAMILNADU ARASU CABLE TV CORPORATION LTD
West Bengal	Techno Bytes Information Pvt. Ltd
West Bengal	Twinstar Industries Ltd.
West Bengal	UT Computers Educational & Welfare Soc

West Bengal	UT of Daman and Diu
West Bengal	United Telecoms Ltd
West Bengal	United Telecoms e-Services Pvt Ltd
West Bengal	Urmila Info solution
West Bengal	Utility Forms Pvt Ltd
West Bengal	VAP INFOSOLUTIONS
West Bengal	VEETECHNOLOGIES PVT. LTD
West Bengal	VISION COMPTECH INTEGRATOR LTD
West Bengal	Vakrangee Softwares Limited
West Bengal	Vayam technologies Ltd
West Bengal	Vedavaag Systems Limited
West Bengal	Virinchi Technologies Ltd
West Bengal	WEBEL TECHNOLOGY LIMITED
West Bengal	Wipro Ltd
West Bengal	Zephyr System Pvt.Ltd.

Time taken: 21.647 seconds, Fetched: 2271 row(s)

## Checkpoint 3

### 8.Top 3 states generating most number of Aadhaar cards.

```
select count(*) as no_of_registrations , state from aadharInfo group by state order by  
no_of_registrations desc limit 3;
```

#### OUTPUT

81776 Bihar

69476 Uttar Pradesh

60485 West Bengal

Time taken: 39.48 seconds, Fetched: 3 row(s)

### 9. Top 3 private agencies generating the most number of Aadhar cards.

```
select count(*) as no_of_registration,enrolmentAgency from aadharInfo group by enrolmentAgency  
order by no_of_registration desc limit 3;
```

#### OUTPUT

100357	CSC SPV
18101 Welfare Society	SRM Education And Social
16972 FINANCES L	SREI INFRASTRUCTURE

Time taken: 39.19 seconds, Fetched: 3 row(s)

### 10. The number of residents providing email, mobile number

```
select count(*) from aadharInfo where residentsProvidingEmail > 0 and  
residentsProvidingMobileNumber>0;
```

#### OUTPUT

16951

Time taken: 22.043 seconds, Fetched: 1 row(s)

### 11. Top 3 districts where enrolment numbers are maximum

```
select count(*) as no_of_registration,district from aadharInfo group by district order by  
no_of_registration desc limit 3;
```

#### OUTPUT

7135 Bardhaman



6894 North 24 Parganas

6078 South 24 Parganas

Time taken: 40.861 seconds, Fetched: 3 row(s)

## 12. No. of Aadhaar cards generated in each state

select count(\*),state from aadharInfo group by state;

### OUTPUT

...

7 Andaman and Nicobar Islands

4540 Andhra Pradesh

632 Arunachal Pradesh

2972 Assam

81776 Bihar

199 Chandigarh

4617 Chhattisgarh

107 Dadra and Nagar Haveli

99 Daman and Diu

7247 Delhi

799 Goa

24146 Gujarat

5138 Haryana

1283 Himachal Pradesh

1331 Jammu and Kashmir

7423 Jharkhand

15755 Karnataka

12378 Kerala

5 Lakshadweep

37360	Madhya Pradesh
19783	Maharashtra
562	Manipur
259	Meghalaya
3172	Mizoram
392	Nagaland
11972	Odisha
12	Others
85	Puducherry
5888	Punjab
28659	Rajasthan
48	Sikkim
21196	Tamil Nadu
3768	Telangana
726	Tripura
69476	Uttar Pradesh
6521	Uttarakhand
60485	West Bengal

Time taken: 19.925 seconds, Fetched: 37 row(s)

## Checkpoint 4

### 13.Data frame using the file and its summary

```
val  
splitData=final_data.map(x=>(x.split(",")(0),x.split(",")(1),x.split(",")(2),x.split(",")(3),x.split(",")(4),x.split(",")
```

```
,"")(5),x.split(",")(6),x.split(",")(7)toInt,x.split(",")(8)toInt,x.split(",")(9)toInt,x.split(",")(10)toInt,x.split(",")(11)toInt));
```

```
val
```

```
aadharDF=splitData.toDF("Registrar","EnrolmentAgency","State","District","SubDistrict","Pincode","Gender","Age","AadharGenerated","EnrolmentRejected","ResidentsProvidingEmail","ResidentsProvidingMobileNumber")
```

```
aadharDF.describe()
```

## OUTPUT

```
res5: org.apache.spark.sql.DataFrame = [summary: string, Age: string, AadharGenerated: string, EnrolmentRejected: string, ResidentsProvidingEmail: string, ResidentsProvidingMobileNumber: string]
```

## 14.Command to see the correlation between “age” and “mobile\_number”

```
select corr(age,residentsProvidingMobileNumber) from aadharInfo;
```

## OUTPUT

```
-0.11754461896889339
```

```
Time taken: 22.423 seconds, Fetched: 1 row(s)
```

## 15. Number of unique pincodes in the data

```
select distinct(pinCode) from aadharInfo;
```

## OUTPUT

```
...
```

```
854325
```

```
854326
```

```
854327
```

```
854328
```

```
854329
```

```
854330
```

```
854331
```

```
854332
```

854333

854334

854335

854336

854337

854338

854339

854340

855101

855102

855105

855106

855107

855108

855113

855114

855115

855116

855117

855456

Others

Time taken: 21.501 seconds, Fetched: 17756 row(s)

## **16. Number of Aadhaar registrations rejected in Uttar Pradesh and Maharashtra**

```
select count(enrolmentrejected),state from aadharInfo where state in("Uttar Pradesh","Maharashtra") and enrolmentrejected=1 group by state;
```

#### OUTPUT

1581    Maharashtra

3892    Uttar Pradesh

Time taken: 23.18 seconds, Fetched: 2 row(s)

## Checkpoint 5

### 17. The top 3 states where the percentage of Aadhaar cards being generated for males is the highest.

```
select state, (sum(aadharGenerated)*100)/(sum(aadharGenerated+enrolmentRejected))as male_numbers from aadharInfo where gender = 'M' group by state order by male_numbers desc limit 3;
```

#### OUTPUT

Andaman and Nicobar Islands      100.0

Others   100.0

Lakshadweep   100.0

Time taken: 40.643 seconds, Fetched: 3 row(s)

### 18. In each of these 3 states, the top 3 districts where the percentage of Aadhaar cards being rejected for females is the highest.

```
select district, (sum(enrolmentRejected)*100)/(sum(aadharGenerated+enrolmentRejected))as female_numbers from aadharInfo where gender = 'F' and state in("Andaman and Nicobar Islands","Others","Lakshadweep") group by district order by female_numbers desc limit 3;
```

#### OUTPUT

Lakshadweep   100.0

South Andaman50.0

North And Middle Andaman 33.33333333333336

Time taken: 41.925 seconds, Fetched: 3 row(s)

### 19. The top 3 states where the percentage of Aadhaar cards being generated for females is the highest

```
select state, (sum(aadhaarGenerated)*100)/(sum(aadhaarGenerated+enrolmentRejected))as  
female_numbers from aadhaarInfo where gender = 'F' group by state order by female_numbers desc  
limit 3;
```

#### OUTPUT

Dadra and Nagar Haveli	100.0
Sikkim	100.0
Others	100.0

### 20. In each of these 3 states, identify the top 3 districts where the percentage of Aadhaar cards being rejected for males is the highest.

```
select district,  
(sum(enrolmentRejected)*100)/(sum(aadhaarGenerated+enrolmentRejected))as  
male_numbers from aadhaarInfo where gender = 'M' and state in("Dadra and Nagar  
Haveli","Sikkim","Others") group by district order by male_numbers desc limit 3;
```

#### OUTPUT

East Sikkim	9.090909090909092
Dadra and Nagar Haveli	3.4482758620689653
West Sikkim	0.0

### 21. The summary of the acceptance percentage of all the Aadhaar cards applications by bucketing the age group into 10 buckets.

```
create table if not exists aadhaarinfo_bucket(registrar string, enrolmentAgency string,state  
string, district string,subDistrict string, pinCode string, gender string,age int,aadhaarGenerated  
int, enrolmentRejected int,residentsProvidingEmail int,residentsProvidingMobileNumber int)
```

```
> clustered by(age) into 10 buckets
> row format delimited fields terminated by ',';
//inserting data from staging table
insert into aadharInfo_bucket select * from aadharInfo;

select((sum(aadharGenerated)/(sum(aadharGenerated+enrolmentRejected)))*100)as
acceptancePercentage from aadharinfo_bucket;
```

### **OUTPUT**

94.81864032289477

Time taken: 20.686 seconds, Fetched: 1 row(s)