

# Project

## 1.Loading data into hdfs—

- `hdfs dfs -put /home/cloudera/Project/aadhar.csv`

## Creating and loading Data managed table for aadhar.csv--

- create table if not exists aadhar(registrar String, enrolmentAgency String ,state String,District String,subDistrict String,Pincode String,gender String, age Int,adhaarGenerated 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 aadhar;

## Command to see the data in the table

- `Select * from aadhar limit 25;`

cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 15 Player (Non-commercial use only)

Player | Applications | Places | System | Thu Aug 8, 10:27 PM | cloudera

cloudera@quickstart:~

File Edit View Search Terminal Help

Time taken: 0.899 seconds, Fetched: 10 row(s)

hive> select \* from aadhar limit 25;

OK

Allahabad Bank	A-Onerealtors Pvt Ltd	Uttar Pradesh	Allahabad	Meja	212303	F	7	1	0	0	1
Allahabad Bank	Asha Security Guard Services	Uttar Pradesh	Sonbhadra	Robertsganj	231213	M	8	1	0	0	0
Allahabad Bank	SGS INDIA PVT LTD	Uttar Pradesh	Sultanpur	Sultanpur	227812	F	13	1	0	0	0
Allahabad Bank	Sri Ramraja Sarkar Lok Kalyan Trust	Uttar Pradesh	Shamli	Shamli	247775	M	6	1	0	0	0
Allahabad Bank	Transmoovers India	Uttar Pradesh	Gorakhpur	Sahjanwa	273001	M	8	1	0	0	0
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Pindra	221101	M	14	1	0	0	1
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Varanasi	221001	M	9	1	0	0	0
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Varanasi	221002	M	4	1	0	0	0
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Varanasi	221002	M	10	0	1	0	0
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Varanasi	221002	M	19	1	0	0	0
Allahabad Bank	Vedavaag Systems Limited	Uttar Pradesh	Bara Banki	Nawabganj	225301	M	8	1	0	0	0
Atalji Janasnehi Directorate	Government of Karnataka	Atalji Janasnehi Directorate	GOK	Assam	Marigaon	Bhura					
Atalji Janasnehi Directorate	Government of Karnataka	Atalji Janasnehi Directorate	GOK	Bihar	Gopalganj	Vijay					
Atalji Janasnehi Directorate	Government of Karnataka	Atalji Janasnehi Directorate	GOK	Karnataka	Bagalkot	B					
Atalji Janasnehi Directorate	Government of Karnataka	Atalji Janasnehi Directorate	GOK	Karnataka	Bagalkot	B					
Atalji Janasnehi Directorate	Government of Karnataka	Atalji Janasnehi Directorate	GOK	Karnataka	Bagalkot	B					

[Cloudera Live... | cloudera | [cloudera@qui... | [Project] | cloudera@quic... | aadhar.csv (~/. | 10:27 PM 8/8/2019]

- `Select * from aadhar1 limit 25;`

cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 15 Player (Non-commercial use only)

Player ▾ | Applications Places System | Thu Aug 8, 10:47 PM cloudera

cloudera@quickstart:~

File Edit View Search Terminal Help

Time taken: 0.073 seconds

```
hive> select * from aadhar1 limit 25;
```

OK

Allahabad Bank	A-Onerealtors Pvt Ltd	Uttar Pradesh	Allahabad	Meja	212303	F	7	1	0	0	1
Allahabad Bank	Asha Security Guard Services	Uttar Pradesh	Sonbhadra	Robertsganj	231213	M	8	1	0	0	0
Allahabad Bank	SGS INDIA PVT LTD	Uttar Pradesh	Sultanpur	Sultanpur	227812	F	13	1	0	0	0
Allahabad Bank	Sri Ramraja Sarkar Lok Kalyan Trust	Uttar Pradesh	Shamli	Shamli	247775	M	6	1	0	0	0
Allahabad Bank	Transmoovers India	Uttar Pradesh	Gorakhpur	Sahjanwa	273001	M	8	1	0	0	0
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Pindra	221101	M	14	1	0	0	1
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Varanasi	221001	M	9	1	0	0	0
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Varanasi	221002	M	4	1	0	0	0
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Varanasi	221002	M	10	0	1	0	0
Allahabad Bank	Transmoovers India	Uttar Pradesh	Varanasi	Varanasi	221002	M	19	1	0	0	0
Allahabad Bank	Vedavaag Systems Limited	Uttar Pradesh	Bara Banki	Nawabganj	225301	M	8	1	0	0	0
Atalji Janasnehi Directorate	Government of Karnataka	Atalji Janasnehi Directorate	GOK	Assam	Marigaon	Bhura					
Atalji Janasnehi Directorate	Government of Karnataka	Atalji Janasnehi Directorate	GOK	Bihar	Gopalganj	Vijay					
Atalji Janasnehi Directorate	Government of Karnataka	Atalji Janasnehi Directorate	GOK	Karnataka	Bagalkot	B					
Atalji Janasnehi Directorate	Government of Karnataka	Atalji Janasnehi Directorate	GOK	Karnataka	Bagalkot	B					
Atalji Janasnehi Directorate	Government of Karnataka	Atalji Janasnehi Directorate	GOK	Karnataka	Bagalkot	B					
Atalji Janasnehi Directorate	Government of Karnataka	Atalji Janasnehi Directorate	GOK	Karnataka	Bagalkot	B					

Project

cloudera@quickstart:~

File Edit View Search Terminal Help

Time taken: 0.091 seconds, Fetched: 25 row(s)

```
hive>
```

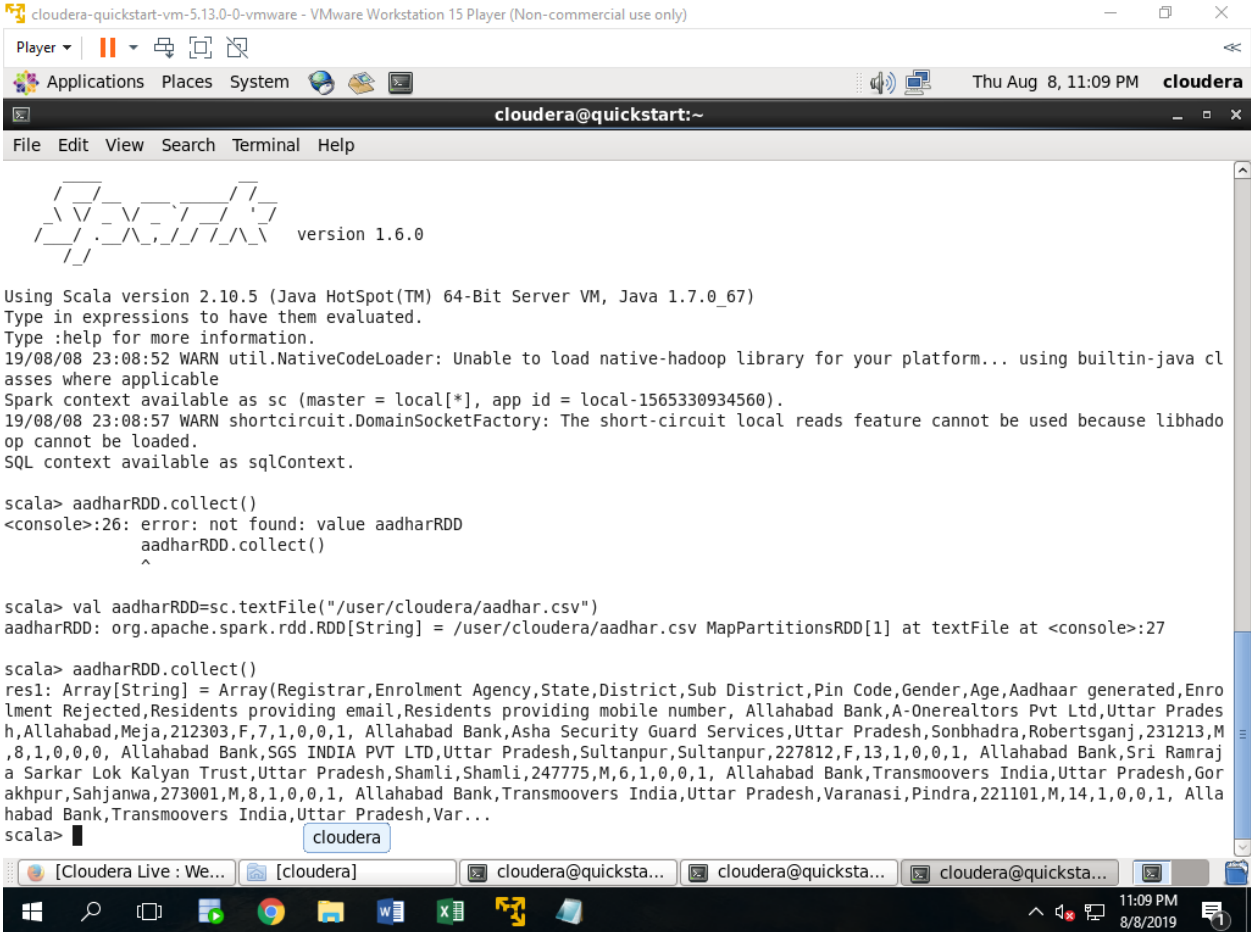
cloudera@quickstart:~

[Cloudera Live... cloudera cloudera@quic... Project cloudera@quic... aadhar.csv (~/...]

10:48 PM 8/8/2019

## Creating Spark DataFrame—

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



- `val remHeader=aadharRDD.first()`
- `val final_data=aadharRDD.filter(row=>row!=(remHeader))`
- `final_data.collect()`
- `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","EnrollmentAgency","State","District","SubDistrict","Pincode`  
`","Gender","Age","AadharGenerated","EnrolmentRejected","ResidentsProvidingEmail","ResidentP`  
`rovidingMobileNumber")`
- `aadharDF.show(25)`

```
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
Email: int, ResidentProvidingMobileNumber: int]

scala> aadharDF.show(25)
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Registrar | EnrollmentAgency | State | District | SubDistrict | Pincode | Gender | Age | AadharGenerated | EnrolmentR |
| ejected | ResidentsProvidingEmail | ResidentProvidingMobileNumber |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Allahabad Bank | A-Onerealtors Pvt... | Uttar Pradesh | Allahabad | Meja | 212303 | F | 7 | 1 |
| 0 | 0 |
| Allahabad Bank | Asha Security Gua... | Uttar Pradesh | Sonbhadra | Robertsganj | 231213 | M | 8 | 1 |
| 0 | 0 |
| Allahabad Bank | SGS INDIA PVT LTD | Uttar Pradesh | Sultanpur | Sultanpur | 227812 | F | 13 | 1 |
| 0 | 0 |
| Allahabad Bank | Sri Ramraja Sarka... | Uttar Pradesh | Shamli | Shamli | 247775 | M | 6 | 1 |
| 0 | 0 |
| Allahabad Bank | Transmoovers India | Uttar Pradesh | Gorakhpur | Sahjanwa | 273001 | M | 8 | 1 |
| 0 | 0 |
| 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 | 0 |
| Allahabad Bank | Transmoovers India | Uttar Pradesh | Varanasi | Varanasi | 221002 | M | 4 | 1 |
| 0 | 0 |
| Allahabad Bank | Transmoovers India | Uttar Pradesh | Varanasi | Varanasi | 221002 | M | 10 | 0 |
| 1 | 0 |
| Allahabad Bank | Transmoovers India | Uttar Pradesh | Varanasi | Varanasi | 221002 | M | 19 | 1 |
| 0 | 0 |
| Allahabad Bank | Vedavaag Systems ... | Uttar Pradesh | Bara Banki | Nawabganj | 225301 | M | 8 | 1 |
| 0 | 0 |
| Atalji Janasnehi ... | Atalji Janasnehi ... | Assam | Mariqaon | Bhuragaon | 782121 | M | 22 | 1 |
| 0 | 0 |

[Cloudera Live: We... [cloudera] cloudera@quicksta... cloudera@quicksta... cloudera@quicksta...
11:45 PM 8/8/2019

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

| Atalji Janasnehi ... | Atalji Janasnehi ... | Bihar | Gopalganj | Vijayeeupur | 841508 | M | 26 | 1 |
| 0 | 0 |
| Atalji Janasnehi ... | Atalji Janasnehi ... | Karnataka | Bagalkot | Badami | 587114 | M | 27 | 1 |
| 0 | 0 |
| Atalji Janasnehi ... | Atalji Janasnehi ... | Karnataka | Bagalkot | Badami | 587155 | F | 2 | 1 |
| 0 | 0 |
| Atalji Janasnehi ... | Atalji Janasnehi ... | Karnataka | Bagalkot | Badami | 587155 | M | 67 | 1 |
| 0 | 0 |
| Atalji Janasnehi ... | Atalji Janasnehi ... | Karnataka | Bagalkot | Badami | 587201 | F | 32 | 1 |
| 0 | 0 |
| Atalji Janasnehi ... | Atalji Janasnehi ... | Karnataka | Bagalkot | Badami | 587203 | M | 27 | 1 |
| 0 | 0 |
| Atalji Janasnehi ... | Atalji Janasnehi ... | Karnataka | Bagalkot | Badami | 587206 | F | 40 | 1 |
| 0 | 0 |
| Atalji Janasnehi ... | Atalji Janasnehi ... | Karnataka | Bagalkot | Badami | 587206 | M | 28 | 1 |
| 0 | 0 |
| Atalji Janasnehi ... | Atalji Janasnehi ... | Karnataka | Bagalkot | Badami | 587206 | M | 44 | 1 |
| 0 | 0 |
| Atalji Janasnehi ... | Atalji Janasnehi ... | Karnataka | Bagalkot | Bagalkot | 587102 | M | 56 | 0 |
| 1 | 0 |
| Atalji Janasnehi ... | Atalji Janasnehi ... | Karnataka | Bagalkot | Bagalkot | 587207 | M | 6 | 1 |
| 0 | 0 |
| Atalji Janasnehi ... | Atalji Janasnehi ... | Karnataka | Bagalkot | Bagalkot | 587207 | M | 73 | 0 |
| 1 | 0 |
| Atalji Janasnehi ... | Atalji Janasnehi ... | Karnataka | Bagalkot | Hungund | 587118 | F | 6 | 1 |
| 0 | 0 |

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 25 rows

scala>
```

2. Describing the schema

- HIVE-describe aadhar;



```

cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 15 Player (Non-commercial use only)
Player | Applications | Places | System | Fri Aug 9, 12:02 AM | cloudera

cloudera@quickstart:~
File Edit View Search Terminal Help
adami 587201 F 32 1 0 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
Atalji Janasnehi Directorate 1 0 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
adami 587203 M 27 1 0 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
Atalji Janasnehi Directorate 1 0 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
adami 587206 F 40 1 0 0 0 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
Atalji Janasnehi Directorate 1 0 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
adami 587206 M 28 1 0 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
Atalji Janasnehi Directorate 1 0 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
adami 587206 M 44 1 0 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
Atalji Janasnehi Directorate 1 0 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
agalkot 587102 M 56 0 1 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
Atalji Janasnehi Directorate 1 0 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
agalkot 587207 M 6 1 0 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
Atalji Janasnehi Directorate 1 0 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
agalkot 587207 M 73 0 1 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot B
Atalji Janasnehi Directorate 1 0 0 1 Atalji Janasnehi Directorate GOK Karnataka Bagalkot H
ungund 587118 F 6 1 0 0 1
Time taken: 0.091 seconds, Fetched: 25 row(s)
hive> describe aadhar;
OK
registrar string
enrolmentagency string
state string
district string
subdistrict string
pincode string
gender string
age int
adhaargenerated int
enrolmentrejected int
residentsprovidingemail int
residentsprovidingmobilenumber int
Time taken: 0.116 seconds, Fetched: 12 row(s)
Current workspace: "Workspace 1"

```

- SPARK-aadharDF.printSchema

```

cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 15 Player (Non-commercial use only)
Player | Applications | Places | System | Fri Aug 9, 12:02 AM | cloudera

cloudera@quickstart:~
File Edit View Search Terminal Help
0| 0| 1|
|Atalji Janasnehi ...|Atalji Janasnehi ...| Karnataka| Bagalkot| Badami| 587206| M| 44| 1|
0| 0| 1|
|Atalji Janasnehi ...|Atalji Janasnehi ...| Karnataka| Bagalkot| Bagalkot| 587102| M| 56| 0|
1| 0| 1|
|Atalji Janasnehi ...|Atalji Janasnehi ...| Karnataka| Bagalkot| Bagalkot| 587207| M| 6| 1|
0| 0| 1|
|Atalji Janasnehi ...|Atalji Janasnehi ...| Karnataka| Bagalkot| Bagalkot| 587207| M| 73| 0|
1| 0| 1|
|Atalji Janasnehi ...|Atalji Janasnehi ...| Karnataka| Bagalkot| Hungund| 587118| F| 6| 1|
0| 0| 1|
+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 25 rows

scala> aadharDF.printSchema
root
|-- Registrar: string (nullable = true)
|-- EnrollmentAgency: 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)
|-- AadhaarGenerated: integer (nullable = false)
|-- EnrolmentRejected: integer (nullable = false)
|-- ResidentsProvidingEmail: integer (nullable = false)
|-- ResidentProvidingMobileNumber: integer (nullable = false)

scala>

```

### 3. Counting the no and name of registrar

- select registrar,count(registrar) from aadhar group by registrar;

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
Women and Child Development Govt. of Jharkhand	39

Time taken: 30.29 seconds, Fetched: 60 row(s)

#### 4. Find the number of states, districts in each state and sub-districts in each district.

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



Andhra Pradesh	13
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

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

Adilabad	41
Agra	6
Ahmadnagar	14
Ahmedabad	9

Aizawl	5
Ajmer	8
Akola	7
Alappuzha	6
Aligarh	5
Alirajpur	2
Allahabad	8
Almora	7
Alwar	12
Ambala	3
Ambedkar Nagar	4
.....	

Time taken: 12.296 seconds, Fetched: 664 row(s)

## 6. Find out the names of private agencies for each state.

- select distinct enrolmentagency,state from aadhar;

SREI INFRASTRUCTURE FINANCES L	Karnataka
SREI INFRASTRUCTURE FINANCES L	Madhya Pradesh
SREI INFRASTRUCTURE FINANCES L	Maharashtra
SREI INFRASTRUCTURE FINANCES L	Meghalaya
SREI INFRASTRUCTURE FINANCES L	Odisha
SREI INFRASTRUCTURE FINANCES L	Punjab
SREI INFRASTRUCTURE FINANCES L	Rajasthan
SREI INFRASTRUCTURE FINANCES L	Tamil Nadu
SREI INFRASTRUCTURE FINANCES L	Telangana
SREI INFRASTRUCTURE FINANCES L	Uttar Pradesh
SREI INFRASTRUCTURE FINANCES L	Uttarakhand
SREI INFRASTRUCTURE FINANCES L	West Bengal
SRM Education And Social Welfare Society	Andhra Pradesh
SRM Education And Social Welfare Society	Arunachal Pradesh
SRM Education And Social Welfare Society	Assam
SRM Education And Social Welfare Society	Bihar
SRM Education And Social Welfare Society	Chhattisgarh
SRM Education And Social Welfare Society	Dadra and Nagar Haveli
SRM Education And Social Welfare Society	Daman and Diu
SRM Education And Social Welfare Society	Delhi
SRM Education And Social Welfare Society	Gujarat
SRM Education And Social Welfare Society	Haryana
SRM Education And Social Welfare Society	Himachal Pradesh

SRM Education And Social Welfare Society	Jammu and Kashmir
SRM Education And Social Welfare Society	Jharkhand
SRM Education And Social Welfare Society	Karnataka
SRM Education And Social Welfare Society	Kerala
SRM Education And Social Welfare Society	Madhya Pradesh
SRM Education And Social Welfare Society	Maharashtra
SRM Education And Social Welfare Society	Manipur
SRM Education And Social Welfare Society	Meghalaya
SRM Education And Social Welfare Society	Nagaland
SRM Education And Social Welfare Society	Odisha
SRM Education And Social Welfare Society	Punjab
SRM Education And Social Welfare Society	Rajasthan
SRM Education And Social Welfare Society	Tamil Nadu
SRM Education And Social Welfare Society	Telangana
SRM Education And Social Welfare Society	Uttar Pradesh
SRM Education And Social Welfare Society	Uttarakhand
SRM Education And Social Welfare Society	West Bengal
SRM Techsol Pvt. Ltd.	Uttar Pradesh
SRR Infotech	Assam
SRR Infotech	Bihar
SRR Infotech	Delhi
SRR Infotech	Haryana
SRR Infotech	Jammu and Kashmir
SRR Infotech	Jharkhand
SRR Infotech	Madhya Pradesh
SRR Infotech	Maharashtra
SRR Infotech	Manipur
SRR Infotech	Nagaland
SRR Infotech	Odisha
SRR Infotech	Punjab
SRR Infotech	Rajasthan
SRR Infotech	Tamil Nadu
SRR Infotech	Uttar Pradesh
SRR Infotech	Uttarakhand
SRR Infotech	West Bengal
SVG Express Services Pvt Ltd	Assam
SVG Express Services Pvt Ltd	Bihar
SVG Express Services Pvt Ltd	Madhya Pradesh
SVG Express Services Pvt Ltd	Rajasthan
SVG Express Services Pvt Ltd	Uttar Pradesh
SVG Express Services Pvt Ltd	Uttarakhand
SVG Express Services Pvt Ltd	West Bengal
Saket Advertising Pvt. Ltd	Bihar
Saket Advertising Pvt. Ltd	Delhi
Saket Advertising Pvt. Ltd	Jharkhand
Saket Advertising Pvt. Ltd	Madhya Pradesh
Saket Advertising Pvt. Ltd	Maharashtra
Saket Advertising Pvt. Ltd	Uttar Pradesh

.....

Time taken: 31.859 seconds, Fetched: 2271 row(s)

## 8. Find top 3 states generating most number of Aadhaar cards?

- select count(\*) as no\_of\_reg,state from aadhar group by state order by no\_of\_reg desc limit 3;

81776 Bihar

69476 Uttar Pradesh

60485 West Bengal

Time taken: 53.541 seconds, Fetched: 3 row(s)

## 9. Find top 3 private agencies generating the most number of Aadhaar cards?

- select count(\*) as no\_of\_reg,enrolmentagency from aadhar group by enrolmentagency order by no\_of\_reg desc limit 3;

100357 CSC SPV

18101 SRM Education And Social Welfare Society

16972 SREI INFRASTRUCTURE FINANCES L

Time taken: 57.888 seconds, Fetched: 3 row(s)

## 10. Find the number of residents providing email, mobile number? (Hint: consider non-zero values.)

- select count(\*) from aadhar where residentsprovidingemail>0 and residentsprovidingmobilenumber>0;

16951

Time taken: 26.083 seconds, Fetched: 1 row(s)

## 11. Find top 3 districts where enrolment numbers are maximum?

- select count(\*) as no\_of\_reg,district from aadhar group by district order by no\_of\_reg desc limit 3;

7135 Bardhaman

6894 North 24 Parganas  
6078 South 24 Parganas  
Time taken: 53.51 seconds, Fetched: 3 row(s)

## 12. Find the no. of Aadhaar cards generated in each state?

- `select count(*),state from aadhar group by state;`

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: 27.521 seconds, Fetched: 37 row(s)

### 13. Create a data frame using the file and provide its summary.

- `val aadharRDD=sc.textFile("/user/cloudera/aadhar.csv")`
- `aadharRDD.collect()`
- `val remHeader=aadharRDD.first()`
- `val final_data=aadharRDD.filter(row=>row!=(remHeader))`
- `final_data.collect()`
- `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","EnrollmentAgency","State","District","SubDistrict","Pincode`  
`", "Gender", "Age", "AadharGenerated", "EnrolmentRejected", "ResidentsProvidingEmail", "ResidentP`  
`rovidingMobileNumber")`
- `aadharDF.describe()`

### 14. Write a command to see the correlation between “age” and “mobile\_number”? (Hint: Consider the percentage of people who have provided the mobile number out of the total applicants)

- `select corr(age,residentsprovidingmobilenumber) from aadhar;`

-0.11754461896889339

Time taken: 32.597 seconds, Fetched: 1 row(s)

### 15. Find the number of unique pincodes in the data?

- select distinct pincode from aadhar;

854109

854112

854113

854114

854115

854116

854117

854201

854202

854203

854204

854205

854206

854301

854302

854303

854304

854305

854311

854312

854315

854316

854317

854318

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: 27.362 seconds, Fetched: 17756 row(s)

16. Find the number of Aadhaar registrations rejected in Uttar Pradesh and Maharashtra?

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

1581    Maharashtra

3892    Uttar Pradesh

Time taken: 25.531 seconds, Fetched: 2 row(s)

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

- select  
state,(sum(adhaargenerated)\*100)/(sum(adhaargenerated+  
enrolmentrejected))as male\_numbers from aadhar where  
gender="M" group by state order by male\_numbers desc  
limit 3;

Andaman and Nicobar Islands 100.0

Others 100.0

Lakshadweep 100.0

Time taken: 25.531 seconds, Fetched: 3 row(s)

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

- select  
district,(sum(adhaargenerated)\*100)/(sum(adhaargene  
rated+enrolmentrejected)) as female\_numbers from  
aadhar where gender="F" and state in("Andaman and  
nicobar islands","Others","Lakshadweep") group by  
district order by female\_numbers desc limit 3;

Lakshadweep 100.0

South Andaman 50.0

North And Andaman 33.333333333333

Time taken: 53.283 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(adhaargenerated)\*100)/(sum(adhaargenerated+  
enrolmentrejected))as male\_numbers from aadhar where  
gender="F" group by state order by male\_numbers desc  
limit 3;

Dadra and Nagar Haveli 100.0

Sikkim 100.0

Others 100.0

Time taken: 57.098 seconds, Fetched: 3 row(s)

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  
district,(sum(adhaargenerated)\*100)/(sum(adhaargene  
rated+enrolmentrejected)) as male\_numbers from  
aadhar where gender="M" and state in("Dadra and  
Nagar Haveli","Others","Sikkim") group by district  
order by male\_numbers desc limit 3;

West Sikkim 9.0909

South Sikkim 3.4482

Others 0.0

Time taken: 54.177 seconds, Fetched: 3 row(s)

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 aadhar\_bucket(registrar string, enrolmentagency string,state string,district string,subdistrict string,pincode string,gender string,age int,aadhargenerated int,enrolmentrejected int,residentsprividngemail int,residentsprovidingmobilenumber int)
- > clustered by(age) into 10 buckets
- > row format delimited fields terminated by ",";

OK

Time taken: 0.101 seconds

//inserting data from staging table

insert into aadhar\_bucket select \* from aadhar;

select(sum(aadhargenerated)/(sum(aadhargeberated+enrolmentrejected)))\*100 as acceptancePercentage from aadhar\_bucket;

94.81864032289477

Time taken: 29.754 seconds, Fetched: 1 row(s)



