```scala
import org.apache.spark.sql._
import org.apache.spark.sql.SQLContext._
import org.apache.spark.sql.hive.HiveContext
import org.apache.spark.sql.hive.HiveContext


import org.apache.spark.sql._
import org.apache.spark.sql.SQLContext._
import org.apache.spark.sql.hive.HiveContext
import org.apache.spark.sql.hive.HiveContext
```

```scala
val uidEnrolmentDF = sqlContext.read.format("com.databricks.spark.csv").option("header", "true").option("inferSchema", "true").load("/FileStore/tables
/0pze7syp1500132853172/UIDAI_ENR_DETAIL_20170711-c2418")
```

uidEnrolmentDF: org.apache.spark.sql.DataFrame = [Registrar: string, Enrolment Agency: string ... 10 more fields]

```scala
uidEnrolmentDF.createOrReplaceTempView("uid_enrolments_detail")
```

```scala
uidEnrolmentDF.printSchema
```

```
root
 |-- Registrar: string (nullable = true)
 |-- Enrolment Agency: string (nullable = true)
 |-- State: string (nullable = true)
 |-- District: string (nullable = true)
 |-- Sub District: string (nullable = true)
 |-- Pin Code: string (nullable = true)
 |-- Gender: string (nullable = true)
 |-- Age: integer (nullable = true)
 |-- Aadhaar generated: integer (nullable = true)
 |-- Enrolment Rejected: integer (nullable = true)
 |-- Residents providing email: integer (nullable = true)
 |-- Residents providing mobile number: integer (nullable = true)
```

```scala
val stateWiseCountDF = sqlContext.sql("""
                                       | SELECT State,
                                       |        SUM(`Aadhaar generated`) as count
                                       | FROM uid_enrolments_detail
                                       | GROUP BY state
                                       | ORDER BY count DESC""")
```

stateWiseCountDF: org.apache.spark.sql.DataFrame = [State: string, count: bigint]

```scala
stateWiseCountDF.show()
```

```
+-----------------+------+
|            State| count|
+-----------------+------+
|    Uttar Pradesh|182712|
|            Bihar| 75723|
|   Madhya Pradesh| 39612|
|        Rajasthan| 36415|
|          Gujarat| 33542|
|      West Bengal| 32455|
|        Karnataka| 26978|
|           Odisha| 24514|
|      Maharashtra| 22678|
|       Tamil Nadu| 21278|
|           Kerala| 10727|
|        Jharkhand|  9514|
|   Andhra Pradesh|  7647|
|        Telangana|  6551|
|Jammu and Kashmir|  4820|
|          Haryana|  4585|
|            Delhi|  4041|
|       Uttarakhand|  2569|
```

```scala
stateWiseCountDF.write.mode("overwrite").saveAsTable("uid_state_wise_count")
```

```scala
val maxEnrolmentAgencyDF = sqlContext.sql("""
                                          | SELECT `Enrolment Agency` as Enrolment_Agency,
                                          |        count(`Aadhaar generated`) as count
                                          | FROM uid_enrolments_detail
                                          | GROUP BY `Enrolment Agency`
                                          | ORDER BY count DESC""".stripMargin)
```

maxEnrolmentAgencyDF: org.apache.spark.sql.DataFrame = [Enrolment_Agency: string, count: bigint]

```scala
maxEnrolmentAgencyDF.show()
```

```
+--------------------+------+
|    Enrolment_Agency| count|
+--------------------+------+
```

```
|           CSC SPV|117009|
|SREI INFRASTRUCTU...| 20452|
|Rajcomp Info Serv...| 12473|
|Karvy Data Manage...| 10474|
|    MPOnline Limited| 10278|
|National Cooperat...|  9586|
|            Akshaya|  9553|
|Vakrangee Softwar...|  7916|
|SRM Education And...|  7423|
|TAMILNADU ARASU C...|  7050|
|Twinstar Industri...|  6478|
|     AISECT Limited|  6183|
|Gujarat Infotech ...|  5608|
|RELIGARE SECURITI...|  5298|
|Abha Systems And ...|  5158|
|      Computer LAB|  4912|
|Electronics Corpo...|  4720|
|  Directorate of ESD|  4492|
```

```
maxEnrolmentAgencyDF.write.mode("overwrite").saveAsTable("uid_agency_wise_count")
```

```scala
val districtWiseGenderCountDF = sqlContext.sql("""
                         | SELECT District,
                         |      count(CASE WHEN Gender='M' THEN 1 END) as male_count,
                         |      count(CASE WHEN Gender='F' THEN 1 END) as FEMALE_count
                         | FROM uid_enrolments_detail
                         | GROUP BY District
                         | ORDER BY male_count DESC, FEMALE_count DESC
                         | LIMIT 10""".stripMargin)
```

districtWiseGenderCountDF: org.apache.spark.sql.DataFrame = [District: string, male_count: bigint ... 1 more field]

```
districtWiseGenderCountDF.write.mode("overwrite").saveAsTable("uid_district_wise_gndr_count")
```

```
districtWiseGenderCountDF.show()
```

```
+-------------------+----------+------------+
|           District|male_count|FEMALE_count|
+-------------------+----------+------------+
|          Bengaluru|      3016|        2795|
|          Allahabad|      1680|        2348|
|            Jaunpur|      1612|        2121|
|           Ghazipur|      1602|        2026|
|              Patna|      1528|        2328|
|   North 24 Parganas|     1470|        1757|
|Thiruvananthapuram |      1357|         220|
|          Ahmedabad|      1340|        2049|
|               Gaya|      1315|        1776|
|         Muzaffarpur|     1288|        1756|
+-------------------+----------+------------+
```

```sql
%sql select * from uid_district_wise_gndr_count
```

District	male_count	FEMALE_count
Bengaluru	3016	2795
Allahabad	1680	2348
Jaunpur	1612	2121
Ghazipur	1602	2026
Patna	1528	2328
North 24 Parganas	1470	1757
Thiruvananthapuram	1357	220
Ahmedabad	1340	2049
Gaya	1315	1776