```
%pyspark
sc
```
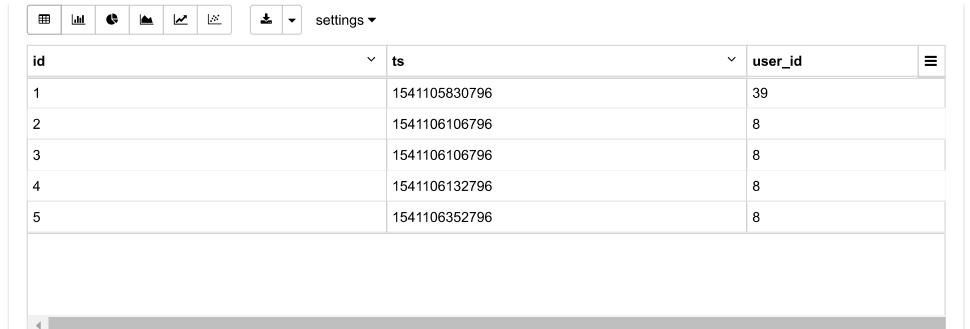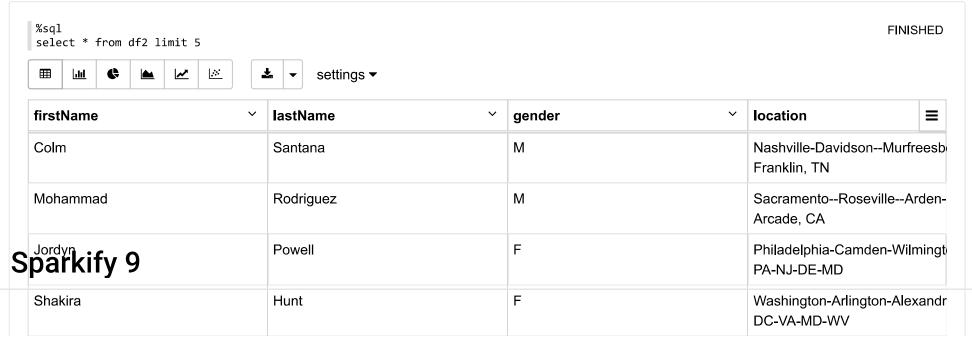
FINISHED

```
<SparkContext master=yarn appName=Zeppelin>
```

Took 0 sec. Last updated by anonymous at May 02 2020, 11:27:50 PM.

```
%pyspark
%sql
CREATE TEMPORARY VIEW df1
USING org.apache.spark.sql.parquet
OPTIONS (
    path "s3a://snell2-s8/output/df1.parquet/part-00000-cde353b1-0575-45dc-9cb5-60d010af3885-c000.snappy.parquet"
)
```

READY

```
%sql
CREATE TEMPORARY VIEW df2
USING org.apache.spark.sql.parquet
OPTIONS (
    path "s3a://snell2-s8/output/df2.parquet/*.parquet"
)
```

FINISHED

Took 3 sec. Last updated by anonymous at May 02 2020, 5:28:46 PM.

```
%sql
CREATE TEMPORARY VIEW df3
USING org.apache.spark.sql.parquet
OPTIONS (
    path "s3a://snell2-s8/output/df3.parquet/part-00000-5673115d-65e0-4170-8012-fb931d6c56a9-c000.snappy.parquet"
)
```

FINISHED

Took 0 sec. Last updated by anonymous at May 02 2020, 5:28:50 PM.

Sparkify 9

```
%sql
select * from df1 limit 5
```

FINISHED

id	ts	user_id	≡
1	1541105830796	39	
2	1541106106796	8	
3	1541106106796	8	
4	1541106132796	8	
5	1541106352796	8	

Took 1 sec. Last updated by anonymous at May 02 2020, 11:22:16 PM.

```
%sql
select * from df2 limit 5
```
FINISHED

firstName	lastName	gender	location	≡
Colm	Santana	M	Nashville-Davidson--Murfreesb Franklin, TN
Mohammad	Rodriguez	M	Sacramento--Roseville--Arden-Arcade, CA
Jordyn	Powell	F	Philadelphia-Camden-Wilmingt PA-NJ-DE-MD
Shakira	Hunt	F	Washington-Arlington-Alexandr DC-VA-MD-WV

Sparkify 9

Sylvie	Cruz	F	Washington-Arlington-Alexandr DC-VA-MD-WV

Took 1 sec. Last updated by anonymous at May 02 2020, 11:22:18 PM.

```
%sql
select * from df3 limit 5
```

FINISHED

id	⌄	artist	≡
1		null	
2		null	
3		Des'ree	
4		null	
5		Mr Oizo	

Took 10 sec. Last updated by anonymous at May 02 2020, 11:42:36 PM.

```
%sql
select artist,Count(artist) AS Count from df3 GROUP BY artist ORDER BY count desc LIMIT 10
```

FINISHED

settings ▾

artist	⌄	Count	≡

Coldplay	58
Kings Of Leon	55
Dwight Yoakam	38
The Black Keys	36
Muse	35
Jack Johnson	35
Florence + The Machine	35

Took 2 sec. Last updated by anonymous at May 02 2020, 11:24:03 PM.

```
%sql                                                                                    FINISHED
select FIRST(artist) AS ArtistName, FROM_UNIXTIME(FIRST(ts)/1000.0) AS Timestamp, FIRST(firstName) AS FirstName , FIRST(lastName) AS LastN
from df1 INNER JOIN df2 ON df1.user_id=df2.userid
INNER JOIN df3 ON df1.id=df3.id
GROUP BY artist ORDER BY COUNT(artist) desc LIMIT 10
```

settings ▾

ArtistName ⌄	Timestamp ⌄	FirstName ☰
Coldplay	2018-11-03 15:59:38	Sara
Kings Of Leon	2018-11-02 11:56:43	Lily
Dwight Yoakam	2018-11-02 16:35:00	Ava
The Black Keys	2018-11-02 10:24:12	Lily
Jack Johnson	2018-11-03 17:28:55	Sara
Muse	2018-11-03 21:21:18	Chloe
Florence + The Machine	2018-11-03 16:10:32	Sara
BiÃ□Â¶rk	2018-11-02 13:44:33	Lily

Sparkify 9

Took 20 sec. Last updated by anonymous at May 02 2020, 11:40:01 PM. (outdated)

```
%sql                                                                    READY
```