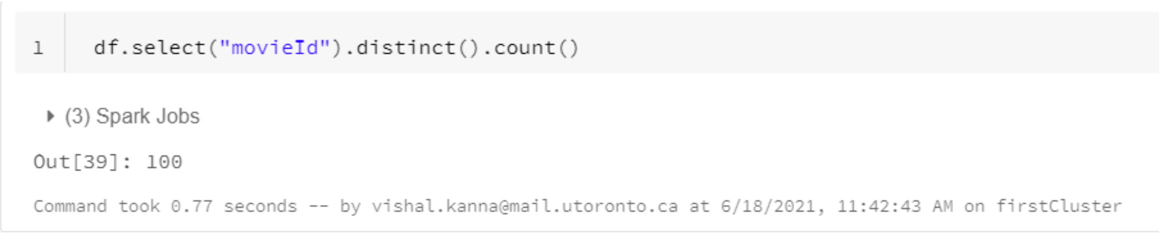
Part b

Question 5

Recommendations-

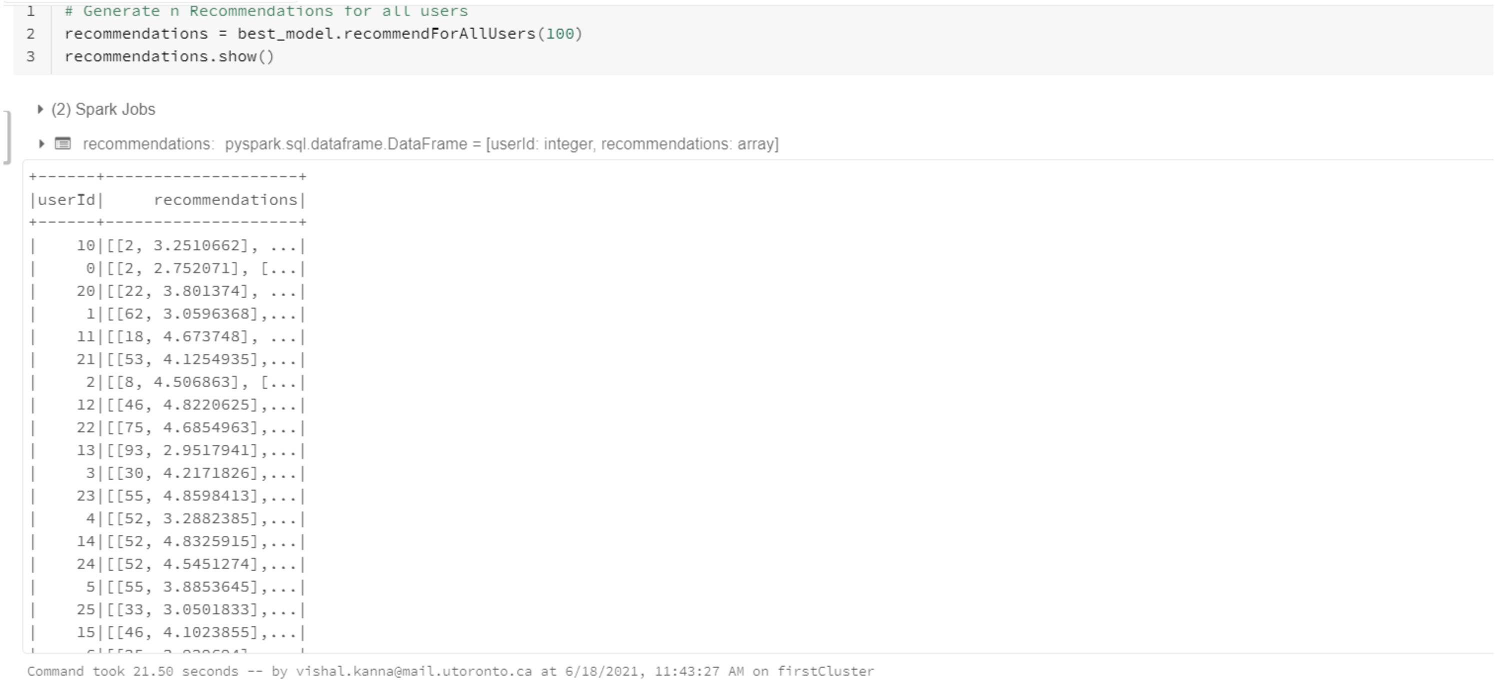


```
df.select("movieId").distinct().count()
```

(3) Spark Jobs

```
Out[39]: 100
```

Command took 0.77 seconds -- by vishal.kanna@mail.utoronto.ca at 6/18/2021, 11:42:43 AM on firstCluster



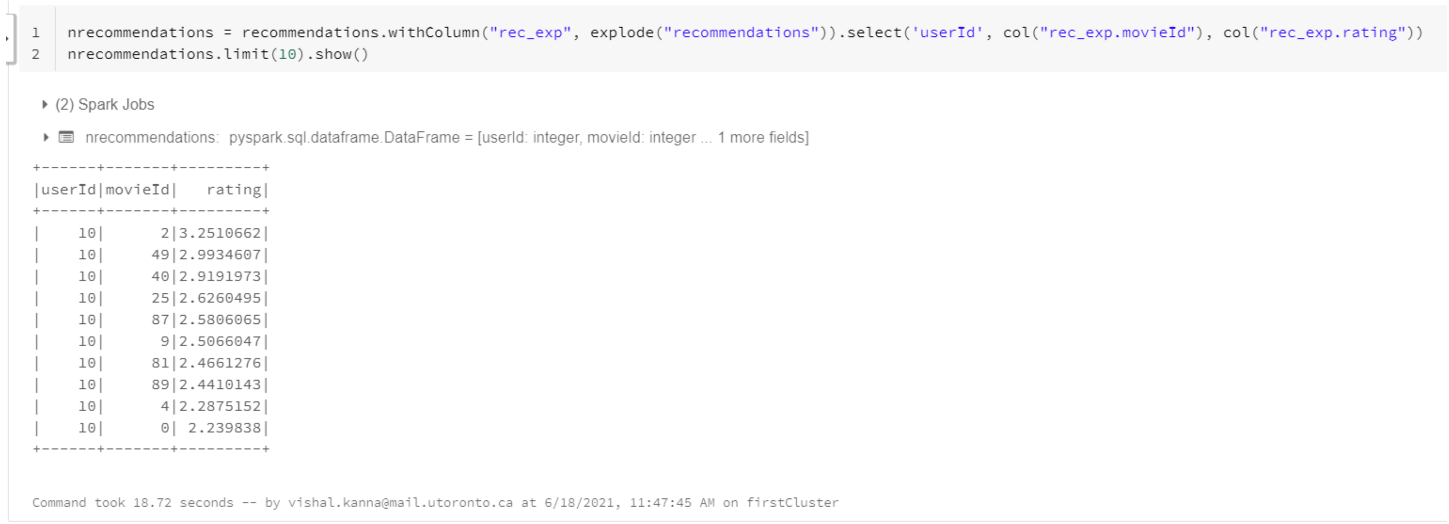
```
# Generate n Recommendations for all users
recommendations = best_model.recommendForAllUsers(100)
recommendations.show()
```

(2) Spark Jobs

recommendations: pyspark.sql.dataframe.DataFrame = [userId: integer, recommendations: array]

```
+------+--------------------+
|userId|     recommendations|
+------+--------------------+
|    10|[[2, 3.2510662], ...|
|     0|[[2, 2.752071], [...|
|    20|[[22, 3.801374], ...|
|     1|[[62, 3.0596368],...|
|    11|[[18, 4.673748], ...|
|    21|[[53, 4.1254935],...|
|     2|[[8, 4.506863], [...|
|    12|[[46, 4.8220625],...|
|    22|[[75, 4.6854963],...|
|    13|[[93, 2.9517941],...|
|     3|[[30, 4.2171826],...|
|    23|[[55, 4.8598413],...|
|     4|[[52, 3.2882385],...|
|    14|[[52, 4.8325915],...|
|    24|[[52, 4.5451274],...|
|     5|[[55, 3.8853645],...|
|    25|[[33, 3.0501833],...|
|    15|[[46, 4.1023855],...|
```

Command took 21.50 seconds -- by vishal.kanna@mail.utoronto.ca at 6/18/2021, 11:43:27 AM on firstCluster



```
nrecommendations = recommendations.withColumn("rec_exp", explode("recommendations")).select('userId', col("rec_exp.movieId"), col("rec_exp.rating"))
nrecommendations.limit(10).show()
```

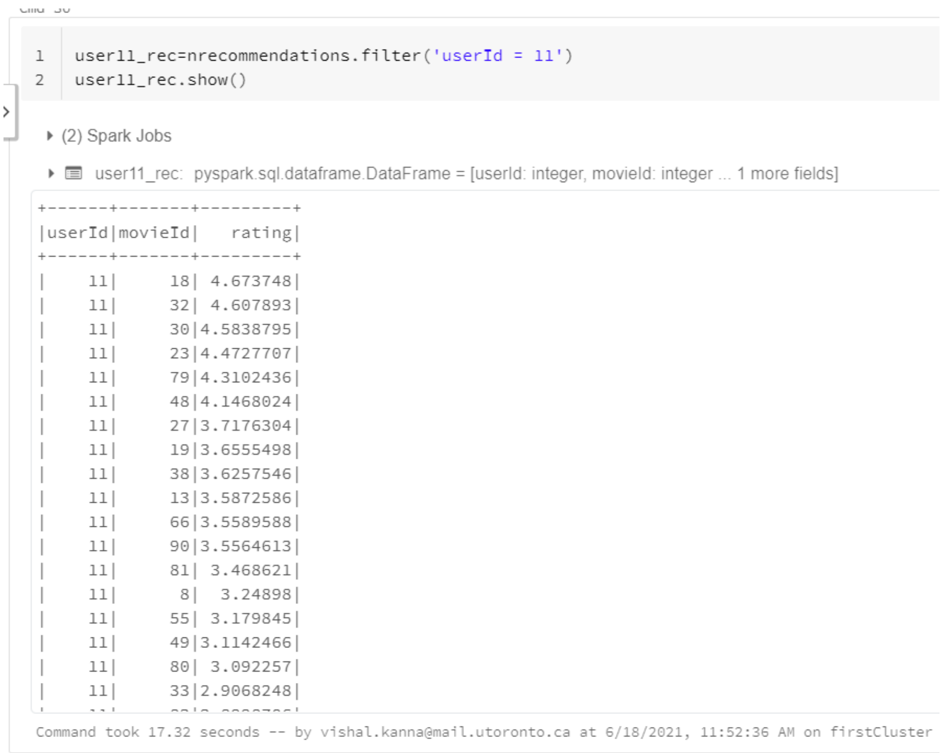
(2) Spark Jobs

nrecommendations: pyspark.sql.dataframe.DataFrame = [userId: integer, movieId: integer ... 1 more fields]

```
+------+-------+---------+
|userId|movieId|   rating|
+------+-------+---------+
|    10|      2|3.2510662|
|    10|     49|2.9934607|
|    10|     40|2.9191973|
|    10|     25|2.6260495|
|    10|     87|2.5806065|
|    10|      9|2.5066047|
|    10|     81|2.4661276|
|    10|     89|2.4410143|
|    10|      4|2.2875152|
|    10|      0| 2.239838|
+------+-------+---------+
```

Command took 18.72 seconds -- by vishal.kanna@mail.utoronto.ca at 6/18/2021, 11:47:45 AM on firstCluster

Top 15 movies recommendations for user id 11-



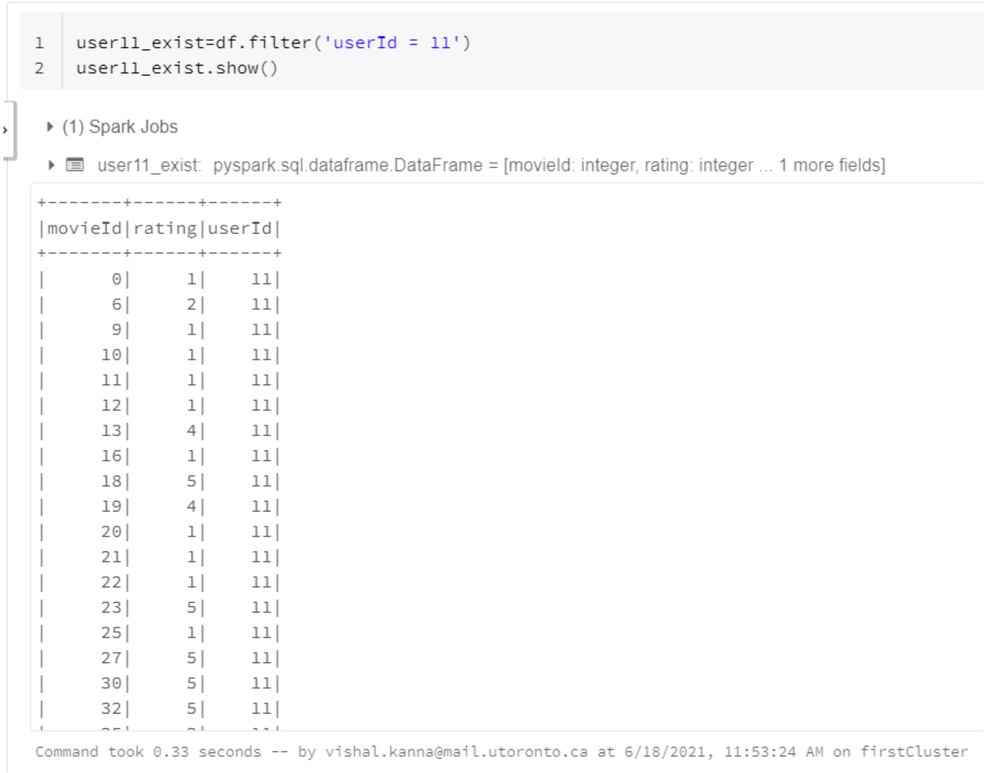
```
user11_rec=nrecommendations.filter('userId = 11')
user11_rec.show()
```

▸ (2) Spark Jobs

▸ user11_rec: pyspark.sql.dataframe.DataFrame = [userId: integer, movieId: integer ... 1 more fields]

```
+------+-------+---------+
|userId|movieId|   rating|
+------+-------+---------+
|    11|     18| 4.673748|
|    11|     32| 4.607893|
|    11|     30|4.5838795|
|    11|     23|4.4727707|
|    11|     79|4.3102436|
|    11|     48|4.1468024|
|    11|     27|3.7176304|
|    11|     19|3.6555498|
|    11|     38|3.6257546|
|    11|     13|3.5872586|
|    11|     66|3.5589588|
|    11|     90|3.5564613|
|    11|     81| 3.468621|
|    11|      8|  3.24898|
|    11|     55| 3.179845|
|    11|     49|3.1142466|
|    11|     80| 3.092257|
|    11|     33|2.9068248|
```

Command took 17.32 seconds -- by vishal.kanna@mail.utoronto.ca at 6/18/2021, 11:52:36 AM on firstCluster



```
user11_exist=df.filter('userId = 11')
user11_exist.show()
```

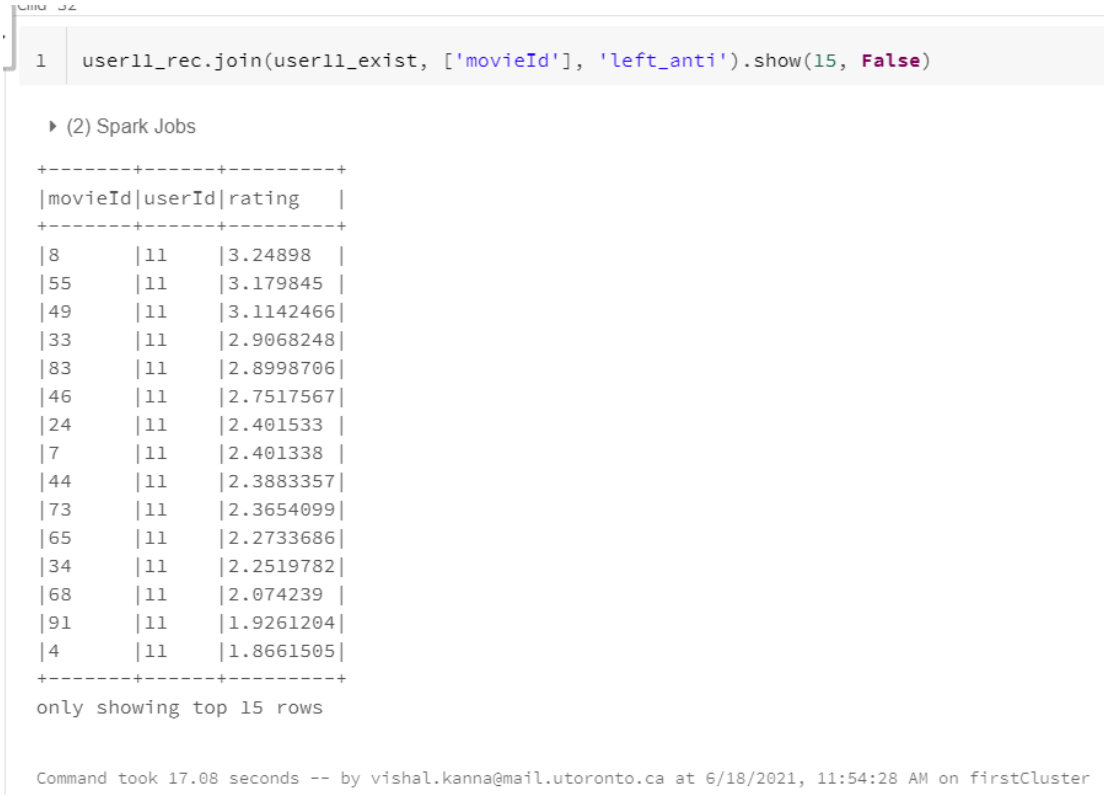
▸ (1) Spark Jobs

▸ user11_exist: pyspark.sql.dataframe.DataFrame = [movieId: integer, rating: integer ... 1 more fields]

```
+-------+------+------+
|movieId|rating|userId|
+-------+------+------+
|      0|     1|    11|
|      6|     2|    11|
|      9|     1|    11|
|     10|     1|    11|
|     11|     1|    11|
|     12|     1|    11|
|     13|     4|    11|
|     16|     1|    11|
|     18|     5|    11|
|     19|     4|    11|
|     20|     1|    11|
|     21|     1|    11|
|     22|     1|    11|
|     23|     5|    11|
|     25|     1|    11|
|     27|     5|    11|
|     30|     5|    11|
|     32|     5|    11|
```

Command took 0.33 seconds -- by vishal.kanna@mail.utoronto.ca at 6/18/2021, 11:53:24 AM on firstCluster

Final recommendations-

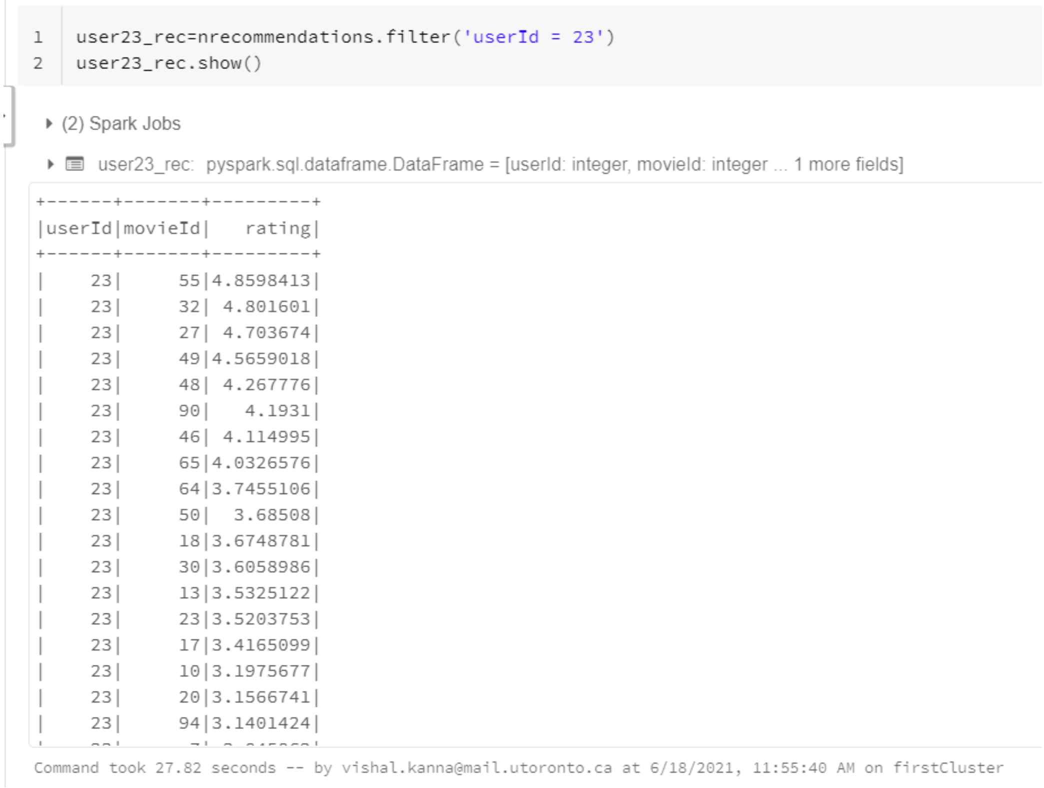


```
user11_rec.join(user11_exist, ['movieId'], 'left_anti').show(15, False)
```

```
+-------+------+---------+
|movieId|userId|rating   |
+-------+------+---------+
|8      |11    |3.24898  |
|55     |11    |3.179845 |
|49     |11    |3.1142466|
|33     |11    |2.9068248|
|83     |11    |2.8998706|
|46     |11    |2.7517567|
|24     |11    |2.401533 |
|7      |11    |2.401338 |
|44     |11    |2.3883357|
|73     |11    |2.3654099|
|65     |11    |2.2733686|
|34     |11    |2.2519782|
|68     |11    |2.074239 |
|91     |11    |1.9261204|
|4      |11    |1.8661505|
+-------+------+---------+
only showing top 15 rows
```

Command took 17.08 seconds -- by vishal.kanna@mail.utoronto.ca at 6/18/2021, 11:54:28 AM on firstCluster

Top 15 movies recommendations for user id 23-

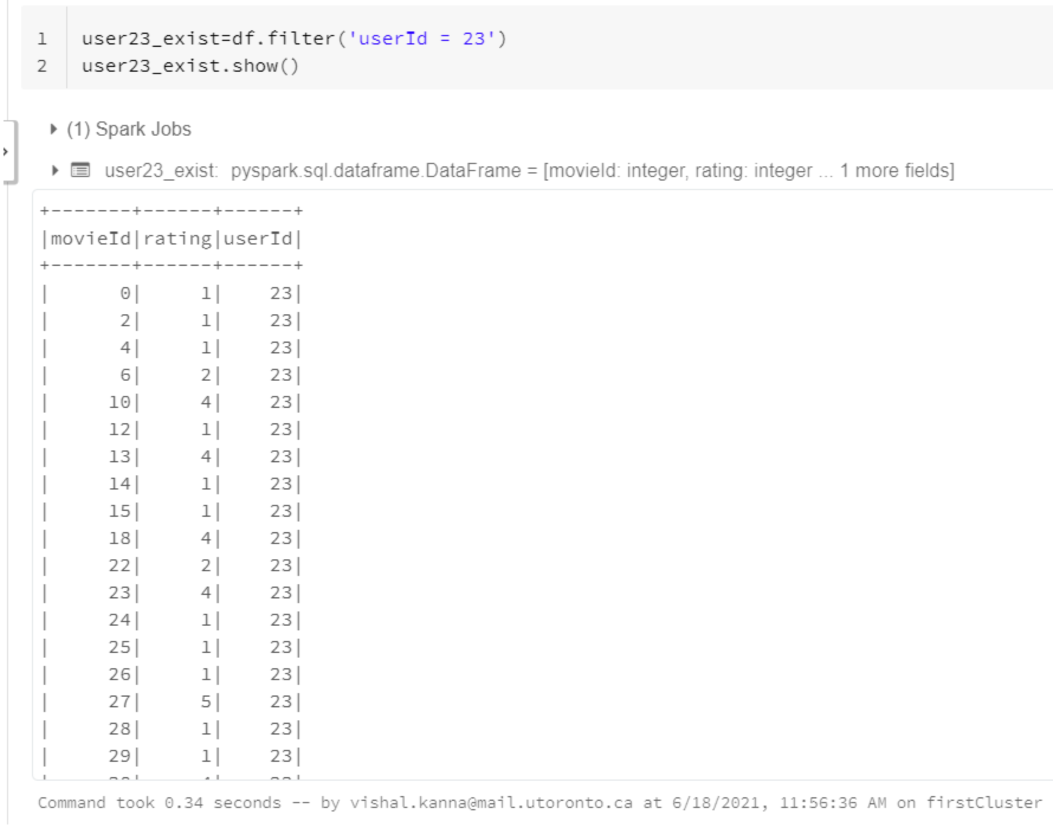


```
user23_rec=nrecommendations.filter('userId = 23')
user23_rec.show()
```

user23_rec: pyspark.sql.dataframe.DataFrame = [userId: integer, movieId: integer ... 1 more fields]

```
+------+-------+---------+
|userId|movieId|   rating|
+------+-------+---------+
|    23|     55|4.8598413|
|    23|     32| 4.801601|
|    23|     27| 4.703674|
|    23|     49|4.5659018|
|    23|     48| 4.267776|
|    23|     90|   4.1931|
|    23|     46| 4.114995|
|    23|     65|4.0326576|
|    23|     64|3.7455106|
|    23|     50|  3.68508|
|    23|     18|3.6748781|
|    23|     30|3.6058986|
|    23|     13|3.5325122|
|    23|     23|3.5203753|
|    23|     17|3.4165099|
|    23|     10|3.1975677|
|    23|     20|3.1566741|
|    23|     94|3.1401424|
```

Command took 27.82 seconds -- by vishal.kanna@mail.utoronto.ca at 6/18/2021, 11:55:40 AM on firstCluster



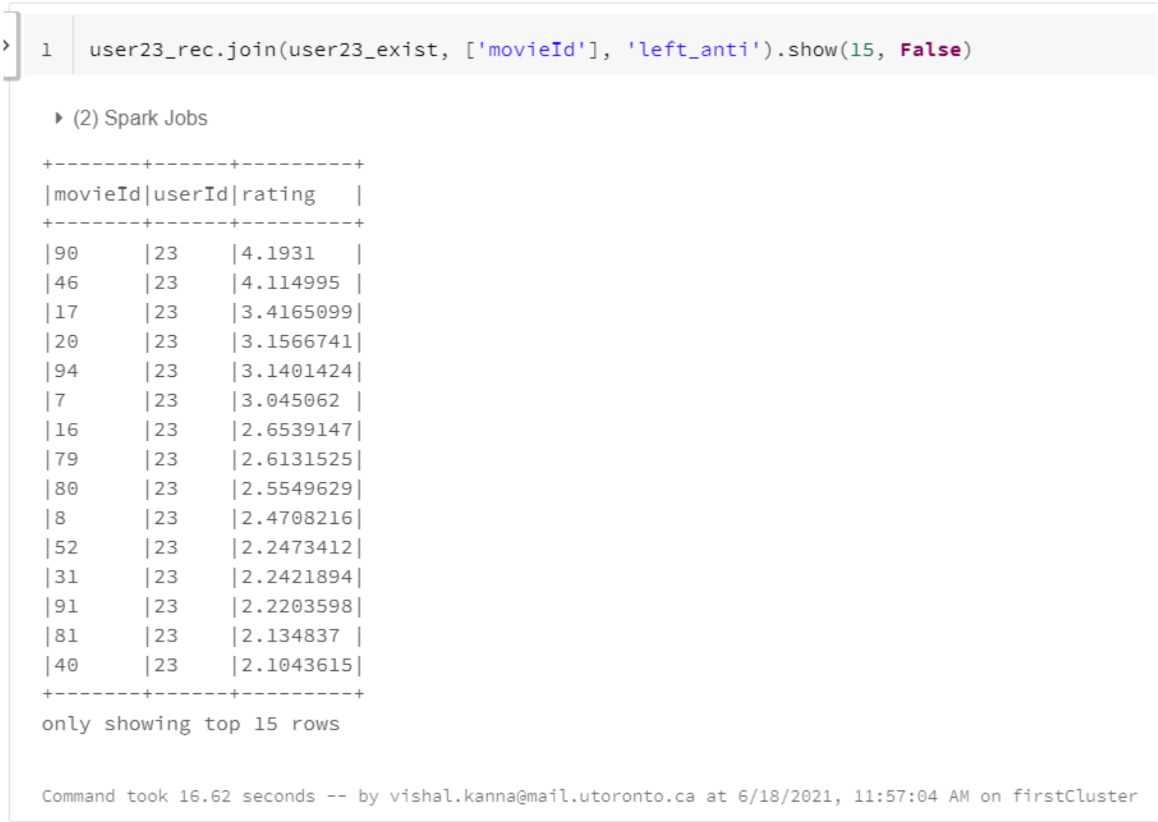
```
user23_exist=df.filter('userId = 23')
user23_exist.show()
```

user23_exist: pyspark.sql.dataframe.DataFrame = [movieId: integer, rating: integer ... 1 more fields]

```
+-------+------+------+
|movieId|rating|userId|
+-------+------+------+
|      0|     1|    23|
|      2|     1|    23|
|      4|     1|    23|
|      6|     2|    23|
|     10|     4|    23|
|     12|     1|    23|
|     13|     4|    23|
|     14|     1|    23|
|     15|     1|    23|
|     18|     4|    23|
|     22|     2|    23|
|     23|     4|    23|
|     24|     1|    23|
|     25|     1|    23|
|     26|     1|    23|
|     27|     5|    23|
|     28|     1|    23|
|     29|     1|    23|
```

Command took 0.34 seconds -- by vishal.kanna@mail.utoronto.ca at 6/18/2021, 11:56:36 AM on firstCluster

Final recommendations-



```
user23_rec.join(user23_exist, ['movieId'], 'left_anti').show(15, False)
```

```
+-------+------+---------+
|movieId|userId|rating   |
+-------+------+---------+
|90     |23    |4.1931   |
|46     |23    |4.114995 |
|17     |23    |3.4165099|
|20     |23    |3.1566741|
|94     |23    |3.1401424|
|7      |23    |3.045062 |
|16     |23    |2.6539147|
|79     |23    |2.6131525|
|80     |23    |2.5549629|
|8      |23    |2.4708216|
|52     |23    |2.2473412|
|31     |23    |2.2421894|
|91     |23    |2.2203598|
|81     |23    |2.134837 |
|40     |23    |2.1043615|
+-------+------+---------+
only showing top 15 rows
```

Command took 16.62 seconds -- by vishal.kanna@mail.utoronto.ca at 6/18/2021, 11:57:04 AM on firstCluster