

```
In [58]: import pyspark
from pyspark.sql import SparkSession
from pyspark.sql.types import StructType, StructField, StringType, IntegerType
from pyspark.sql.types import ArrayType, DoubleType, BooleanType
from pyspark.sql.functions import col, array_contains
```

```
spark = SparkSession.builder.appName('SparkByExamples.com').getOrCreate()
```

```
df = spark.read.csv(r"C:\Users\SaikumarGanji\Downloads\complaints.csv")
```

```
df.show()
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
-+
|_c0|_c1|_c2|_c3|_c4|_c5|_c6|_c7|_c8|_c9|_c10|_c11|_c12|_c13|
+-----+-----+-----+-----+-----+-----+-----+-----+
-+
|Complaint ID|Product|Sub-product|Issue|Sub-issue|State|ZIP code|Submitted via|Date received|Date sent to company|Company|Company response|Timely response?|Consumer dispute d?|
|1114245|Debt collection|Medical|Disclosure verifi...|Not given enough ...|FL|32219|Web|11/13/2014|11/13/2014|Choice Recovery, Inc|Closed with expla...|Yes|NA|
|1114488|Debt collection|Medical|Disclosure verifi...|Right to dispute ...|TX|75006|Web|11/13/2014|11/13/2014|Expert Global Sol...|In progress|Yes|NA|
|1114255|Bank account or s...|Checking account|Deposits and with...|NA|NY|11102|Web|11/13/2014|11/13/2014|FNIS (Fidelity Na...|In progress|Yes|NA|
|1115106|Debt collection|Other (phone, hea...|Communication tac...|Frequent or repea...|GA|31721|Web|11/13/2014|11/13/2014|Expert Global Sol...|In progress|Yes|NA|
|1115890|Credit reporting|NA|Incorrect informa...|Information is no...|FL|33461|Web|11-12-2014|11/13/2014|TransUnion|In progress|Yes|NA|
|1114180|Debt collection|Credit card|Cont'd attempts c...|Debt is not mine|CA|95035|Web|11-12-2014|11/13/2014|Expert Global Sol...|In progress|Yes|NA|
|1114124|Consumer loan|Vehicle loan|Managing the loan...|NA|MI|48430|Fax|11-12-2014|11/13/2014|Ally Financial Inc|In progress|Yes|NA|
|1113211|Debt collection|Credit card|False statements ...|Indicated committ...|MS|39056|
```

Web	11-12-2014	11-12-2014	Premium Asset Ser...	In progress	Yes			NA
	1112443	Consumer loan	Installment loan	Managing the loan...	NA	CA	94509	
Web	11-12-2014	11-12-2014	Ally Financial Inc	In progress	Yes			NA
	1112618	Debt collection	Medical	Disclosure verifi...	Not given enough ...	FL	34743	
Web	11-12-2014	11-12-2014	Enhanced Recovery...	Closed with non-m...	Yes			NA
	1111987	Credit reporting	NA	Credit reporting ...	No notice of inve...	NJ	8852	
Web	11-12-2014	11-12-2014	TransUnion	Closed with expla...	Yes			NA
	1113072	Credit reporting	NA	Incorrect informa...	Account status	NY	11514	
Web	11-12-2014	11-12-2014	TransUnion	Closed with expla...	Yes			NA
	1112733	Debt collection	Medical	Communication tac...	Used obscene/prof...	WA	98034	
Web	11-12-2014	11-12-2014	Renton Collection...	Closed with expla...	Yes			NA
	1112115	Mortgage	Conventional fixe...	Loan modification...	NA	NJ	8527	
Web	11-12-2014	11-12-2014	Seterus	Closed with expla...	Yes			NA
	1112705	Credit reporting	NA	Unable to get cre...	Problem getting r...	CA	90716	
Web	11-12-2014	11-12-2014	TransUnion	Closed with expla...	Yes			NA
	1113116	Credit reporting	NA	Incorrect informa...	Account status	CT	6242	
Web	11-12-2014	11-12-2014	TransUnion	Closed with expla...	Yes			NA
	1112181	Debt collection	Credit card	Cont'd attempts c...	Debt is not mine	FL	34481	
Web	11-12-2014	11-12-2014	Commonwealth Fina...	Closed with expla...	Yes			NA
	1113153	Debt collection	Other (phone, hea...	Cont'd attempts c...	Debt was paid	GA	30345	
Web	11-12-2014	11-12-2014	EOS Holdings, Inc	Closed with expla...	Yes			NA
	1112011	Debt collection	Medical	Cont'd attempts c...	Debt resulted fro...	SC	29360	
Web	11-12-2014	11-12-2014	Revenue Recovery ...	Closed with expla...	Yes			NA

```

+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+

```

-+

only showing top 20 rows

```
In [59]: import pyspark
from pyspark.sql import SparkSession
df=spark.read.csv(r'C:\\Users\\SaikumarGanji\\Downloads\\Complaints.csv')
df.show()
df.count()
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
-+
|      _c0|      _c1|      _c2|      _c3|      _c4|      _c5|      _c6|
_c7|      _c8|      _c9|      _c10|      _c11|      _c12|      _c13|
+-----+-----+-----+-----+-----+-----+-----+-----+
-+
|Complaint ID|      Product|      Sub-product|      Issue|      Sub-issue|State|ZIP code|Submit
ted via|Date received|Date sent to company|      Company|      Company response|Timely response?|Consumer dispute
d?|
|      1114245|      Debt collection|      Medical|Disclosure verifi...|Not given enough ...|      FL|      32219|
Web|      11/13/2014|      11/13/2014|Choice Recovery, Inc|Closed with expla...|      Yes|      NA|
|      1114488|      Debt collection|      Medical|Disclosure verifi...|Right to dispute ...|      TX|      75006|
Web|      11/13/2014|      11/13/2014|Expert Global Sol...|      In progress|      Yes|      NA|
|      1114255|Bank account or s...|      Checking account|Deposits and with...|      NA|      NY|      11102|
Web|      11/13/2014|      11/13/2014|FNIS (Fidelity Na...|      In progress|      Yes|      NA|
|      1115106|      Debt collection|Other (phone, hea...|Communication tac...|Frequent or repea...|      GA|      31721|
Web|      11/13/2014|      11/13/2014|Expert Global Sol...|      In progress|      Yes|      NA|
|      1115890|      Credit reporting|      NA|Incorrect informa...|Information is no...|      FL|      33461|
Web|      11-12-2014|      11/13/2014|      TransUnion|      In progress|      Yes|      NA|
|      1114180|      Debt collection|      Credit card|Cont'd attempts c...|      Debt is not mine|      CA|      95035|
Web|      11-12-2014|      11/13/2014|Expert Global Sol...|      In progress|      Yes|      NA|
|      1114124|      Consumer loan|      Vehicle loan|Managing the loan...|      NA|      MI|      48430|
Fax|      11-12-2014|      11/13/2014|      Ally Financial Inc|      In progress|      Yes|      NA|
|      1113211|      Debt collection|      Credit card|False statements ...|Indicated committ...|      MS|      39056|
Web|      11-12-2014|      11-12-2014|Premium Asset Ser...|      In progress|      Yes|      NA|
|      1112443|      Consumer loan|      Installment loan|Managing the loan...|      NA|      CA|      94509|
Web|      11-12-2014|      11-12-2014|      Ally Financial Inc|      In progress|      Yes|      NA|
|      1112618|      Debt collection|      Medical|Disclosure verifi...|Not given enough ...|      FL|      34743|
Web|      11-12-2014|      11-12-2014|Enhanced Recovery...|Closed with non-m...|      Yes|      NA|
|      1111987|      Credit reporting|      NA|Credit reporting ...|No notice of inve...|      NJ|      8852|
Web|      11-12-2014|      11-12-2014|      TransUnion|Closed with expla...|      Yes|      NA|
|      1113072|      Credit reporting|      NA|Incorrect informa...|      Account status|      NY|      11514|
Web|      11-12-2014|      11-12-2014|      TransUnion|Closed with expla...|      Yes|      NA|
```

	1112733	Debt collection	Medical	Communication tac...	Used obscene/prof...	WA	98034	
Web	11-12-2014	11-12-2014	Renton Collection...	Closed with expla...	Yes		NA	
	1112115	Mortgage	Conventional fixe...	Loan modification...	NA	NJ	8527	
Web	11-12-2014	11-12-2014	Seterus	Closed with expla...	Yes		NA	
	1112705	Credit reporting	NA	Unable to get cre...	Problem getting r...	CA	90716	
Web	11-12-2014	11-12-2014	TransUnion	Closed with expla...	Yes		NA	
	1113116	Credit reporting	NA	Incorrect informa...	Account status	CT	6242	
Web	11-12-2014	11-12-2014	TransUnion	Closed with expla...	Yes		NA	
	1112181	Debt collection	Credit card	Cont'd attempts c...	Debt is not mine	FL	34481	
Web	11-12-2014	11-12-2014	Commonwealth Fina...	Closed with expla...	Yes		NA	
	1113153	Debt collection	Other (phone, hea...	Cont'd attempts c...	Debt was paid	GA	30345	
Web	11-12-2014	11-12-2014	EOS Holdings, Inc	Closed with expla...	Yes		NA	
	1112011	Debt collection	Medical	Cont'd attempts c...	Debt resulted fro...	SC	29360	
Web	11-12-2014	11-12-2014	Revenue Recovery ...	Closed with expla...	Yes		NA	

```

+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-+
only showing top 20 rows

```

Out[59]: 312913

```
In [72]: df1=spark.sql("select count(_c0) from complaints where _c11 = 'In progress'")
df1.show()
```

```

+-----+
|count(_c0)|
+-----+
|      2612|
+-----+

```

In [73]:

```

from pyspark.sql.types import IntegerType
df=df.withColumn('_c0',df['_c0'].cast(IntegerType()))
df.groupBy('_c10').max('_c0').show()

```

```

+-----+-----+
|          _c10|max(_c0)|
+-----+-----+
|Praxis Financial ...| 1083515|
|Worldwide Process...| 1076364|
|Lincoln Mortgage ...| 1071017|
|Law Office of Joe...| 1064107|
|Array Services Gr...| 1058623|
|      Loan To Learn| 1035595|
|      PlusFour, Inc|  998737|
|      Security Finance| 1091472|
|Capital Markets C...| 1079608|
|Law Office of Har...| 1059214|
|Holloway & Moxley...| 1055073|
|  Toll Brothers Inc| 1056117|
|Brazos Higher Edu...| 1046185|
|Collection Bureau...| 1071486|
|GLOBAL SOLUTION B...| 1048376|
|  Stillman Law Office| 1025998|
|Platinum Mortgage...| 1026337|
|Metropolitan Home...| 1020230|
|Homebridge Financ...| 1012938|
|Louisiana Recover...| 1010424|
+-----+-----+
only showing top 20 rows

```

```
In [71]: df.createOrReplaceTempView("complaints")
df3=spark.sql("select _c10 from complaints where complaints._c11='Closed with explanation' and complaints._c12 ='Yes'").show()
```

```
+-----+
|          _c10|
+-----+
|Choice Recovery, Inc|
|          TransUnion|
|          TransUnion|
|Renton Collection...|
|          Seterus|
|          TransUnion|
|          TransUnion|
|Commonwealth Fina...|
|      EOS Holdings, Inc|
|Revenue Recovery ...|
|          TransUnion|
|GLA Collection Co...|
|          Wells Fargo|
|          TransUnion|
|          TransUnion|
|          TransUnion|
|Stellar Recovery Inc|
|          TransUnion|
|Convergent Resour...|
|Southwest Recover...|
+-----+
only showing top 20 rows
```

```
In [79]: df=spark.sql("select max(_c10) from complaints where _c12='Yes'")
df.show()
```

```
+-----+
|  max(_c10)|
+-----+
|iServe Trust|
+-----+
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```