

```
In [21]: import findspark
findspark.init()

In [22]: import pyspark
from pyspark.sql import SparkSession

In [23]: spark = SparkSession.Builder().master("local[1]").appName("SparkByExamples.com").getOrCreate()

In [24]: df=spark.read.options(header='True',inferSchema='True').csv(r'C:\Users\VinayTantri\Downloads\complaints.csv')
df.show()
df.count()
```

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+										
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+										
Complaint_ID	Product	Sub-product	Issue		Sub-issue	State	ZIP code	Submitted via	Date received	Date s
ent to company	Company	Company_response	Timely_response	Consumer	disputed?					
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+										
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						
1111995	Credit reporting		NA	Credit reporting ...	Problem with stat...	CA	92809	Web	11/12/2014	
11/12/2014	TransUnion	Closed with expla...	Yes	NA						
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+										
only showing top 20 rows										

Out[24]: 312912

```
In [25]: df.createOrReplaceTempView("facts")
df= spark.sql("select count(Complaint_ID) from facts where Company_response='In progress' ")
df.show()

+-----+
|count(Complaint_ID)|
+-----+
|                2612|
+-----+
```

```
In [26]: spark.sql("select company, count(issue) as no_of_issue from facts group by company, issue order by no_of_issue desc limit 1").show()

+-----+-----+
|      company|no_of_issue|
+-----+-----+
|Bank of America|       20038|
+-----+-----+
```

```
In [35]: spark.sql("select company from facts where company_response='Closed with explanation' and timely_response='Yes' group by company").show()

+-----+
|      company|
+-----+
|Praxis Financial ...|
|Worldwide Process...|
|Law Office of Joe...|
|      Loan To Learn|
|Array Services Gr...|
|      PlusFour, Inc|
|Investigation & R...|
|F&L Marketing Ent...|
|Greenlight Financial|
|Lincoln Mortgage ...|
|      Security Finance|
|Holloway & Moxley...|
|      Toll Brothers Inc|
|Brazos Higher Edu...|
|Capital Markets C...|
|Law Office of Har...|
|      BAM Financial, LLC|
|Richard A Weidel ...|
|Collection Bureau...|
|GLOBAL SOLUTION B...|
+-----+
only showing top 20 rows
```

```
In [31]: spark.sql("select company from facts where timely_response='Yes' group by company,timely_response").show()

+-----+
|      company|
+-----+
|JA Cambece Law Of...|
|      BlueChip Financial|
|FEDERAL PACIFIC C...|
|Royal United Mort...|
|Suttell & Hammer PS|
|      RFNA, LP|
|Blatt, Hasenmille...|
|Midway Resolution...|
|      TrueAccord Corp|
|The McHughes Law ...|
|Habitat for Humanity|
|Professional Reco...|
|Investigation & R...|
|      RCO Legal, PS|
|      BluFi Lending|
|Old Republic Nati...|
|      Americash|
|Turning Point Sol...|
|Credit Bureau of ...|
|P&B Capital Group...|
+-----+
only showing top 20 rows
```

In []: