databricksExercise_9

```
sparkdata = spark.read.format("csv")\
          .option("header",True) \
          .option("inferschema",True)\
          .load("/FileStore/tables/Spark_Data.csv")
from pyspark.sql import functions as F
from pyspark.sql import *
sparkdata.createOrReplaceTempView("sparkdata")
spark.sql("select * from sparkdata")
```

```
Out[2]: DataFrame[InvoiceNo: string, StockCode: string, Description: string, Q
uantity: int, InvoiceDate: string, UnitPrice: double, CustomerID: int, Countr
y: string]
```

```
%sql select to_date(InvoiceDate, 'dd/MM/yyyy') as Invoice_date from sparkdata
```

	Invoice_date
1	2010-01-12
2	2010-01-12
3	2010-01-12
4	2010-01-12
5	2010-01-12
6	2010-01-12
7	2010-01-12
8	2010-01-12

Showing the first 1000 rows.

```
%sql
create global temp view indesepkc as
select to_date(InvoiceDate, 'dd/MM/yyyy') as
Invoice_date,Description,abs(sum(Quantity)) as quant_sum
from sparkdata
where Description like 'damaged'or Description like 'missing' or Description
like 'throw away'or Description like 'thrown away'or Description like
'unsaleable'or Description like 'destroyed'
group by Invoice_date,Description
order by Description ASC ;
```

OK

	Invoice_date	Damaged_Products_By_Date
1	null	2254
2	2011-01-07	8
3	2011-03-06	58
4	2011-03-10	29
5	2011-03-11	40
6	2011-04-02	2212
7	2011-07-11	81
8	2011-08-06	140

Showing all 12 rows.