

databricksExercise10 - Assessment1

```
import pyspark
from pyspark.sql.types import StructType, StringType, IntegerType, DoubleType, FloatType, DateType
```

```
ex5_schema = StructType() \
    .add("InvoiceNo",StringType(),False) \
    .add("StockCode",StringType(),False) \
    .add("Description",StringType(),True) \
    .add("Quantity",IntegerType(),True) \
    .add("InvoiceDate",StringType(),True) \
    .add("UnitPrice",DoubleType(),True) \
    .add("CustomerID",IntegerType(),True) \
    .add("Country",StringType(),True)
```

```
kaggldata = spark.read.format("csv") \
    .option("header",True) \
    .schema(ex5_schema) \
    .load("FileStore/tables/data.csv")
```

```
kaggldata.printSchema()
kaggldata.show(5)
```

```
root
|-- InvoiceNo: string (nullable = true)
|-- StockCode: string (nullable = true)
|-- Description: string (nullable = true)
|-- Quantity: integer (nullable = true)
|-- InvoiceDate: string (nullable = true)
|-- UnitPrice: double (nullable = true)
|-- CustomerID: integer (nullable = true)
|-- Country: string (nullable = true)
```

```
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+
|InvoiceNo|StockCode|          Description|Quantity|  InvoiceDate|UnitPrice|Cu
stomerID|      Country|
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+
|   536365|   85123A|WHITE HANGING HEA...|      6|12/1/2010 8:26|    2.55|
17850|United Kingdom|
|   536365|   71053|WHITE METAL LANTERN|      6|12/1/2010 8:26|    3.39|
17850|United Kingdom|
|   536365|   84406B|CREAM CUPID HEART...|      8|12/1/2010 8:26|    2.75|
17850|United Kingdom|
```

```
| 536365| 84029G|KNITTED UNION FLA...| 6|12/1/2010 8:26| 3.39|
17850|United Kingdom|
| 536365| 84029E|RED WOOLLY HOTTIE...| 6|12/1/2010 8:26| 3.39|
17850|United Kingdom|
+-----+-----+-----+-----+-----+-----+
-----+-----+
only showing top 5 rows
```

```
from pyspark.sql import *
df = kaggledata.createOrReplaceTempView("data")
spark.sql("select * from data").show(10)
```

```
+-----+-----+-----+-----+-----+-----+
-----+-----+
|InvoiceNo|StockCode|          Description|Quantity|  InvoiceDate|UnitPrice|Cu
stomerID|      Country|
+-----+-----+-----+-----+-----+-----+
-----+-----+
| 536365| 85123A|WHITE HANGING HEA...| 6|12/1/2010 8:26| 2.55|
17850|United Kingdom|
| 536365| 71053| WHITE METAL LANTERN| 6|12/1/2010 8:26| 3.39|
17850|United Kingdom|
| 536365| 84406B|CREAM CUPID HEART...| 8|12/1/2010 8:26| 2.75|
17850|United Kingdom|
| 536365| 84029G|KNITTED UNION FLA...| 6|12/1/2010 8:26| 3.39|
17850|United Kingdom|
| 536365| 84029E|RED WOOLLY HOTTIE...| 6|12/1/2010 8:26| 3.39|
17850|United Kingdom|
| 536365| 22752|SET 7 BABUSHKA NE...| 2|12/1/2010 8:26| 7.65|
17850|United Kingdom|
| 536365| 21730|GLASS STAR FROSTE...| 6|12/1/2010 8:26| 4.25|
17850|United Kingdom|
| 536366| 22633|HAND WARMER UNION...| 6|12/1/2010 8:28| 1.85|
17850|United Kingdom|
| 536366| 22632|HAND WARMER RED P...| 6|12/1/2010 8:28| 1.85|
17850|United Kingdom|
| 536367| 84879|ASSORTED COLOUR B...| 32|12/1/2010 8:34| 1.69|
13047|United Kingdom|
+-----+-----+-----+-----+-----+-----+
-----+-----+
only showing top 10 rows
```

```
%sql select * from data where CustomerId is NULL
```



1	536414	22139	null	56
2	536544	21773	DECORATIVE ROSE BATHROOM BOTTLE	1
3	536544	21774	DECORATIVE CATS BATHROOM BOTTLE	2
4	536544	21786	POLKADOT RAIN HAT	4
5	536544	21787	RAIN PONCHO RETROSPOT	2
6	536544	21790	VINTAGE SNAP CARDS	9
7	536544	21791	VINTAGE HEADS AND TAILS CARD GAME	2
8	536544	21801	CHRISTMAS TREE DECORATION WITH BELL	10

Showing the first 1000 rows.



```
%sql select * from data where Quantity < 0
```

	InvoiceNo ▲	StockCode ▲	Description ▲	Quar
1	C536379	D	Discount	-1
2	C536383	35004C	SET OF 3 COLOURED FLYING DUCKS	-1
3	C536391	22556	PLASTERS IN TIN CIRCUS PARADE	-12
4	C536391	21984	PACK OF 12 PINK PAISLEY TISSUES	-24
5	C536391	21983	PACK OF 12 BLUE PAISLEY TISSUES	-24
6	C536391	21980	PACK OF 12 RED RETROSPOT TISSUES	-24
7	C536391	21484	CHICK GREY HOT WATER BOTTLE	-12
8	C536391	22557	PLASTERS IN TIN VINTAGE PAISLEY	-12

Showing the first 1000 rows.



```
%sql select * from data where CustomerId is NULL and Quantity < 0
```

	InvoiceNo ▲	StockCode ▲	Description ▲	Quan
1	536589	21777	null	-10
2	536764	84952C	null	-38
3	536996	22712	null	-20
4	536997	22028	null	-20
5	536998	85067	null	-6
6	537000	21414	null	-22
7	537001	21653	null	-6

8	537003	85126	null	-2
---	--------	-------	------	----

Showing the first 1000 rows.



	InvoiceNo ▲	StockCode ▲	Description ▲	Quantit
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6
2	536365	71053	WHITE METAL LANTERN	6
3	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8
4	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6
5	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6
6	536365	22752	SET 7 BABUSHKA NESTING BOXES	2
7	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6
8	536366	22633	HAND WARMER UNION JACK	6

Showing the first 1000 rows.

