

```
In [1]: from pyspark.sql import SparkSession
```

```
In [2]: spark = SparkSession.builder.appName("SparkSQL").getOrCreate()
```

```
24/10/13 20:22:00 WARN Utils: Your hostname, sundharakumar-HP-348-G7 resolves to a loopb
ack address: 127.0.1.1; using 192.168.0.114 instead (on interface wlo1)
24/10/13 20:22:00 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLev
el).
24/10/13 20:22:00 WARN NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
24/10/13 20:22:01 WARN Utils: Service 'SparkUI' could not bind on port 4040. Attempting
port 4041.
```

```
In [3]: people = spark.read.option("header", "true").option("inferSchema", "true").csv("file:///ho
```

```
In [4]: people.printSchema()
```

```
root
 |-- Id: integer (nullable = true)
 |-- Name: string (nullable = true)
 |-- age: integer (nullable = true)
 |-- friends: integer (nullable = true)
```

```
In [5]: people.filter(people.age<21).show()
```

```
+---+-----+---+-----+
| Id|   Name|age|friends|
+---+-----+---+-----+
| 21|   Miles| 19|    268|
| 48|    Nog| 20|      1|
| 52|Beverly| 19|    269|
| 54|   Brunt| 19|      5|
| 60|Geordi| 20|    100|
| 73|   Brunt| 20|    384|
|106|Beverly| 18|    499|
|115|   Dukat| 18|    397|
|133|   Quark| 19|    265|
|136|   Will| 19|    335|
|225|   Elim| 19|    106|
|304|   Will| 19|    404|
|327| Julian| 20|      63|
|341|   Data| 18|    326|
|349|Kasidy| 20|    277|
|366|  Keiko| 19|    119|
|373|   Quark| 19|    272|
|377|Beverly| 18|    418|
|404|Kasidy| 18|      24|
|409|    Nog| 19|    267|
+---+-----+---+-----+
only showing top 20 rows
```

```
In [ ]:
```

```
In [ ]:
```