Spark SQL Quiz

```
+----------+-----------+-----------+
|   name   | birthYear | specialty |
+----------+-----------+-----------+
| Sheldon  |      1953 | physics   |
| Leonard  |      1643 | math      |
| Howard   |      1965 | scientist |
+----------+-----------+-----------+
```

Let's create a Dork class and make a list of the dorks in the table above. Use the list of Dork objects to make a Dataset of the dorks.

```scala
case class Dork(name: String, birthYear: Int, specialty: String)
val dorks = List(
  Dork("Sheldon", 1953, "physics"),
  Dork("Leonard", 1643, "math"),
  Dork("Howard", 1965, "engineer")
)
var dorkDs: org.apache.spark.sql.Dataset[Dork] = spark.createDataset(dorks)
```

1. **Display the contents of the dorkDs Dataset.**

   Solution: The show method is used to display the contents of a Dataset.

   dorkDs.show()

2. **Print the schema of the dorkDs**

Solution: dorkDs.printSchema()

3. **Create a youngDorkDs Dataset that contains all the dorks that were born after 1900.**

   Solution: val youngDorkDs = dorkDs.filter {

     dorkDs("birthYear") > 1900

   }

4. **Sort the dorkDs by the name column.**

   Solution: val sortedDorkDs = dorkDs.orderBy("name").

   You can achieve the same result the SQL way.

   dorkDs.createOrReplaceTempView("dorks")

   val sortedDorkDs = spark.sql("select * from dorks order by name")

5. **What type of object does dorkDs("name") return? Explain the result.**

   Solution: dorkDs("name") returns a org.apache.spark.sql.Column object.