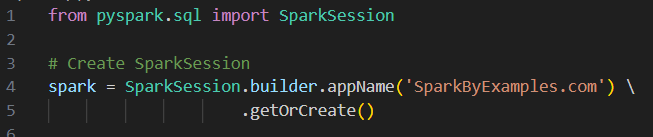
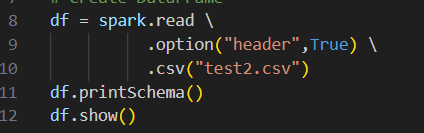
**Pyspark SQL:**

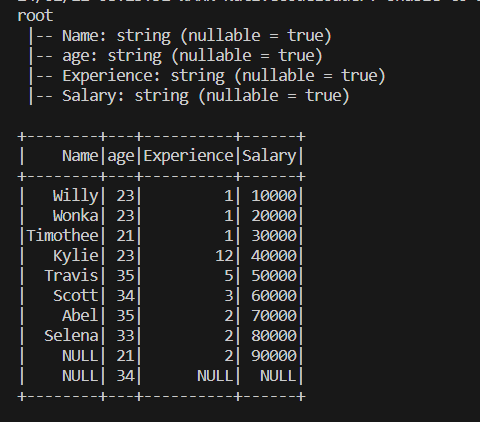
The pyspark.sql is a module in PySpark that is used to perform SQL-like operations on the data stored in memory. You can either leverage using programming API to query the data or use the ANSI SQL queries similar to RDBMS. You can also mix both, for example, use API on the result of an SQL query.

****

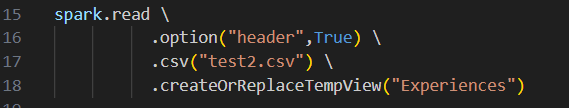
**Creating a data frame**

****

**Output:**

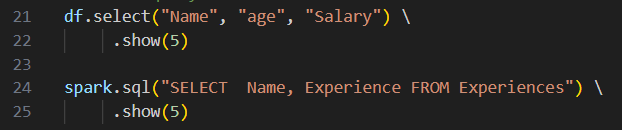
****

A view has to be created in for the data

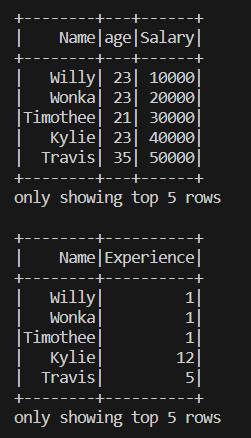
****

**SELECT query:**

We can perform select operation on the data using the view or without the view.

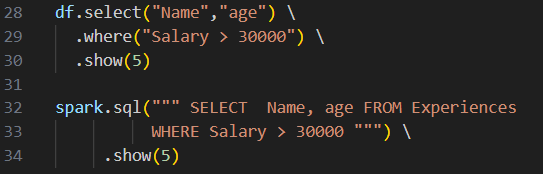
****

**Output:**

****

**WHERE :**

We can filter the data using the view or without the view.

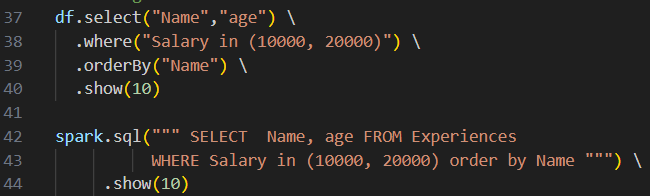
****

**Output:**

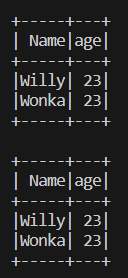
****

**Sorting the data:**

We can sort the data in two ways.

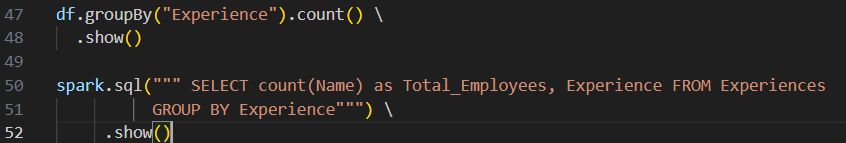
****

**Output:**

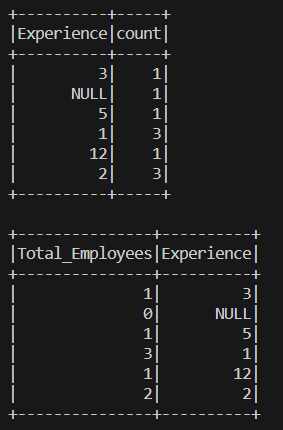
****

**Group by aggregations:**

The [groupBy().count()](https://sparkbyexamples.com/pyspark/pyspark-groupby-count-explained/) is used to perform the group by on DataFrame.

****

**Output:**

****