In [ ]:
```
import pyspark
```

In [ ]:
```
import pandas as pd
pd.read_csv("DataSet.csv")
```

Out[ ]:

	First Name	Last Name	Age
0	Billy	Banks	56
1	Tailer	Davis	25
2	Bob	Dillon	45

PYSPARK IMPORT

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

CREATE/BUILD SPARK SESSION VARIABLE

In [ ]:
```
spark=SparkSession.builder.appName('Practice').getOrCreate()
```

In [ ]:
```
spark
```

Out[ ]:

SparkSession - in-memory

SparkContext

Spark UI

Version	`v3.3.0`
Master	`local[*]`
AppName	`Practice`

In [ ]:
```
df_spark= spark.read.csv("DataSet.csv")
```

In [ ]:
```
type(df_spark)
```

Out[ ]:
```
pyspark.sql.dataframe.DataFrame
```

MAKES THE FIRST ROW IN DATASET THE HEADER

In [ ]:
```
df_spark= spark.read.option('header', 'True').csv('DataSet.csv')
```

DISPLAYS DATA SET (MORE STRUCTURED VIEW)

In [ ]:
```
df_spark.show()
```

```
+----------+---------+---+
|First Name|Last Name|Age|
+----------+---------+---+
|    Billy |    Banks| 56|
|    Tailer|   Davis | 25|
|      Bob |   Dillon| 45|
+----------+---------+---+
```

SHOWS MORE INFORMATION ON THE SCHEMA FOR DATASET (COLUMN NAMES, DATATYPES)

In [ ]:
```
df_spark.printSchema()
```

```
root
 |-- First Name: string (nullable = true)
 |-- Last Name: string (nullable = true)
 |-- Age: string (nullable = true)
```

CALLS TOP 3 ROWS FROM DATASET

In [ ]:
```
df_spark.head(3)
```

Out[ ]:
```
[Row(First Name='Billy ', Last Name='Banks', Age='56'),
 Row(First Name='Tailer', Last Name='Davis ', Age='25'),
 Row(First Name='Bob ', Last Name='Dillon', Age='45')]
```