Exercise 3: JSON data in Databricks

1. **Import JSON data**

df = spark.read.json("examples/src/main/resources/people.json")

1. **Displays the content of the DataFrame to stdout**

df.show()

1. **Print the schema in a tree format**

df.printSchema()

1. **Select only the "name" column**

df.select("name").show()

1. **Select everybody, but increment the age by 1**

df.select(df['name'], df['age'] + 1).show()

1. **Select people older than 21**

df.filter(df['age'] > 21).show()

1. **Count people by age**

df.groupBy("age").count().show()