sparksql-onu-jupyter

November 14, 2025

```python
df=spark.read.csv('s3://usernamedatalake/datasets/onu/hdi/
 ↪',inferSchema=True,header=True)
```

```python
df.show(10)
```

```python
df.createOrReplaceTempView("hdi")
```

```python
dfsql = spark.sql("select country, lifeex from hdi where lifeex > 80")
```

```python
dfsql.show()
```

```python
#target directory
write_uri='s3://<bucket/dir>/df_csv'
dfsql.coalesce(1).write.format("csv").option("header","true").save(write_uri)
```