

## 6.Read\_and\_WriteSCV

March 4, 2025

```
[14]: from pyspark.sql import SparkSession
import json
```

```
#importar este CVS
import csv
```

```
[15]: # Criar uma sessão Spark
spark = SparkSession.builder.appName("CSVExample").getOrCreate()
```

```
[16]: # Read um CVS
csv_filename = "Aula 2/CSV/review.csv"
df = spark.read.csv(csv_filename, header=True, inferSchema=True)
```

```
[17]: # Mostrar os dados
df.show()
```

```
+-----+-----+-----+-----+
---+-----+
|          App|  Translated_Review|          Sentiment|
Sentiment_Polarity|Sentiment_Subjectivity|
+-----+-----+-----+-----+
---+-----+
|10 Best Foods for...|"I like eat delic...| also ""Best Befo...|
Positive|                  1.0|
|10 Best Foods for...|This help eating ...|          Positive|
0.25|    0.28846153846153844|
|10 Best Foods for...|          nan|          nan|
nan|          nan|
|10 Best Foods for...|Works great espec...|          Positive|
0.4|          0.875|
|10 Best Foods for...|          Best idea us|          Positive|
1.0|          0.3|
|10 Best Foods for...|          Best way|          Positive|
1.0|          0.3|
|10 Best Foods for...|          Amazing|          Positive|
0.60000000000000001|          0.9|
|10 Best Foods for...|          nan|          nan|
nan|          nan|
```

```
|10 Best Foods for...|Looking forward app,| Neutral|
0.0| 0.0|
|10 Best Foods for...|It helpful site !...| Neutral|
0.0| 0.0|
|10 Best Foods for...| good you.| Positive|
0.7| 0.6000000000000001|
|10 Best Foods for...|Useful informatio...| Positive|
0.2| 0.1|
|10 Best Foods for...|Thank you! Great ...| Positive|
0.75| 0.875|
|10 Best Foods for...|Greatest ever Com...| Positive|
0.9921875| 0.8666666666666667|
|10 Best Foods for...|Good health...| Positive|
0.5499999999999999| 0.5111111111111112|
|10 Best Foods for...| nan| nan|
nan| nan|
|10 Best Foods for...|Health It's impor...| Positive|
0.45| 1.0|
|10 Best Foods for...|Mrs sunita bhati ...| Positive|
0.6| 0.6666666666666666|
|10 Best Foods for...|Very Useful in di...|
Positive|0.29500000000000004| 0.1|
|10 Best Foods for...| One greatest apps.| Positive|
1.0| 1.0|
+-----+-----+-----+-----+
---+-----+
only showing top 20 rows
```

```
[18]: # Criar uma sessão Spark
spark = SparkSession.builder.appName("CSVExample").getOrCreate()

# Criar um ficheiro CSV
data = [
    [1, "Alice", 25],
    [2, "Bob", 30],
    [3, "Carlos", 22],
    [4, "Diana", 28],
    [5, "Eva", 35],
    [6, "Fernando", 40]
]
```

```
[19]: # Cria um CVS
csv_filename = "dados.csv"
with open(csv_filename, mode="w", newline="", encoding="utf-8") as file:
    writer = csv.writer(file)
    writer.writerow(["id", "nome", "idade"])
```

```
writer.writerows(data)
```

```
[20]: # Carregar o CSV com PySpark
df = spark.read.option("header", "true").csv(csv_filename, inferSchema=True)
```

```
[21]: # Mostrar os dados
df.show()
```

```
+---+-----+-----+
| id|    nome|idade|
+---+-----+-----+
|  1|   Alice|   25|
|  2|    Bob|   30|
|  3|  Carlos|   22|
|  4|   Diana|   28|
|  5|    Eva|   35|
|  6|Fernando|   40|
+---+-----+-----+
```

```
[22]: df.show(3)
```

```
+---+-----+-----+
| id|    nome|idade|
+---+-----+-----+
|  1|   Alice|   25|
|  2|    Bob|   30|
|  3|  Carlos|   22|
+---+-----+-----+
only showing top 3 rows
```

```
[23]: # Mostrar o top 5
df.limit(5).show()
```

```
+---+-----+-----+
| id|    nome|idade|
+---+-----+-----+
|  1|   Alice|   25|
|  2|    Bob|   30|
|  3|  Carlos|   22|
|  4|   Diana|   28|
|  5|    Eva|   35|
+---+-----+-----+
```

```
[24]: from pyspark.sql import SparkSession
import json
```

```
[25]: # Criar uma sessão Spark
spark = SparkSession.builder.appName("APP_JSONExample").getOrCreate()

# Criar um ficheiro JSON
data = [
    {"id": 1, "nome": "Alice", "idade": 25},
    {"id": 2, "nome": "Bob", "idade": 30},
    {"id": 3, "nome": "Carlos", "idade": 22},
    {"id": 4, "nome": "Diana", "idade": 28},
    {"id": 5, "nome": "Eva", "idade": 35},
    {"id": 6, "nome": "Fernando", "idade": 40}
]
```

```
[26]: json_filename = "Aula 2/Json/dados.json"
with open(json_filename, "w") as f:
    json.dump(data, f)
```

```
[ ]:
```