

5_DataFrame_carros

March 4, 2025

```
[1]: # Importar bibliotecas
from pyspark.sql import SparkSession

# Criar sessão do Spark
spark = SparkSession.builder \
    .appName("Carros_Analysis") \
    .getOrCreate()
```

```
[2]: # Criar dados dos carros
dados = [
    (1, 'Tesla Model S', 'Tesla', 'Elétrico', 2022, 120000, 3.2),
    (2, 'BMW M3', 'BMW', 'Gasolina', 2021, 85000, 4.1),
    (3, 'Audi e-tron', 'Audi', 'Elétrico', 2023, 95000, 4.5),
    (4, 'Mercedes C-Class', 'Mercedes', 'Gasolina', 2020, 60000, 5.0),
    (5, 'Toyota Corolla', 'Toyota', 'Híbrido', 2019, 30000, 8.0)
]

# Def
```

```
[3]: colunas = ["id_carro", "modelo", "marca", "tipo", "ano", "preco",
    ↪ "aceleracao_0_100"]
```

```
[4]: # Criar DataFrame
df = spark.createDataFrame(dados, colunas)
```

```
[10]: df.show()
```

id_carro	modelo	marca	tipo	ano	preco	aceleracao_0_100
1	Tesla Model S	Tesla	Elétrico	2022	120000	3.2
2	BMW M3	BMW	Gasolina	2021	85000	4.1
3	Audi e-tron	Audi	Elétrico	2023	95000	4.5
4	Mercedes C-Class	Mercedes	Gasolina	2020	60000	5.0
5	Toyota Corolla	Toyota	Híbrido	2019	30000	8.0

```
[6]: # Criar uma tabela temporária
df.createOrReplaceTempView("carros")
```

```
[9]: spark.sql("""
      SELECT * FROM carros
      WHERE tipo = 'Elétrico'
      """).show()
```

id_carro	modelo	marca	tipo	ano	preco	aceleracao_0_100
1	Tesla Model S	Tesla	Elétrico	2022	120000	3.2
3	Audi e-tron	Audi	Elétrico	2023	95000	4.5

```
[11]: df.show()
```

id_carro	modelo	marca	tipo	ano	preco	aceleracao_0_100
1	Tesla Model S	Tesla	Elétrico	2022	120000	3.2
2	BMW M3	BMW	Gasolina	2021	85000	4.1
3	Audi e-tron	Audi	Elétrico	2023	95000	4.5
4	Mercedes C-Class	Mercedes	Gasolina	2020	60000	5.0
5	Toyota Corolla	Toyota	Híbrido	2019	30000	8.0

```
[12]: spark.sql("""
      SELECT * FROM carros
      ORDER BY preco DESC
      """).show()
```

id_carro	modelo	marca	tipo	ano	preco	aceleracao_0_100
1	Tesla Model S	Tesla	Elétrico	2022	120000	3.2
3	Audi e-tron	Audi	Elétrico	2023	95000	4.5
2	BMW M3	BMW	Gasolina	2021	85000	4.1
4	Mercedes C-Class	Mercedes	Gasolina	2020	60000	5.0
5	Toyota Corolla	Toyota	Híbrido	2019	30000	8.0

```
[13]: df.show()
```

id_carro	modelo	marca	tipo	ano	preco	aceleracao_0_100
----------	--------	-------	------	-----	-------	------------------

id_carro	modelo	marca	tipo	ano	preco	aceleracao_0_100	
	1	Tesla Model S	Tesla	Elétrico	2022	120000	3.2
	2	BMW M3	BMW	Gasolina	2021	85000	4.1
	3	Audi e-tron	Audi	Elétrico	2023	95000	4.5
	4	Mercedes C-Class	Mercedes	Gasolina	2020	60000	5.0
	5	Toyota Corolla	Toyota	Híbrido	2019	30000	8.0

```
[16]: spark.sql("""
        SELECT marca, modelo FROM carros
        ORDER BY preco DESC
        LIMIT 1
        """).show()
```

```
+-----+-----+
|marca|      modelo|
+-----+-----+
|Tesla|Tesla Model S|
+-----+-----+
```

```
[ ]: # 1. Liste todos os carros elétricos disponíveis.
      # 2. Ordene os carros pelo preço em ordem decrescente.
      # 3. Calcule a média de aceleração (0-100 km/h) por tipo de combustível.
      # 4. Identifique a marca com o carro mais caro.
      # 5. Filtre os carros que tenham um tempo de aceleração inferior a 5 segundos.
```