

ProyectoFinal



FINISHED ⓘ ⚙️ 🔒 default ▾

Carga del csv de articulos a un dataframe

Took 0 seconds. Last updated by anonymous at July 24 2025, 12:22:36 AM.

SPARK JOB FINISHED ⓘ ⚙️ 🔒

```
val dfArticulos = spark.read
  .option("inferSchema", "true")
  .option("header", "true")
  .option("delimiter", ",")
  .csv("/home/pepinillopepe/Downloads/ProyectoPAC4/ProyectoFinal/ProyectoFinal/articulos.csv")
```

dfArticulos: **org.apache.spark.sql.DataFrame** = [AÑO: int, TIPO: string ... 10 more fields]

Took 1 second. Last updated by anonymous at July 24 2025, 5:03:45 PM.

FINISHED ⓘ ⚙️ 🔒

Mostrar el esquema de articulos

Took 0 seconds. Last updated by anonymous at July 24 2025, 12:30:35 AM.

ERROR ⓘ ⚙️ 🔒

```
dfArticulos.printSchema
```

```
java.lang.RuntimeException
    at org.apache.zeppelin.interpreter.remote.PooledRemoteClient.callRemoteFunction(PooledRemoteClient.java:100)
    at org.apache.zeppelin.interpreter.remote.RemoteInterpreterProcess.callRemoteFunction(RemoteInterpreterProcess.java:100)
    at org.apache.zeppelin.interpreter.remote.RemoteInterpreter.interpret(RemoteInterpreter.java:100)
    at org.apache.zeppelin.notebook.Paragraph.jobRun(Paragraph.java:474)
    at org.apache.zeppelin.notebook.Paragraph.jobRun(Paragraph.java:68)
    at org.apache.zeppelin.scheduler.Job.run(Job.java:187)
    at org.apache.zeppelin.scheduler.AbstractScheduler.runJob(AbstractScheduler.java:100)
    at org.apache.zeppelin.scheduler.RemoteScheduler$JobRunner.run(RemoteScheduler$JobRunner.java:100)
    at java.base/java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1166)
    at java.base/java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:638)
    at java.base/java.lang.Thread.run(Thread.java:840)
```

Took 20 minutes. Last updated by anonymous at July 24 2025, 9:36:23 PM.

FINISHED    

Mostrar las filas totales de articulos



Took 1 second. Last updated by anonymous at July 24 2025, 12:10:20 PM.

 SPARK JOB FINISHED    

```
println(s"Total de filas: ${dfArticulos.count}")
```

Total de filas: 87073



Took 0 seconds. Last updated by anonymous at July 24 2025, 1:28:32 PM.

FINISHED    

Carga del csv de docentes a un data frame



Took 0 seconds. Last updated by anonymous at July 24 2025, 12:28:41 AM.

 SPARK JOB FINISHED    

```
val dfDocentes = spark.read
  .option("inferSchema", "true")
  .option("header", "true")
  .option("delimiter", ",")
  .csv("/home/pepinillopepe/Downloads/ProyectoPAC4/ProyectoFinal/ProyectoFinal/docentes.csv")
```

dfDocentes: **org.apache.spark.sql.DataFrame** = [AÑO: int, CODIGO_IES: int ... 12 more fields]



Took 0 seconds. Last updated by anonymous at July 24 2025, 5:03:48 PM.

FINISHED    

Mostrar el esquema de docentes



Took 0 seconds. Last updated by anonymous at July 24 2025, 12:32:01 AM.

FINISHED    

```
dfDocentes.printSchema
```

```
root
|-- AÑO: integer (nullable = true)
|-- CODIGO_IES: integer (nullable = true)
|-- NOMBRE_IES: string (nullable = true)
|-- SEXO: string (nullable = true)
|-- DISCAPACIDAD: string (nullable = true)
|-- ETNIA: string (nullable = true)
|-- NACIONALIDAD: string (nullable = true)
|-- NIVEL_FORMACION: string (nullable = true)
|-- RELACION_IES: string (nullable = true)
|-- TIEMPO_DEDICACION: string (nullable = true)
```



```
|-- NIVEL: string (nullable = true)
|-- GRADO: string (nullable = true)
|-- TOTAL: integer (nullable = true)
|-- id_universidad: integer (nullable = true)
```

Took 0 seconds. Last updated by anonymous at July 24 2025, 1:28:38 PM.

FINISHED    

Mostrar las filas totales de Docentes



Took 0 seconds. Last updated by anonymous at July 24 2025, 12:11:22 PM.

 SPARK JOB FINISHED    

```
println(s"Total de filas: ${dfDocentes.count}")
```

Total de filas: 38447



Took 1 second. Last updated by anonymous at July 24 2025, 1:28:41 PM.

FINISHED    

Carga del csv de matriculas a un data frame



Took 0 seconds. Last updated by anonymous at July 24 2025, 12:29:07 AM.

 SPARK JOB FINISHED    

```
val dfMatriculas = spark.read
  .option("inferSchema", "true")
  .option("header", "true")
  .option("delimiter", ",")
  .csv("/home/pepinillopepe/Downloads/ProyectoPAC4/ProyectoFinal/ProyectoFinal/matr
```

dfMatriculas: **org.apache.spark.sql.DataFrame** = [AÑO: int, NOMBRE_IES: string ... 18 mc



Took 1 second. Last updated by anonymous at July 24 2025, 5:03:52 PM.

FINISHED    

Mostrar el esquema de matriculas



Took 0 seconds. Last updated by anonymous at July 24 2025, 12:29:57 AM.

FINISHED    

```
dfMatriculas.printSchema
```

root

```
|-- AÑO: integer (nullable = true)
|-- NOMBRE_IES: string (nullable = true)
```



```
|-- TIPO_FINANCIAMIENTO: string (nullable = true)
|-- NOMBRE_CARRERA: string (nullable = true)
|-- NIVEL_FORMACION: string (nullable = true)
|-- MODALIDAD: string (nullable = true)
|-- CAMPO_AMPLIO: string (nullable = true)
|-- CAMPO_ESPECIFICO: string (nullable = true)
|-- CAMPO_DETALLADO: string (nullable = true)
|-- TIPO_SEDE: string (nullable = true)
|-- PROVINCIA_SEDE: string (nullable = true)
|-- CANTON_SEDE: string (nullable = true)
|-- SEXO: string (nullable = true)
|-- ETNIA: string (nullable = true)
|-- PUEBLO_NACIONALIDAD: string (nullable = true)
|-- DISCAPACIDAD: string (nullable = true)
|-- PROVINCIA_RESIDENCIA: string (nullable = true)
|-- CANTON_RESIDENCIA: string (nullable = true)
|-- TOTAL: integer (nullable = true)
|-- id_universidad: integer (nullable = true)
```

Took 0 seconds. Last updated by anonymous at July 24 2025, 1:29:36 PM.

FINISHED    

Mostrar las filas totales de Matriculas



Took 0 seconds. Last updated by anonymous at July 24 2025, 12:11:52 PM.

 SPARK JOB FINISHED    

```
println(s"Total de filas: ${dfMatriculas.count}")
```

Total de filas: 141833



Took 0 seconds. Last updated by anonymous at July 24 2025, 1:29:39 PM.

FINISHED    

Carga del csv de Universidades a un data frame



Took 0 seconds. Last updated by anonymous at July 24 2025, 1:15:05 PM.

 SPARK JOB FINISHED    

```
val dfUniversidades = spark.read
  .option("inferSchema", "true")
  .option("header", "true")
  .option("delimiter", ",")
  .csv("/home/pepinillopepe/Downloads/ProyectoPAC4/ProyectoFinal/ProyectoFinal/uni
```

dfUniversidades: **org.apache.spark.sql.DataFrame** = [id_universidad: int, nombre: string



Took 0 seconds. Last updated by anonymous at July 24 2025, 5:03:55 PM.

FINISHED    

Mostrar el esquema de Universidades



Took 0 seconds. Last updated by anonymous at July 24 2025, 1:15:55 PM.

FINISHED    

```
dfUniversidades.printSchema
```

```
root
|-- id_universidad: integer (nullable = true)
|-- nombre: string (nullable = true)
```



Took 0 seconds. Last updated by anonymous at July 24 2025, 1:29:45 PM.

FINISHED    

Mostrar las filas totales de Universidades



Took 0 seconds. Last updated by anonymous at July 24 2025, 1:16:25 PM.

 SPARK JOB FINISHED    

```
println(s"Total de filas: ${dfUniversidades.count}")
```

Total de filas: 67



Took 1 second. Last updated by anonymous at July 24 2025, 4:45:30 PM.

FINISHED    

Preguntas:



Took 2 seconds. Last updated by anonymous at July 24 2025, 4:57:52 PM.

FINISHED    

```
dfDocentes.createOrReplaceTempView("docentes")
dfMatriculas.createOrReplaceTempView("matriculas")
dfArticulos.createOrReplaceTempView("articulos")
dfUniversidades.createOrReplaceTempView("universidades")
```

Took 1 second. Last updated by anonymous at July 24 2025, 5:03:59 PM.

FINISHED    

¿Cual es el nivel de formación de los docentes en las universidades?



Took 0 seconds. Last updated by anonymous at July 24 2025, 5:38:13 PM.

FINISHED    

```
%jdbc
-- Muestra el nivel de formación por docentes de cada universidad
SELECT
  u.nombre AS universidad,
  d.NIVEL_FORMACION,
  SUM(d.TOTAL) AS total_docentes
FROM docentes d
JOIN universidades u ON d.id_universidad = u.id_universidad
GROUP BY u.nombre, d.NIVEL_FORMACION
ORDER BY total_docentes DESC
```



Setting ▼

universidad ▼	NIVEL_FORMACION ▼
UNIVERSIDAD DE GUAYAQUIL	CUARTO_NIVEL
PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR	CUARTO_NIVEL
UNIVERSIDAD CENTRAL DEL ECUADOR	CUARTO_NIVEL
UNIVERSIDAD TECNICA PARTICULAR DE LOJA	CUARTO_NIVEL
UNIVERSIDAD DE LAS AMERICAS	CUARTO_NIVEL
UNIVERSIDAD CATOLICA DE SANTIAGO DE GUAYAQUIL	CUARTO_NIVEL
UNIVERSIDAD LAICA ELOY ALFARO DE MANABI	CUARTO_NIVEL
UNIVERSIDAD DE LAS FUERZAS ARMADAS	CUARTO_NIVEL
UNIVERSIDAD DE CUENCA	CUARTO_NIVEL
UNIVERSIDAD POLITECNICA SALESIANA	CUARTO_NIVEL



Export ▼

< 1 2 3 4 5 ... 17 > 10 / page ▼



Took 0 seconds. Last updated by anonymous at July 24 2025, 5:04:08 PM. (outdated)

FINISHED







¿Cual es el promedio de mujeres y hombres matriculados en cada provincia?






...



Took 0 seconds. Last updated by anonymous at July 24 2025, 5:42:28 PM.

FINISHED



```
%jdbc
-- Muestra el promedio de matriculados por sexo de cada provincia
SELECT
  m.PROVINCIA_SEDE,
  m.SEXO,
  AVG(m.TOTAL) AS promedio_matriculados
FROM matriculas m
GROUP BY m.PROVINCIA_SEDE, m.SEXO
ORDER BY promedio_matriculados DESC
```





Setting ▾

PROVINCIA_SEDE ▾	SEXO ▾
ESMERALDAS	MUJER
GALAPAGOS	MUJER
EL ORO	MUJER
SANTA ELENA	MUJER
BOLIVAR	MUJER
PICHINCHA	MUJER
SANTA ELENA	HOMBRE
PICHINCHA	HOMBRE
EL ORO	HOMBRE
SANTO DOMINGO DE LOS TSACHILAS	MUJER

Export ▾

< 1 2 3 4 5 > 10 / page ▾

...

Took 1 second. Last updated by anonymous at July 24 2025, 5:11:00 PM. (outdated)





FINISHED








¿Cuál es el total de artículos que tiene cada universidad?



Took 0 seconds. Last updated by anonymous at July 24 2025, 5:42:59 PM.

FINISHED



```
%jdbc
-- Muestra cuantos articulos tiene cada universidad
SELECT
  u.nombre AS universidad,
  a.`BASE DATOS INDEXADA`,
  COUNT(*) AS total_articulos
FROM articulos a
JOIN universidades u ON a.id_universidad = u.id_universidad
GROUP BY u.nombre, a.`BASE DATOS INDEXADA`
ORDER BY total_articulos DESC
```





Setting ▾

universidad ▾	BASE DATOS INDEXADA ▾
UNIVERSIDAD SAN FRANCISCO DE QUITO	SCOPUS
ESCUELA SUPERIOR POLITECNICA DE CHIMBORAZO	LATIN_INDEX
UNIVERSIDAD DE GUAYAQUIL	LATIN_INDEX
ESCUELA POLITECNICA NACIONAL	SCOPUS
UNIVERSIDAD TECNICA DE MANABI	LATIN_INDEX
UNIVERSIDAD TECNICA PARTICULAR DE LOJA	SCOPUS
UNIVERSIDAD REGIONAL AUTÓNOMA DE LOS ANDES	LATIN_INDEX
UNIVERSIDAD DE LAS FUERZAS ARMADAS (ESPE)	SCOPUS
UNIVERSIDAD TECNICA DE AMBATO	LATIN_INDEX
UNIVERSIDAD REGIONAL AUTÓNOMA DE LOS ANDES	SCOPUS

Export ▾

<

1

2

3

4

5

...

58

>

10 / page ▾

Took 0 seconds. Last updated by anonymous at July 24 2025, 5:24:17 PM. (outdated)

FINISHED    

¿Cuál es la cantidad de matriculas por etnia que hay?

Took 0 seconds. Last updated by anonymous at July 24 2025, 5:43:41 PM.

FINISHED    

```
%jdbc
-- Muestra las matriculas por etnias
SELECT
  ETNIA,
  SUM(TOTAL) AS total_matriculados
FROM matriculas
GROUP BY ETNIA
ORDER BY total_matriculados DESC
```



Took 0 seconds. Last updated by anonymous at July 24 2025, 4:47:38 PM. (outdated)

FINISHED    

¿Cuántos docentes hombres y cuántas docentes mujeres hay en las universidades por año?

Took 0 seconds. Last updated by anonymous at July 24 2025, 5:44:51 PM.

FINISHED    

```
%jdbc
-- Mostrar un conteo de todos los docentes hombres y mujeres ordenados por año
SELECT
  `AÑO`,
  ...
```

