```
# Instalar SDK Java 8

!apt-get install openjdk-8-jdk-headless -qq > /dev/null

# Descargar Spark 3.2.2

!wget -q https://archive.apache.org/dist/spark/spark-3.2.3/spark-3.2.3-bin-hadoop3.2.tgz

# Descomprimir el archivo descargado de Spark

!tar xf spark-3.2.3-bin-hadoop3.2.tgz

# Establecer las variables de entorno

import os

os.environ["JAVA_HOME"] = "/usr/lib/jvm/java-8-openjdk-amd64"
os.environ["SPARK_HOME"] = "/content/spark-3.2.3-bin-hadoop3.2"

# Instalar la librería findspark

!pip install -q findspark

# Instalar pyspark

!pip install -q pyspark
```

```
         ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 281.4/281.4 MB 4.9 MB/s eta 0:00:00
      Preparing metadata (setup.py) ... done
         ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 199.7/199.7 KB 18.0 MB/s eta 0:00:00
      Building wheel for pyspark (setup.py) ... done
```

```
#Inicio una sesion de spark

import findspark
findspark.init()
from pyspark.sql import SparkSession

spark = SparkSession.builder.getOrCreate()
sc = spark.sparkContext
```

```
# Crear DataFrames con la funcion .toDF

rdd = sc.parallelize([item for item in range(10)]).map(lambda x: (x, x ** 2))

rdd.collect()

df = rdd.toDF(['numero', 'cudrado'])

df.printSchema()

df.show()
```

```
root
 |-- numero: long (nullable = true)
 |-- cudrado: long (nullable = true)

+------+-------+
|numero|cudrado|
+------+-------+
|     0|      0|
|     1|      1|
|     2|      4|
|     3|      9|
|     4|     16|
|     5|     25|
|     6|     36|
|     7|     49|
|     8|     64|
|     9|     81|
+------+-------+
# Crear un DataFrame a partir de un RDD con schema

rdd1 = sc.parallelize([(1, 'Jose', 35.5), (2, 'Teresa', 54.3), (3, 'Katia', 12.7)])

from pyspark.sql.types import StructType, StructField, IntegerType, StringType, DoubleType



# Primera vía para crear la estructura

esquema1 = StructType(
    [
     StructField('id', IntegerType(), True),
     StructField('nombre', StringType(), True),
     StructField('saldo', DoubleType(), True)
    ]
)


# Segunda vía

esquema2 = "`id` INT, `nombre` STRING, `saldo` DOUBLE"

df1 = spark.createDataFrame(rdd1, schema=esquema1)

df1.printSchema()

df1.show()

df2 = spark.createDataFrame(rdd1, schema=esquema2)

df2.printSchema()

df2.show()
```

```
root
 |-- id: integer (nullable = true)
 |-- nombre: string (nullable = true)
 |-- saldo: double (nullable = true)

+---+------+-----+
| id|nombre|saldo|
+---+------+-----+
|  1|  Jose| 35.5|
|  2|Teresa| 54.3|
|  3| Katia| 12.7|
+---+------+-----+

root
 |-- id: integer (nullable = true)
 |-- nombre: string (nullable = true)
 |-- saldo: double (nullable = true)

+---+------+-----+
| id|nombre|saldo|
+---+------+-----+
|  1|  Jose| 35.5|
|  2|Teresa| 54.3|
|  3| Katia| 12.7|
+---+------+-----+
```

```
# Crear un DataFrame a partir de un rango de números

spark.range(5).toDF('id').show()

spark.range(3, 15).toDF('id').show()

spark.range(0, 20, 2).toDF('id').show()
```

```
+---+
| id|
+---+
|  0|
|  1|
|  2|
|  3|
|  4|
+---+

+---+
| id|
+---+
|  3|
|  4|
|  5|
|  6|
|  7|
|  8|
|  9|
| 10|
| 11|
| 12|
| 13|
| 14|
+---+

+---+
| id|
+---+
|  0|
|  2|
|  4|
|  6|
|  8|
| 10|
| 12|
| 14|
| 16|
| 18|
+---+
```

```
#Antes de nada creo una carpeta nueva, llamada data, en la que cargare los txt
#Se tiene que hacer siempre que se actualice el colab
```

```
# Crear un DataFrame mediante la lectura de un archivo de texto

df = spark.read.text('./data/dataTXT.txt')

df.show()
```

```
+--------------------+
|               value|
+--------------------+
|Estamos en el cur...|
|En este capítulo ...|
|En esta sección e...|
|y en este ejemplo...|
+--------------------+
```

```
df.show(truncate=False)
```

```
+-----------------------------------------------------------------------+
|value                                                                  |
+-----------------------------------------------------------------------+
|Estamos en el curso de pyspark                                         |
|En este capítulo estamos estudiando el API SQL de Saprk                |
|En esta sección estamos creado dataframes a partir de fuentes de datos,|
|y en este ejemplo creamos un dataframe a partir de un texto plano      |
+-----------------------------------------------------------------------+
```

```
# Crear un DataFrame mediante la lectura de un archivo csv

df1 = spark.read.csv('./data/dataCSV.csv')

df1.show()
```

```
+-----------+-------------+--------------------+--------------------+-----------+--------------------+--------------------+-------
|        _c0|          _c1|                 _c2|                 _c3|        _c4|                 _c5|                 _c6|    _c7
+-----------+-------------+--------------------+--------------------+-----------+--------------------+--------------------+-------
|   video_id|trending_date|               title|       channel_title|category_id|        publish_time|                tags|  views
|2kyS6SvSYSE|     17.14.11|WE WANT TO TALK A...|        CaseyNeistat|         22|2017-11-13T17:13:...|     SHANtell martin| 748374
|1ZAPwfrtAFY|     17.14.11|The Trump Preside...|     LastWeekTonight|         24|2017-11-13T07:30:...|"last week tonigh...|2418783
|5qpjK5DgCt4|     17.14.11|Racist Superman |...|        Rudy Mancuso|         23|2017-11-12T19:05:...|"racist superman"...|3191434
|puqaWrEC7tY|     17.14.11|Nickelback Lyrics...|Good Mythical Mor...|         24|2017-11-13T11:00:...|"rhett and link"|...| 343168
|d380meD0W0M|     17.14.11|I Dare You: GOING...|            nigahiga|         24|2017-11-12T18:01:...|"ryan"|"higa"|"hi...|2095731
|gHZ1Qz0KiKM|     17.14.11|2 Weeks with iPho...|            iJustine|         28|2017-11-13T19:07:...|"ijustine"|"week ...| 119180
|39idVpFF7NQ|     17.14.11|Roy Moore & Jeff ...| Saturday Night Live|         24|2017-11-12T05:37:...|"SNL"|"Saturday N...|2103417
|nc99ccSXST0|     17.14.11|5 Ice Cream Gadge...|  CrazyRussianHacker|         28|2017-11-12T21:50:...|"5 Ice Cream Gadg...| 817732
|jr9QtXwC9vc|     17.14.11|The Greatest Show...|    20th Century Fox|          1|2017-11-13T14:00:...|"Trailer"|"Hugh J...| 826059
|TUmyygCMMGA|     17.14.11|Why the rise of t...|                 Vox|         25|2017-11-13T13:45:...|"vox.com"|"vox"|"...| 256426
|9wRQljFNDW8|     17.14.11|Dion Lewis' 103-Y...|                 NFL|         17|2017-11-13T02:05:...|"NFL"|"Football"|...|  81377
|VifQlJit6A0|     17.14.11|(SPOILERS) 'Shiva...|                 amc|         24|2017-11-13T03:00:...|"The Walking Dead...| 104578
```

```
|5E4ZBSInqUU|     17.14.11|Marshmello - Bloc...|          marshmello|         10|2017-11-13T17:00:...|"marshmello"|"blo...| 687582
|GgVmn66oK_A|     17.14.11|Which Countries A...|       NowThis World|         25|2017-11-12T14:00:...|"nowthis"|"nowthi...| 544770
|TaTleo4cOs8|     17.14.11|SHOPPING FOR NEW ...|     The king of DIY|         15|2017-11-12T18:30:...|"shopping for new...| 207532
|kgaO45SyaO4|     17.14.11|    The New SpotMini|      BostonDynamics|         28|2017-11-13T20:09:...|"Robots"|"Boston ...|  75752
|ZAQs-ctOqXQ|     17.14.11|One Change That W...|             Cracked|         23|2017-11-12T17:00:...|"pacific rim"|"pa...| 295639
|YVfyYrEmzgM|     17.14.11|How does your bod...|              TED-Ed|         27|2017-11-13T16:00:...|"TED"|"TED-Ed"|"T...|  78044
|eNSN6qet1kE|     17.14.11|HomeMade Electric...|         PeterSripol|         28|2017-11-13T15:30:...|"ultralight"|"air...|  97007
+-----------+-------------+--------------------+--------------------+-----------+--------------------+--------------------+-------
only showing top 20 rows
```

```
df1 = spark.read.option('header', 'true').csv('./data/dataCSV.csv')

df1.show()
```

```
+-----------+-------------+--------------------+--------------------+-----------+--------------------+--------------------+-------
|   video_id|trending_date|               title|       channel_title|category_id|        publish_time|                tags|  views
+-----------+-------------+--------------------+--------------------+-----------+--------------------+--------------------+-------
|2kyS6SvSYSE|     17.14.11|WE WANT TO TALK A...|        CaseyNeistat|         22|2017-11-13T17:13:...|     SHANtell martin| 748374
|1ZAPwfrtAFY|     17.14.11|The Trump Preside...|     LastWeekTonight|         24|2017-11-13T07:30:...|"last week tonigh...|2418783
|5qpjK5DgCt4|     17.14.11|Racist Superman |...|        Rudy Mancuso|         23|2017-11-12T19:05:...|"racist superman"...|3191434
|puqaWrEC7tY|     17.14.11|Nickelback Lyrics...|Good Mythical Mor...|         24|2017-11-13T11:00:...|"rhett and link"|...| 343168
|d380meD0W0M|     17.14.11|I Dare You: GOING...|            nigahiga|         24|2017-11-12T18:01:...|"ryan"|"higa"|"hi...|2095731
|gHZ1Qz0KiKM|     17.14.11|2 Weeks with iPho...|            iJustine|         28|2017-11-13T19:07:...|"ijustine"|"week ...| 119180
|39idVpFF7NQ|     17.14.11|Roy Moore & Jeff ...| Saturday Night Live|         24|2017-11-12T05:37:...|"SNL"|"Saturday N...|2103417
|nc99ccSXST0|     17.14.11|5 Ice Cream Gadge...|  CrazyRussianHacker|         28|2017-11-12T21:50:...|"5 Ice Cream Gadg...| 817732
|jr9QtXwC9vc|     17.14.11|The Greatest Show...|    20th Century Fox|          1|2017-11-13T14:00:...|"Trailer"|"Hugh J...| 826059
|TUmyygCMMGA|     17.14.11|Why the rise of t...|                 Vox|         25|2017-11-13T13:45:...|"vox.com"|"vox"|"...| 256426
|9wRQljFNDW8|     17.14.11|Dion Lewis' 103-Y...|                 NFL|         17|2017-11-13T02:05:...|"NFL"|"Football"|...|  81377
|VifQlJit6A0|     17.14.11|(SPOILERS) 'Shiva...|                 amc|         24|2017-11-13T03:00:...|"The Walking Dead...| 104578
|5E4ZBSInqUU|     17.14.11|Marshmello - Bloc...|          marshmello|         10|2017-11-13T17:00:...|"marshmello"|"blo...| 687582
|GgVmn66oK_A|     17.14.11|Which Countries A...|       NowThis World|         25|2017-11-12T14:00:...|"nowthis"|"nowthi...| 544770
|TaTleo4cOs8|     17.14.11|SHOPPING FOR NEW ...|     The king of DIY|         15|2017-11-12T18:30:...|"shopping for new...| 207532
|kgaO45SyaO4|     17.14.11|    The New SpotMini|      BostonDynamics|         28|2017-11-13T20:09:...|"Robots"|"Boston ...|  75752
|ZAQs-ctOqXQ|     17.14.11|One Change That W...|             Cracked|         23|2017-11-12T17:00:...|"pacific rim"|"pa...| 295639
|YVfyYrEmzgM|     17.14.11|How does your bod...|              TED-Ed|         27|2017-11-13T16:00:...|"TED"|"TED-Ed"|"T...|  78044
|eNSN6qet1kE|     17.14.11|HomeMade Electric...|         PeterSripol|         28|2017-11-13T15:30:...|"ultralight"|"air...|  97007
|B5HORANmzHw|     17.14.11|Founding An Inbre...|             SciShow|         27|2017-11-12T22:00:...|"SciShow"|"scienc...| 223871
+-----------+-------------+--------------------+--------------------+-----------+--------------------+--------------------+-------
only showing top 20 rows
```

```
# Leer un archivo de texto con un delimitador diferente y con cabecera

df2 = spark.read.option('header', 'true').option('delimiter', '|').csv('./data/dataTab.txt')

df2.show()
```

```
+----+----+----------+-----+
|pais|edad|     fecha|color|
+----+----+----------+-----+
|  MX|  23|2021-02-21| rojo|
|  CA|  56|2021-06-10| azul|
|  US|  32|2020-06-02|verde|
+----+----+----------+-----+
```

```
# Crear un DataFrame a partir de un json proporcionando un schema

from pyspark.sql.types import StructType, StructField, StringType, IntegerType, DateType #importo las clases que voy a necisitar

json_schema =  StructType(
    [
     StructField('color', StringType(), True),
     StructField('edad', IntegerType(), True),
     StructField('fecha', DateType(), True),
     StructField('pais', StringType(), True)
    ]
)

df4 = spark.read.schema(json_schema).json('./data/dataJSON.json')

df4.show()
```

```
+-----+----+----------+----+
|color|edad|     fecha|pais|
+-----+----+----------+----+
| rojo|null|2021-02-21|  MX|
| azul|null|2021-06-10|  CA|
|verde|null|2020-06-02|  US|
+-----+----+----------+----+
```

```
df4.printSchema()
```

```
root
 |-- color: string (nullable = true)
 |-- edad: integer (nullable = true)
 |-- fecha: date (nullable = true)
 |-- pais: string (nullable = true)
```

```
# Crear un DataFrame a partir de un archivo parquet

df5 = spark.read.parquet('./data/dataPARQUET.parquet')

df5.show()
```

```
+-----------+-------------+--------------------+--------------------+-----------+--------------------+--------------------+-------
|   video_id|trending_date|               title|       channel_title|category_id|        publish_time|                tags|  views
+-----------+-------------+--------------------+--------------------+-----------+--------------------+--------------------+-------
|2kyS6SvSYSE|     17.14.11|WE WANT TO TALK A...|        CaseyNeistat|         22|2017-11-13T17:13:...|     SHANtell martin| 748374
|1ZAPwfrtAFY|     17.14.11|The Trump Preside...|     LastWeekTonight|         24|2017-11-13T07:30:...|"last week tonigh...|2418783
|5qpjK5DgCt4|     17.14.11|Racist Superman |...|        Rudy Mancuso|         23|2017-11-12T19:05:...|"racist superman"...|3191434
|puqaWrEC7tY|     17.14.11|Nickelback Lyrics...|Good Mythical Mor...|         24|2017-11-13T11:00:...|"rhett and link"|...| 343168
|d380meD0W0M|     17.14.11|I Dare You: GOING...|            nigahiga|         24|2017-11-12T18:01:...|"ryan"|"higa"|"hi...|2095731
|gHZ1Qz0KiKM|     17.14.11|2 Weeks with iPho...|            iJustine|         28|2017-11-13T19:07:...|"ijustine"|"week ...| 119180
|39idVpFF7NQ|     17.14.11|Roy Moore & Jeff ...| Saturday Night Live|         24|2017-11-12T05:37:...|"SNL"|"Saturday N...|2103417
|nc99ccSXST0|     17.14.11|5 Ice Cream Gadge...|  CrazyRussianHacker|         28|2017-11-12T21:50:...|"5 Ice Cream Gadg...| 817732
|jr9QtXwC9vc|     17.14.11|The Greatest Show...|    20th Century Fox|          1|2017-11-13T14:00:...|"Trailer"|"Hugh J...| 826059
|TUmyygCMMGA|     17.14.11|Why the rise of t...|                 Vox|         25|2017-11-13T13:45:...|"vox.com"|"vox"|"...| 256426
|9wRQljFNDW8|     17.14.11|Dion Lewis' 103-Y...|                 NFL|         17|2017-11-13T02:05:...|"NFL"|"Football"|...|  81377
|VifQlJit6A0|     17.14.11|(SPOILERS) 'Shiva...|                 amc|         24|2017-11-13T03:00:...|"The Walking Dead...| 104578
|5E4ZBSInqUU|     17.14.11|Marshmello - Bloc...|          marshmello|         10|2017-11-13T17:00:...|"marshmello"|"blo...| 687582
|GgVmn66oK_A|     17.14.11|Which Countries A...|       NowThis World|         25|2017-11-12T14:00:...|"nowthis"|"nowthi...| 544770
|TaTleo4cOs8|     17.14.11|SHOPPING FOR NEW ...|     The king of DIY|         15|2017-11-12T18:30:...|"shopping for new...| 207532
|kgaO45SyaO4|     17.14.11|    The New SpotMini|      BostonDynamics|         28|2017-11-13T20:09:...|"Robots"|"Boston ...|  75752
|ZAQs-ctOqXQ|     17.14.11|One Change That W...|             Cracked|         23|2017-11-12T17:00:...|"pacific rim"|"pa...| 295639
|YVfyYrEmzgM|     17.14.11|How does your bod...|              TED-Ed|         27|2017-11-13T16:00:...|"TED"|"TED-Ed"|"T...|  78044
|eNSN6qet1kE|     17.14.11|HomeMade Electric...|         PeterSripol|         28|2017-11-13T15:30:...|"ultralight"|"air...|  97007
|B5HORANmzHw|     17.14.11|Founding An Inbre...|             SciShow|         27|2017-11-12T22:00:...|"SciShow"|"scienc...| 223871
+-----------+-------------+--------------------+--------------------+-----------+--------------------+--------------------+-------
only showing top 20 rows
```

```
# Otra alternativa para leer desde una fuente de datos parquet en este caso

df6 = spark.read.format('parquet').load('./data/dataPARQUET.parquet')

df6.printSchema()
```

```
root
 |-- video_id: string (nullable = true)
 |-- trending_date: string (nullable = true)
 |-- title: string (nullable = true)
 |-- channel_title: string (nullable = true)
 |-- category_id: string (nullable = true)
 |-- publish_time: string (nullable = true)
 |-- tags: string (nullable = true)
 |-- views: string (nullable = true)
 |-- likes: string (nullable = true)
 |-- dislikes: string (nullable = true)
 |-- comment_count: string (nullable = true)
 |-- thumbnail_link: string (nullable = true)
 |-- comments_disabled: string (nullable = true)
 |-- ratings_disabled: string (nullable = true)
 |-- video_error_or_removed: string (nullable = true)
 |-- description: string (nullable = true)
```

```
# Primera alternativa para referirnos a las columnas

df5.select('title').show()
```

```
+--------------------+
|               title|
+--------------------+
|WE WANT TO TALK A...|
|The Trump Preside...|
|Racist Superman |...|
|Nickelback Lyrics...|
|I Dare You: GOING...|
|2 Weeks with iPho...|
|Roy Moore & Jeff ...|
|5 Ice Cream Gadge...|
|The Greatest Show...|
|Why the rise of t...|
|Dion Lewis' 103-Y...|
|(SPOILERS) 'Shiva...|
|Marshmello - Bloc...|
|Which Countries A...|
|SHOPPING FOR NEW ...|
|    The New SpotMini|
|One Change That W...|
|How does your bod...|
|HomeMade Electric...|
|Founding An Inbre...|
+--------------------+
only showing top 20 rows
```

```
# Segunda alternativa para referirnos a las columnas

from pyspark.sql.functions import col

df5.select(col('title')).show()
```

```
+--------------------+
|               title|
+--------------------+
|WE WANT TO TALK A...|
|The Trump Preside...|
|Racist Superman |...|
|Nickelback Lyrics...|
|I Dare You: GOING...|
|2 Weeks with iPho...|
|Roy Moore & Jeff ...|
|5 Ice Cream Gadge...|
|The Greatest Show...|
|Why the rise of t...|
|Dion Lewis' 103-Y...|
|(SPOILERS) 'Shiva...|
|Marshmello - Bloc...|
|Which Countries A...|
|SHOPPING FOR NEW ...|
|    The New SpotMini|
|One Change That W...|
|How does your bod...|
|HomeMade Electric...|
|Founding An Inbre...|
+--------------------+
only showing top 20 rows
```

```
# select

df·=·spark.read.parquet('./data/datos.parquet')

df.printSchema()

from pyspark.sql.functions import col

df.select(col('video_id')).show()

df.select('video_id', 'trending_date').show()
```

```
 |-- likes: integer (nullable = true)
 |-- dislikes: integer (nullable = true)
 |-- comment_count: integer (nullable = true)
 |-- thumbnail_link: string (nullable = true)
 |-- comments_disabled: string (nullable = true)
 |-- ratings_disabled: string (nullable = true)
 |-- video_error_or_removed: string (nullable = true)
 |-- description: string (nullable = true)

+-----------+
|nc99ccSXST0|
|jr9QtXwC9vc|
|TUmyygCMMGA|
|9wRQljFNDW8|
|VifQlJit6A0|
|5E4ZBSInqUU|
|GgVmn66oK_A|
|TaTleo4cOs8|
|kgaO45SyaO4|
|ZAQs-ctOqXQ|
|YVfyYrEmzgM|
|eNSN6qet1kE|
|B5HORANmzHw|
+-----------+
only showing top 20 rows

+-----------+-------------+
|   video_id|trending_date|
+-----------+-------------+
|2kyS6SvSYSE|     17.14.11|
|1ZAPwfrtAFY|     17.14.11|
|5qpjK5DgCt4|     17.14.11|
|puqaWrEC7tY|     17.14.11|
|d380meD0W0M|     17.14.11|
|gHZ1Qz0KiKM|     17.14.11|
|39idVpFF7NQ|     17.14.11|
|nc99ccSXST0|     17.14.11|
|jr9QtXwC9vc|     17.14.11|
|TUmyygCMMGA|     17.14.11|
|9wRQljFNDW8|     17.14.11|
|VifQlJit6A0|     17.14.11|
|5E4ZBSInqUU|     17.14.11|
|GgVmn66oK_A|     17.14.11|
|TaTleo4cOs8|     17.14.11|
|kgaO45SyaO4|     17.14.11|
|ZAQs-ctOqXQ|     17.14.11|
|YVfyYrEmzgM|     17.14.11|
|eNSN6qet1kE|     17.14.11|
|B5HORANmzHw|     17.14.11|
```

```
#select con una expresión nueva
df.select(
    col('likes'),
    col('dislikes'),
    (col('likes') - col('dislikes')).alias('aceptacion')
).show()
```

```
+------+--------+----------+
| likes|dislikes|aceptacion|
+------+--------+----------+
| 57527|    2966|     54561|
| 97185|    6146|     91039|
|146033|    5339|    140694|
| 10172|     666|      9506|
|132235|    1989|    130246|
|  9763|     511|      9252|
| 15993|    2445|     13548|
| 23663|     778|     22885|
|  3543|     119|      3424|
| 12654|    1363|     11291|
|   655|      25|       630|
|  1576|     303|      1273|
|114188|    1333|    112855|
|  7848|    1171|      6677|
|  7473|     246|      7227|
|  9419|      52|      9367|
|  8011|     638|      7373|
|  5398|      53|      5345|
| 11963|      36|     11927|
|  8421|     191|      8230|
+------+--------+----------+
only showing top 20 rows
```

```
# selectExpr

df.selectExpr('likes', 'dislikes', '(likes - dislikes) as aceptacion').show()
df.selectExpr("count(distinct(video_id)) as videos").show()
```

```
+------+--------+----------+
| likes|dislikes|aceptacion|
+------+--------+----------+
| 57527|    2966|     54561|
| 97185|    6146|     91039|
|146033|    5339|    140694|
| 10172|     666|      9506|
|132235|    1989|    130246|
|  9763|     511|      9252|
| 15993|    2445|     13548|
| 23663|     778|     22885|
|  3543|     119|      3424|
| 12654|    1363|     11291|
|   655|      25|       630|
|  1576|     303|      1273|
|114188|    1333|    112855|
|  7848|    1171|      6677|
|  7473|     246|      7227|
|  9419|      52|      9367|
|  8011|     638|      7373|
|  5398|      53|      5345|
| 11963|      36|     11927|
|  8421|     191|      8230|
+------+--------+----------+
only showing top 20 rows

+------+
|videos|
+------+
|  6837|
+------+
```

```
# filter
from pyspark.sql.functions import col

df.show()
```

```
+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
|   video_id|trending_date|               title|       channel_title|category_id|       publish_time|                tags|  views|
+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
|2kyS6SvSYSE|     17.14.11|WE WANT TO TALK A...|        CaseyNeistat|         22|2017-11-13 17:13:01|     SHANtell martin| 748374|
|1ZAPwfrtAFY|     17.14.11|The Trump Preside...|     LastWeekTonight|         24|2017-11-13 07:30:00|"last week tonigh...|2418783|
|5qpjK5DgCt4|     17.14.11|Racist Superman |...|        Rudy Mancuso|         23|2017-11-12 19:05:24|"racist superman"...|3191434|1
|puqaWrEC7tY|     17.14.11|Nickelback Lyrics...|Good Mythical Mor...|         24|2017-11-13 11:00:04|"rhett and link"|...| 343168|
|d380meD0W0M|     17.14.11|I Dare You: GOING...|            nigahiga|         24|2017-11-12 18:01:41|"ryan"|"higa"|"hi...|2095731|1
|gHZ1Qz0KiKM|     17.14.11|2 Weeks with iPho...|            iJustine|         28|2017-11-13 19:07:23|"ijustine"|"week ...| 119180|
|39idVpFF7NQ|     17.14.11|Roy Moore & Jeff ...| Saturday Night Live|         24|2017-11-12 05:37:17|"SNL"|"Saturday N...|2103417|
|nc99ccSXST0|     17.14.11|5 Ice Cream Gadge...|  CrazyRussianHacker|         28|2017-11-12 21:50:37|"5 Ice Cream Gadg...| 817732|
|jr9QtXwC9vc|     17.14.11|The Greatest Show...|    20th Century Fox|          1|2017-11-13 14:00:23|"Trailer"|"Hugh J...| 826059|
|TUmyygCMMGA|     17.14.11|Why the rise of t...|                 Vox|         25|2017-11-13 13:45:16|"vox.com"|"vox"|"...| 256426|
|9wRQljFNDW8|     17.14.11|Dion Lewis' 103-Y...|                 NFL|         17|2017-11-13 02:05:26|"NFL"|"Football"|...|  81377|
|VifQlJit6A0|     17.14.11|(SPOILERS) 'Shiva...|                 amc|         24|2017-11-13 03:00:00|"The Walking Dead...| 104578|
|5E4ZBSInqUU|     17.14.11|Marshmello - Bloc...|          marshmello|         10|2017-11-13 17:00:00|"marshmello"|"blo...| 687582|1
|GgVmn66oK_A|     17.14.11|Which Countries A...|        NowThis World|         25|2017-11-12 14:00:00|"nowthis"|"nowthi...| 544770|
|TaTleo4cOs8|     17.14.11|SHOPPING FOR NEW ...|     The king of DIY|         15|2017-11-12 18:30:01|"shopping for new...| 207532|
|kgaO45SyaO4|     17.14.11|    The New SpotMini|      BostonDynamics|         28|2017-11-13 20:09:58|"Robots"|"Boston ...|  75752|
|ZAQs-ctOqXQ|     17.14.11|One Change That W...|             Cracked|         23|2017-11-12 17:00:05|"pacific rim"|"pa...| 295639|
|YVfyYrEmzgM|     17.14.11|How does your bod...|              TED-Ed|         27|2017-11-13 16:00:07|"TED"|"TED-Ed"|"T...|  78044|
|eNSN6qet1kE|     17.14.11|HomeMade Electric...|          PeterSripol|         28|2017-11-13 15:30:17|"ultralight"|"air...|  97007|
|B5HORANmzHw|     17.14.11|Founding An Inbre...|             SciShow|         27|2017-11-12 22:00:01|"SciShow"|"scienc...| 223871|
+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
only showing top 20 rows
```

```
# filter por id
df.filter(col('video_id') == '2kyS6SvSYSE').show()
```

```
+-----------+-------------+--------------------+-------------+-----------+-------------------+---------------+-------+-----+------
|   video_id|trending_date|               title|channel_title|category_id|       publish_time|           tags|  views|likes|dislike
+-----------+-------------+--------------------+-------------+-----------+-------------------+---------------+-------+-----+------
|2kyS6SvSYSE|     17.14.11|WE WANT TO TALK A...| CaseyNeistat|         22|2017-11-13 17:13:01|SHANtell martin| 748374|57527|    296
|2kyS6SvSYSE|     17.15.11|WE WANT TO TALK A...| CaseyNeistat|         22|2017-11-13 17:13:01|SHANtell martin|2188590|88099|    715
|2kyS6SvSYSE|     17.16.11|WE WANT TO TALK A...| CaseyNeistat|         22|2017-11-13 17:13:01|SHANtell martin|2325233|91111|    754
|2kyS6SvSYSE|     17.17.11|WE WANT TO TALK A...| CaseyNeistat|         22|2017-11-13 17:13:01|SHANtell martin|2400741|92831|    768
|2kyS6SvSYSE|     17.18.11|WE WANT TO TALK A...| CaseyNeistat|         22|2017-11-13 17:13:01|SHANtell martin|2468267|94303|    786
|2kyS6SvSYSE|     17.19.11|WE WANT TO TALK A...| CaseyNeistat|         22|2017-11-13 17:13:01|SHANtell martin|2524854|95587|    789
|2kyS6SvSYSE|     17.20.11|WE WANT TO TALK A...| CaseyNeistat|         22|2017-11-13 17:13:01|SHANtell martin|2564903|96321|    797
+-----------+-------------+--------------------+-------------+-----------+-------------------+---------------+-------+-----+------
```

```
## filter con where
df1 = spark.read.parquet('./data/datos.parquet').where(col('trending_date') != '17.14.11')

df1.show()
```

```
+--------------------+--------------------+--------------------+--------------------+--------------------+--------------------+----
|            video_id|       trending_date|               title|       channel_title|         category_id|       publish_time|
+--------------------+--------------------+--------------------+--------------------+--------------------+--------------------+----
```

```
|\nCook with confi...|             recipes|              videos| and restaurant g...| dining destinations|                null|
|\nVogue is the au...|      culture trends|     beauty coverage|              videos|     celebrity style|                null|
|\nWIRED is where ...| WIRED explores t...|          innovation| and culture.\n\n...|                null|                null|
|         YvfYK0EEhK4|            17.15.11|Brent Pella - Why...|         Brent Pella|                  23|2017-11-14 15:32:51|"spir
|         cxMvzK2OQTw|            17.15.11|Cards Against Hum...|Cards Against Hum...|                  23|2017-11-14 16:43:11|
|         bAkEd8r7Nnw|            17.15.11|Slow Mo Katana Sw...|    The Slow Mo Guys|                  24|2017-11-14 18:31:20|"slom
|         ItYOdWRo0JY|            17.15.11|Selling My iPhone...|          TechSmartt|                  28|2017-11-14 00:45:15|"iphc
|         5530I_pYjbo|            17.15.11|How I Trained My ...|         JunsKitchen|                  26|2017-11-14 12:06:56|"how'
|         dUMH6DVYskc|            17.15.11|Ten-Year-Old's Fa...|     Attaullah Malik|                  28|2017-11-14 12:02:49|"iPhc
|         gjXrm2Q-te4|            17.15.11|Jimmy Fallon Pays...|The Tonight Show ...|                  23|2017-11-14 04:55:24|"Jimm
|         9SK1I0V6U5c|            17.15.11|Lie Detector | An...|        Anwar Jibawi|                  23|2017-11-14 18:01:03|"lie 
|         VsYmwBOYfW8|            17.15.11|Mean Tweets – Jim...|   Jimmy Kimmel Live|                  23|2017-11-14 08:30:01|"jimm
|         kgaO45SyaO4|            17.15.11|    The New SpotMini|      BostonDynamics|                  28|2017-11-13 20:09:58|"Robc
|         CtBca6H6Teg|            17.15.11|Honest Trailers -...|      Screen Junkies|                   1|2017-11-14 18:00:01|"batm
|         pcWKpGzhgq4|            17.15.11|Jason Momoa Shows...|The Graham Norton...|                  24|2017-11-14 12:13:35|"Grah
|         L3br0klRqF4|            17.15.11|Watch live: Sessi...|     Washington Post|                  25|2017-11-14 21:51:20|"brea
|         9kF9xY74h-E|            17.15.11|Spilling Tea Abou...|         Dolan Twins|                  23|2017-11-14 21:00:47|"Dola
|         p2EY93WfBQk|            17.15.11|Is the Morphe 350...|         Jackie Aina|                  26|2017-11-14 21:03:47|"morp
|         xL_qpDkF5A8|            17.15.11|American Crime St...|           TV Promos|                  24|2017-11-15 01:25:18|"Dona
|         eQVhAN7-IAw|            17.15.11|The Making of a S...|        Taylor Swift|                  10|2017-11-15 03:43:47|"DIRE
+--------------------+--------------------+--------------------+--------------------+--------------------+-------------------+-----
only showing top 20 rows
```

```
df2 = spark.read.parquet('./data/datos.parquet').where(col('likes') > 5000)

df2.filter((col('trending_date') != '17.14.11') & (col('likes') > 7000)).show()

df2.filter(col('trending_date') != '17.14.11').filter(col('likes') > 7000).show()
```

```
+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
|   video_id|trending_date|               title|       channel_title|category_id|       publish_time|                tags|  views|
+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
|YvfYK0EEhK4|     17.15.11|Brent Pella - Why...|         Brent Pella|         23|2017-11-14 15:32:51|"spirit airlines"...| 462490|
|bAkEd8r7Nnw|     17.15.11|Slow Mo Katana Sw...|    The Slow Mo Guys|         24|2017-11-14 18:31:20|"slomo"|"slow"|"m...|1525400|
|ItYOdWRo0JY|     17.15.11|Selling My iPhone...|          TechSmartt|         28|2017-11-14 00:45:15|"iphone x"|"iphon...|1836419|
|5530I_pYjbo|     17.15.11|How I Trained My ...|         JunsKitchen|         26|2017-11-14 12:06:56|"how"|"trained"|"...| 977285|
|gjXrm2Q-te4|     17.15.11|Jimmy Fallon Pays...|The Tonight Show ...|         23|2017-11-14 04:55:24|"Jimmy Fallon"|"T...|1611093|
|9SK1I0V6U5c|     17.15.11|Lie Detector | An...|        Anwar Jibawi|         23|2017-11-14 18:01:03|"lie detector"|"a...| 983365|
|VsYmwBOYfW8|     17.15.11|Mean Tweets – Jim...|   Jimmy Kimmel Live|         23|2017-11-14 08:30:01|"jimmy"|"jimmy ki...|2765121|
|kgaO45SyaO4|     17.15.11|    The New SpotMini|      BostonDynamics|         28|2017-11-13 20:09:58|"Robots"|"Boston ...|3701763|
|CtBca6H6Teg|     17.15.11|Honest Trailers -...|      Screen Junkies|          1|2017-11-14 18:00:01|"batman"|"screenj...| 829132|
|9kF9xY74h-E|     17.15.11|Spilling Tea Abou...|         Dolan Twins|         23|2017-11-14 21:00:47|"Dolan Twins"|"Sp...| 891912|1
|p2EY93WfBQk|     17.15.11|Is the Morphe 350...|         Jackie Aina|         26|2017-11-14 21:03:47|"morphe"|"morphe ...| 175539|
|eQVhAN7-IAw|     17.15.11|The Making of a S...|        Taylor Swift|         10|2017-11-15 03:43:47|"DIRECTV"|"DIRECT...| 125645|
|2kyS6SvSYSE|     17.15.11|WE WANT TO TALK A...|        CaseyNeistat|         22|2017-11-13 17:13:01|     SHANtell martin|2188590|
|Y9nDagqKL7Q|     17.15.11|venting online ab...|               ProZD|          1|2017-11-13 20:47:49|"venting online"|...| 394709|
|DbJ2s_g1oDc|     17.15.11|10 LIFE HACKS YOU...|               REACT|         24|2017-11-13 20:00:01|"easy life hacks"...| 863116|
|1ZAPwfrtAFY|     17.15.11|The Trump Preside...|      LastWeekTonight|         24|2017-11-13 07:30:00|"last week tonigh...|4632016|1
|gHZ1Qz0KiKM|     17.15.11|2 Weeks with iPho...|            iJustine|         28|2017-11-13 19:07:23|"ijustine"|"week ...| 758998|
|jr9QtXwC9vc|     17.15.11|The Greatest Show...|    20th Century Fox|          1|2017-11-13 14:00:23|"Trailer"|"Hugh J...|2671756|
|l4bAoNAx2uo|     17.15.11|American Things E...|The Infographics ...|         27|2017-11-13 17:11:30|"American Things ...| 419965|
|puqaWrEC7tY|     17.15.11|Nickelback Lyrics...|Good Mythical Mor...|         24|2017-11-13 11:00:04|"rhett and link"|...| 772235|
+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
only showing top 20 rows

+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
|   video_id|trending_date|               title|       channel_title|category_id|       publish_time|                tags|  views|
+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
|YvfYK0EEhK4|     17.15.11|Brent Pella - Why...|         Brent Pella|         23|2017-11-14 15:32:51|"spirit airlines"...| 462490|
|bAkEd8r7Nnw|     17.15.11|Slow Mo Katana Sw...|    The Slow Mo Guys|         24|2017-11-14 18:31:20|"slomo"|"slow"|"m...|1525400|
|ItYOdWRo0JY|     17.15.11|Selling My iPhone...|          TechSmartt|         28|2017-11-14 00:45:15|"iphone x"|"iphon...|1836419|
|5530I_pYjbo|     17.15.11|How I Trained My ...|         JunsKitchen|         26|2017-11-14 12:06:56|"how"|"trained"|"...| 977285|
|gjXrm2Q-te4|     17.15.11|Jimmy Fallon Pays...|The Tonight Show ...|         23|2017-11-14 04:55:24|"Jimmy Fallon"|"T...|1611093|
|9SK1I0V6U5c|     17.15.11|Lie Detector | An...|        Anwar Jibawi|         23|2017-11-14 18:01:03|"lie detector"|"a...| 983365|
|VsYmwBOYfW8|     17.15.11|Mean Tweets – Jim...|   Jimmy Kimmel Live|         23|2017-11-14 08:30:01|"jimmy"|"jimmy ki...|2765121|
|kgaO45SyaO4|     17.15.11|    The New SpotMini|      BostonDynamics|         28|2017-11-13 20:09:58|"Robots"|"Boston ...|3701763|
|CtBca6H6Teg|     17.15.11|Honest Trailers -...|      Screen Junkies|          1|2017-11-14 18:00:01|"batman"|"screenj...| 829132|
|9kF9xY74h-E|     17.15.11|Spilling Tea Abou...|         Dolan Twins|         23|2017-11-14 21:00:47|"Dolan Twins"|"Sp...| 891912|1
|p2EY93WfBQk|     17.15.11|Is the Morphe 350...|         Jackie Aina|         26|2017-11-14 21:03:47|"morphe"|"morphe ...| 175539|
|eQVhAN7-IAw|     17.15.11|The Making of a S...|        Taylor Swift|         10|2017-11-15 03:43:47|"DIRECTV"|"DIRECT...| 125645|
|2kyS6SvSYSE|     17.15.11|WE WANT TO TALK A...|        CaseyNeistat|         22|2017-11-13 17:13:01|     SHANtell martin|2188590|
|Y9nDagqKL7Q|     17.15.11|venting online ab...|               ProZD|          1|2017-11-13 20:47:49|"venting online"|...| 394709|
|DbJ2s_g1oDc|     17.15.11|10 LIFE HACKS YOU...|               REACT|         24|2017-11-13 20:00:01|"easy life hacks"...| 863116|
|1ZAPwfrtAFY|     17.15.11|The Trump Preside...|      LastWeekTonight|         24|2017-11-13 07:30:00|"last week tonigh...|4632016|1
|gHZ1Qz0KiKM|     17.15.11|2 Weeks with iPho...|            iJustine|         28|2017-11-13 19:07:23|"ijustine"|"week ...| 758998|
|jr9QtXwC9vc|     17.15.11|The Greatest Show...|    20th Century Fox|          1|2017-11-13 14:00:23|"Trailer"|"Hugh J...|2671756|
|l4bAoNAx2uo|     17.15.11|American Things E...|The Infographics ...|         27|2017-11-13 17:11:30|"American Things ...| 419965|
|puqaWrEC7tY|     17.15.11|Nickelback Lyrics...|Good Mythical Mor...|         24|2017-11-13 11:00:04|"rhett and link"|...| 772235|
+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
only showing top 20 rows
```

```
# Transformaciones - funciones distinct y dropDuplicates
```

```
# distinct

df_sin_duplicados = df.distinct()

print('El conteo del dataframe original es {}'.format(df.count()))
print('El conteo del dataframe sin duplicados es {}'.format(df_sin_duplicados.count()))
```

```
El conteo del dataframe original es 48137
El conteo del dataframe sin duplicados es 41428
```

```
# función dropDuplicates

dataframe = spark.createDataFrame([(1, 'azul', 567), (2, 'rojo', 487), (1, 'azul', 345), (2, 'verde', 783)]).toDF('id', 'color', 'import

dataframe.show()

dataframe.dropDuplicates(['id', 'color']).show()
```

```
+---+-----+-------+
| id|color|importe|
+---+-----+-------+
|  1| azul|    567|
|  2| rojo|    487|
|  1| azul|    345|
|  2|verde|    783|
+---+-----+-------+

+---+-----+-------+
| id|color|importe|
+---+-----+-------+
|  1| azul|    567|
|  2| rojo|    487|
|  2|verde|    783|
+---+-----+-------+
```

```
# Transformaciones - funciones sort y orderBy

#Antes de nada creo mi df borrando duplicados

from pyspark.sql.functions import col

df = (spark.read.parquet('./data/datos.parquet')
    .select(col('likes'), col('views'), col('video_id'), col('dislikes'))
    .dropDuplicates(['video_id'])
)

df.show()
```

```
+------+-------+--------------------+--------+
| likes|  views|            video_id|dislikes|
+------+-------+--------------------+--------+
| 63995|1525400|         bAkEd8r7Nnw|     896|
|   427|   9036|         eijd-yjXY9E|      14|
|  4145| 318249|         npcqBt_e4k0|     110|
|  6669| 203615|         LeWtF5y9-6Q|     136|
|  2166| 104499|         GhcqN2FDAnA|    1066|
| 10834| 160196|         v_CMMWCN5nQ|     162|
| 36068| 962042|         R8WBN3fJmwM|     845|
|   982|  36848|         oKuPJ7zF0_k|       6|
| 26482| 713615|         B3JFSL8AA70|    2443|
|275632|2822642|         f6Egj7ncOi8|    1444|
| 23922| 321885|         8gE6cek7F30|     317|
|    70|  13670|         EdkK29-TWJk|       1|
|  1131| 120802|         8szK9FBpdPI|      92|
| 12355| 294080|         6gFj1XJ6b5o|      80|
|  null|   null|\nhttp://www.Mast...|    null|
| 12070| 233766|         wOFuVNiAJQQ|     117|
| 21067| 210371|         PpElRBQ-yGc|     135|
|  4609| 363194|         q11UD-6XT-8|     955|
|   188|  31145|         IzQwbRdh5Ts|       1|
|  2184|  74090|         IfdihPR__WI|      47|
+------+-------+--------------------+--------+
only showing top 20 rows
```

```
# sort

df.sort('likes').show()
```

```
from pyspark.sql.functions import desc

df.sort(desc('likes')).show() #ordenando por la columna likes
```

```
+-----+-----+--------------------+--------+
|likes|views|            video_id|dislikes|
+-----+-----+--------------------+--------+
| null| null|\nFor more videos...|    null|
| null| null|\nFashion Editor:...|    null|
| null| null|\nAccess Hollywoo...|    null|
| null| null|\nStill haven’t s...|    null|
| null| null|\nhttps://www.you...|    null|
| null| null|Horror Outro ► ht...|    null|
| null| null|\nChapped lips ar...|    null|
| null| null|\nRoar: https://w...|    null|
| null| null|\nThe leading int...|    null|
| null| null|             \nToday|    null|
| null| null|\nONE STRANGE ROC...|    null|
| null| null|\nSNAPCHAT: fishi...|    null|
| null| null|\nInstagram: http...|    null|
| null| null|\nInstagram.com/w...|    null|
| null| null|\n5050 State Hwy....|    null|
| null| null|\nSIGN UP FOR BRA...|    null|
| null| null|\nJames Ambler an...|    null|
| null| null|\nhttp://www.Mast...|    null|
| null| null|\nEver After Tuto...|    null|
| null| null|          \nEvelin 7|    null|
+-----+-----+--------------------+--------+
only showing top 20 rows

+-------+--------+-----------+--------+
|  likes|   views|   video_id|dislikes|
+-------+--------+-----------+--------+
|3880071|39349927|7C2z4GqqS5E|   72707|
|2055137|13945717|kTlv5_Bs8aw|   23888|
|2050527|10695328|OK3GJ0WIQ8s|   14711|
|1956202|10666323|p8npDG2ulKQ|   13966|
|1735895|37736281|6ZfuNTqbHE8|   21969|
|1634124|33523622|2Vv-BfVoq4g|   21082|
|1572997| 7518332|kX0vO4vlJuU|    8113|
|1437859| 5884233|D_6QmL6rExk|    6390|
|1405355|31648454|VYOjWnS4cMY|   51547|
|1401915| 5275672|8O_MwlZ2dEg|    6268|
|1386616|15873034|ffxKSjUwKdU|   40714|
|1366736|16884972|J2HytHu5VBI|   59930|
|1290509| 6416697|2tDKp41nrw8|    4358|
|1207457|13754992|_5d-sQ7Fh5M|  280675|
|1167488| 8041970|oWjxSkJpxFU|  147643|
|1149185|24782158|FlsCjmMhFmw|  483924|
|1111592|38873543|i0p1bmr0EmE|   96407|
|1065777|14089954|dfnCAmr569k|   47839|
| 983693|14820746|tCXGJQYZ9JA|   44254|
| 975715|19716689|QwievZ1Tx-8|    9118|
+-------+--------+-----------+--------+
only showing top 20 rows
```

```
# función orderBy (es una función más relacional)

df.orderBy(col('views')).show()

df.orderBy(col('views').desc()).show()

dataframe = spark.createDataFrame([(1, 'azul', 568), (2, 'rojo', 235), (1, 'azul', 456), (2, 'azul', 783)]).toDF('id', 'color', 'importe

dataframe.show()

dataframe.orderBy(col('color').desc(), col('importe')).show()
```

```
+-----+-----+--------------------+--------+
|likes|views|            video_id|dislikes|
+-----+-----+--------------------+--------+
| null| null|\nIMDB - http://w...|    null|
| null| null|\nThis is the fir...|    null|
| null| null|\nAccess Hollywoo...|    null|
| null| null|\nStill haven’t s...|    null|
| null| null|\nhttps://www.you...|    null|
| null| null|          \nEvelin 7|    null|
| null| null|Horror Outro ► ht...|    null|
| null| null|\nChapped lips ar...|    null|
| null| null|\nRoar: https://w...|    null|
| null| null|\nThe leading int...|    null|
| null| null|             \nToday|    null|
| null| null|\nONE STRANGE ROC...|    null|
| null| null|\nSNAPCHAT: fishi...|    null|
| null| null|\nInstagram: http...|    null|
| null| null|\nInstagram.com/w...|    null|
| null| null|\n5050 State Hwy....|    null|
| null| null|\nFor more videos...|    null|
| null| null|\nJames Ambler an...|    null|
| null| null|\nFashion Editor:...|    null|
| null| null|\nEver After Tuto...|    null|
+-----+-----+--------------------+--------+
only showing top 20 rows

+-------+--------+-----------+--------+
|  likes|   views|   video_id|dislikes|
+-------+--------+-----------+--------+
| 609101|48431654|-BQJo3vK8O8|   52259|
|3880071|39349927|7C2z4GqqS5E|   72707|
|1111592|38873543|i0p1bmr0EmE|   96407|
|1735895|37736281|6ZfuNTqbHE8|   21969|
|1634124|33523622|2Vv-BfVoq4g|   21082|
|1405355|31648454|VYOjWnS4cMY|   51547|
| 850362|27973210|u9Mv98Gr5pY|   26541|
|1149185|24782158|FlsCjmMhFmw|  483924|
| 641546|24421448|U9BwWKXjVaI|   16517|
| 587326|23758250|1J76wN0TPI4|   18799|
|      0|20921796|BhIEIO0vaBE|       0|
| 975715|19716689|QwievZ1Tx-8|    9118|
| 511753|18639195|rRr1qiJRsXk|   15606|
| 754791|18195959|rRzxEiBLQCA|   65326|
| 399200|18184886|vn9mMeWcgoM|   17473|
| 787419|17158531|n1WpP7iowLc|   43420|
|1366736|16884972|J2HytHu5VBI|   59930|
|1386616|15873034|ffxKSjUwKdU|   40714|
| 278743|15006579|yDiXQl7grPQ|   13599|
| 983693|14820746|tCXGJQYZ9JA|   44254|
+-------+--------+-----------+--------+
only showing top 20 rows

+---+-----+-------+
| id|color|importe|
+---+-----+-------+
|  1| azul|    568|
|  2| rojo|    235|
|  1| azul|    456|
```

```
# funcion limit

top_10 = df.orderBy(col('views').desc()).limit(10)

top_10.show()
```

```
+-------+--------+-----------+--------+
|  likes|   views|   video_id|dislikes|
+-------+--------+-----------+--------+
| 609101|48431654|-BQJo3vK8O8|   52259|
|3880071|39349927|7C2z4GqqS5E|   72707|
|1111592|38873543|i0p1bmr0EmE|   96407|
|1735895|37736281|6ZfuNTqbHE8|   21969|
|1634124|33523622|2Vv-BfVoq4g|   21082|
|1405355|31648454|VYOjWnS4cMY|   51547|
| 850362|27973210|u9Mv98Gr5pY|   26541|
|1149185|24782158|FlsCjmMhFmw|  483924|
| 641546|24421448|U9BwWKXjVaI|   16517|
| 587326|23758250|1J76wN0TPI4|   18799|
+-------+--------+-----------+--------+
```

```
# Transformaciones - funciones withColumn y withColumnRenamed

# withColumn

from pyspark.sql.functions import col

df_valoracion = df.withColumn('valoracion', col('likes') - col('dislikes'))

df_valoracion.printSchema()

df_valoracion1 = (df.withColumn('valoracion', col('likes') - col('dislikes'))
                    .withColumn('res_div', col('valoracion') % 10)
)

df_valoracion1.printSchema()

df_valoracion1.select(col('likes'), col('dislikes'), col('valoracion'), col('res_div')).show()
```

```
root
 |-- likes: integer (nullable = true)
 |-- views: integer (nullable = true)
 |-- video_id: string (nullable = true)
 |-- dislikes: integer (nullable = true)
 |-- valoracion: integer (nullable = true)
```

```
root
 |-- likes: integer (nullable = true)
 |-- views: integer (nullable = true)
 |-- video_id: string (nullable = true)
 |-- dislikes: integer (nullable = true)
 |-- valoracion: integer (nullable = true)
 |-- res_div: integer (nullable = true)

+------+--------+----------+-------+
| likes|dislikes|valoracion|res_div|
+------+--------+----------+-------+
| 63995|     896|     63099|      9|
|   427|      14|       413|      3|
|  4145|     110|      4035|      5|
|  6669|     136|      6533|      3|
|  2166|    1066|      1100|      0|
| 10834|     162|     10672|      2|
| 36068|     845|     35223|      3|
|   982|       6|       976|      6|
| 26482|    2443|     24039|      9|
|275632|    1444|    274188|      8|
| 23922|     317|     23605|      5|
|    70|       1|        69|      9|
|  1131|      92|      1039|      9|
| 12355|      80|     12275|      5|
|  null|    null|      null|   null|
| 12070|     117|     11953|      3|
| 21067|     135|     20932|      2|
|  4609|     955|      3654|      4|
|   188|       1|       187|      7|
|  2184|      47|      2137|      7|
+------+--------+----------+-------+
only showing top 20 rows
```

```
# withColumnRenamed

df_renombrado = df.withColumnRenamed('video_id', 'id')

df_renombrado.printSchema()

df_error = df.withColumnRenamed('nombre_que_no_existe', 'otro_nombre') #como esa columna no existe no hace nada, tampoco sale ningún err

df_error.printSchema()
```

```
root
 |-- likes: integer (nullable = true)
 |-- views: integer (nullable = true)
 |-- id: string (nullable = true)
 |-- dislikes: integer (nullable = true)

root
 |-- likes: integer (nullable = true)
 |-- views: integer (nullable = true)
 |-- video_id: string (nullable = true)
 |-- dislikes: integer (nullable = true)
```

```
# Transformaciones - funciones drop, sample y randomSplit

df = spark.read.parquet('./data/datos.parquet')

# drop

df.printSchema()

df_util = df.drop('comments_disabled')

df_util.printSchema()

df_util = df.drop('comments_disabled', 'ratings_disabled', 'thumbnail_link')

df_util.printSchema()

df_util = df.drop('comments_disabled', 'ratings_disabled', 'thumbnail_link', 'cafe')

df_util.printSchema()
```

```
 |-- category_id: string (nullable = true)
 |-- comment_count: integer (nullable = true)
 |-- thumbnail_link: string (nullable = true)
 |-- comments_disabled: string (nullable = true)
 |-- ratings_disabled: string (nullable = true)
 |-- video_error_or_removed: string (nullable = true)
 |-- description: string (nullable = true)

root
 |-- video_id: string (nullable = true)
 |-- trending_date: string (nullable = true)
 |-- title: string (nullable = true)
 |-- channel_title: string (nullable = true)
 |-- category_id: string (nullable = true)
 |-- publish_time: timestamp (nullable = true)
 |-- tags: string (nullable = true)
 |-- views: integer (nullable = true)
 |-- likes: integer (nullable = true)
 |-- dislikes: integer (nullable = true)
 |-- comment_count: integer (nullable = true)
 |-- thumbnail_link: string (nullable = true)
 |-- ratings_disabled: string (nullable = true)
 |-- video_error_or_removed: string (nullable = true)
 |-- description: string (nullable = true)

root
 |-- video_id: string (nullable = true)
 |-- trending_date: string (nullable = true)
 |-- title: string (nullable = true)
 |-- channel_title: string (nullable = true)
 |-- category_id: string (nullable = true)
 |-- publish_time: timestamp (nullable = true)
 |-- tags: string (nullable = true)
 |-- views: integer (nullable = true)
 |-- likes: integer (nullable = true)
 |-- dislikes: integer (nullable = true)
 |-- comment_count: integer (nullable = true)
 |-- video_error_or_removed: string (nullable = true)
 |-- description: string (nullable = true)

root
 |-- video_id: string (nullable = true)
 |-- trending_date: string (nullable = true)
 |-- title: string (nullable = true)
 |-- channel_title: string (nullable = true)
 |-- category_id: string (nullable = true)
 |-- publish_time: timestamp (nullable = true)
 |-- tags: string (nullable = true)
 |-- views: integer (nullable = true)
 |-- likes: integer (nullable = true)
 |-- dislikes: integer (nullable = true)
 |-- comment_count: integer (nullable = true)
 |-- video_error_or_removed: string (nullable = true)
```

```
# sample

df_muestra = df.sample(0.8)

num_filas = df.count()
num_filas_muestra = df_muestra.count()

print('El 80% de filas del dataframe original es {}'.format(num_filas - (num_filas*0.2)))
print('El numero de filas del dataframe muestra es {}'.format(num_filas_muestra))


#Si quisieramos replicar esa misma muestra utilizamos el parámetro seed como en los siguentes ejemplos
df_muestra = df.sample(fraction=0.8, seed=1234)
df_muestra = df.sample(withReplacement=True, fraction=0.8, seed=1234) #Si queremos que tome esas filas del dataframe con remplazo
```

```
El 80% de filas del dataframe original es 38509.6
El numero de filas del dataframe muestra es 38397
```

```
# randomSplit

train, test = df.randomSplit([0.8, 0.2], seed=1234)

train, validation, test = df.randomSplit([0.6, 0.2, 0.2], seed=1234)

print(train.count())

validation.count()

print(test.count())
```

```
28808
9631
```

```
# Trabajo con datos incorrectos o faltantes
df = spark.read.parquet('./data/datos.parquet')

df.count() #cuantos registros tiene el df
```

```
48137
```

```
df.na.drop().count() #Número de registros no null
```

```
40379
```

```
df.na.drop('any').count() #Mismo resultado
```

```
40379
```

```
df.dropna().count() #Mismo resultado
```

```
40379
```

```
df.na.drop(subset=['views']).count() #Número de registros no null en views
```

```
40949
```

```
df.na.drop(subset=['views', 'dislikes']).count() #Número de registros no null en views y en dislikes
```

```
40949
```

```
from pyspark.sql.functions import col

df.orderBy(col('views')).select(col('views'), col('likes'), col('dislikes')).show()

df.fillna(0).orderBy(col('views')).select(col('views'), col('likes'), col('dislikes')).show() #sustituir null por 0

df.fillna(0, subset=['likes', 'dislikes']).orderBy(col('views')).select(col('views'), col('likes'), col('dislikes')).show() #sustituir n
```

```
| null| null|    null|
| null| null|    null|
| null| null|    null|
| null| null|    null|
| null| null|    null|
| null| null|    null|
| null| null|    null|
| null| null|    null|
+-----+-----+--------+
only showing top 20 rows

+-----+-----+--------+
|views|likes|dislikes|
+-----+-----+--------+
|    0|    0|       0|
|    0|    0|       0|
|    0|    0|       0|
|    0|    0|       0|
|    0|    0|       0|
|    0|    0|       0|
|    0|    0|       0|
|    0|    0|       0|
|    0|    0|       0|
|views|likes|dislikes|
+-----+-----+--------+
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
| null|    0|       0|
```

```
# Acciones sobre un dataframe en Spark SQL
df = spark.read.parquet('./data/datos.parquet')
```

```
# show

df.show()

df.show(5)

df.show(5, truncate=False)
```

```
+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
|   video_id|trending_date|               title|       channel_title|category_id|       publish_time|                tags|  views|
+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
|2kyS6SvSYSE|     17.14.11|WE WANT TO TALK A...|        CaseyNeistat|         22|2017-11-13 17:13:01|     SHANtell martin| 748374|
|1ZAPwfrtAFY|     17.14.11|The Trump Preside...|     LastWeekTonight|         24|2017-11-13 07:30:00|"last week tonigh...|2418783|
|5qpjK5DgCt4|     17.14.11|Racist Superman |...|        Rudy Mancuso|         23|2017-11-12 19:05:24|"racist superman"...|3191434|1
|puqaWrEC7tY|     17.14.11|Nickelback Lyrics...|Good Mythical Mor...|         24|2017-11-13 11:00:04|"rhett and link"|...| 343168|
|d380meD0W0M|     17.14.11|I Dare You: GOING...|            nigahiga|         24|2017-11-12 18:01:41|"ryan"|"higa"|"hi...|2095731|1
|gHZ1Qz0KiKM|     17.14.11|2 Weeks with iPho...|            iJustine|         28|2017-11-13 19:07:23|"ijustine"|"week ...| 119180|
|39idVpFF7NQ|     17.14.11|Roy Moore & Jeff ...| Saturday Night Live|         24|2017-11-12 05:37:17|"SNL"|"Saturday N...|2103417|
|nc99ccSXST0|     17.14.11|5 Ice Cream Gadge...| CrazyRussianHacker|         28|2017-11-12 21:50:37|"5 Ice Cream Gadg...| 817732|
|jr9QtXwC9vc|     17.14.11|The Greatest Show...|    20th Century Fox|          1|2017-11-13 14:00:23|"Trailer"|"Hugh J...| 826059|
|TUmyygCMMGA|     17.14.11|Why the rise of t...|                 Vox|         25|2017-11-13 13:45:16|"vox.com"|"vox"|"...| 256426|
|9wRQljFNDW8|     17.14.11|Dion Lewis' 103-Y...|                 NFL|         17|2017-11-13 02:05:26|"NFL"|"Football"|...|  81377|
|VifQlJit6A0|     17.14.11|(SPOILERS) 'Shiva...|                 amc|         24|2017-11-13 03:00:00|"The Walking Dead...| 104578|
|5E4ZBSInqUU|     17.14.11|Marshmello - Bloc...|          marshmello|         10|2017-11-13 17:00:00|"marshmello"|"blo...| 687582|1
|GgVmn66oK_A|     17.14.11|Which Countries A...|       NowThis World|         25|2017-11-12 14:00:00|"nowthis"|"nowthi...| 544770|
|TaTleo4cOs8|     17.14.11|SHOPPING FOR NEW ...|     The king of DIY|         15|2017-11-12 18:30:01|"shopping for new...| 207532|
|kgaO45SyaO4|     17.14.11|    The New SpotMini|     BostonDynamics|         28|2017-11-13 20:09:58|"Robots"|"Boston ...|  75752|
|ZAQs-ctOqXQ|     17.14.11|One Change That W...|             Cracked|         23|2017-11-12 17:00:05|"pacific rim"|"pa...| 295639|
|YVfyYrEmzgM|     17.14.11|How does your bod...|              TED-Ed|         27|2017-11-13 16:00:07|"TED"|"TED-Ed"|"T...|  78044|
|eNSN6qet1kE|     17.14.11|HomeMade Electric...|         PeterSripol|         28|2017-11-13 15:30:17|"ultralight"|"air...|  97007|
|B5HORANmzHw|     17.14.11|Founding An Inbre...|             SciShow|         27|2017-11-12 22:00:01|"SciShow"|"scienc...| 223871|
+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
only showing top 20 rows

+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
|   video_id|trending_date|               title|       channel_title|category_id|       publish_time|                tags|  views|
+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
|2kyS6SvSYSE|     17.14.11|WE WANT TO TALK A...|        CaseyNeistat|         22|2017-11-13 17:13:01|     SHANtell martin| 748374|
|1ZAPwfrtAFY|     17.14.11|The Trump Preside...|     LastWeekTonight|         24|2017-11-13 07:30:00|"last week tonigh...|2418783|
|5qpjK5DgCt4|     17.14.11|Racist Superman |...|        Rudy Mancuso|         23|2017-11-12 19:05:24|"racist superman"...|3191434|1
|puqaWrEC7tY|     17.14.11|Nickelback Lyrics...|Good Mythical Mor...|         24|2017-11-13 11:00:04|"rhett and link"|...| 343168|
|d380meD0W0M|     17.14.11|I Dare You: GOING...|            nigahiga|         24|2017-11-12 18:01:41|"ryan"|"higa"|"hi...|2095731|1
+-----------+-------------+--------------------+--------------------+-----------+-------------------+--------------------+-------+-
only showing top 5 rows

+-----------+-------------+--------------------------------------------------------------+---------------------+-----------+-------
|video_id   |trending_date|title                                                         |channel_title        |category_id|publish
+-----------+-------------+--------------------------------------------------------------+---------------------+-----------+-------
|2kyS6SvSYSE|17.14.11     |WE WANT TO TALK ABOUT OUR MARRIAGE                            |CaseyNeistat         |22         |2017-11
|1ZAPwfrtAFY|17.14.11     |The Trump Presidency: Last Week Tonight with John Oliver (HBO)|LastWeekTonight      |24         |2017-11
|5qpjK5DgCt4|17.14.11     |Racist Superman | Rudy Mancuso, King Bach & Lele Pons         |Rudy Mancuso         |23         |2017-11
|puqaWrEC7tY|17.14.11     |Nickelback Lyrics: Real or Fake?                              |Good Mythical Morning|24         |2017-11
|d380meD0W0M|17.14.11     |I Dare You: GOING BALD!?                                      |nigahiga             |24         |2017-11
+-----------+-------------+--------------------------------------------------------------+---------------------+-----------+-------
only showing top 5 rows
```

```
# take
#La salida será una lista

df.take(1)
```

```
[Row(video_id='2kyS6SvSYSE', trending_date='17.14.11', title='WE WANT TO TALK ABOUT OUR MARRIAGE', channel_title='CaseyNeistat',
category_id='22', publish_time=datetime.datetime(2017, 11, 13, 17, 13, 1), tags='SHANtell martin', views=748374, likes=57527,
dislikes=2966, comment_count=15954, thumbnail_link='https://i.ytimg.com/vi/2kyS6SvSYSE/default.jpg', comments_disabled='False',
ratings_disabled='False', video_error_or_removed='False', description="SHANTELL'S CHANNEL -
https://www.youtube.com/shantellmartin\\nCANDICE - https://www.lovebilly.com\\n\\nfilmed this video in 4k on this --
http://amzn.to/2sTDnRZ\\nwith this lens -- http://amzn.to/2rUJOmD\\nbig drone - http://tinyurl.com/h4ft3oy\\nOTHER GEAR ---
http://amzn.to/2o3GLX5\\nSony CAMERA http://amzn.to/2nOBmnv\\nOLD CAMERA; http://amzn.to/2o2cQBT\\nMAIN LENS;
http://amzn.to/2od5gBJ\\nBIG SONY CAMERA; http://amzn.to/2nrdJRO\\nBIG Canon CAMERA; http://tinyurl.com/jn4q4vz\\nBENDY TRIPOD
THING; http://tinyurl.com/gw3ylz2\\nYOU NEED THIS FOR THE BENDY TRIPOD; http://tinyurl.com/j8mzzua\\nWIDE LENS;
http://tinyurl.com/jkfcm8t\\nMORE EXPENSIVE WIDE LENS; http://tinyurl.com/zrdgtou\\nSMALL CAMERA;
http://tinyurl.com/hrrzhor\\nMICROPHONE; http://tinyurl.com/zefm4jy\\nOTHER MICROPHONE; http://tinyurl.com/jxgpj86\\nOLD DRONE
(cheaper but still great);http://tinyurl.com/zcfmnmd\\n\\nfollow me; on http://instagram.com/caseyneistat\\non
https://www.facebook.com/cneistat\\non https://twitter.com/CaseyNeistat\\n\\namazing intro song by
https://soundcloud.com/discoteeth\\n\\nad disclosure.  THIS IS NOT AN AD.  not selling or promoting anything.  but samsung did
produce the Shantell Video as a 'GALAXY PROJECT' which is an initiative that enables creators like Shantell and me to make
projects we might otherwise not have the opportunity to make.  hope that's clear.  if not ask in the comments and i'll answer any
specifics.")]
```

```
# head

df.head(1)
```

```
[Row(video_id='2kyS6SvSYSE', trending_date='17.14.11', title='WE WANT TO TALK ABOUT OUR MARRIAGE', channel_title='CaseyNeistat',
category_id='22', publish_time=datetime.datetime(2017, 11, 13, 17, 13, 1), tags='SHANtell martin', views=748374, likes=57527,
dislikes=2966, comment_count=15954, thumbnail_link='https://i.ytimg.com/vi/2kyS6SvSYSE/default.jpg', comments_disabled='False',
ratings_disabled='False', video_error_or_removed='False', description="SHANTELL'S CHANNEL -
https://www.youtube.com/shantellmartin\\nCANDICE - https://www.lovebilly.com\\n\\nfilmed this video in 4k on this --
http://amzn.to/2sTDnRZ\\nwith this lens -- http://amzn.to/2rUJOmD\\nbig drone - http://tinyurl.com/h4ft3oy\\nOTHER GEAR ---
http://amzn.to/2o3GLX5\\nSony CAMERA http://amzn.to/2nOBmnv\\nOLD CAMERA; http://amzn.to/2o2cQBT\\nMAIN LENS;
http://amzn.to/2od5gBJ\\nBIG SONY CAMERA; http://amzn.to/2nrdJRO\\nBIG Canon CAMERA; http://tinyurl.com/jn4q4vz\\nBENDY TRIPOD
THING; http://tinyurl.com/gw3ylz2\\nYOU NEED THIS FOR THE BENDY TRIPOD; http://tinyurl.com/j8mzzua\\nWIDE LENS;
http://tinyurl.com/jkfcm8t\\nMORE EXPENSIVE WIDE LENS; http://tinyurl.com/zrdgtou\\nSMALL CAMERA;
http://tinyurl.com/hrrzhor\\nMICROPHONE; http://tinyurl.com/zefm4jy\\nOTHER MICROPHONE; http://tinyurl.com/jxgpj86\\nOLD DRONE
(cheaper but still great);http://tinyurl.com/zcfmnmd\\n\\nfollow me; on http://instagram.com/caseyneistat\\non
https://www.facebook.com/cneistat\\non https://twitter.com/CaseyNeistat\\n\\namazing intro song by
https://soundcloud.com/discoteeth\\n\\nad disclosure.  THIS IS NOT AN AD.  not selling or promoting anything.  but samsung did
produce the Shantell Video as a 'GALAXY PROJECT' which is an initiative that enables creators like Shantell and me to make
projects we might otherwise not have the opportunity to make.  hope that's clear.  if not ask in the comments and i'll answer any
specifics.")]
```

```
# collect
#Riesgo para la memoria

df.select('likes').collect()
```

```
 Row(likes=223821),
 Row(likes=39816),
 Row(likes=6729),
 Row(likes=28309),
 Row(likes=12594),
 Row(likes=None),
 Row(likes=None),
 Row(likes=None),
 Row(likes=None),
 Row(likes=135292),
 Row(likes=88541),
 Row(likes=281095),
 Row(likes=None),
 Row(likes=None),
 Row(likes=None),
 Row(likes=None),
```

```
# Escritura de DataFrames

df1 = df.repartition(2) #creo 2 particiones

df1.write.format('csv').option('sep', '|').save('./output/csv1') #cambio el separador,selecciono la carpeta donde guardar en csv de sali
```

```
#Si quisiera devolverlo a una sola partición
df1.coalesce(1).write.format('csv').option('sep', '|').save('./output/csv2')
```

```
#para guardar el dataframe particionando los datos

df.printSchema()#df original

df.select('comments_disabled').distinct().show() #me centro en la columna comment_disabled

from pyspark.sql.functions import col

df_limpio = df.filter(col('comments_disabled').isin('True', 'False')) #Realizo un filtrado para limpiarlo

df_limpio.write.partitionBy('comments_disabled').parquet('./output/parquet') #los guardo particionando los datos por la columna comment_
#crea tantas particiones como valores diferentes tenga la columna seleccionada
```

```
root
 |-- video_id: string (nullable = true)
 |-- trending_date: string (nullable = true)
 |-- title: string (nullable = true)
 |-- channel_title: string (nullable = true)
 |-- category_id: string (nullable = true)
 |-- publish_time: timestamp (nullable = true)
 |-- tags: string (nullable = true)
 |-- views: integer (nullable = true)
 |-- likes: integer (nullable = true)
 |-- dislikes: integer (nullable = true)
 |-- comment_count: integer (nullable = true)
 |-- thumbnail_link: string (nullable = true)
 |-- comments_disabled: string (nullable = true)
 |-- ratings_disabled: string (nullable = true)
 |-- video_error_or_removed: string (nullable = true)
 |-- description: string (nullable = true)

+-----------------+
|comments_disabled|
+-----------------+
|            False|
|             null|
| sports and more.|
|          Wiz Kid|
|             True|
|         farfalle|
+-----------------+
```

```
# Persistencia de DataFrames


df = spark.createDataFrame([(1, 'a'), (2, 'b'), (3, 'c')], ['id', 'valor'])

df.show()

df.persist() #persistir en memorio

df.unpersist() #para retirar en memoria

df.cache() #Almacenar sólo en memoria
```

```
from pyspark.storagelevel import StorageLevel

df.persist(StorageLevel.DISK_ONLY) #almacenar solo en disco

df.persist(StorageLevel.MEMORY_AND_DISK) #tanto memoria como disco
```

```
+---+-----+
| id|valor|
+---+-----+
|  1|    a|
|  2|    b|
|  3|    c|
+---+-----+

DataFrame[id: bigint, valor: string]
```

✓ 0 s completado a las 12:17