

Introduccion

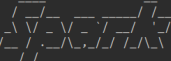
(<https://www.kaggle.com/datasets/hosammhmdali/video-game-sales-2024?resource=download>) donde se muestran las ventas e información de videojuegos.

Primeros pasos

```

hadoopnamenode:~$ pyspark
Python 3.12.3 (main, Jan 17 2025, 18:03:48) [GCC 13.3.0] on linux
Type "help", "copyright", "credits" or "license()" for more information.
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
25/02/20 18:21:01 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
25/02/20 18:21:02 WARN Utils: Service 'SparkUI' could not bind on port 4040. Attempting port 4041.
25/02/20 18:21:02 WARN Utils: Service 'SparkUI' could not bind on port 4041. Attempting port 4042.
25/02/20 18:21:02 WARN Utils: Service 'SparkUI' could not bind on port 4042. Attempting port 4043.
Welcome to

```



```

version 3.5.4

Using Python version 3.12.3 (main, Jan 17 2025 18:03:48)
spark context Web UI available at http://namenode:4043
spark context available as 'sc' (master = local[*], app id = local-1740072062411).
sparkSession available as 'spark'
>>> from pyspark.sql.functions import col, sum, count
>>> df = spark.read.csv("archivos/videogames.csv", header=True, inferSchema=True)
>>> df.show()

```

	ing	title	console	genre	publisher	developer	critic_score	total_sales	na_sales	jp_sales	pa_sales	other_sales	release_date	last_update
/games/boxart/ful...	Grand Theft Auto V	P53	Action	Rockstar Games	Rockstar North		9.4	20.32	6.37	0.99	9.85	3.12	2013-09-17	NULL
/games/boxart/ful...	Grand Theft Auto V	P54	Action	Rockstar Games	Rockstar North		9.7	19.39	6.06	0.6	9.71	3.02	2014-11-18	2018-01-03
/games/boxart/827...	Grand Theft Auto...	P52	Action	Rockstar Games	Rockstar North		9.6	16.15	8.41	0.47	5.49	1.78	2002-10-28	NULL
/games/boxart/ful...	Grand Theft Auto V	X360	Action	Rockstar Games	Rockstar North		NULL	15.86	9.06	0.06	5.33	1.42	2013-09-17	NULL
/games/boxart/ful...	Call of Duty: Bla...	P54	Shooter	Activision	Treyarch		8.1	15.09	6.18	0.41	6.05	2.44	2015-11-06	2018-01-14
/games/boxart/ful...	Call of Duty: Mod...	X360	Shooter	Activision	Infinity Ward		8.7	14.82	9.07	0.13	4.29	1.33	2011-11-08	NULL
/games/boxart/ful...	Call of Duty: Bla...	X360	Shooter	Activision	Treyarch		8.8	14.74	9.76	0.11	3.73	1.14	2010-11-09	NULL
/games/boxart/ful...	Red Dead Redempti...	P54	Action-Adventure	Rockstar Games	Rockstar Games		9.8	13.94	5.26	0.21	6.21	2.26	2018-10-26	2018-11-02
/games/boxart/ful...	Call of Duty: Bla...	X360	Shooter	Activision	Treyarch		8.4	13.86	8.27	0.07	4.32	1.2	2012-11-13	2018-04-07
/games/boxart/ful...	Call of Duty: Bla...	P53	Shooter	Activision	Treyarch		8.0	13.8	4.99	0.65	5.88	2.28	2012-11-13	2018-04-07
/games/boxart/ful...	Call of Duty: Mod...	X360	Shooter	Activision	Infinity Ward		9.5	13.53	8.54	0.08	3.63	1.28	2009-11-10	NULL
/games/boxart/ful...	Call of Duty: WWII	P54	Shooter	Activision	Sledgehammer Games		8.1	13.4	4.67	0.4	6.21	2.12	2017-11-03	2017-12-31
/games/boxart/ful...	Call of Duty: Mod...	P53	Shooter	Activision	Infinity Ward		8.8	13.35	5.54	0.49	5.78	1.54	2011-11-08	NULL
/games/boxart/357...	Grand Theft Auto III	P52	Action	Rockstar Games	DMA Design		9.5	13.1	6.99	0.3	4.51	1.3	2001-10-23	NULL
/games/boxart/ful...	Call of Duty: Bla...	P53	Shooter	Activision	Treyarch		8.7	12.67	6.01	0.48	4.4	1.78	2001-11-09	NULL
/games/boxart/ful...	FIFA 18	P54	Sports	EA Sports	EA Vancouver		8.3	11.8	1.27	0.15	8.64	1.73	2017-09-29	2018-04-02
/games/boxart/ful...	Grand Theft Auto IV	X360	Action	Rockstar Games	Rockstar North		10.0	11.99	1.6	0.12	3.11	1.04	2008-04-29	NULL
/games/boxart/ful...	FIFA 17	P54	Sports	Electronic Arts	EA Canada		8.9	10.94	2.6	0.12	7.95	1.61	2016-09-27	2018-03-28
/games/boxart/ful...	Call of Duty: Mod...	P53	Shooter	Activision	Infinity Ward		9.5	10.61	4.99	0.38	3.66	1.59	2009-11-10	NULL
/games/boxart/ful...	Grand Theft Auto IV	P53	Action	Rockstar Games	Rockstar North		10.0	10.57	4.79	0.44	3.73	1.62	2008-04-29	NULL

```

only showing top 20 rows

```

Juegos con mas ventas (sumando todas las consolas)

```
>>> df.groupby("title").agg(sum("total_sales").alias("Total_Sales")).orderBy(col("Total_Sales").desc()).show(10)
+-----+-----+
|          title|      Total_Sales|
+-----+-----+
| Grand Theft Auto V|          64.29|
|Call of Duty: Bla...|30.990000000000002|
|Call of Duty: Mod...|          30.71|
|Call of Duty: Bla...|29.590000000000003|
|Call of Duty: Ghosts|28.800000000000004|
|Call of Duty: Bla...|          26.72|
|Call of Duty: Mod...|          25.02|
|      Minecraft|24.009999999999998|
| Grand Theft Auto IV|          22.53|
|Call of Duty: Adv...|          21.78|
+-----+-----+
only showing top 10 rows
```

Consolas con más juegos vendidos (suma de ventas por consola)

```
>>> df.groupBy("console").agg(sum("total_sales").alias("Total_Sales")).orderBy(col("Total_Sales").desc()).show(10)
+-----+-----+
|console|    Total_Sales|
+-----+-----+
|    PS2|1027.7599999999927|
|   X360| 859.7899999999987|
|    PS3| 839.6999999999981|
|     PS| 546.2499999999987|
|    PS4| 539.9199999999955|
|    Wii|459.43999999999835|
|     DS|458.16999999999416|
|   XOne| 268.9599999999991|
|    PSP|245.29000000000013|
|     XB|232.0500000000058|
+-----+-----+
only showing top 10 rows
```

Géneros más populares de 2018 (el de más ventas en ese año)

```
>>> df.filter(col("year") == 2018) \
...     .groupBy("genre").agg(sum("total_sales").alias("Total_Sales")) \
...     .orderBy(col("Total_Sales").desc()) \
...     .show(3)
+-----+-----+
|      genre|    Total_Sales|
+-----+-----+
|Action-Adventure|35.10000000000001|
|      Sports|          27.16|
|      Shooter|21.17000000000001|
+-----+-----+
only showing top 3 rows
```