

```
%sh
python -m pip install kaggle
```

FINISHED

Defaulting to user installation because normal site-packages is not writeable

Requirement already satisfied: kaggle in ./local/lib/python3.6/site-packages (1.5.12)

Requirement already satisfied: python-slugify in ./local/lib/python3.6/site-packages (from kaggle) (5.0.2)

Requirement already satisfied: tqdm in ./local/lib/python3.6/site-packages (from kaggle) (4.62.3)

Requirement already satisfied: urllib3 in /usr/lib/python3/dist-packages (from kaggle) (1.22)

Requirement already satisfied: certifi in /usr/lib/python3/dist-packages (from kaggle) (2018.1.18)

Requirement already satisfied: requests in /usr/lib/python3/dist-packages (from kaggle) (2.18.4)

Requirement already satisfied: python-dateutil in ./local/lib/python3.6/site-packages (from kaggle) (2.8.2)

Requirement already satisfied: six>=1.10 in /usr/lib/python3/dist-packages (from kaggle) (1.11.0)

Requirement already satisfied: text-unidecode>=1.3 in ./local/lib/python3.6/site-packages (from python-slugify->kaggle) (1.3)

Took 4 sec. Last updated by anonymous at November 28 2021, 12:35:38 PM.

```
%sh
export KAGGLE_USERNAME=igortolstikov
export KAGGLE_KEY=fb99c3ad6bf931513798d91567033035
mkdir -p /home/hduser/videogame
cd /home/hduser/videogame
kaggle datasets files gregorut/videogamesales
kaggle datasets download gregorut/videogamesales
```

FINISHED

```
name      size  creationDate
-----
vgsales.csv  1MB  2016-10-26 09:10:49

Downloading videogamesales.zip to /home/hduser/videogame
100%#####| 381k/381k [00:00<00:00, 5.79MB/s]
```

Took 3 sec. Last updated by anonymous at November 28 2021, 12:47:56 PM.

```
%sh
cd /home/hduser/videogame
unzip videogamesales.zip
rm videogamesales.zip
ls -la
```

FINISHED

Archive: videogamesales.zip

inflating: vgsales.csv

total 1340

drwxr-xr-x 2 hduser hduser 4096 Nov 28 09:51 .

drwxr-xr-x 1 hduser hduser 4096 Nov 28 09:47 ..

-rw-r--r-- 1 hduser hduser 1355781 Sep 19 2019 vgsales.csv

Took 0 sec. Last updated by anonymous at November 28 2021, 12:51:33 PM.

```
%sh
hdfs dfs -put /home/hduser/videogame /user/hduser
hdfs dfs -ls /user/hduser/videogame
```

FINISHED

Found 1 items

-rw-r--r-- 1 hduser supergroup 1355781 2021-11-28 10:07 /user/hduser/videogame/vgsales.csv

Took 4 sec. Last updated by anonymous at November 28 2021, 1:07:27 PM.

```
%sh
hdfs dfs -cat /user/hduser/videogame/vgsales.csv | head
```

FINISHED

Rank,Name,Platform,Year,Genre,Publisher,NA_Sales,EU_Sales,JP_Sales,Other_Sales,Global_Sales

1,WiiSports: Unable to write to output stream.

Sports,Wii,2006,Sports,Nintendo,41.49,29.02,3.77,8.46,82.74

2,Super Mario Bros.,NES,1985,Platform,Nintendo,29.08,3.58,6.81,0.77,40.24

3,Mario Kart Wii,Wii,2008,Racing,Nintendo,15.85,12.88,3.79,3.31,35.82

4,Wii Sports Resort,Wii,2009,Sports,Nintendo,15.75,11.01,3.28,2.96,33

5,Pokemon Red/Pokemon Blue,GB,1996,Role-Playing,Nintendo,11.27,8.89,10.22,1.31,37

6,Tetris,GB,1989,Puzzle,Nintendo,23.2,2.26,4.22,0.58,30.26

7,New Super Mario Bros.,DS,2006,Platform,Nintendo,11.38,9.23,6.5,2.9,30.01

8,Wii Play,Wii,2006,Misc,Nintendo,14.03,9.2,2.93,2.85,29.02

9,New Super Mario Bros. Wii,Wii,2009,Platform,Nintendo,14.59,7.06,4.7,2.26,28.62

Took 2 sec. Last updated by anonymous at November 28 2021, 1:11:46 PM.

```
%spark.sql
CREATE TABLE if not exists videogame
using csv
options (
  path "/user/hduser/videogame/vgsales.csv",
  header true,
  interSchema true
);
```

SPARK JOB FINISHED

Took 59 sec. Last updated by anonymous at November 28 2021, 1:42:17 PM. (outdated)

```
%spark.sql
SELECT * FROM videogame LIMIT 10;
```

SPARK JOB FINISHED

settings

Rank	名前	プラットフォーム	年	ジャンル	出版者	NA_Sales		
1	Wii Sports	Wii	2006	Sports	Nintendo	41.49		
2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08		
3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85		
4	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.75		
5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27		
6	Tetris	GB	1989	Puzzle	Nintendo	23.2		
7	New Super Mario Bros.	DS	2006	Platform	Nintendo	11.38		

Took 1 sec. Last updated by anonymous at November 28 2021, 1:44:28 PM. (outdated)

3. Выведите самую продаваемую игру за всё время

READY

```
%spark.sql
SELECT * FROM videogame ORDER BY Rank asc LIMIT 1;
```

SPARK JOB FINISHED

settings

Rank	名前	プラットフォーム	年	ジャンル	出版者	NA_Sales		
1	Wii Sports	Wii	2006	Sports	Nintendo	41.49		

Took 0 sec. Last updated by anonymous at November 28 2021, 4:35:30 PM. (outdated)

4. Какая платформа самая популярная в каждом регионе (NA, EU, JP)?

READY

```
%spark.sql
WITH x as (
  SELECT
    Platform,
    sum(NA_Sales) NA_Sum,
    sum(EU_Sales) EU_Sum,
    sum(JP_Sales) JP_Sum
  FROM videogame
  GROUP BY Platform
)
(
  SELECT 'NA' as region, Platform
  FROM x
  ORDER BY NA_Sum desc
  LIMIT 1
)
union
(
  SELECT 'EU' as region, Platform
  FROM x
  ORDER BY EU_Sum desc
  LIMIT 1
)
union
(
  SELECT 'JP' as region, Platform
  FROM x
  ORDER BY JP_Sum desc
  LIMIT 1
)
)
```

SPARK JOB FINISHED

settings

region	プラットフォーム		
JP	DS		
EU	PS3		
NA	X360		

Took 5 sec. Last updated by anonymous at November 28 2021, 4:06:26 PM. (outdated)

5. Какой жанр популярен больше всего в каждом регионе (NA, EU, JP)?

READY

```
%spark.sql
WITH x as (
  SELECT
    Genre,
    sum(NA_Sales) NA_Sum,
    sum(EU_Sales) EU_Sum,
    sum(JP_Sales) JP_Sum
  FROM videogame
  GROUP BY Genre
)
(
  SELECT 'NA' as region, Genre
  FROM x
  ORDER BY NA_Sum desc
  LIMIT 1
)
union
(
  SELECT 'EU' as region, Genre
  FROM x
  ORDER BY EU_Sum desc
  LIMIT 1
)
union
(
  SELECT 'JP' as region, Genre
  FROM x
  ORDER BY JP_Sum desc
  LIMIT 1
)
)
```

SPARK JOB FINISHED

settings

region	ジャンル		
JP	Role-Playing		
NA	Action		
EU	Action		

Took 4 sec. Last updated by anonymous at November 28 2021, 4:14:14 PM. (outdated)

6. Выведите самый популярный жанр на каждый год

READY

```
%spark.sql
WITH x as (
  SELECT
    Genre,
    Year,
    sum(Global_Sales) Global_Sales,
    row_number() over(partition by Year ORDER BY sum(Global_Sales) desc) as row_number
  FROM videogame
  WHERE Year != 'N/A'
  GROUP BY Year, Genre
  ORDER BY Year, Global_Sales desc
)
SELECT Year, Genre
FROM x
WHERE row_number == 1;
```

SPARK JOB FINISHED

settings

Year	ジャンル		
1980	Shooter		
1981	Action		
1982	Puzzle		
1983	Platform		
1984	Shooter		
1985	Platform		
1986	Action		
1987	Fighting		
1988	Platform		
1989	Puzzle		
1990	Platform		
1991	Platform		
1992	Fighting		
1993	Platform		
1994	Platform		
1995	Platform		
1996	Role-Playing		
1997	Racing		
1998	Sports		
1999	Role-Playing		
2000	Sports		
2001	Action		
2002	Action		
2003	Action		
2004	Action		
2005	Action		
2006	Sports		
2007	Action		
2008	Action		

Took 3 sec. Last updated by anonymous at November 28 2021, 4:32:19 PM. (outdated)

%spark.sql

READY