

[Strona główna](#) / [Moje kursy](#) / [WliT](#) / [Informatyka](#) / [Stacjonarne](#) / [I stopień](#) / [Semestr 7 \[WliT-Inf-st-I\]](#)
/ [Przetwarzanie maszynowych danych - Big Data 23/24-Z](#) / [7. Spark SQL - DataFrame](#)
/ [Spark - DataFrame \(Python\) - zadania - zarejestruj swoje rozwiązania tutaj](#)

Rozpoczęto piątek, 24 listopada 2023, 09:15

Stan Ukończone

Ukończono środa, 29 listopada 2023, 21:50

Wykorzystany czas 5 dni 12 godzin

Pytanie **1**

Poprawnie

Punkty maks.: 7,00

Zad 09

Na początek coś prostego. Wyświetl trzy najlepiej ocenione gry wydane w roku 2016 na platformę PC.

Rozwiąż zadanie samodzielnie, a następnie spróbuj rozwiązać go wykorzystując poniższy szablon. Wypełnij w nim brakujące fragmenty:

```
gameInfosDF. filter (col("platform") == "PC").\n\n    filter (col("release_year") == "2016").\n\n    orderBy ( col ("score"). desc ().\n\n    limit (3). select ("title","score")
```

Pytanie **2**

Poprawnie

Punkty maks.: 6,00

Zad 10

Określ dla każdej oceny opisowej (score_phrase) minimalną i maksymalną ocenę liczbową. Wyniki posortuj rosnąco pod względem minimalnej oceny liczbowej.

Rozwiąż zadanie samodzielnie, a następnie spróbuj rozwiązać go wykorzystując poniższy szablon. Wypełnij w nim brakujące fragmenty:

```
gameInfosDF. groupBy ("score_phrase").\n\n    agg (max(col("score")). alias ("max"), min (col("score")).alias("min")).\n\n    select ("score_phrase","max","min").\n\n    orderBy ("min")
```

Pytanie 3

Częściowo poprawnie

Punkty maks.: 16,00

Zad 11 a

To może coś trudniejszego. Wyznacz liczbę oraz średnią ocenę gier wydawanych w poszczególnych latach począwszy od roku 2000 na poszczególne platformy. Nie analizuj wszystkich platform – ogranicz je tylko do tych, dla których liczba wszystkich recenzji gier biorąc pod uwagę wszystkie lata przekroczyła 500.

Rozwiąż to zadanie na dwa sposoby:

a. Za pomocą DataFrame API

Rozwiąż zadanie samodzielnie, a następnie spróbuj rozwiązać go wykorzystując poniższy szablon. Wypełnij w nim brakujące fragmenty:

```
platformsOver500 = gameInfosDF.\
    groupBy (expr("platform as platformName")).\
    agg (count (gameInfoDF._c0).alias("how_many")).\
    select ("platformName", "how_many").\
    filter (col("how_many") > 500)

platformsOver500.\
    join (gameInfoDF, platformsOver500["platformName"] == gameInfosDF.platform).\
    groupBy("release_year", "platform").\
    agg (avg (col("score")).alias("avg"), count("_c0").alias("count")).\
    select ("platform", "release_year", "avg", "count").\
    filter (col("release_year") > 1999).\
    orderBy(col("release_year"), col("platform"))
```

Pytanie 4

Poprawnie

Punkty maks.: 9,00

Zad 11 b

b. Za pomocą SQL (po zarejestrowaniu źródeł danych jako tymczasowych perspektyw).

Rozwiąż zadanie samodzielnie, a następnie spróbuj rozwiązać go wykorzystując poniższy szablon. Wypełnij w nim brakujące fragmenty:

```
gameInfosDF.createOrReplaceTempView("gameInfos")
```

```
spark.sql("""
SELECT      ✓   gi1.release_year, gi1.platform, count(*), avg(gi1.score)
FROM        ✓   gameInfos gi1 JOIN
(SELECT platform, count(*) c FROM      ✓   gameInfos
GROUP BY    ✓   platform HAVING      ✓   c      ✓   >= 500) gi2
ON (gi1.platform =   gi2      ✓   .platform)
GROUP BY gi1.release_year, gi1.platform
HAVING     ✓   gi1.release_year >=   2000      ✓
ORDER BY   ✓   gi1.release_year, gi1.platform """)
)
```

Pytanie 5

Poprawnie

Punkty maks.: 1,00

Zad 17

Oblicz sumę ludności wszystkich Państw na rok 2010. W sytuacji gdy w danym kraju nie przeprowadzono badania w roku 2010 wykorzystaj najnowsze z badań wcześniejszych.

Zmodyfikuj teraz Twoje zapytanie, tak aby obliczyć sumę ludności wszystkich Państw na rok 2006

Jako rozwiązanie zamieść uzyskany w ten sposób wynik.

Odpowiedź: 6070573393 ✓

Pytanie **6**

Poprawnie

Punkty maks.: 1,00

Zad 18

Było ciężko? Nie wierzę. Zatem teraz już będzie z górki. Podaj trzy kraje o największej gęstości zaludnienia **na** rok 2010.

Dokonaj teraz modyfikacji swojego zapytania, tak aby poznać **ósm kraj pod względem gęstości zaludnienia na rok 2010.**

W odpowiedzi zarejestruj jego nazwę.

Odpowiedź: Gaza Strip

Pytanie **7**

Poprawnie

Punkty maks.: 1,00

Zad 19

Podaj trzy kraje o największym procencie ludności żyjącym w miastach powyżej 50 000 mieszkańców w roku 2010.

Dokonaj teraz modyfikacji swojego zapytania, tak aby poznać **dziewiąty kraj o największym procencie ludności żyjącym w miastach powyżej 50 000 mieszkańców w roku 2010.**

W odpowiedzi zarejestruj jego nazwę.

Odpowiedź: Taiwan

