

PO12: Przetwarzanie Big Data 22/23-Z

Rozpoczęto	wtorek, 13 grudnia 2022, 21:11
Stan	Ukończone
Ukończono	wtorek, 13 grudnia 2022, 22:53
Wykorzystany czas	1 godzina 41 min.
Punkty	6,77/7,00
Ocena	9,67 pkt. na 10,00 pkt. możliwych do uzyskania (96,73%)

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 %>% where(✓ ($"platform" === "PC").  
  where(✓ ($"release_year"=== "2016").  
  orderBy(✓ (gameInfosDF %>% ✓ ("score"). desc ✓ ).  
  limit(✓ (3), select(✓ ("title","score").show()
```

Pytanie 1
Poprawnie
Punkty: 1,00 z 1,00
[Oflaguj pytanie](#)

Nawigacja w teście

Discussion

Zakończ przegląd

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:

```
gameInfoDF.
  groupBy    ✓ ("score_phrase").
  agg        ✓ (max($"score").max      ✗ ("max"), min      ✓ ($"score").as("min")).
  select     ✓ ("score_phrase","max","min").
  orderBy    ✓ ("min").show()
```

Pytanie 2

Częściowo poprawnie

Punkty: 0,83 z 1,00

[Oflaguj pytanie](#)

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:

```
val platformsOver500 = gameInfosDF.  
  rollup ❌ (expr("platform as platformName ✅")).  
  agg ✅ ("_c0" -> " count ✅").  
  select ✅ ("platformName", "count(_c0)").  
  where ✅ ("$_count(_c0)" > 500)  
  
val toTemp = platformsOver500.  
  join ✅ ( gameInfosDF ✅ , platformsOver500(" platformName ✅ ") === gameInfosDF(" platform ✅ ")).  
  groupBy ✅ (" release_year ✅ ", "platform").  
  agg ✅ ( avg ✅ ("score").as("avg"), count("$_c0").as("count")).  
  select ✅ ("platform", "release_year", "avg", "count").  
  where ✅ ("release_year" > 1999).show()
```

Pytanie 3

Częściowo poprawnie

Punkty: 0,94 z 1,00

[Oflaguj pytanie](#)

Zad 11 b

b.Za pomocą SQL (po zarejestrowaniu źródeł danych jako tymczasowych perspektyw). Wykorzystaj interfejs notatnika Zeppelina do zaprezentowania wyniku za pomocą wykresu (zmień interpreter za pomocą wyrażenia %sql umieszczonego w pierwszej linii paragrafu).

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

```
gameInfosDF.createOrReplaceTempView("gameInfos")

val result = 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 """)
)

result.createOrReplaceTempView("resultTable" ✓)

result.show()

%sql
select * from resultTable
```

Pytanie 4

Poprawnie

Punkty: 1,00 z 1,00

[Oflaguj pytanie](#)

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 2009

Jako rozwiązanie zamieść uzyskany w ten sposób wynik.

Odpowiedź: 6175723236

Pytanie 5

Poprawnie

Punkty: 1,00 z 1,00

 Oflaguj pytanie

Zad 18

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

Dokonaj teraz modyfikacji swojego zapytania, tak aby poznać **czwarty** kraj pod względem gęstości zaludnienia w roku 2010.

W odpowiedzi zarejestruj jego nazwę.

Odpowiedź: Hong Kong

Pytanie 6
Poprawnie
Punkty: 1,00 z 1,00
[Oflaguj pytanie](#)

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 **czwarty** kraj o największym procencie ludności żyjącym w miastach powyżej 50 000 mieszkańców w roku 2010.

W odpowiedzi zarejestruj jego nazwę.

Odpowiedź: Singapore ✓

Pytanie 7

Poprawnie

Punkty: 1,00 z 1,00

 Oflaguj pytanie

Poprawna odpowiedź to: Singapore

Zakończ przegląd