

SQL do zaawansowanych analiz danych – część 2.

Funkcje analityczne

Materiały laboratoryjne

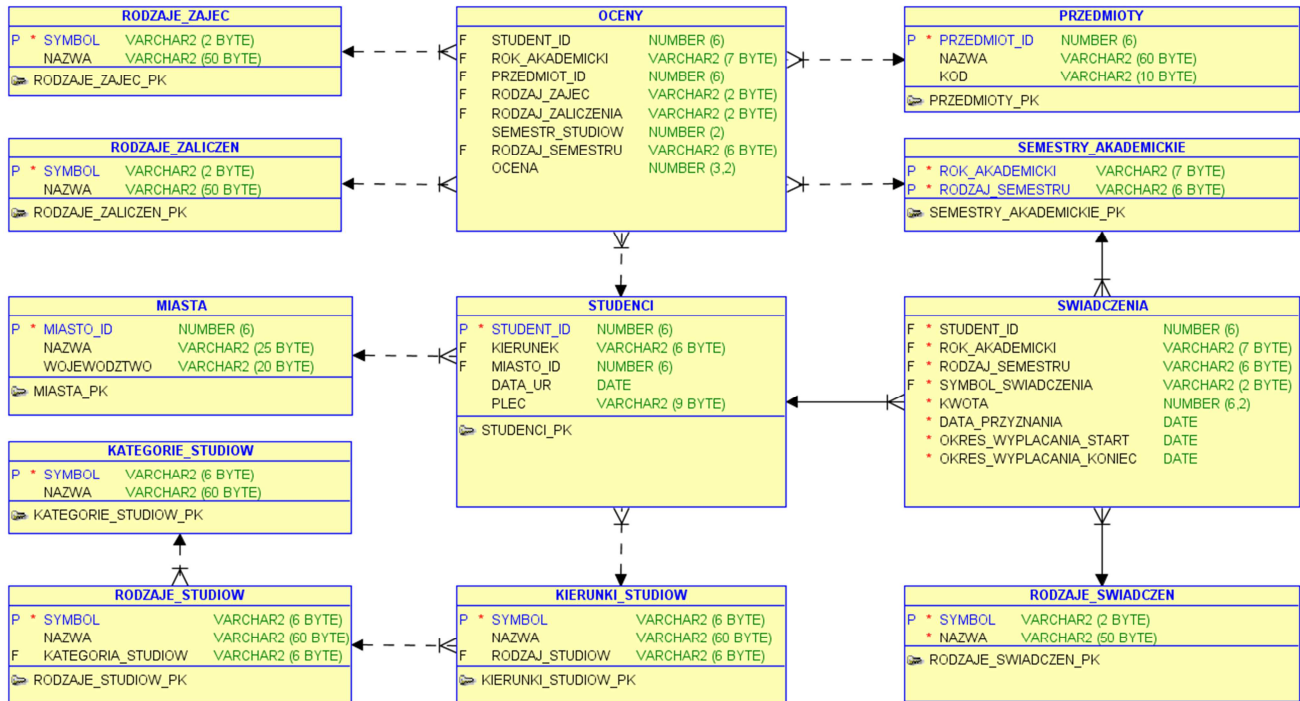
Bartosz Bębel

Politechnika Poznańska, Instytut Informatyki

Wprowadzenie do ćwiczenia

W niniejszym ćwiczeniu student nabierze praktycznych umiejętności w postugiwaniu się poleceniami SQL wykorzystującymi funkcje analityczne.

Wszystkie zadania będą realizowane na danych przedstawionego poniżej schematu relacyjnego.



Ćwiczenia należy realizować w bazie *dblab01*. Twoje konto bazodanowe posiada nazwę utworzoną z dodania do Twojego numeru indeksu przedrostka "BIO" (np. *BIO12345* dla numeru indeksu *12345*). Początkowe hasło jest takie samo jak nazwa użytkownika. Używamy w ćwiczeniu narzędziem jest *Oracle SQL Developer*.

Ćwiczenie 1. Funkcje rankingowe

W niniejszym ćwiczeniu będziemy wykorzystywać funkcje rankingowe: *RANK*, *DENSE_RANK*, *CUME_DIST*, *PERCENT_RANK*, *ROW_NUMBER* i *NTILE*.

1. Zbuduj zapytanie, które dla każdego kierunku studiów znajdzie liczbę studiujących na nim studentów. Następnie utwórz ranking kierunków ze względu na liczbę studentów.

KIERUNEK	LICZBA	RANKING
Pedagogika (stacjonarne)	1487	1
Turystyka i rekreacja (stacjonarne)	1128	2
Ekonomia (stacjonarne)	1040	3
Wychowanie fizyczne (stacjonarne)	1024	4
Turystyka i rekreacja (niestacjonarne)	1007	5
Pedagogika (niestacjonarne)	1007	5
Wychowanie fizyczne (niestacjonarne)	910	7
Ekonomia (niestacjonarne)	869	8
Elektrotechnika (niestacjonarne)	852	9
Rolnictwo (stacjonarne)	832	10
Elektrotechnika (stacjonarne)	697	11
Mechanika i budowa maszyn (niestacjonarne)	566	12
Mechanika i budowa maszyn (stacjonarne)	533	13
Rolnictwo (niestacjonarne)	334	14
Wychowanie fizyczne z edukacją zdrowotną (niestacjonarne)	176	15
Edukacja artystyczna (stacjonarne)	153	16

16 wierszy zostało wybranych.

2. Zmodyfikuj zapytanie z p. 1 w taki sposób, aby w zbiorze wynikowym pojawiła się dodatkowa kolumna pokazująca ranking gęsty. Zwróć uwagę na różnice w obu rankingach.

KIERUNEK	LICZBA	RANK	RANK_GESTY
Pedagogika (stacjonarne)	1487	1	1
Turystyka i rekreacja (stacjonarne)	1128	2	2
Ekonomia (stacjonarne)	1040	3	3
Wychowanie fizyczne (stacjonarne)	1024	4	4
Turystyka i rekreacja (niestacjonarne)	1007	5	5
Pedagogika (niestacjonarne)	1007	5	5
Wychowanie fizyczne (niestacjonarne)	910	7	6
Ekonomia (niestacjonarne)	869	8	7
Elektrotechnika (niestacjonarne)	852	9	8
Rolnictwo (stacjonarne)	832	10	9
Elektrotechnika (stacjonarne)	697	11	10
Mechanika i budowa maszyn (niestacjonarne)	566	12	11
Mechanika i budowa maszyn (stacjonarne)	533	13	12
Rolnictwo (niestacjonarne)	334	14	13
Wychowanie fizyczne z edukacją zdrowotną (niestacjonarne)	176	15	14
Edukacja artystyczna (stacjonarne)	153	16	15

16 wierszy zostało wybranych.

3. Zmodyfikuj zapytanie z poprzedniego punktu w taki sposób, aby otrzymać dane jedynie trzech pierwszych kierunków w rankingu (weź pod uwagę ranking zwykły).

KIERUNEK	LICZBA	RANK
Pedagogika (stacjonarne)	1487	1
Turystyka i rekreacja (stacjonarne)	1128	2
Ekonomia (stacjonarne)	1040	3

4. Dokonaj kolejnej modyfikacji zapytania, tym razem chcemy uzyskać informacje o pięciu najmniej licznych kierunkach (ponownie użyj zwykłego rankingu).

KIERUNEK	LICZBA	RANK
Edukacja artystyczna (stacjonarne)	153	1
Wychowanie fizyczne z edukacją zdrowotną (niestacjonarne)	176	2
Rolnictwo (niestacjonarne)	334	3
Mechanika i budowa maszyn (stacjonarne)	533	4
Mechanika i budowa maszyn (niestacjonarne)	566	5

5. Przekształć ranking, uzyskany w zadaniu 1., w ranking procentowy (użyj funkcji *PERCENT_RANK*). Ogranicz wynik do dwóch pozycji po przecinku.

KIERUNEK	LICZBA	RANKING_PROG
Pedagogika (stacjonarne)	1487	0
Turystyka i rekreacja (stacjonarne)	1128	6.67
Ekonomia (stacjonarne)	1040	13.33
Wychowanie fizyczne (stacjonarne)	1024	20
Turystyka i rekreacja (niestacjonarne)	1007	26.67
Pedagogika (niestacjonarne)	1007	26.67
Wychowanie fizyczne (niestacjonarne)	910	40
Ekonomia (niestacjonarne)	869	46.67
Elektrotechnika (niestacjonarne)	852	53.33
Rolnictwo (stacjonarne)	832	60
Elektrotechnika (stacjonarne)	697	66.67
Mechanika i budowa maszyn (niestacjonarne)	566	73.33
Mechanika i budowa maszyn (stacjonarne)	533	80
Rolnictwo (niestacjonarne)	334	86.67
Wychowanie fizyczne z edukacją zdrowotną (niestacjonarne)	176	93.33
Edukacja artystyczna (stacjonarne)	153	100

16 wierszy zostało wybranych.

6. Zmodyfikuj zapytanie z punktu poprzedniego w taki sposób, aby otrzymać informacje o kierunkach, które lokują się w 25% najliczniej obsadzonych kierunków.

KIERUNEK	LICZBA	RANKING_PROG
Pedagogika (stacjonarne)	1487	0
Turystyka i rekreacja (stacjonarne)	1128	6.67
Ekonomia (stacjonarne)	1040	13.33
Wychowanie fizyczne (stacjonarne)	1024	20

7. Dodaj do wyniku zadania 6. kolumnę wyliczającą percentyle (funkcja *CUME_DIST*). Porównaj wyniki uzyskane w kolumnach *RANKING_PROC* i *PERCENTYL*.

KIERUNEK	LICZBA	R_PROC	PERC
Pedagogika (stacjonarne)	1487	0	6.25
Turystyka i rekreacja (stacjonarne)	1128	6.67	12.5
Ekonomia (stacjonarne)	1040	13.33	18.75
Wychowanie fizyczne (stacjonarne)	1024	20	25
Turystyka i rekreacja (niestacjonarne)	1007	26.67	37.5
Pedagogika (niestacjonarne)	1007	26.67	37.5
Wychowanie fizyczne (niestacjonarne)	910	40	43.75
Ekonomia (niestacjonarne)	869	46.67	50
Elektrotechnika (niestacjonarne)	852	53.33	56.25
Rolnictwo (stacjonarne)	832	60	62.5
Elektrotechnika (stacjonarne)	697	66.67	68.75
Mechanika i budowa maszyn (niestacjonarne)	566	73.33	75
Mechanika i budowa maszyn (stacjonarne)	533	80	81.25
Rolnictwo (niestacjonarne)	334	86.67	87.5
Wychowanie fizyczne z edukacją zdrowotną (niestacjonarne)	176	93.33	93.75
Edukacja artystyczna (stacjonarne)	153	100	100

16 wierszy zostało wybranych.

8. Podaj hipotetyczną pozycję w rankingu kierunku, na którym studiowałoby dokładnie 1000 studentów. Użyj rankingu zwykłego.

```
KIERUNEK_1000_STUDENTOW
-----
                          7
```

9. Przydziel każdej pozycji w rankingu kierunków z punktu 1. unikalny numer porządkowy (wykorzystaj funkcję *ROW_NUMBER*). Porównaj numer porządkowy rekordu z pozycją w rankingu.

KIERUNEK	LICZBA	RANKING	NUMER_PORZ
Pedagogika (stacjonarne)	1487	1	1
Turystyka i rekreacja (stacjonarne)	1128	2	2
Ekonomia (stacjonarne)	1040	3	3
Wychowanie fizyczne (stacjonarne)	1024	4	4
Turystyka i rekreacja (niestacjonarne)	1007	5	5
Pedagogika (niestacjonarne)	1007	5	6
Wychowanie fizyczne (niestacjonarne)	910	7	7
Ekonomia (niestacjonarne)	869	8	8
Elektrotechnika (niestacjonarne)	852	9	9
Rolnictwo (stacjonarne)	832	10	10
Elektrotechnika (stacjonarne)	697	11	11
Mechanika i budowa maszyn (niestacjonarne)	566	12	12
Mechanika i budowa maszyn (stacjonarne)	533	13	13
Rolnictwo (niestacjonarne)	334	14	14
Wychowanie fizyczne z edukacją zdrowotną (niestacjonarne)	176	15	15
Edukacja artystyczna (stacjonarne)	153	16	16

16 wierszy zostało wybranych.

10. Masz podzielić kierunki na cztery "koszyki" w zależności od ich pozycji w rankingu zbudowanym wg liczby studiujących studentów. W każdym koszyku powinno znaleźć się tyle samo kierunków (liczby kierunków w poszczególnych koszykach mogą się różnić o co najwyżej 1).

KIERUNEK	LICZBA	KOSZYK
Pedagogika (stacjonarne)	1487	Koszyk 1
Turystyka i rekreacja (stacjonarne)	1128	Koszyk 1
Ekonomia (stacjonarne)	1040	Koszyk 1
Wychowanie fizyczne (stacjonarne)	1024	Koszyk 1
Turystyka i rekreacja (niestacjonarne)	1007	Koszyk 2
Pedagogika (niestacjonarne)	1007	Koszyk 2
Wychowanie fizyczne (niestacjonarne)	910	Koszyk 2
Ekonomia (niestacjonarne)	869	Koszyk 2
Elektrotechnika (niestacjonarne)	852	Koszyk 3
Rolnictwo (stacjonarne)	832	Koszyk 3
Elektrotechnika (stacjonarne)	697	Koszyk 3
Mechanika i budowa maszyn (niestacjonarne)	566	Koszyk 3
Mechanika i budowa maszyn (stacjonarne)	533	Koszyk 4
Rolnictwo (niestacjonarne)	334	Koszyk 4
Wychowanie fizyczne z edukacją zdrowotną (niestacjonarne)	176	Koszyk 4
Edukacja artystyczna (stacjonarne)	153	Koszyk 4

16 wierszy zostało wybranych.

11. Zbuduj ponownie ranking kierunku studiów wg liczby studiujących studentów, tym razem osobno dla każdej z płci.

KIERUNEK	PLEC	LICZBA	RANKING
Pedagogika (stacjonarne)	kobieta	1336	1
Pedagogika (niestacjonarne)	kobieta	865	2
Ekonomia (stacjonarne)	kobieta	715	3
...			
Rolnictwo (niestacjonarne)	kobieta	86	11
Wychowanie fizyczne z edukacją zdrowotną (niestacjonarne)	kobieta	58	12
Elektrotechnika (stacjonarne)	kobieta	42	13
Mechanika i budowa maszyn (stacjonarne)	kobieta	11	14
Elektrotechnika (niestacjonarne)	kobieta	9	15
Mechanika i budowa maszyn (niestacjonarne)	kobieta	1	16
Elektrotechnika (niestacjonarne)	mężczyzna	843	1
Wychowanie fizyczne (stacjonarne)	mężczyzna	703	2
Wychowanie fizyczne (niestacjonarne)	mężczyzna	701	3
Elektrotechnika (stacjonarne)	mężczyzna	655	4
Mechanika i budowa maszyn (niestacjonarne)	mężczyzna	565	5
...			
Pedagogika (stacjonarne)	mężczyzna	151	13
Pedagogika (niestacjonarne)	mężczyzna	142	14
Wychowanie fizyczne z edukacją zdrowotną (niestacjonarne)	mężczyzna	118	15
Edukacja artystyczna (stacjonarne)	mężczyzna	66	16

32 wierszy zostało wybranych.

Ćwiczenie 2. Funkcje okna

W niniejszym ćwiczeniu zapoznamy się z funkcjami okna.

1. Policz, ile ocen zdobył student o identyfikatorze 1344. Pomiń puste oceny (rekordy tabeli *OCENY* z pustą wartością kolumny *OCENA*).

```
LICZBA_OCEN
-----
      144
```

2. Poznaj szczegółowy rozkład liczby ocen, zdobytych przez studenta 1344, z podziałem na semestry studiów. Wylicz też narastającą liczbę ocen, zdobytych przez tego studenta, od początku studiów.

```
SEMESTR_STUDIOW LICZBA_OCEN_W_SEMESTRZE LICZBA_OCEN_OD_POCZATKU
-----
          1                12                12
          2                46                58
          3                14                72
          4                12                84
          5                17               101
          6                16               117
          7                13               130
          8                14               144
```

8 wierszy zostało wybranych.

3. W wyniku zadania 2. zastąp kolumnę *LICZBA_OCEN_OD_POCZATKU* kolumną, która pokaże liczbę ocen zdobytą w roku. Rok jest tu rozumiany jako dwa semestry – bieżący i poprzedni.

```
SEMESTR_STUDIOW LICZBA_OCEN_W_SEMESTRZE LICZBA_OCEN_W_ROKU
-----
          1                12                12
          2                46                58
          3                14                60
          4                12                26
          5                17                29
          6                16                33
          7                13                29
          8                14                27
```

8 wierszy zostało wybranych.

4. Tym razem chcemy poddać analizie świadczenia przyznane studentom. Podaj zapytanie, które pokaże sumaryczną kwotę przyznanych studentom świadczeń w rozbiciu na kolejne lata kalendarzowe (pobierz rok z daty przyznania studentowi świadczenia, użyj funkcji *EXTRACT(YEAR FROM <data>)*).

ROK	RAZEM_W_ROKU
2000	420479
2001	1301545
2002	2313999
2003	3032863
2004	3299393
2005	3515095
2006	3570723
2007	1764089

8 wierszy zostało wybranych.

5. Dodaj do poprzedniego wyniku kolumnę, która pokaże udział procentowy kwoty świadczeń przyznanych w danym roku do całkowitej sumy przyznanych świadczeń. Do wyznaczenia całkowitej sumy przyznanych świadczeń użyj funkcji okna (o zakresie obejmującym wszystkie rekordy okna).

ROK	RAZEM_W_ROKU	PROCENT_CALOSCI
2000	420479	2.19
2001	1301545	6.77
2002	2313999	12.04
2003	3032863	15.78
2004	3299393	17.17
2005	3515095	18.29
2006	3570723	18.58
2007	1764089	9.18

8 wierszy zostało wybranych.

6. Rozwiąż poprzednie zadanie, tym razem użyj funkcji *LAST_VALUE* na wyniku będącym półkostką danych (użyj klauzuli *ROLLUP*). Wskazówka: zbuduj najpierw zapytanie wyliczające półkostkę danych, następnie dodaj do zapytania funkcję analityczną, w której do wyliczenia procentu będziesz brał/a wartość dla podsumowania półkostki.

ROK	RAZEM_W_ROKU	PROCENT_CALOSCI
2000	420479	2.19
2001	1301545	6.77
2002	2313999	12.04
2003	3032863	15.78
2004	3299393	17.17
2005	3515095	18.29
2006	3570723	18.58
2007	1764089	9.18
RAZEM	19218186	100

9 wierszy zostało wybranych.

7. Tym razem interesują nas szczegółowe kwoty świadczeń, przyznanych studentom w 2000 roku. Na początek zbuduj zapytanie, które dla każdego kierunku wyświetli sumę świadczeń dla studentów tego kierunku w 2000 roku oraz pokaże sumaryczną kwotę świadczeń przyznanych w 2000 roku.

SYMBOL_KIERUNKU	SUMA_DLA_KIERUNKU
NLK2	55070
NLK3	47765
NLK4	27064
NLK5	44497
NZK3	25880
SLK1	10500
SLK3	57583
SLK4	42548
SLK5	39032
SZK1	26465
SZK2	19168
SZK3	24907
RAZEM	420479

13 wierszy zostało wybranych.

Następne zapytanie ma wyświetlić identyfikator studenta wraz z kwotą przyznanego świadczenia, symbol kierunku, na którym studiuje student, sumaryczną kwotę przyznanych w 2000 roku studentom świadczeń na tym samym kierunku co aktualny student (wskazówka: zastosuj partycjonowanie danych, na których będzie działała funkcja sumująca przyznane kwoty, wg kierunku studiów studentów), jaki procent sumy świadczeń dla kierunku stanowi świadczenie studenta oraz jaki procent całkowitej sumy przyznanych w 2000 roku świadczeń stanowi świadczenie studenta. Przy wyliczeniach zastosuj funkcje okna.

STUDENT_ID	KWOTA	KIERUNEK	SUMA_DLA_KIERUNKU	PROCENT_KIERUNKU	PROCENT_CALOSCI
365	750	NLK5	44497	1.69	0.18
366	1000	NLK5	44497	2.25	0.24
366	196	NLK5	44497	0.44	0.05
367	750	NLK5	44497	1.69	0.18
369	1000	NLK5	44497	2.25	0.24
370	750	NLK5	44497	1.69	0.18
371	1000	NLK5	44497	2.25	0.24
371	126	NLK5	44497	0.28	0.03
...					
383	750	NLK5	44497	1.69	0.18
385	1000	NLK5	44497	2.25	0.24
388	119	NLK5	44497	0.27	0.03
388	1000	NLK5	44497	2.25	0.24
390	1000	SLK1	10500	9.52	0.24
391	750	SLK1	10500	7.14	0.18
392	1000	SLK1	10500	9.52	0.24
393	750	SLK1	10500	7.14	0.18
...					
2478	1500	NLK4	27064	5.54	0.36
9728	102	NLK4	27064	0.38	0.02
9728	500	NLK4	27064	1.85	0.12
9910	750	NLK2	55070	1.36	0.18

819 wierszy zostało wybranych.

Ćwiczenie 3. Funkcje raportujące

Bieżące ćwiczenie zapozna studenta z funkcjami raportującymi.

1. Zbuduj zapytanie, które dla każdego studenta, studiującego na kierunku rodzaju studiów o symbolu *SL* wypisze jego identyfikator, symbol kierunku, na jakim studiuje oraz wyliczy jego średnią ocen (zastosuj zwykłe grupowanie).

```
STUDENT_ID KIERUNEK SREDNIA_STUDENTA
-----
377 SLK1 4.24
380 SLK1 4.51
384 SLK1 4.57
390 SLK1 4.96
391 SLK1 4.52
392 SLK1 4.99
393 SLK1 4.6
395 SLK1 4.19
396 SLK1 4.34
397 SLK1 4.49
398 SLK1 4.64
```

...

3115 wierszy zostało wybranych.

Wyświetl również maksymalne średnie ocen studentów na poszczególnych kierunkach rodzaju *SL*.

```
SYMBOL MAKS_SREDNIA_KIER
-----
SLK1 4.99
SLK2 4.79
SLK3 4.96
SLK4 4.84
SLK5 4.5
```

2. Rozbuduj zapytanie z punktu 1. o kolumnę pokazującą maksymalną średnią ocen wśród studentów na kierunku, na którym studiuje bieżący student.

STUDENT_ID	KIERUNEK	SREDNIA_STUDENTA	MAKS_DLA_KIERUNKU
377	SLK1	4.24	4.99
380	SLK1	4.51	4.99
384	SLK1	4.57	4.99
390	SLK1	4.96	4.99
391	SLK1	4.52	4.99
392	SLK1	4.99	4.99
393	SLK1	4.6	4.99
395	SLK1	4.19	4.99
396	SLK1	4.34	4.99
397	SLK1	4.49	4.99
398	SLK1	4.64	4.99
399	SLK1	4.07	4.99
400	SLK1	4.4	4.99
401	SLK1	3.69	4.99
402	SLK1	4.51	4.99
403	SLK1	3.89	4.99
404	SLK1	4.81	4.99
405	SLK1	4.5	4.99
406	SLK1	3.95	4.99
407	SLK1	4.67	4.99
408	SLK1	4.2	4.99
409	SLK1	3.67	4.99
410	SLK1	4.32	4.99
411	SLK1	4.73	4.99
412	SLK1	4.62	4.99
459	SLK5	3.78	4.5
462	SLK5	3.34	4.5
467	SLK5	3.53	4.5
473	SLK3	3.89	4.96
480	SLK5	3.86	4.5
483	SLK5	3.31	4.5

...

3 115 wierszy zostało wybranych.

3. Ponownie rozszerz zapytanie z punktu poprzedniego, tym razem w taki sposób, aby uzyskać kolumnę wyświetlającą maksymalną średnią ocen wśród wszystkich studentów studiujących na kierunkach pod rodzajem studiów o symbol SL (w tym wypadku będzie to najwyższa średnia wśród wszystkich wyszukanych studentów).

STUDENT_ID	KIERUNEK	SREDNIA_STUDENTA	MAKS_DLA_KIERUNKU	MAKS_DLA_RODZAJU
377	SLK1	4.24	4.99	4.99
380	SLK1	4.51	4.99	4.99
384	SLK1	4.57	4.99	4.99
390	SLK1	4.96	4.99	4.99
391	SLK1	4.52	4.99	4.99
392	SLK1	4.99	4.99	4.99
393	SLK1	4.6	4.99	4.99
395	SLK1	4.19	4.99	4.99
396	SLK1	4.34	4.99	4.99
397	SLK1	4.49	4.99	4.99
398	SLK1	4.64	4.99	4.99
399	SLK1	4.07	4.99	4.99
400	SLK1	4.4	4.99	4.99
401	SLK1	3.69	4.99	4.99
402	SLK1	4.51	4.99	4.99
403	SLK1	3.89	4.99	4.99
404	SLK1	4.81	4.99	4.99
405	SLK1	4.5	4.99	4.99
406	SLK1	3.95	4.99	4.99
407	SLK1	4.67	4.99	4.99
408	SLK1	4.2	4.99	4.99
409	SLK1	3.67	4.99	4.99
410	SLK1	4.32	4.99	4.99
411	SLK1	4.73	4.99	4.99
412	SLK1	4.62	4.99	4.99
459	SLK5	3.78	4.5	4.99
462	SLK5	3.34	4.5	4.99
467	SLK5	3.53	4.5	4.99
473	SLK3	3.89	4.96	4.99
480	SLK5	3.86	4.5	4.99
483	SLK5	3.31	4.5	4.99
490	SLK5	3.33	4.5	4.99
491	SLK5	4.33	4.5	4.99
492	SLK5	3.5	4.5	4.99

...

3 115 wierszy zostało wybranych.

4. Ponownie wykonaj zadanie 7. z poprzedniego ćwiczenia, tym razem zastosuj funkcje raportujące.

STUDENT_ID	KWOTA	KIERUNEK	SUMA_DLA_KIERUNKU	PROCENT_KIERUNKU	PROCENT_CALosci
365	750	NLK5	44497	1.69	0.18
366	1000	NLK5	44497	2.25	0.24
366	196	NLK5	44497	0.44	0.05
367	750	NLK5	44497	1.69	0.18
369	1000	NLK5	44497	2.25	0.24
370	750	NLK5	44497	1.69	0.18
371	1000	NLK5	44497	2.25	0.24
371	126	NLK5	44497	0.28	0.03
372	750	NLK5	44497	1.69	0.18
373	750	NLK5	44497	1.69	0.18
374	1000	NLK5	44497	2.25	0.24
375	750	NLK5	44497	1.69	0.18
382	1000	NLK5	44497	2.25	0.24
383	187	NLK5	44497	0.42	0.04
383	750	NLK5	44497	1.69	0.18
385	1000	NLK5	44497	2.25	0.24
388	119	NLK5	44497	0.27	0.03
388	1000	NLK5	44497	2.25	0.24
390	1000	SLK1	10500	9.52	0.24
391	750	SLK1	10500	7.14	0.18
392	1000	SLK1	10500	9.52	0.24
393	750	SLK1	10500	7.14	0.18
...					
2478	1500	NLK4	27064	5.54	0.36
9728	102	NLK4	27064	0.38	0.02
9728	500	NLK4	27064	1.85	0.12
9910	750	NLK2	55070	1.36	0.18

819 wierszy zostało wybranych.

5. Przygotuj raport, który policzy ilu studentów studiuje na poszczególnych kierunkach studiów.

KIERUNEK	STUD_NA_KIERUNKU
NLK1	869
NLK2	1007
NLK3	1007
NLK4	910
NLK5	176
NZK1	566
NZK2	334
NZK3	852
SLK1	153
SLK2	1040
SLK3	1487
SLK4	1128
SLK5	1024
SZK1	697
SZK2	533
SZK3	832

16 wierszy zostało wybranych.

6. Rozszerz zapytanie z poprzedniego zadania tak, aby otrzymać symbol rodzaju studiów dla kierunku oraz procent liczby studentów danego kierunku w liczbie studentów na danym rodzaju studiów.

KIERUNEK	STUD_NA_KIERUNKU	RS	%RS
NLK1	869	NL	21.89
NLK2	1007	NL	25.37
NLK3	1007	NL	25.37
NLK4	910	NL	22.93
NLK5	176	NL	4.43
NZK1	566	NZ	32.31
NZK2	334	NZ	19.06
NZK3	852	NZ	48.63
SLK1	153	SL	3.17
SLK2	1040	SL	21.52
SLK3	1487	SL	30.77
SLK4	1128	SL	23.34
SLK5	1024	SL	21.19
SZK1	697	SZ	33.8
SZK2	533	SZ	25.85
SZK3	832	SZ	40.35

16 wierszy zostało wybranych.

7. Dokonaj kolejnej modyfikacji zapytania, tym razem chcemy dodatkowo uzyskać informacje o udziale procentowym studentów danego kierunku w liczbie studentów kategorii studiów oraz w całkowitej liczbie studentów.

KIERUNEK	STUD_NA_KIERUNKU	RS	%RS	KAT	%KAT	%CALOSCI
NLK1	869	NL	21.89	N	15.19	6.89
NLK2	1007	NL	25.37	N	17.6	7.98
NLK3	1007	NL	25.37	N	17.6	7.98
NLK4	910	NL	22.93	N	15.91	7.21
NLK5	176	NL	4.43	N	3.08	1.4
NZK1	566	NZ	32.31	N	9.89	4.49
NZK2	334	NZ	19.06	N	5.84	2.65
NZK3	852	NZ	48.63	N	14.89	6.75
SLK1	153	SL	3.17	S	2.22	1.21
SLK2	1040	SL	21.52	S	15.09	8.24
SLK3	1487	SL	30.77	S	21.57	11.79
SLK4	1128	SL	23.34	S	16.36	8.94
SLK5	1024	SL	21.19	S	14.85	8.12
SZK1	697	SZ	33.8	S	10.11	5.53
SZK2	533	SZ	25.85	S	7.73	4.23
SZK3	832	SZ	40.35	S	12.07	6.6

16 wierszy zostało wybranych.

Ćwiczenie 4. Pozostałe funkcje analityczne

W niniejszym ćwiczeniu zapoznamy się z pozostałymi funkcjami analitycznymi: *LAG* i *LEAD*, *LISTAGG*, *FIRST* i *LAST*, *PERCENTILE_DESC* i *PERCENTILE_CONT* oraz *WIDTH_BUCKET*.

1. Dla studenta o identyfikatorze 1344 wylicz średnie ocen, jakie uzyskał w poszczególnych latach akademickich. Uzupełnij wynik o wskazanie, o ile w danym roku wzrosła/zmała średnia w stosunku do roku poprzedniego (użyj funkcji *LAG*).

ROK	SREDNIA_W_ROKU	PRZYROST_SPADEK
2001/02	3.148148148	0
2002/03	3	-0.1481481481
2003/04	3.21875	0.21875
2004/05	2.942307692	-0.2764423077
2005/06	3.166666667	0.2243589744
2006/07	3.796296296	0.6296296296

6 wierszy zostało wybranych

2. Ponownie zanalizuj świadczenia przyznane studentom. Podaj zapytanie, które pokaże sumaryczną kwotę oraz liczbę przyznanych studentom świadczeń w rozbiciu na kolejne lata kalendarzowe (pobierz rok z daty przyznania studentowi świadczenia, użyj funkcji *EXTRACT(YEAR FROM <data>)*). Dodaj kolumnę, która pokaże procentowy wzrost/spadek kwoty i liczby przyznanych w danym roku kalendarzowym świadczeń w stosunku do roku poprzedniego.

ROK	SUMA_W_ROKU	ZMIANA KWOT%	LICZBA_W_ROKU	ZMIANA LICZBY%
2000	420479	0	819	0
2001	1301545	209.54	2414	194.75
2002	2313999	77.79	3912	62.05
2003	3032863	31.07	5092	30.16
2004	3299393	8.79	5704	12.02
2005	3515095	6.54	6338	11.12
2006	3570723	1.58	6676	5.33
2007	1764089	-50.6	3284	-50.81

8 wierszy zostało wybranych.

3. Ponownie zanalizujemy średnie studentów, tym razem używając funkcji *FIRST* i *LAST*. Na początek zbuduj zapytanie, które dla każdego studenta wypisze jego identyfikator, symbol kierunku, na jakim studiuje oraz wyliczy jego średnią ocen (zastosuj zwykłe grupowanie).

```
STUDENT_ID KIERUNEK SREDNIA_STUDENTA
-----
2992 NLK1 2.92
3019 NLK1 3.23
3323 NLK1 3.45
3355 NLK1 4.5
3356 NLK1 4.24
3357 NLK1 3.95
3358 NLK1 3.86
3359 NLK1 3.68
3360 NLK1 4.12
3361 NLK1 3.44
3363 NLK1 3.62
3364 NLK1 3.98
3365 NLK1 3.19
...
```

8345 wierszy zostało wybranych.

4. Rozbuduj zapytanie z poprzedniego punktu w taki sposób, aby dla każdego studenta wyliczyło jego średnią ocen, wyświetliło jego identyfikator i kierunek studiów oraz średnią najlepszego studenta na danym kierunku i identyfikator studenta z najlepszą średnią na kierunku (użyj funkcji *FIRST* jako agregującej funkcji raportującej).

```
STUDENT_ID KIER SRED_STUD STUD_Z_MAKS_SRED_KIER MAKS_SRED_NA_KIER
-----
2992 NLK1 2.92 6926 4.68
3019 NLK1 3.23 6926 4.68
3323 NLK1 3.45 6926 4.68
3355 NLK1 4.5 6926 4.68
3356 NLK1 4.24 6926 4.68
3357 NLK1 3.95 6926 4.68
3358 NLK1 3.86 6926 4.68
3359 NLK1 3.68 6926 4.68
3360 NLK1 4.12 6926 4.68
3361 NLK1 3.44 6926 4.68
3363 NLK1 3.62 6926 4.68
3364 NLK1 3.98 6926 4.68
3365 NLK1 3.19 6926 4.68
...
15273 NLK1 4.53 6926 4.68
1162 NLK2 4.41 2174 4.79
2043 NLK2 3.77 2174 4.79
...
13653 NLK2 3.73 2174 4.79
1479 NLK3 3.6 3589 4.6
1578 NLK3 3.44 3589 4.6
2251 NLK3 3.45 3589 4.6
2252 NLK3 4.07 3589 4.6
...
```

8345 wierszy zostało wybranych.

5. Do wyniku poprzedniego zapytania dodaj ostatnią kolumnę, która wyświetli maksymalną średnią ocen wśród wszystkich studentów.

STUDENT_ID	KIER	SRED_STUD	STUD_Z	MAKS_SRED_KIER	MAKS_SRED_NA_KIER	MAKS_SREDNIA
2992	NLK1	2.92		6926	4.68	5
3019	NLK1	3.23		6926	4.68	5
3323	NLK1	3.45		6926	4.68	5
3355	NLK1	4.5		6926	4.68	5
3356	NLK1	4.24		6926	4.68	5
3357	NLK1	3.95		6926	4.68	5
3358	NLK1	3.86		6926	4.68	5
3359	NLK1	3.68		6926	4.68	5
3360	NLK1	4.12		6926	4.68	5
3361	NLK1	3.44		6926	4.68	5
3363	NLK1	3.62		6926	4.68	5
3364	NLK1	3.98		6926	4.68	5
...						

8345 wierszy zostało wybranych.

6. Zbuduj zapytanie, które dla każdego kierunku studiów wyświetli jego symbol oraz symbol rodzaju studiów, do którego dany kierunek należy.

RODZAJ_STUDIOW	KIERUNEK
NL	NLK1
NL	NLK2
NL	NLK3
NL	NLK4
NL	NLK5
NZ	NZK1
NZ	NZK2
NZ	NZK3
SL	SLK1
SL	SLK2
SL	SLK3
SL	SLK4
SL	SLK5
SZ	SZK1
SZ	SZK2
SZ	SZK3

16 wierszy zostało wybranych.

Następnie przekształć zapytanie w taki sposób, aby symbole w ramach rodzaju studiów były wyświetlone w postaci listy.

RODZAJ_STUDIOW	KIERUNKI
NL	NLK1, NLK2, NLK3, NLK4, NLK5
NZ	NZK1, NZK2, NZK3
SL	SLK1, SLK2, SLK3, SLK4, SLK5
SZ	SZK1, SZK2, SZK3

7. Zbuduj ranking kierunków ze względu na sumaryczne kwoty przyznanych studentom na kierunkach świadczeń. Następnie wyznacz medianę powstałego rankingu. Porównaj wyniki funkcji odwróconych percentyli: dyskretną i ciągłą.

KIERUNEK	SUMA_SWIADCZEN	RANKING
SLK3	3461020	1
NLK2	2595930	2
SLK4	2151778	3
SLK2	1684795	4
NLK3	1483361	5
SZK3	1400502	6
SLK5	1172863	7
NLK1	1024122	8
NLK4	808704	9
SZK2	757514	10
NZK3	745724	11
SZK1	712081	12
SLK1	443684	13
NLK5	357500	14
NZK1	282062	15
NZK2	136546	16

16 wierszy zostało wybranych.

MEDIANA_DYSKRETNIA	MEDIANA_CIAGLA
1024122	916413

8. Zbuduj zapytanie, które ponownie wyliczy średnią ocen studenta. Następnie uzupełnij zapytanie o wyrażenie, które przydzieli każdego studenta do jednej z czterech kategorii w zależności od uzyskanej średniej:

- kategoria 1. – średnie większe bądź równe 2 i mniejsze od 3 (również studenci bez średniej – brak ocen),
- kategoria 2. – średnie większe bądź równe 3 i mniejsze od 4,
- kategoria 3. – średnie większe bądź równe 4 i mniejsze od 5,
- kategoria 4. – średnie równe 5.

Użyj funkcji *WIDTH_BUCKET*.

STUDENT_ID	SREDNIA	KATEGORIA
2478	5	4
392	4.99	3
390	4.96	3
1015	4.96	3
5829	4.95	3
7773	4.93	3
1079	4.88	3
1003	4.88	3
...		
10222	4	3
3461	4	3
9121	4	3
12094	3.99	2
2152	3.99	2
8765	3.99	2
...		
1459	3	2
1864	3	2
10699	2.99	1
5760	2.99	1
6314	2.99	1
847	2.99	1
...		
2351	2	1
2390	2	1
2366		1
9289		1
...		

8345 wierszy zostało wybranych.