

### Cel ćwiczenia:

1. Nabycie umiejętności wykorzystania funkcji analitycznych do oceny danych, ich opisu oraz przetwarzania

### Środki:

1. Zbiór danych Sales History (SH).
2. Serwer baz danych Oracle Enterprise Edition 11g.
3. Aplikacja SQL Developer.
4. Strony internetowe:
  - a. [http://docs.oracle.com/cd/E11882\\_01/server.112/e41084/functions004.htm#SQLRF06174](http://docs.oracle.com/cd/E11882_01/server.112/e41084/functions004.htm#SQLRF06174)
  - b. [http://docs.oracle.com/cd/E11882\\_01/server.112/e25554/analysis.htm#DWHSG8755](http://docs.oracle.com/cd/E11882_01/server.112/e25554/analysis.htm#DWHSG8755)

### Przebieg:

1. Używając funkcji analitycznych:
  - a. Zaprezentuj sumaryczną liczbę sprzedanych produktów (na podstawie quantity\_sold) zarejestrowanych w tabeli SALES wraz z informacją, jaka jest pozycja danego miesiąca w rankingu liczby sprzedanych sztuk.
  - b. Wyświetl również kolumnę prezentującą sumaryczną kwotę na jaką dokonano sprzedaży (na podstawie kolumny amount\_sold) wraz z informacją, jaka jest pozycja danego miesiąca w rankingu sumarycznej kwoty sprzedaży.
  - c. Posortuj dane po miesiącach.
  - d. Na podstawie wyniku łatwo możesz zauważyć, które miesiące były najlepsze ze względu na liczbę sprzedanych sztuk, a które ze względu na sumaryczną kwotę.
  - e. Fragment wyniku:

MIESIAC	SUM_LICZBA_SPRZED	POZYCJA_SLS	SUM_KWOTA_SPRZED	POZYCJA_SKS
1998-01	15132	40	2277420.49	12
1998-02	14307	44	2372690.87	7
1998-03	14248	45	1830572.64	38
1998-04	11818	47	1975978.3	27
1998-05	12309	46	1748287.23	40
1998-06	11631	48	1869728.61	36
1998-07	16257	39	1932282.28	30
1998-08	17199	36	1972532.63	28
1998-09	17059	37	2167008.19	15
1998-10	19878	24	2236464.53	13
1998-11	14612	42	1959664.22	29
1998-12	14384	43	1741284.96	41
1999-01	20437	20	2077439.76	22

2. Używając funkcji analitycznych:

- a. Zaprezentuj sumaryczną kwotę na jaką dokonano sprzedaży (na podstawie kolumny amount\_sold)
- b. Obok wyświetl kwotę z poprzedniego miesiąca.
- c. Oblicz procentową różnicę zmian pomiędzy bieżącym a poprzednim miesiącem.
- d. Posortuj dane po miesiącach.
- e. Na podstawie wyniku łatwo możesz zauważyć jaka jest dynamika zmian w sprzedaży, oraz czy dominują wzrosty, czy spadki.
- f. Fragment wyniku:

MIESIAC	SUM_KWOTA_SPRZED	SUM_KWOTA_SPRZED_W_POP_MIES	PROCENT_ZMIAN
1998-01	2277420.49		
1998-02	2372690.87	2277420.49	4.18
1998-03	1830572.64	2372690.87	-22.85
1998-04	1975978.3	1830572.64	7.94
1998-05	1748287.23	1975978.3	-11.52
1998-06	1869728.61	1748287.23	6.95
1998-07	1932282.28	1869728.61	3.35
1998-08	1972532.63	1932282.28	2.08
1998-09	2167008.19	1972532.63	9.86
1998-10	2236464.53	2167008.19	3.21
1998-11	1959664.22	2236464.53	-12.38
1998-12	1741284.96	1959664.22	-11.14
1999-01	2077439.76	1741284.96	19.3

3. Używając funkcji analitycznych:

- a. zaprezentuj sumaryczną liczbę sprzedanych produktów (na podstawie quantity\_sold) zarejestrowanych w tabeli SALES wraz z informacją, jaka jest pozycja danego kwartału w rankingu sprzedaży.
- b. Wyświetl procentowy udział danego kwartału w całym roku (pamiętaj o sformatowaniu wartości w kolumnie, np "24.43%")
- c. Fragment wyniku:

KWARTAL	SUM_LICZBA_SPRZED	POZYCJA_SLS	UDZIAL_W_ROKU
1998-1	43687	15	24.43%
1998-2	35758	16	20.00%
1998-3	50515	13	28.25%
1998-4	48874	14	27.33%
1999-1	64186	4	25.89%
1999-2	54233	12	21.87%
1999-3	67138	2	27.08%
1999-4	62388	6	25.16%
2000-1	62197	7	26.73%

4. Analiza danych sprzedażowych:

- a) Wgraj do swojego schematu skrypt: „ED\_LAB\_03-dane.sql”
- b) Wykorzystując funkcję analityczną i np. wyrażenia regularne wyświetl następującą listę:

PRODUCENT	LISTA
1 Beko	CNA365E30ZGB, CNA400E30ZDX, CNE520E23ZDX, CNA365E20ZX, CNE520EE0ZGW...
2 Bosh	KGF, KGN, KGN36XW24E, KGN39XI36, KGN36VL25E, KGN36XL35, KGN39XW24E,...
3 Gorenje	ONRK193CO, ORK193BK
4 Indesit	LI7, LI8, LR6S2W, LI8, LR6, LR8, LI70, XI9
5 Lg	GBB60NSYFE, GBB59PZFFB, GBB60NSYQE, GBB530NSCQE
6 Samsung	RB41J7839S4, RB31FERNCSS, RB33J3230WW, RB30J3230SA, RB37K63612C, RB...
7 Whirlpool	BSNF8999PB

5. Korzystając z funkcji NTILE dokonaj podział produktów pod kątem pojemności oraz zużycia energii: grupa 1 - najgorsza, 2 - średnia, 3 - najlepsza
- a. Zaobserwuj, że np. lodówka o id = 6 jest w drugiej grupie pojemności (przeciętna) ale w trzeciej grupie energetycznej (a nie drugiej). Takich przypadków jest więcej.

ID	MODEL	KLASA	POJEMNOSC	NTILE_POJ	ENERGIA	NTILE_ENE
1	5 Indesit LR6 S1 X	A+	271	1	266	3
2	7 Indesit LI70 FF1 X	A+	274	1	310	1
3	32 Bosh KGN33NL20	A+	279	1	279	3
4	27 Bosh KGN36XW24E	A+	287	1	280	3
5	29 Bosh KGN36VL25E	A+	287	1	280	3
6	12 Samsung RB30J3230SA	A+	311	2	272	3
7	31 Bosh KGN39XW24E	A+	315	2	309	2
8	33 Bosh KGN39LB20E	A+	315	2	309	2
9	34 Bosh KGN39XL24E	A+	315	2	309	2
10	11 Samsung RB33J3230WW	A+	328	2	280	2
11	6 Indesit LR8 S1 X	A+	339	3	309	2
12	22 Beko CNA365E20ZX	A+	365	3	336	1
13	13 Samsung RB37K63612C/EF	A+	367	3	314	1
14	8 Indesit XI9 T10 X H	A+	369	3	324	1
15	21 Beko CNE520E23ZDX	A+	520	3	403	1

6. Zmodyfikuj polecenie z poprzedniego zadania, przypisując wagi do wartości w kolumnach i wylicz wzór: klasyfikacja = ntile\_poj waga 0.6 + ntile\_ene waga 0.4
- a. Posortuj zbiór wyników po kolumnie klasyfikacja malejąco.
- b. Jak można zauważyć, jest kilka lodówek z trzech klas (A+++, A++, A+), które szczególnie zasługują na uwagę.

ID	MODEL	KLASA	POJEMNOSC	ENERGIA	NTILE_POJ	NTILE_ENE	KLASYFIKACJA
1	24 Beko RCNA365E40X	A+++	365	176	3	2	2,6
2	18 Samsung RB37J5345SS	A++	367	259	3	2	2,6
3	6 Indesit LR8 S1 X	A+	339	309	3	2	2,6
4	16 Samsung RB37J5029SS	A+++	365	173	3	2	2,6
5	12 Samsung RB30J3230SA	A+	311	272	2	3	2,4
6	35 Bosh KGN36VI32	A++	319	239	2	3	2,4
7	41 LG GBB60NSYQE	A+++	343	143	2	3	2,4
8	44 LG GBB530NSCQE	A+++	343	133	2	3	2,4
9	20 Beko CNA400E30ZDX	A++	400	273	3	1	2,2
10	22 Beko CNA365E20ZX	A+	365	336	3	1	2,2
11	13 Samsung RB37K63612C/EF	A+	367	314	3	1	2,2
12	8 Indesit XI9 T10 X H	A+	369	324	3	1	2,2
13	21 Beko CNE520E23ZDX	A+	520	403	3	1	2,2
14	23 Beko CNE520EE0ZGW	A+++	520	211	3	1	2,2
15	36 Bosh KGN39LB35	A++	366	273	3	1	2,2

7. Ponieważ w poprzednim zadaniu klasyfikacja końcowa opierała się na NTILE, nie było dostatecznego rozróżnienia pomiędzy obiektami.
- Zamień funkcję NTILE na RANK, a po uzyskaniu wyniku porównaj go z wynikami z poprzedniego zadania.
  - Zwróć uwagę, że produkt, który by na pierwszym miejscu w poprzednim zadaniu, w obecnym zestawieniu jest na miejscu aż 16-tym.
  - Jest to m.in. spowodowane drobnymi różnicami w parametrach (np. pojemność jest o 2 cm<sup>2</sup> mniejsza a więc rank\_poj o 3 pozycje niższy).
  - Czy Twoim zdaniem zestawienie z poprzedniego zadania, czy to z bieżącego bardziej oddaje istotę wartości produktu dla klienta?

ID	MODEL	KLASA	POJEMNOSC	ENERGIA	RANK_POJ	RANK_ENE	KLASYFIKACJA
1	18 Samsung RB37J5345SS	A++	367	259	14	7	11,2
2	11 Samsung RB33J3230WW	A+	328	280	10	10	10
3	16 Samsung RB37J5029SS	A+++	365	173	11	8	9,8
4	8 Indesit XI9 T10 X H	A+	369	324	14	3	9,6
5	44 LG GBB530NSQEQ	A+++	343	133	6	15	9,6
6	13 Samsung RB37K63612C/EF	A+	367	314	13	4	9,4
7	17 Samsung RB33J8035SR/EF	A++	335	255	9	10	9,4
8	23 Beko CNE520EE0ZGW	A+++	520	211	15	1	9,4
9	20 Beko CNA400E30ZDX	A++	400	273	15	1	9,4
10	21 Beko CNE520E23ZDX	A+	520	403	15	1	9,4
11	35 Bosh KGN36VI32	A++	319	239	6	14	9,2
12	15 Samsung RB36J8799S4/EF	A+++	350	173	10	8	9,2
13	41 LG GBB60NSYQEQ	A+++	343	143	6	14	9,2
14	12 Samsung RB30J3230SA	A+	311	272	6	14	9,2
15	9 Samsung RB41J7839S4/EF	A+++	406	191	13	3	9
16	24 Beko RCNA365E40X	A+++	365	176	11	6	9
17	6 Indesit LR8 S1 X	A+	339	309	11	6	9
18	40 BEKO CNE520EE0ZGW	A+++	454	211	14	1	8,8

8. W celu ustalenie, czy istnieje współzależność pomiędzy zużyciem energii elektrycznej a pojemnością lodówki w danej klasie energetycznej, oblicz za pomocą odpowiedniej funkcji korelacje pomiędzy tymi dwiema wartościami dla poszczególnych klas. Zmienną objaśniają będzie zużycie energii a zmienną objaśnianą pojemność.
- Jako dodatkowe ćwiczenie możesz wykorzystać inne funkcje korelacji (Spearman'a, Kendall'a) poza podstawową: Pearson'a. Czy potrafisz objaśnić ich działanie?

KLASA	CORR_PEARSON
1 A+	0,9028904789496816272269716475815969008209
2 A++	0,7750107738590336645100790465090720842052
3 A+++	0,7502196057683461979467194748310569231836