**Hurtownie Danych - laboratorium**

**Lista 2**

***Podstawy SQL: PIVOT i CASE***

***Analiza jakości i profilowanie danych***

**Wstęp teoretyczny**

Polecenie PIVOT umożliwia prezentowanie wyniku zapytania w postaci tabeli przestawnej.

Stosuje się następującą składnię polecenia PIVOT:

|  |
| --- |
| **SELECT** <non-pivoted column>,  [first pivoted column] AS <column name>,  [second pivoted column] AS <column name>,  ...  [last pivoted column] AS <column name>  **FROM**  (<SELECT query that produces the data>)  **AS** <alias for the source query>  **PIVOT**  (  <aggregation function>(<column being aggregated>)  **FOR**  [<column that contains the values that will become column headers>]  IN ( [first pivoted column], [second pivoted column],  ... [last pivoted column])  ) **AS** <alias for the pivot table>  <optional ORDER BY clause>; |

Polecenie CASE sprawdza określony warunek w postaci wyrażenia logicznego i zwraca jeden z możliwych wyników. Wynikiem jest pierwszy przypadek, dla którego spełniony jest warunek.

|  |
| --- |
| *Searched CASE expression:*  **CASE**  **WHEN** Boolean\_expression THEN result\_expression [ ...n ]  [ **ELSE** else\_result\_expression ]  **END** |

Możliwe jest również porównanie wyrażenia input\_expression z kolejnymi warunkami when\_expression w celu określenia wyniku result\_expression.

|  |
| --- |
| *Simple CASE expression:*  **CASE** input\_expression  **WHEN** when\_expression THEN result\_expression [ ...n ]  [ **ELSE** else\_result\_expression ]  **END** |

Źródła:

<https://docs.microsoft.com/en-us/sql/t-sql/queries/from-using-pivot-and-unpivot?view=sql-server-ver15>

<https://docs.microsoft.com/en-us/sql/t-sql/language-elements/case-transact-sql?view=sql-server-ver15>

**Zadania do wykonania**

Baza danych: **AdventureWorks**

**Zad. 1. Ekstrakcja danych**

1. Utworzyć tabelę Sprzedaz zawierającą dane dotyczące liczby sprzedanych produktów przez pracowników firmy AdventureWorks w poszczególnych latach zgodnie z podanym schematem:

Sprzedaz(pracID, prodID, "Nazwa produktu", Rok, Liczba).

Wykonać to samo zapytanie przy użyciu polecenia pivot:

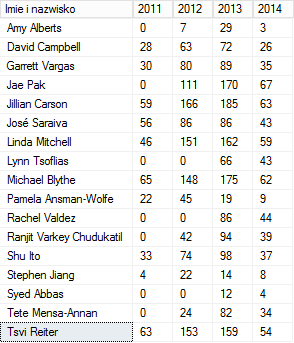
* 1. w kolejnych kolumnach wyświetlić lata;
  2. w kolejnych kolumnach wyświetlić 5 najlepszych produktów.

1. Utworzyć zestawienie, które dla poszczególnych miesięcy i lat przedstawi informację o liczbie różnych klientów. Przygotuj zapytanie z i bez użycia polecenia pivot.



*Przykładowe rozwiązanie*

1. Utworzyć zestawienie zawierające w wierszach imiona i nazwiska sprzedawców,   
   a w kolumnach kolejne lata. Wartością będzie liczba obsłużonych transakcji.



*Przykładowe rozwiązanie*

Wyświetlić tylko tych sprzedawców, którzy pracowali przez wszystkie 4 lata.

1. Zdefiniować zapytanie wyznaczające sumę kwot sprzedaży towarów oraz liczbę różnych produktów w zamówieniach w poszczególnych latach, miesiącach, dniach.



*Przykładowe rozwiązanie*

1. Wykorzystując polecenie CASE przygotować podsumowania do zestawienia z poprzedniego zadania tak, aby sumowane były kwoty zamówień oraz obliczana liczba różnych produktów dla poszczególnych miesięcy i dni tygodnia.

*Uwaga: Pamiętaj o wybraniu właściwego atrybutu funkcji* datepart *tak, aby zgadzała się nazwa dnia tygodnia.*

|  |
| --- |
| case  when Dzień = 1 then 'Poniedziałek'  when Dzień = 2 then 'Wtorek'  when Dzień = 3 then 'Środa'  when Dzień = 4 then 'Czwartek'  when Dzień = 5 then 'Piątek'  when Dzień = 6 then 'Sobota'  when Dzień = 7 then 'Niedziela'  end |

*Przykład wykorzystania polecenia CASE.*



*Przykładowe rozwiązanie*

1. Przygotować zestawienie, w którym dla wybranych klientów przygotujemy kartę lojalnościową:
   1. srebrną, jeśli klient wykonał co najmniej 5 transakcji w sklepie;
   2. złotą, jeśli klient wykonał co najmniej 2 transakcje, której kwota przekraczała 150% średniej wartości zamówień w bazie (łączna kwota przekraczała 300% i każda z transakcji przekraczała 150%);
   3. platynową, jeśli klient co roku spełniał warunki otrzymania karty złotej.

Schemat wynikowej tabeli powinien wyglądać następująco:

KartyLojalnosciowe(Imie, Nazwisko, Liczba transakcji, Łączna kwota transakcji, Kolor karty)



*Przykładowe rozwiązanie*

**Zad. 2. Analiza danych**

1. Przedstaw wyniki zadania 1 w postaci tabel i wykresów przestawnych w programie MS Excel. Zinterpretuj wyniki.
2. Przygotuj 5 dodatkowych tabel/wykresów, które pokażą ciekawe zależności w bazie AdventureWorks przy użyciu narzędzia Power BI lub Tableau.

**Zad. 3. Ocena jakości danych – profilowanie danych**

1. Przeanalizować, scharakteryzować i ocenić dane znajdujące się w plikach dane\_1.txt oraz dane\_2.csv wykorzystując profilowanie danych pakietu Visual Studio: projekt *Integration Services Project*. Z menu bocznego *SSIS Toolbox* należy:
   1. wybrać bloczek *Data Profiling Task* (przeciągnąć na kanwę projektu),
   2. określić dane w sekcji *Destination*,
   3. skonfigurować *Quick Profile*,
   4. uruchomić pakiet (Run).
2. Aby obejrzeć wynik należy użyć *Data Profile Viewer* lub ponownie edytując bloczek *Data Profiling Task* użyć opcji *Open Profile Viewer*.

**Rozwiązania:**

**Wnioski:**

***Uwaga!***

* Sprawozdanie, bez wniosków podsumowujących aspekt zagadnień analizowanych na zajęciach laboratoryjnych i zawartych w sprawozdaniu, jest automatycznie oceniane negatywnie!