Aleksander Stepaniuk 272644  
Hurtownie Danych – wykład 3 zadania

Zad 1.

*Jaka jest łączna suma transakcji (SalesOrderHeader.Subtotal) w poszczególnych latach dla poszczególnych dni tygodnia*.

A screenshot of a computer

AI-generated content may be incorrect.

A computer screen shot of a program

AI-generated content may be incorrect.

Wnioski:

Sprzedaż rosła w latach 2011-2013, osiągając szczyt w 2013, po czym pozornie spadła w 2014 (tylko pozornie, bo w praktyce dane na rok 2014 są niekompletne i są jedynie do czerwca). W 2011 w sprzedaży dominowały soboty i poniedziałki, od 2012 większe przychody zaczęły generować niedziele. W 2013 nastąpił wyjątkowy wzrost sprzedaży w środy (łącznie ponad 9mln), co może wskazywać na skuteczną strategię promocyjną. Jednak dane te nie wydają się pokazywać wiele przydatnych informacji przy takim zestawieniu, ponieważ sam wpływ dnia tygodnia na sprzedaż wydaje się nie być tak prosto przewidywalny.

Zad 2.

*Zaproponuj podział klientów na 3 rozłączne grupy wiekowe (0-50, 51-60, 61+),   
a) ilu różnych klientów dokonało zakupów w kolejnych miesiącach roku w każdej z grup?   
b) ilu klientów w poszczególnych grupach dokonało zakupu dokładnie raz?*

A screenshot of a computer

AI-generated content may be incorrect.

A computer screen shot of a program

AI-generated content may be incorrect.

A computer screen shot of a program

AI-generated content may be incorrect.

Wnioski:

Klienci tego sklepu nie należą do najmłodszych. Znaczna ich część należy do kategorii wiekowej 61+, a także jest bardzo niewielu klientów w kategorii wiekowej 0-50 (większość z nich jest bliżej górnej granicy). Widać więc że sklep ten jest skierowany raczej do starszej klienteli.

Zad 3.

*Zestawienie produktów, których sprzedaje się miesięcznie co najmniej 20 sztuk. Dla każdego produktu podaj jego kategorie. Jeżeli warto użyć CTE to porównaj efektywność rozwiązania z wersją i bez CTE.*

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

Wnioski:

W tym konkretnym przypadku oba zapytania są tak samo wydajne, ponieważ silnik SQL Server zoptymalizuje je w ten sam sposób. CTE byłoby lepsze, gdybyśmy chcieli się wielokrotnie odwoływać do tymczasowej tabeli w ramach tego samego zapytania. CTE może być preferowane w złożonych zapytaniach dla lepszej czytelności kodu.

Zad 4.

*Przygotuj zestawienie, w którym przeanalizujesz, ilu jest różnych klientów dla każdej płci w kolejnych miesiącach (05.2011 – 06.2014). Jak procentowo rozkłada się ich udział w całkowitej wartości sprzedaży (Sales.SalesOrderHeader.TotalDue)?*

A screenshot of a computer

AI-generated content may be incorrect.A computer screen shot of a program

AI-generated content may be incorrect.

Wnioski:

Dla każdej płci jest bardzo podobna ilość klientów (około 50%) i nie widać tutaj żadnego trendu na którąś z płci. Widać standardowy nagły wzrost w wielkości liczb w roku 2013 (tak jak w poprzednich zadaniach spowodowany ten wzrost jest lepszym dla firmy okresem sprzedażowym). Warto dodać że dużo osób ma wpisaną płeć jako NULL, więc dane zostały przefiltrowane, tak żeby osoby bez wpisanej płci wykluczyć z zestawienia dla zwiększenia jego czytelności.

Zad 5.

A screenshot of a computer

AI-generated content may be incorrect.

Dane są w formacie, który łatwo przełożyć na czytelny wykres:  
A screenshot of a computer

AI-generated content may be incorrect.

A screen shot of a computer program

AI-generated content may be incorrect.

Wnioski:

W kategorii Accessories dominują kaski – mają one zarówno największy udział ilościowy (31,55%), jak i wartośćowy (38,05%), inne produkty takie jak butelki czy bagażniki prezentują znaczne różnice między udziałem ilościowym a wartością i są daleko z tyłu za kaskami.

W kategorii Bikes dominuje segment rowerów szosowych, który odpowiada za ponad połowę sprzedaży pod względem ilości i blisko połowy wartości, z rowerami górskimi zajmującymi drugie miejsce.

W kategorii Clothing najważniejsze są koszulki (jerseys), które osiągają największe udziały zarówno ilościowe (30,83%), jak i wartość (35,47%), zaś akcesoria takie jak czapki, mimo wyższego udziału ilościowego, generują znacznie mniejszą wartość.

W kategorii Components największy udział mają ramy – zarówno górskie, jak i szosowe, które łącznie stanowią ponad 47% sprzedaży ilościowej i aż 72% wartości sprzedaży, co podkreśla ich kluczowe znaczenie w strukturze sprzedaży.

Zad 6.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

Wnioski:

Jedyna waluta w jakiej występują wszystkie opłaty to USD. W każdym z Territory jest ona obowiązującą walutą.