

Hurtownie danych

Lista 6 – Hierarchie, partycje, prosta analiza danych



Aleksander stepaniuk

nr. indeksu: 272644

Politechnika Wrocławska, Informatyka Stosowana

**Rozwiązania:**

Zadanie 1.

a)

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

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

b)

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

AI-generated content may be incorrect.

c)

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a black screen

AI-generated content may be incorrect.

d)

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Zad 2.

a)

A screenshot of a computer

AI-generated content may be incorrect.

b)

A screenshot of a computer

AI-generated content may be incorrect.

c)

1. Suma wartości sprzedaży rowerów w poszczególnych miesiącach

A screenshot of a computer

AI-generated content may be incorrect.

2. Maksymalny rabat udzielony na rowery w poszczególnych miesiącach

A screenshot of a computer

AI-generated content may be incorrect.

3. Ilość różnych klientów w poszczególnych terytoriach

A screenshot of a computer

AI-generated content may be incorrect.

4. Maksymalna ilość zamówionych produktów jednego rodzaju na przestrzeni lat i kwartałów

A screenshot of a graph

AI-generated content may be incorrect.

5. Ilość różnych klientów w poszczególnych dniach tygodnia

A screenshot of a computer

AI-generated content may be incorrect.

Zad 3.

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

AI-generated content may be incorrect.

Zad 4.

a)

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

b)

A screenshot of a computer

AI-generated content may be incorrect.

A screen shot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**Wnioski:**

**Wykres 1:**

Widać, że największa sprzedaż rowerów jest w marcu, jest to w końcu pierwszy miesiąc wiosny, a także sezonu rowerowego. Wysoka sprzedaż jest również w miesiącach taki jak styczeń, maj, październik co może być spowodowane na przykład promocjami noworocznymi, promocjami na majówkę, albo wyprzedażą na koniec sezonu. Najmniejsza sprzedaż jest w lutym oraz kwietniu, co może być spowodowaną zimną lub zmienną pogodą.

**Wykres 2:**

Widać, że największe rabaty udzielane są w miesiącach ciepłych na początku sezonu rowerowego czyli podczas wiosny (marzec, kwiecień, maj) – później maksymalne rabaty stopniowo się zmniejszają i utrzymują stały poziom od września do lutego. Widać tutaj wyraźnie strategię rabatową sklepu rowerowego, który to w sezonie rowerowym ustawia wysokie promocje na rowery, a poza sezonem te promocje są o wiele mniejsze i wynoszą maksymalnie 5%.

**Wykres 3:**

Widać, że jest o wiele więcej unikalnych klientów w Australii, czy poszczególnych regionach zachodnich Ameryki, niż w pozostałych regionach – pozwala to zobaczyć skalę rynków w poszczególnych regionach. Widać też wyraźnie, że Ameryka w znacznej wielkości posiada klientów na zachodzie, gdzie zarówno centrum jak i wschód zawiera jedynie pojedyncze przypadki, znacznie odstające od reszty.

**Wykres 4:**

Widać znaczący pik w kwartale drugim oraz trzecim nałożony jednocześnie z ogólnym wzrostem i rozwojem sprzedaży sklepu w latach 2012-2013. Najlepszym kwartałem okazał się pierwszy kwartał 2013 roku, gdzie było historycznie duże zamówienie.

**Wykres 5:**

Widać, że ilość różnych klientów ma zdecydowany pik w poniedziałek oraz najniższą swoją wartość osiąga w piątek. Jest to nieintuicyjne i może być spowodowane tym, że poniedziałek jest pierwszym dniem roboczym, więc ludzie po weekendzie decydują się zrobić zakupy w sklepie rowerowym.

**Zad 4:**

Podział kostki na partycje pozwala szybciej przetwarzać kostkę oraz poprawia wydajność zapytań, ponieważ jeśli zapytanie nie wymaga danych z innych partycji nie ma potrzeby żeby przeszukiwać pozostałe partycje. Pierwszy sposób na tworzenie partycji jest lepszy, kiedy dane źródłowe już są podzielone na tabele w pewien sposób (na przykład podzielone w czasie na równe odstępy jak kwartały czy lata) – nie istnieje wtedy potrzeba tworzenia nowych tabel w bazie ani ich łączenia. Jeśli jednak dane nie są podzielone (czyli istnieje jedna tabela faktów) to wtedy lepiej się sprawdzi sposób drugi, ponieważ nie trzeba wtedy tworzyć nowych tabel w bazie.