

HURTOWNIE DANYCH

Laboratorium 4

Maciej Kopiński 254578

Zad 1. Przygotowanie schematu

W bazie danych należy utworzyć schemat, którego nazwa będzie odpowiadać nazwisku wykonującego ćwiczenie (zapisać zapytanie tworzące ten schemat).

Zad 2. Tworzenie tabel wymiarów i tabeli faktów

W nowo utworzonym schemacie utworzyć tabele wymiarów: klienta, produktu i sprzedawcy (zapisać skrypt `CREATE TABLE`), opisane w następujących schematach:

- DIM_CUSTOMER (CustomerID, FirstName, LastName, Title, City, TerritoryName, CountryRegionCode, Group)
- DIM_PRODUCT (ProductID, Name, ListPrice, Color, SubCategoryName, CategoryName, Weight, Size, IsPurchased)
- DIM SALESPERSON (SalesPersonID, FirstName, LastName, Title, Gender, CountryRegionCode, Group)

oraz tabelę faktów:

- FACT_SALES (ProductID, CustomerID, SalesPersonID, OrderDate, ShipDate, OrderQty, UnitPrice, UnitPriceDiscount, LineTotal)

Uwaga 1: Proszę sprawdzić, dlaczego niektóre atrybuty mają wartość NULL. Czy zawsze wynika to z braku danych?

Uwaga 2: Kolumny OrderDate oraz ShipDate przechowują dane typu całkowitego, gdzie cztery pierwsze cyfry oznaczają rok, dwie następne miesiąc, a dwie ostatnie dzień. Do pobrania poszczególnych części daty użyć funkcji datepart.

Uwaga 3: Atrybuty OrderQty oraz LineTotal będą miarami w naszej hurtowni.

Zad. 3. Wypełnianie danych – denormalizacja źródłowej bazy

Wypełnić nowoutworzone tabele danymi znajdującymi się w tabelach źródłowych. Do wypełnienia użyć instrukcji `INSERT INTO`. Proszę sprawdzić liczbę skopiowanych rekordów.

Uwaga 1. Do tabeli DIM_PRODUCT należy także skopiować produkty, które nie mają przypisanej podkategorii.

Uwaga 2. Do tabeli FACT_SALES należy skopiować również transakcje, które nie mają sprzedawcy.

Zad. 4. Więzy integralności

1. Dodać integralność referencyjną i klucze główne do tabel już zdefiniowanych.
2. Przygotować instrukcję INSERT INTO, która sprawdzi poprawność integralności referencyjnej oraz klucze główne.

Zad. 5. Tworzenie kostki

Należy utworzyć projekt Analysis Services, w którym zostanie przygotowana kostka zawierająca utworzone wymiary (klienta, produktu i sprzedawcy) oraz tabelę faktów.

Używając Visual Studio utworzyć projekt typu Analysis Services Project (Menu File -> New Project)

- a) Dodać źródło danych (Solution Explorer -> Data Sources -> New Data Source), które będzie wskazywało na bazę danych, która przechowuje tabele faktów i wymiarów. Przeanalizować opcje związane z bezpieczeństwem dostępu do danych.
- b) Utworzyć nowy widok źródła danych (Solution Explorer -> Data Source Views -> New Data Source View). Dodać wcześniej utworzone tabele.
- c) Utworzyć nową kostkę za pomocą asystenta (Solution Explorer -> Cubes -> New Cube):
 - o Wybrać utworzenie kostki na podstawie istniejących tabel (Use existing tables)
 - o Wybrać, utworzony w poprzednim punkcie, widok źródła danych
 - o Jako tabelę faktów (Measure group tables) wybrać FACT_SALES
 - o Na stronie dotyczącej miar wybrać OrderQty, UnitPriceDiscount, Line Total. Zastanowić się nad użytecznością wybranych miar. Dlaczego nie wszystkie atrybuty tabeli FACT_SALES mogą być użyte jako miary?
 - o Na stronie dotyczącej wymiarów wybrać wszystkie tabele z przedrostkiem DIM.

Po utworzeniu kostki dokonać edycji wymiarów (Solution Explorer -> Dimensions lub zakładka Cube Structure -> Dimensions).

Dla każdego z wymiarów zdefiniować potrzebne atrybuty. Przykładowo wymiar produkt powinien zawierać: Nazwę, Cenę, Kolor, Podkategorię i Kategorię. W przypadku tabeli Produkt należy zmienić jej definicję i sposób ładowania, tak aby zawierała ona zarówno kategorię jak i podkategorię (jeśli nie zrobiło się tego już podczas pracy z kreatorem dodawania nowej kostki).

Zad. 6. Uruchomienie kostki

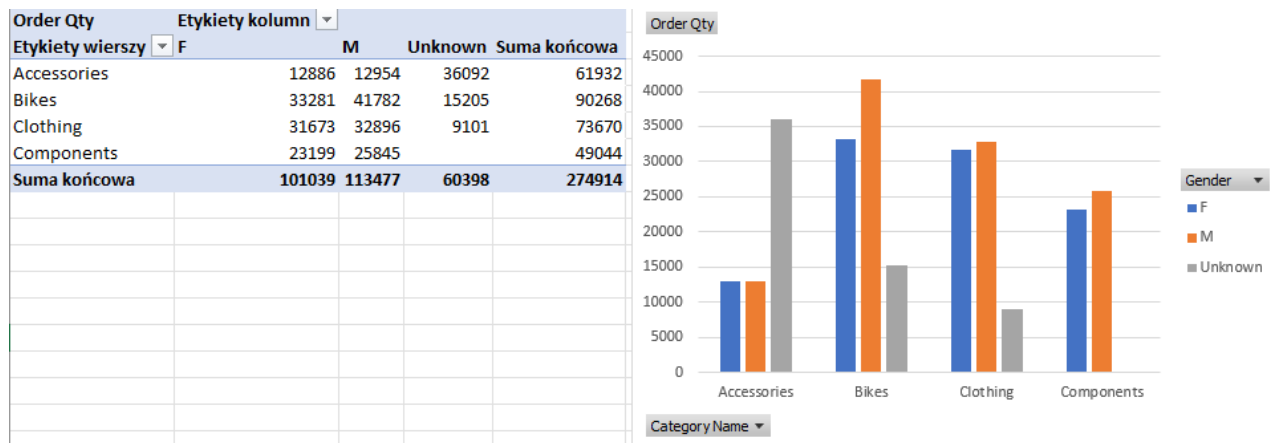
- a) Sprawdzić ustawienia projektu (Solution Explorer -> Nazwa projektu -> Properties).
- b) Sprawdzić ustawienia konfiguracji wdrożenia (Configuration properties -> Deployment -> Target) projektu – Server: localhost, Database: PierwszyProjektKostek
- c) Wdrożyć projekt (Menu Build -> Deploy). W przypadku błędu:
 - w konfiguracji wdrożenia (Configuration properties -> Deployment -> Target) przekleić nazwę serwera (z ekranu wyboru serwera w MSSQL Management Studio);
 - klikając dwukrotnie na utworzony Data Source, sprawdzić działanie innych opcji „Impersonation information” (VS powinien być uruchamiany z konta o uprawnieniach administratora);
- d) Przetworzyć kostkę (Przycisk Process).

Zad. 7. Proste raporty

Połączyć się z programem MS Excel i przygotować raporty (tabele i wykresy przestawne), w których zostaną ujęte ciekawe zależności pomiędzy danymi.

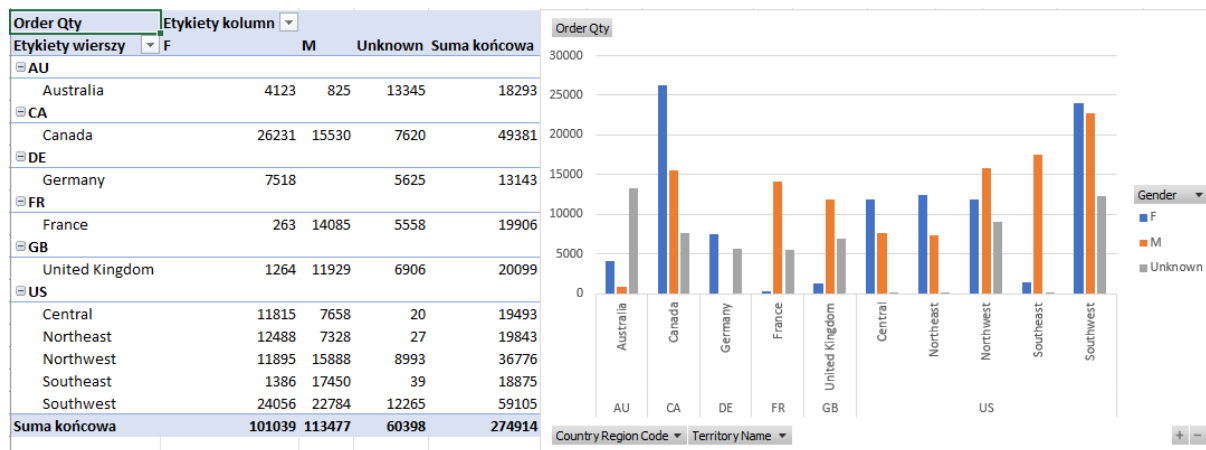
Wnioski:

Wykres liczby sprzedanych produktów dla danej kategorii w zależności od płci



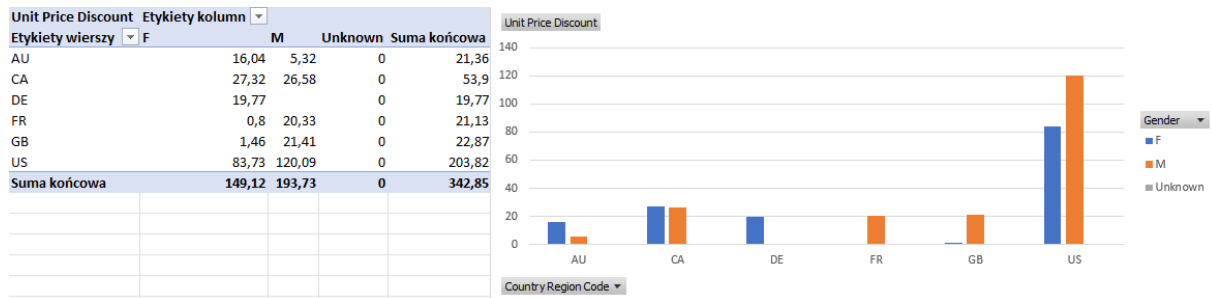
Z powyższego wykresu widać, że akcesoria kupowane są głównie od sprzedawców bez określonej płci, co oznacza, że były kupowane przez internet.

Wykres liczby sprzedanych produktów dla kodu kraju oraz nazwy terytorium w zależności od płci sprzedawcy



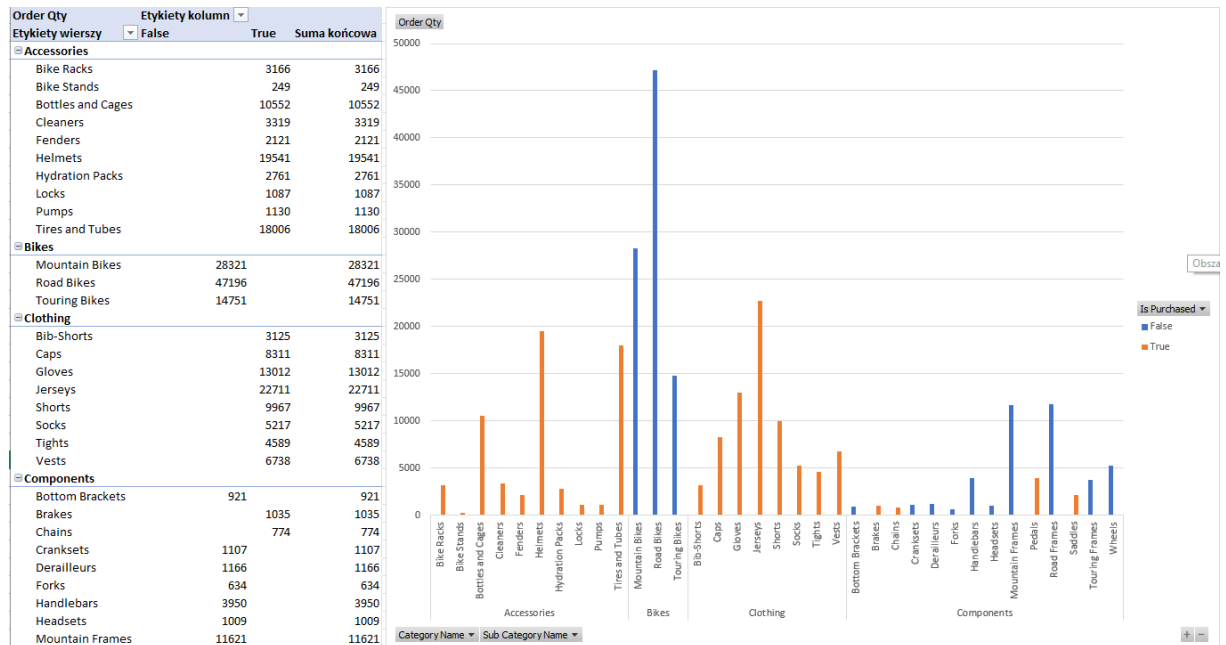
W ponad połowie regionów kobiety sprzedają więcej produktów niż mężczyźni, chociaż dla Francji, Wielkiej Brytanii oraz w południowo wschodniej części Stanów Zjednoczonych kobiety sprzedają o wiele mniej.

Wykres sumy zniżek dla kodów krajów w zależności od płci sprzedawcy



Na produkty kupowane przez internet nie są udzielane zniżki.

Wykres liczby sprzedanych produktów w zależności od ich pochodzenia (kupione/wyprodukowane)



Produkty z wszystkich podkategorii akcesoriów oraz ubrań są kupowane przez sklep. Rowery z wszystkich podkategorii są produkowane. Większość komponentów jest produkowana, za wyjątkiem komponentów z podkategorii: hamulce, łańcuchy, pedały oraz siodełka.