Hurtownie danych  
Laboratorium Czw 11:15  
  
Lista 4  
  
Kajetan Pynka 254495

**Zad 1**

CREATE SCHEMA Pynka;

**Zad 2**

CREATE TABLE Pynka.DIM\_CUSTOMER (

    CustomerID INT NOT NULL,

    FirstName NVARCHAR(50),

    LastName NVARCHAR(50),

    Title NVARCHAR(8),

    City NVARCHAR(30),

    TerritoryName NVARCHAR(50),

    CountryRegionCode NVARCHAR(3),

    [Group] NVARCHAR(50)

);

CREATE TABLE Pynka.DIM\_PRODUCT (

    ProductID INT NOT NULL,

    Name NVARCHAR(50) NOT NULL,

    ListPrice MONEY NOT NULL,

    Color NVARCHAR(15),

    SubCategoryName NVARCHAR(50),

    CategoryName NVARCHAR(50),

    Weight DECIMAL(8,2),

    Size NVARCHAR(5),

    IsPurchased BIT NOT NULL

);

CREATE TABLE Pynka.DIM\_SALESPERSON (

    SalesPersonID INT NOT NULL,

    FirstName NVARCHAR(50) NOT NULL,

    LastName NVARCHAR(50) NOT NULL,

    Title NVARCHAR(8),

    Gender NCHAR(1) NOT NULL,

    CountryRegionCode NVARCHAR(3),

    [Group] NVARCHAR(50)

);

CREATE TABLE Pynka.FACT\_SALES (

    ProductID INT NOT NULL,

    CustomerID INT NOT NULL,

    SalesPersonID INT,

    OrderDate INT NOT NULL,

    ShipDate INT NOT NULL,

    OrderQty SMALLINT NOT NULL,

    UnitPrice MONEY NOT NULL,

    UnitPriceDiscount MONEY NOT NULL,

    LineTotal NUMERIC(38,6) NOT NULL

);

Wszelkie typy danych oraz obligatoryjność zostały zaczerpnięte z oryginalnych danych (za wyjątkiem dat, które zostały sklejone w liczby całkowite). Klucze główne i obce dodaję dopiero w ramach zadania czwartego.

**Zad 3**

INSERT INTO Pynka.DIM\_CUSTOMER

SELECT DISTINCT C.CustomerID, P.FirstName, P.LastName, P.Title, MIN(A.City), MIN(ST.Name), MIN(ST.CountryRegionCode), MIN(ST.[Group])

    FROM Sales.Customer C

    JOIN Person.Person P ON P.BusinessEntityID=C.PersonID

    LEFT JOIN Person.BusinessEntityAddress BEA ON BEA.BusinessEntityID=P.BusinessEntityID

    LEFT JOIN Person.Address A ON A.AddressID=BEA.AddressID

    LEFT JOIN Sales.SalesTerritory ST ON ST.TerritoryID=C.TerritoryID

    GROUP BY C.CustomerID, P.FirstName, P.LastName, P.Title;

INSERT INTO Pynka.DIM\_PRODUCT

SELECT DISTINCT P.ProductID, P.Name, P.ListPrice, P.Color, PSC.Name, PC.Name, P.Weight, P.Size,

    CASE

        WHEN (SELECT COUNT(\*) FROM Sales.SalesOrderDetail SOD WHERE SOD.ProductID=P.ProductID) > 0 THEN 1 ELSE 0

    END "IsPurchased"

    FROM Production.Product P

    LEFT JOIN Production.ProductSubcategory PSC ON PSC.ProductSubcategoryID=P.ProductSubcategoryID

    LEFT JOIN Production.ProductCategory PC ON PC.ProductCategoryID=PSC.ProductCategoryID

    WHERE (SELECT COUNT(\*) FROM Sales.SalesOrderDetail SOD WHERE SOD.ProductID=P.ProductID) > 0;

INSERT INTO Pynka.DIM\_SALESPERSON

SELECT DISTINCT SP.BusinessEntityID, P.FirstName, P.LastName, P.Title, E.Gender, ST.CountryRegionCode, ST.[Group]

    FROM Sales.SalesPerson SP

    LEFT JOIN Person.Person P ON P.BusinessEntityID=SP.BusinessEntityID

    LEFT JOIN HumanResources.Employee E ON E.BusinessEntityID=SP.BusinessEntityID

    LEFT JOIN Sales.SalesTerritory ST ON ST.TerritoryID=SP.TerritoryID;

INSERT INTO Pynka.FACT\_SALES

SELECT DISTINCT SOD.ProductID, SOH.CustomerID, SOH.SalesPersonID,

    CAST(FORMAT(SOH.OrderDate, 'yyyy') + FORMAT(SOH.OrderDate, 'MM') + FORMAT(SOH.OrderDate, 'dd') AS INT),

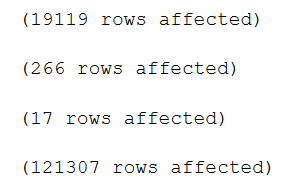
    CAST(FORMAT(SOH.ShipDate, 'yyyy') + FORMAT(SOH.ShipDate, 'MM') + FORMAT(SOH.ShipDate, 'dd') AS INT),

    SOD.OrderQty, SOD.UnitPrice, SOD.UnitPriceDiscount, SOD.LineTotal

    FROM Sales.SalesOrderHeader SOH

    LEFT JOIN Sales.SalesOrderDetail SOD ON SOD.SalesOrderID=SOH.SalesOrderID;

Liczba wstawionych rekordów (po kolei dla: DIM\_CUSTOMER, DIM\_PRODUCT, DIM\_SALESPERSON, FACT\_SALES):

****

Ignoruję klientów bez podanych danych osobistych oraz produkty, które nie zostały ani razu sprzedane.

**Zad 4.1**

ALTER TABLE Pynka.DIM\_CUSTOMER

ADD CONSTRAINT dim\_cust\_pk PRIMARY KEY (CustomerID);

ALTER TABLE Pynka.DIM\_PRODUCT

ADD CONSTRAINT dim\_prod\_pk PRIMARY KEY (ProductID);

ALTER TABLE Pynka.DIM\_SALESPERSON

ADD CONSTRAINT dim\_salespers\_pk PRIMARY KEY (SalesPersonID);

ALTER TABLE Pynka.FACT\_SALES

ADD CONSTRAINT fact\_sales\_fk\_prod FOREIGN KEY (ProductID) REFERENCES Pynka.DIM\_PRODUCT(ProductID);

ALTER TABLE Pynka.FACT\_SALES

ADD CONSTRAINT fact\_sales\_fk\_cust FOREIGN KEY (CustomerID) REFERENCES Pynka.DIM\_CUSTOMER(CustomerID);

ALTER TABLE Pynka.FACT\_SALES

ADD CONSTRAINT fact\_sales\_fk\_salesper FOREIGN KEY (SalesPersonID) REFERENCES Pynka.DIM\_SALESPERSON(SalesPersonID);

**Zad 4.2**

INSERT INTO Pynka.DIM\_CUSTOMER VALUES(11242, 'Josh', 'Test', NULL, 'Baltimore', 'Central', 'US', 'North America');

INSERT INTO Pynka.DIM\_PRODUCT VALUES(710, 'TEST', 123.40, 'Silver', NULL, NULL, 3.50, 40, 1);

INSERT INTO Pynka.DIM\_SALESPERSON VALUES(280, 'John', 'Brooks', 'Mr.', 'M', 'US', 'North America');

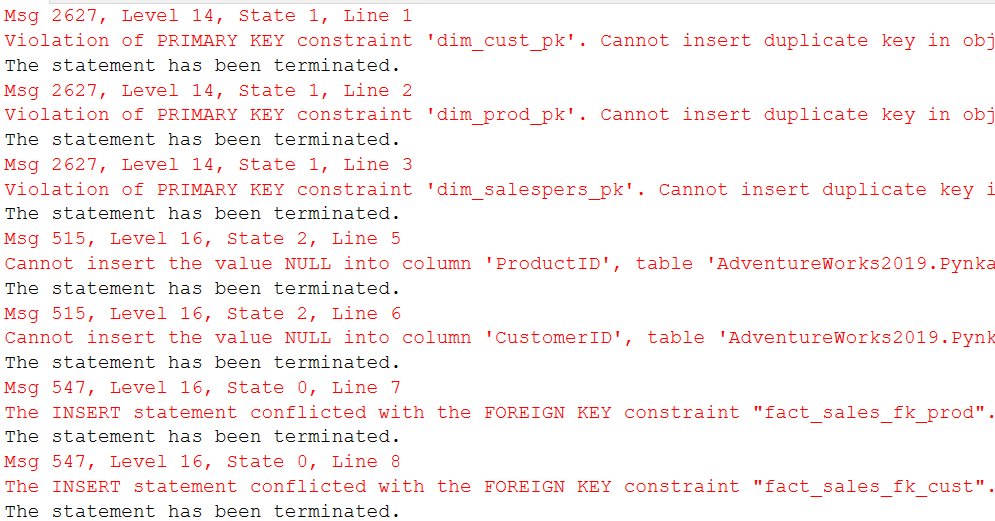
INSERT INTO Pynka.FACT\_SALES VALUES(NULL, 11185, NULL, 20160220, 20160323, 2, 34.90, 0.00, 34.90);

INSERT INTO Pynka.FACT\_SALES VALUES(707, NULL, NULL, 20160123, 20160131, 2, 34.90, 0.00, 34.90);

INSERT INTO Pynka.FACT\_SALES VALUES(1500, 11185, NULL, 20150404, 20150416, 2, 34.90, 0.00, 34.90);

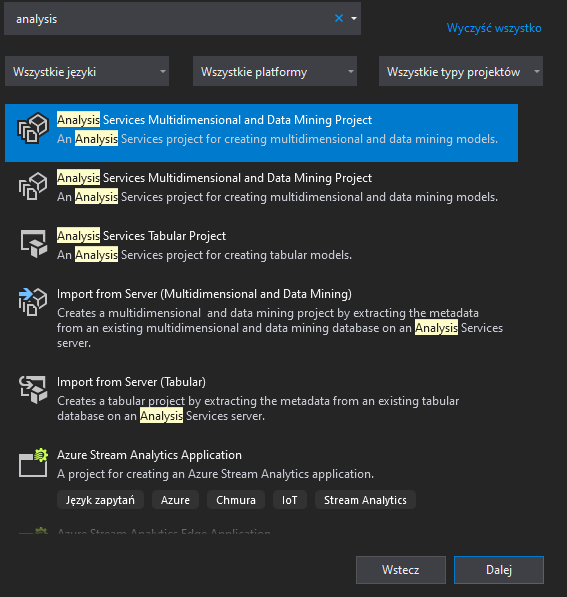
INSERT INTO Pynka.FACT\_SALES VALUES(707, 40000, NULL, 20130712, 20130720, 2, 34.90, 0.00, 34.90);

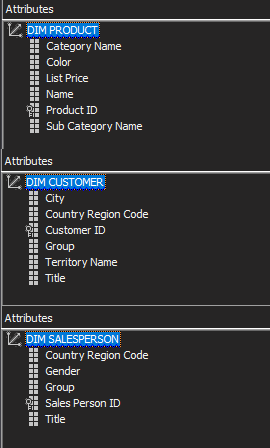
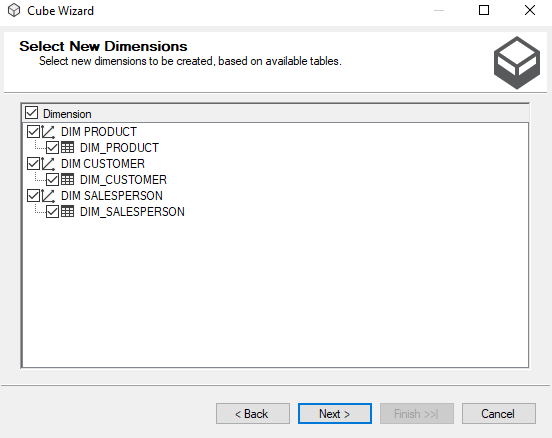
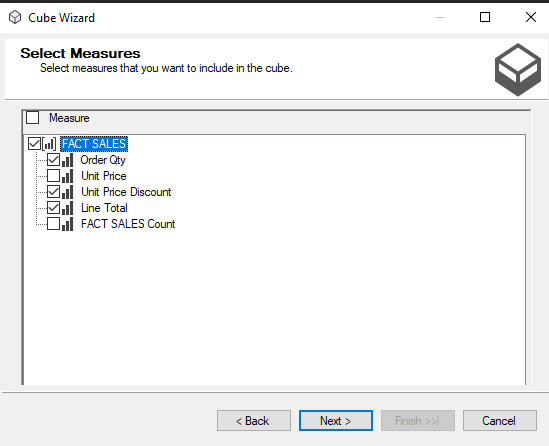
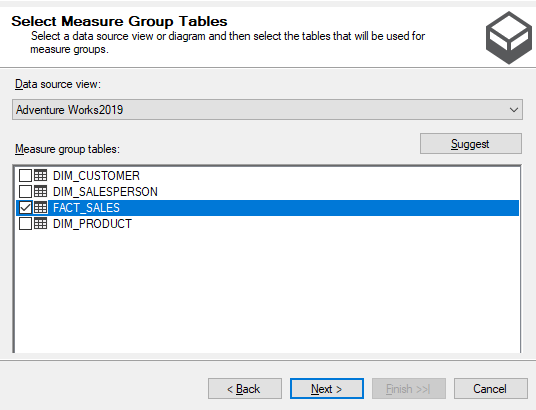
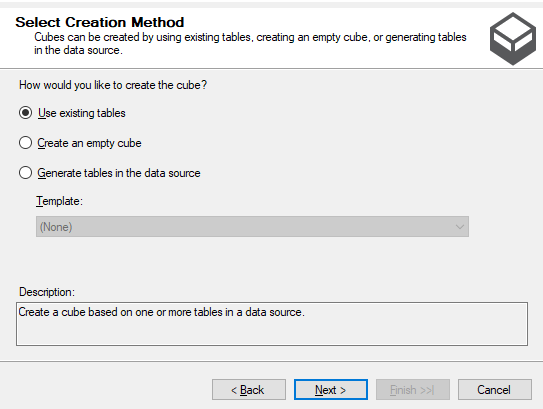
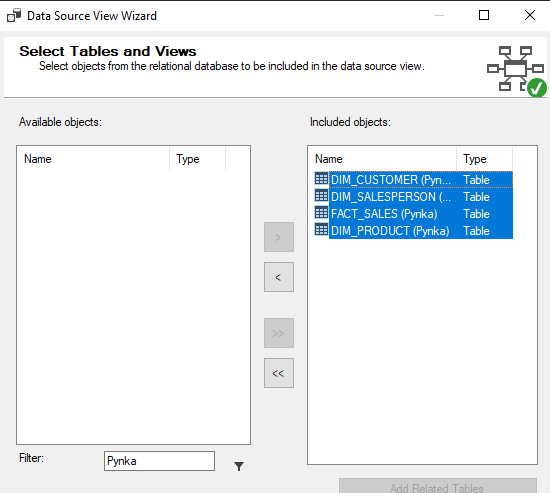
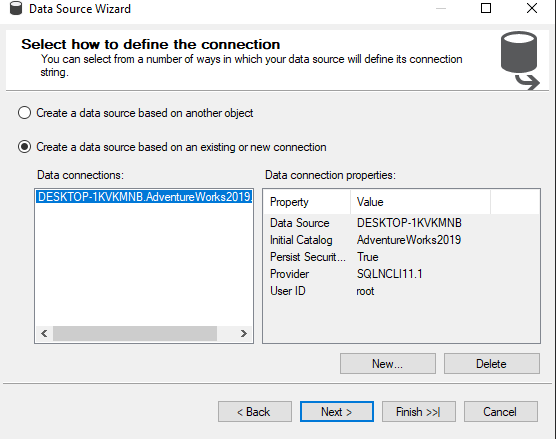
Komunikaty zwrócone przez SZBD (dla powyższych zapytań INSERT INTO):

****

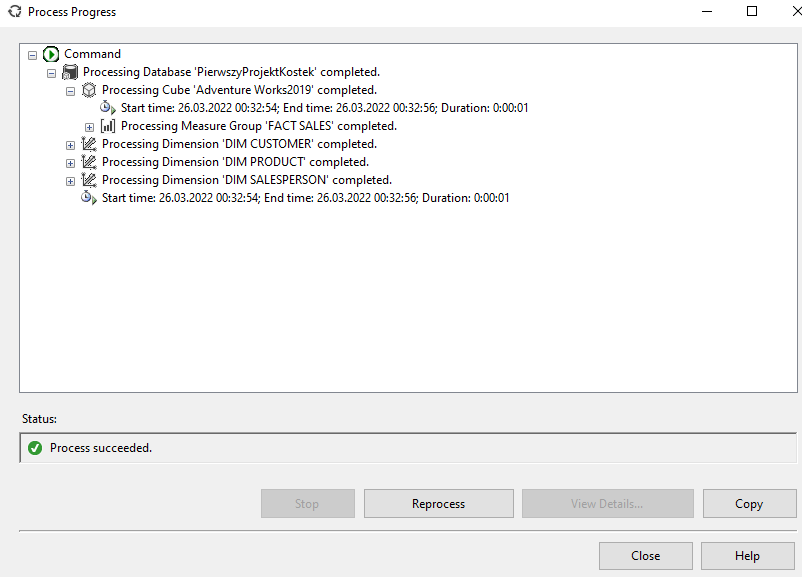
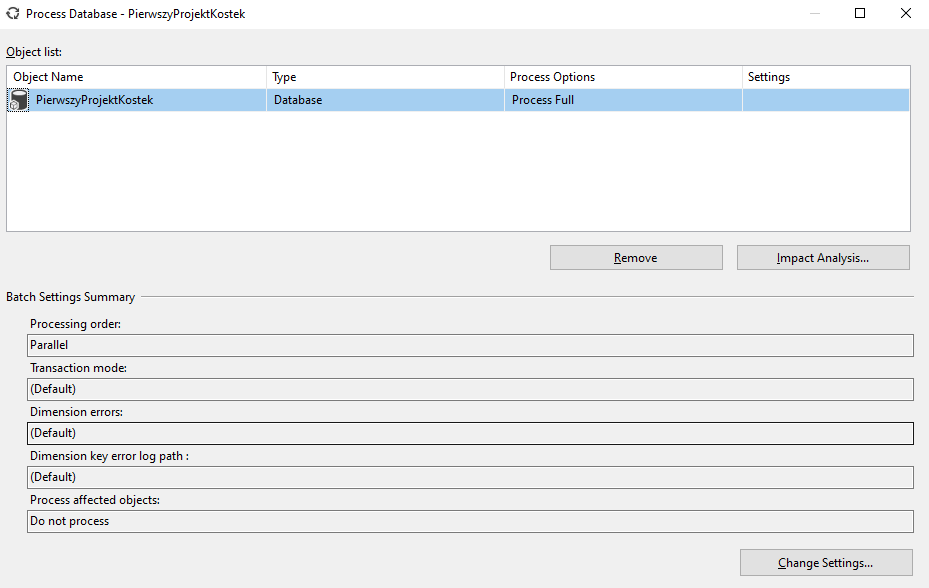
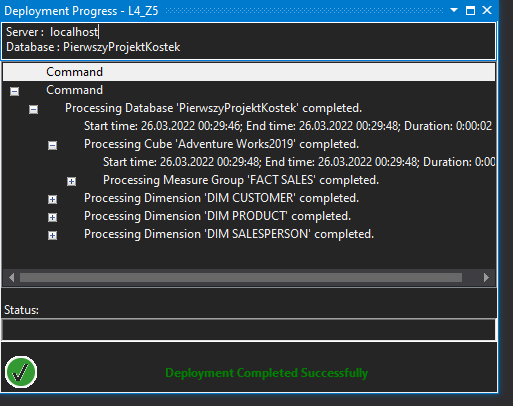
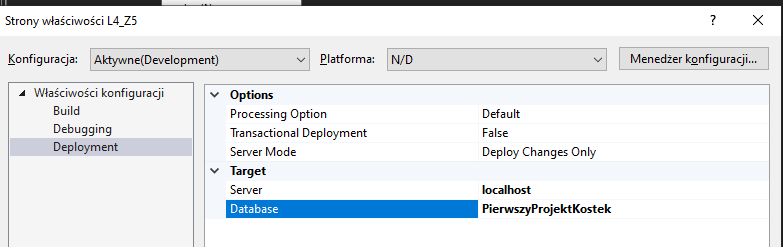
**Wniosek:** Wszystkie ograniczenia zostały poprawnie nałożone.

**Zad 5**

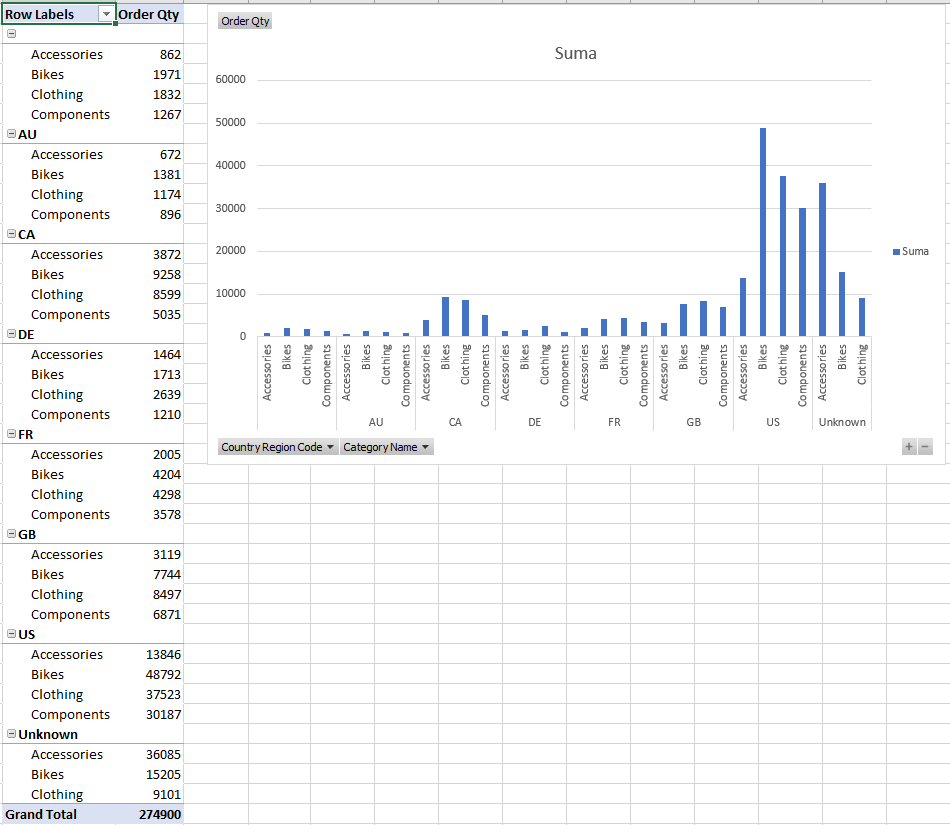
****

****

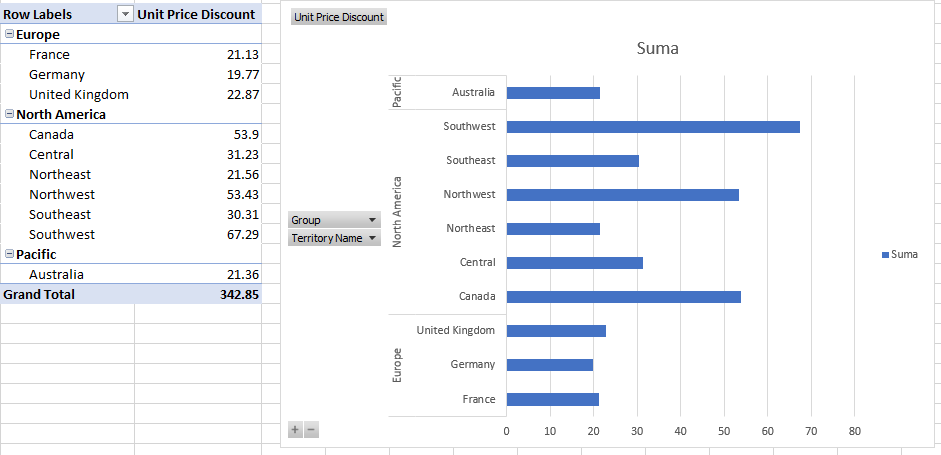
**Zad 6**

****

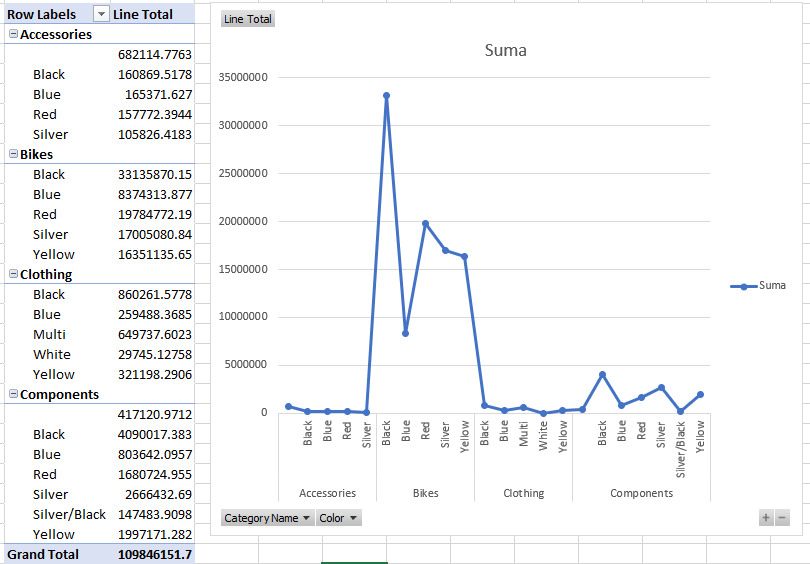
**Zad 7**

****

**Wniosek:** Najwięcej sztuk produktów sprzedawanych jest w Stanach Zjednoczonych, z czego najrzadziej kupowane są akcesoria. W przypadku nieznanego regionu natomiast akcesoria stanowią znaczną część. W pozostałych regionach ogólnie najrzadziej kupowane są akcesoria, najczęściej rowery i ubrania.

****

**Wniosek:** Na podstawie tych danych można zastanawiać się nad użytecznością miary UnitPriceDiscount. Wiadomo, że sumarycznie najwięcej przecen pojawiło się na zachodzie Stanów Zjednoczonych i Kanadzie. Nie uwzględnia to jednak faktu, że w tych regionach było sprzedawanych najwięcej produktów (być może wypadałoby to uśrednić). Większy sens wydaje się mieć liczenie przecen (gdy UnitPriceDiscount > 0) niż sumowanie procentów.

****

**Wniosek:** W każdej kategorii najlepiej sprzedają się produkty w kolorze czarnym (jeśli zignorujemy produkty bez oznaczenia koloru). Prawdopodobnie najlepszym posunięciem byłoby wycofanie niebieskich rowerów i zwiększenie sprzedaży rowerów w pozostałych kolorach (ponieważ to rowery przynoszą ogólnie największe zyski).

**Wnioski:**

* Sporządzenie tabel wymiarów i faktów jest stosunkowo proste i sprowadza się do zwykłych zapytań SQL’owych. Istnieją dedykowane narzędzia i rozszerzenia, które pozwolą nam następnie utworzyć kostkę w bardzo prosty sposób.
* Przetworzoną kostkę możemy wykorzystać do analizowania posiadanych danych ze względu na dowolne wybrane przez nas wymiary i dowolne miary. Pozwala to wysuwać wnioski co do powiązania pewnych wartości i zależności występujących w hurtowni.