

Hurtownie danych – Sprawozdanie z zadania 6.

PWr. WIZ, Informatyka, Data: 06.04.2020

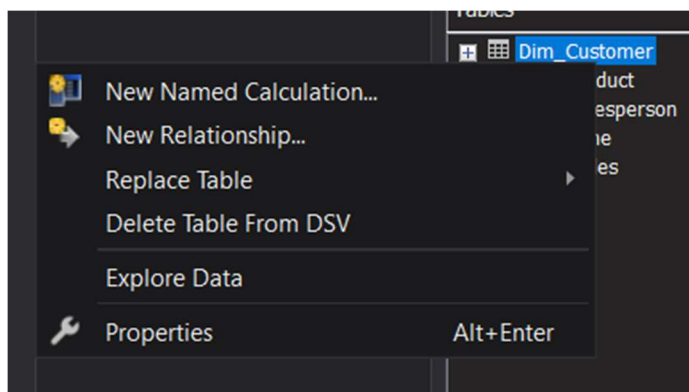
| | | |
|----------|---------------------------------------------------------------------------------|-------|
| Student | Email: 242493@student.pwr.edu.pl | Ocena |
| Indeks | <u>242493</u> | |
| Imię | <u>Arkadiusz</u> | |
| Nazwisko | <u>Rasz</u> | |

Spis treści

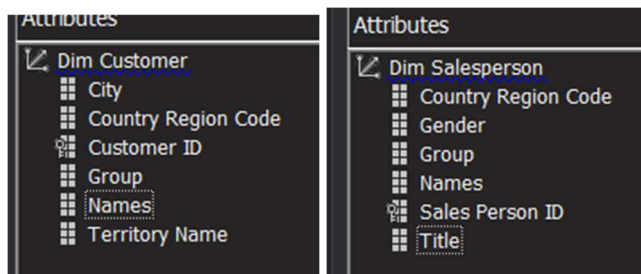
| | |
|-----------------------------------------------------|----|
| Zad. 1. Modyfikacja wymiarów i tabeli faktów | 2 |
| Zad. 2. Przegląd danych i tworzenie zestawień | 5 |
| Przykładowe tabele oraz wykresy | 6 |
| Zad. 3. Partycje..... | 9 |
| Zasilanie partycji za pomocą zapytań SQL | 9 |
| Zasilanie partycji za pomocą osobnych tabel | 10 |
| Zad. 4..... | 10 |
| Przykładowe wykresy:..... | 12 |

Zad. 1. Modyfikacja wymiarów i tabeli faktów

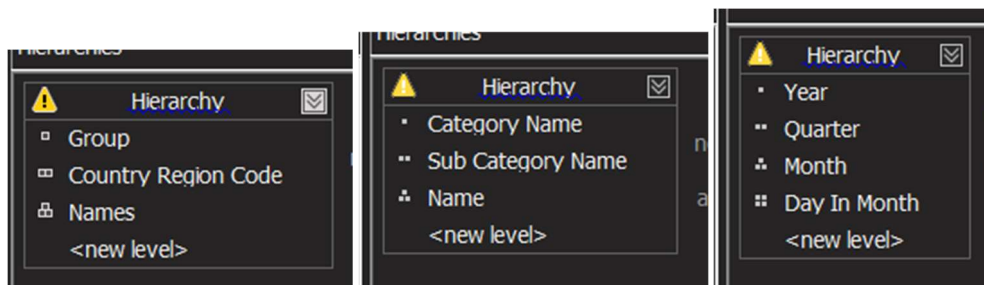
W wymiarach CUSTOMER oraz SALESPERSON dodano nowe atrybuty za pomocą nazwanej kalkulacji:



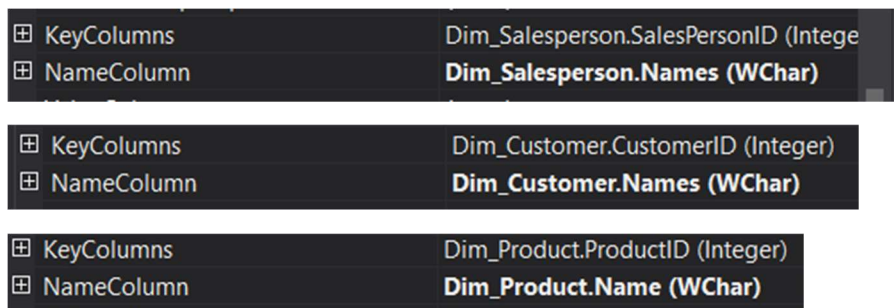
Zmodyfikowano atrybuty w wymiarach:



Utworzono hierarchie:



Do kluczy dodano nazwy:



Klucz tabeli Dim_Time pominięto.

Utworzono nowe miary:

New Measure

Usage: Distinct count

Source table: Fact_Sales

Source column:

- ProductID
- CustomerID
- SalesPersonID
- OrderDate
- ShipDate
- OrderQty
- UnitPrice
- UnitPriceDiscount
- LineTotal

☒ Show all columns

OK Cancel Help

New Measure

Usage: Distinct count

Source table: Fact_Sales

Source column:

- ProductID
- CustomerID
- SalesPersonID
- OrderDate
- ShipDate
- OrderQty
- UnitPrice
- UnitPriceDiscount
- LineTotal

☒ Show all columns

OK Cancel Help

New Measure

Usage: Maximum

Source table: Fact_Sales

Source column:

- OrderQty
- UnitPrice
- UnitPriceDiscount
- LineTotal

☐ Show all columns

OK Cancel Help

New Measure

Usage: Maximum

Source table: Fact_Sales

Source column:

- OrderQty
- UnitPrice
- UnitPriceDiscount
- LineTotal

☐ Show all columns

OK Cancel Help

New Measure

Usage: Distinct count

Source table: Fact_Sales

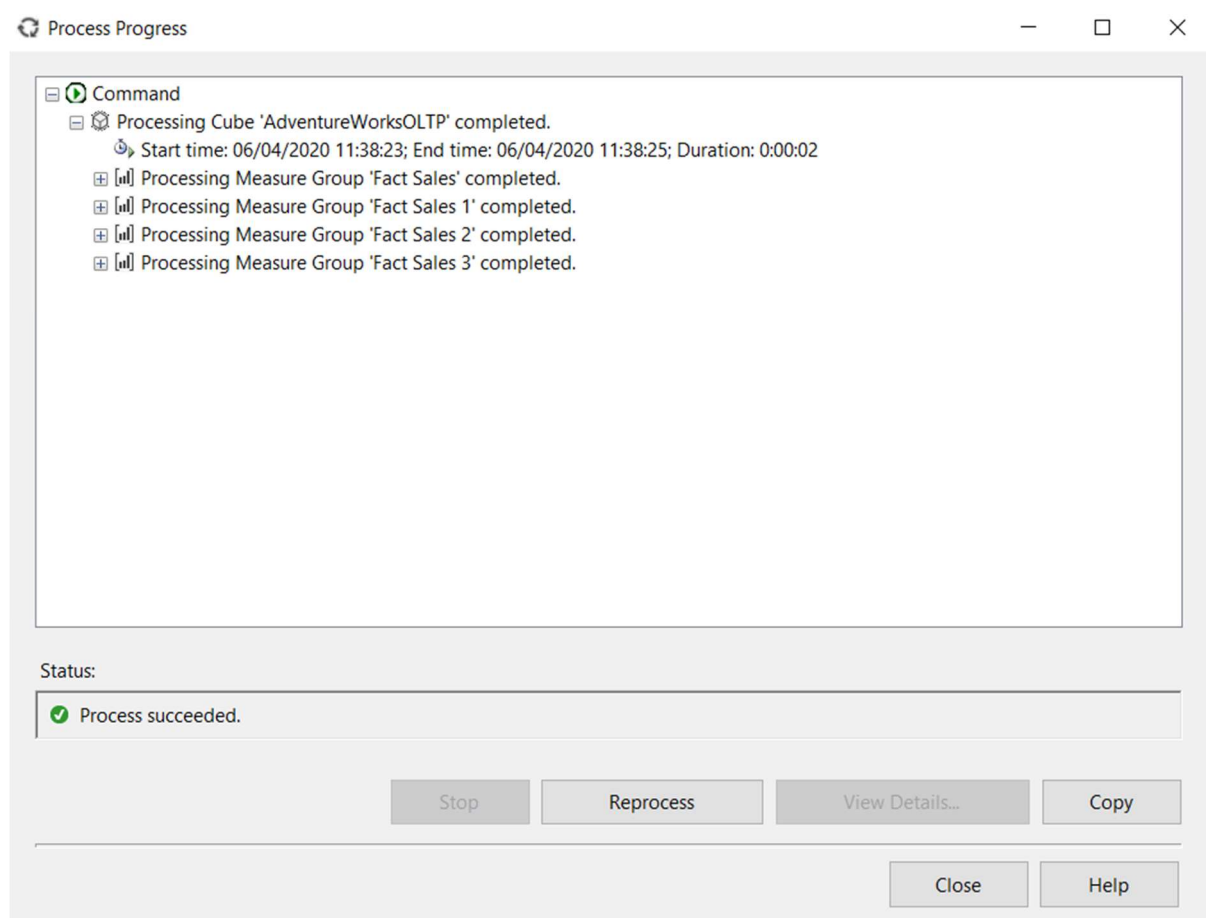
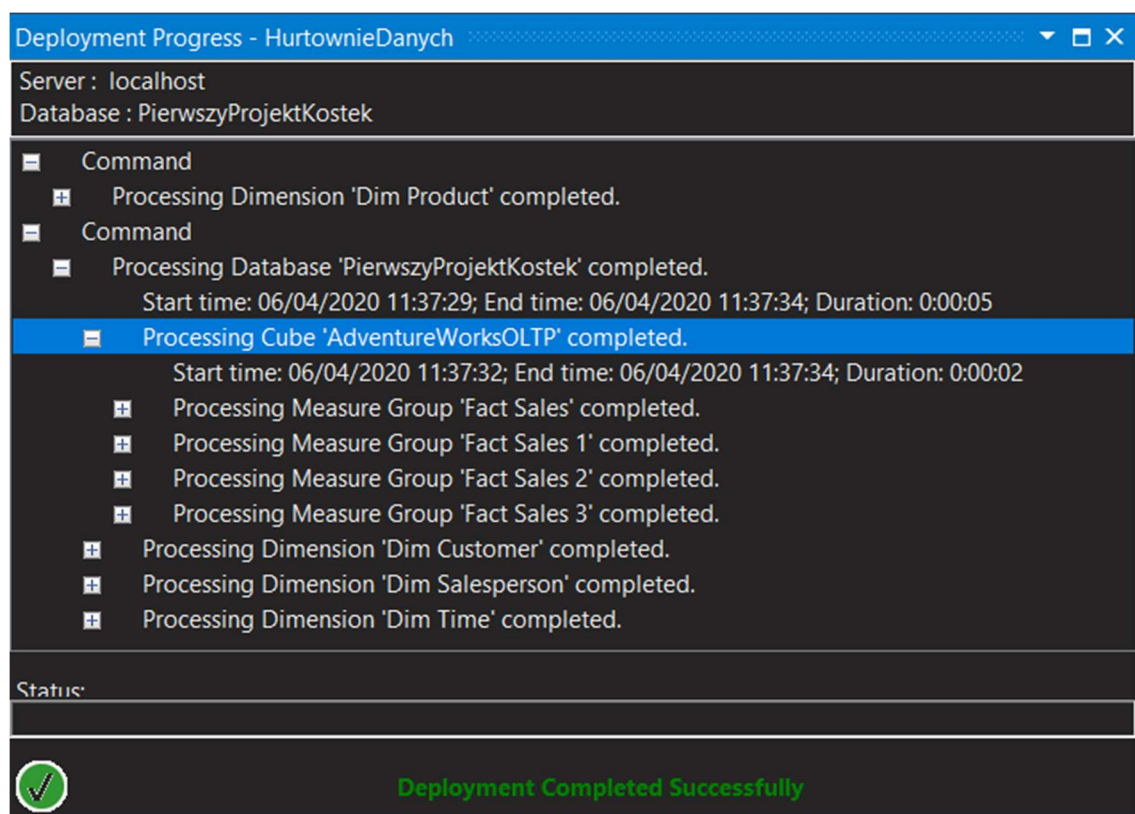
Source column:

- ProductID
- CustomerID
- SalesPersonID
- OrderDate
- ShipDate
- OrderQty
- UnitPrice
- UnitPriceDiscount
- LineTotal

☒ Show all columns

OK Cancel Help

Kostkę wdrożono oraz przeprocesowano:



Zad. 2. Przegląd danych i tworzenie zestawień

Sprawdzić, czy dane zapisane w kostce zgadzają się z danymi zapisanymi w tabelach, przeciągając za pomocą myszy.

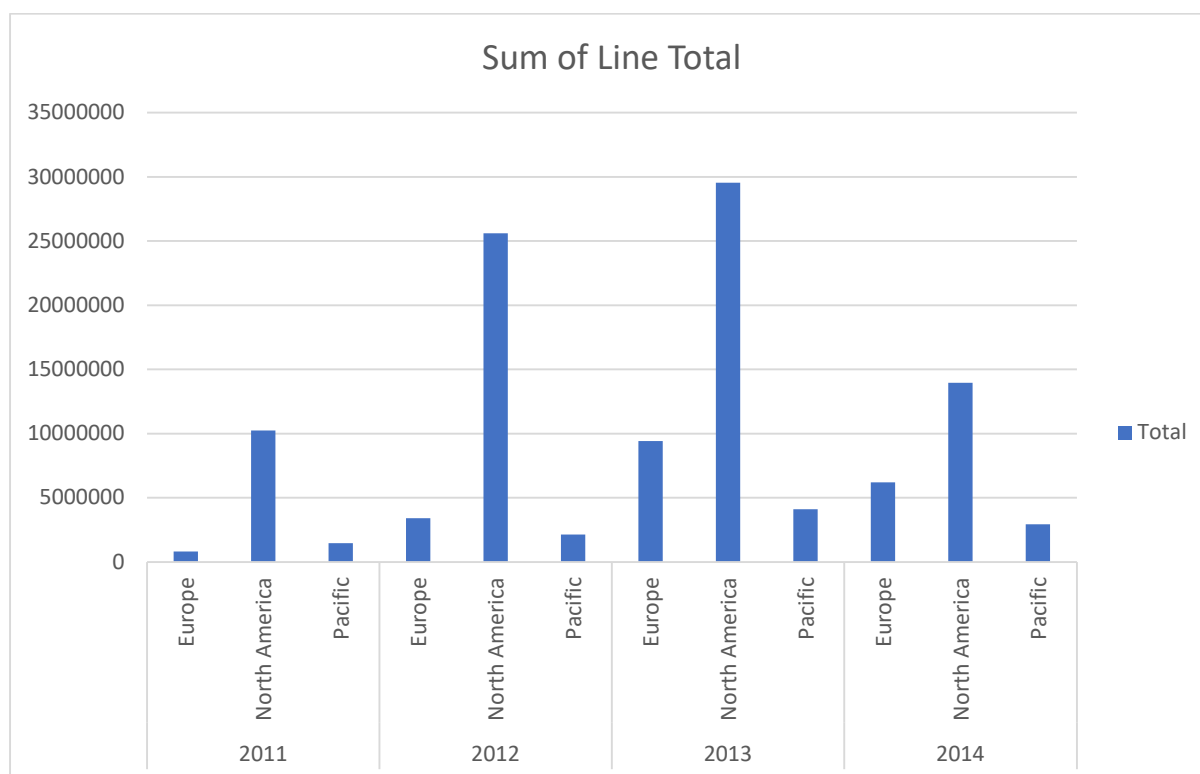
Przetestować możliwości przeglądarki (Browser) – operator wyboru danych (Operator), wyrażenia filtrujące dane (Filter Expression)

The screenshot displays the Microsoft SQL Server Data Tools (SSDT) environment. On the left, the 'Metadata' pane shows the 'AdventureWorksLT' database structure. The 'Dim Customer' dimension is expanded, revealing its hierarchy: City, Country Region Code, Customer ID, Group, Names, Territory Name, and Hierarchy. Below this, other dimensions like 'Dim Product', 'Dim Salesperson', and 'Order Date' are listed. The main area shows the 'Table Designer' for the 'FactSales' table. The table has three columns: 'Territory Name', 'Weekday', and 'Line Total'. The 'Table Properties' tab is active, showing the table name and other metadata. The 'Columns' tab shows the column definitions. The 'Table Designer' also includes a 'Table Properties' section with a 'Table Name' field set to 'FactSales'.

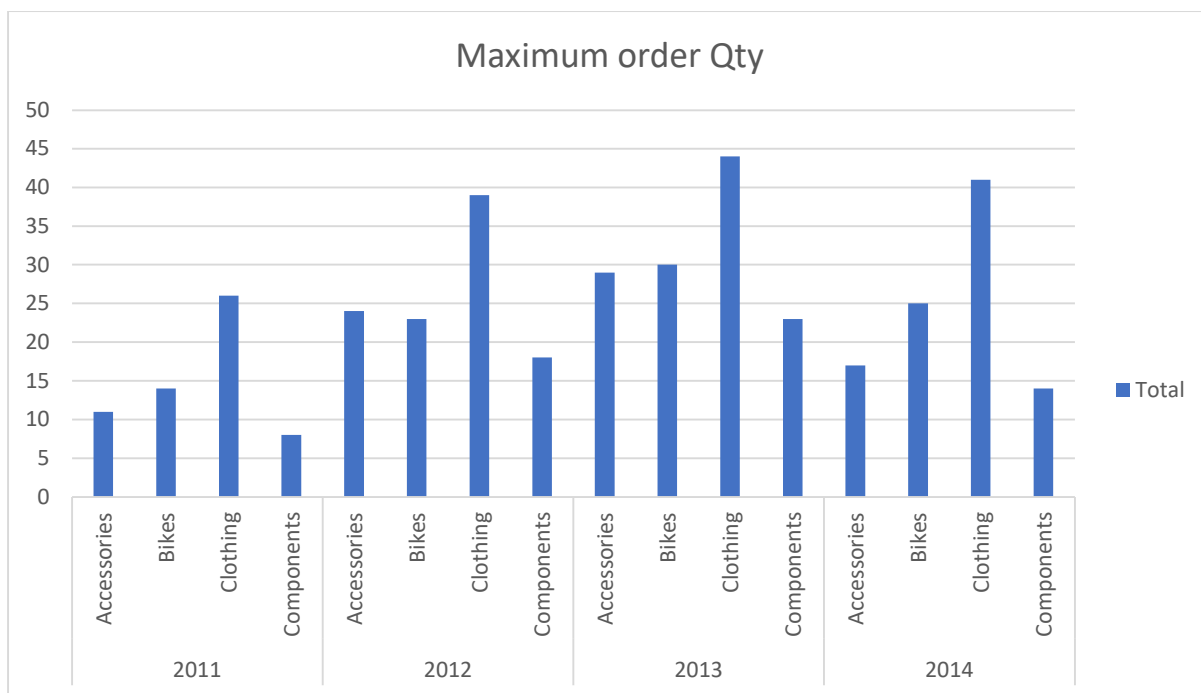
The screenshot displays the Microsoft SQL Server Data Tools (SSDT) interface for the AdventureWorksOLTP database. The left-hand pane shows the 'Measures' folder expanded, listing various measures such as 'Line Total', 'Maximum Order Qty', 'Maximum Unit Price Discount', 'Order Qty', 'Unit Price Discount', and several 'Fact Sales' folders. The right-hand pane shows the 'Dimension' table with columns: Dimension, Hierarchy, Operator, Filter Expression, and Param... The 'Ship Date' dimension is selected, showing its hierarchy as 'Ship Date.Hierarchy' and operator as 'Equal'. Below the table, there is a link that says 'Click to execute the query.'

Przykładowe tabele oraz wykresy

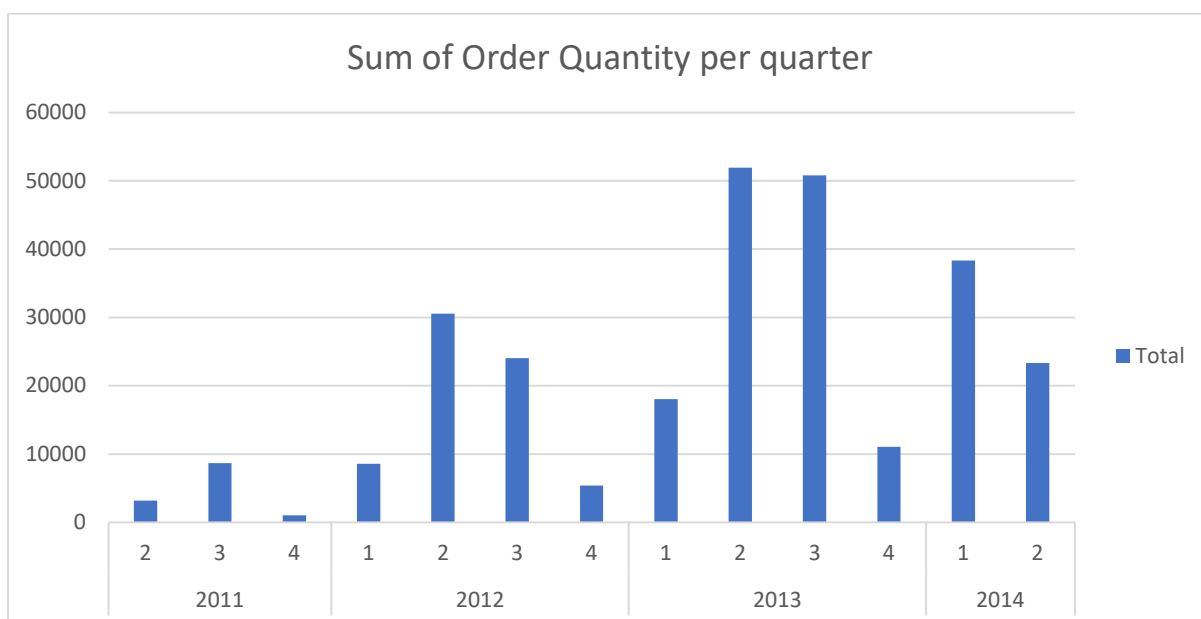
| Row Labels | Line Total | Utworzenie hierarchii pozwoliło na dalsze grupowanie danych. Kolejne elementy hierarchii mogą być dowolnie rozwijane. |
|-----------------|-------------|-----------------------------------------------------------------------------------------------------------------------|
| [-] Europe | | |
| + 000 | 10907642.01 | |
| + DE | 2894312.338 | |
| + FR | 2644017.714 | |
| + GB | 3391712.211 | |
| + North America | 79353361.16 | |
| [-] Pacific | | |
| + 000 | 1594335.375 | |
| [-] AU | | |
| Aaron Diaz | 6029.57 | |
| Aaron Hill | 35.96 | |
| Aaron Powell | 4.99 | |
| Abby Kapoor | 751.34 | |
| Abby Martinez | 54.98 | |



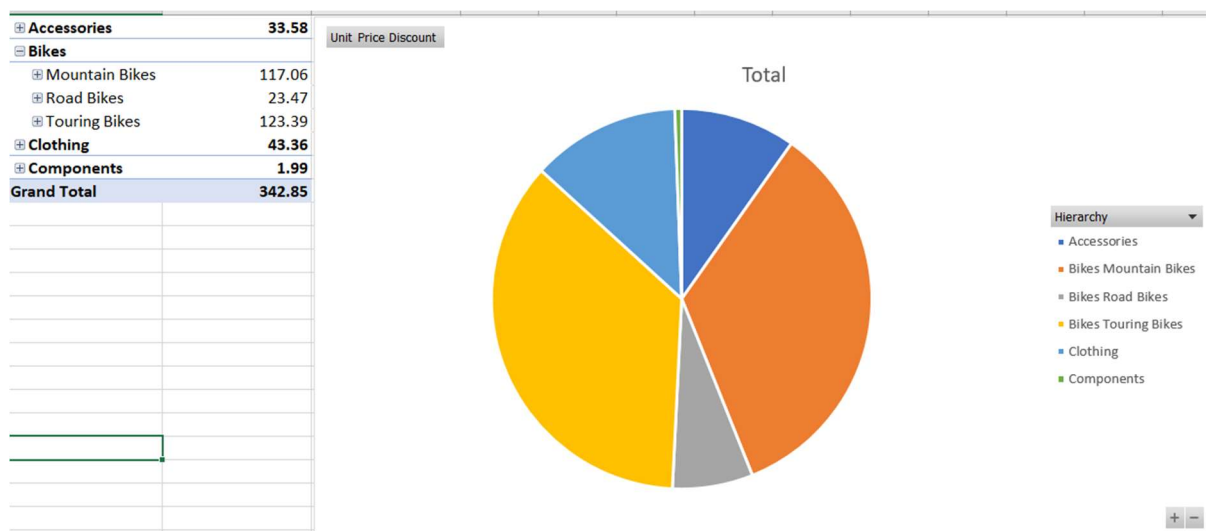
Rysunek 1. Suma sprzedaży z podziałem na rok oraz regiony sprzedaży



Rysunek 2. Maksymalna liczba zamówionych produktów w jednym zamówieniu z podziałem na lata oraz kategorie



Rysunek 3. Liczba zamówień z podziałem na kwartały



Rysunek 4. Suma obniżek cenowych w zależności od kategorii

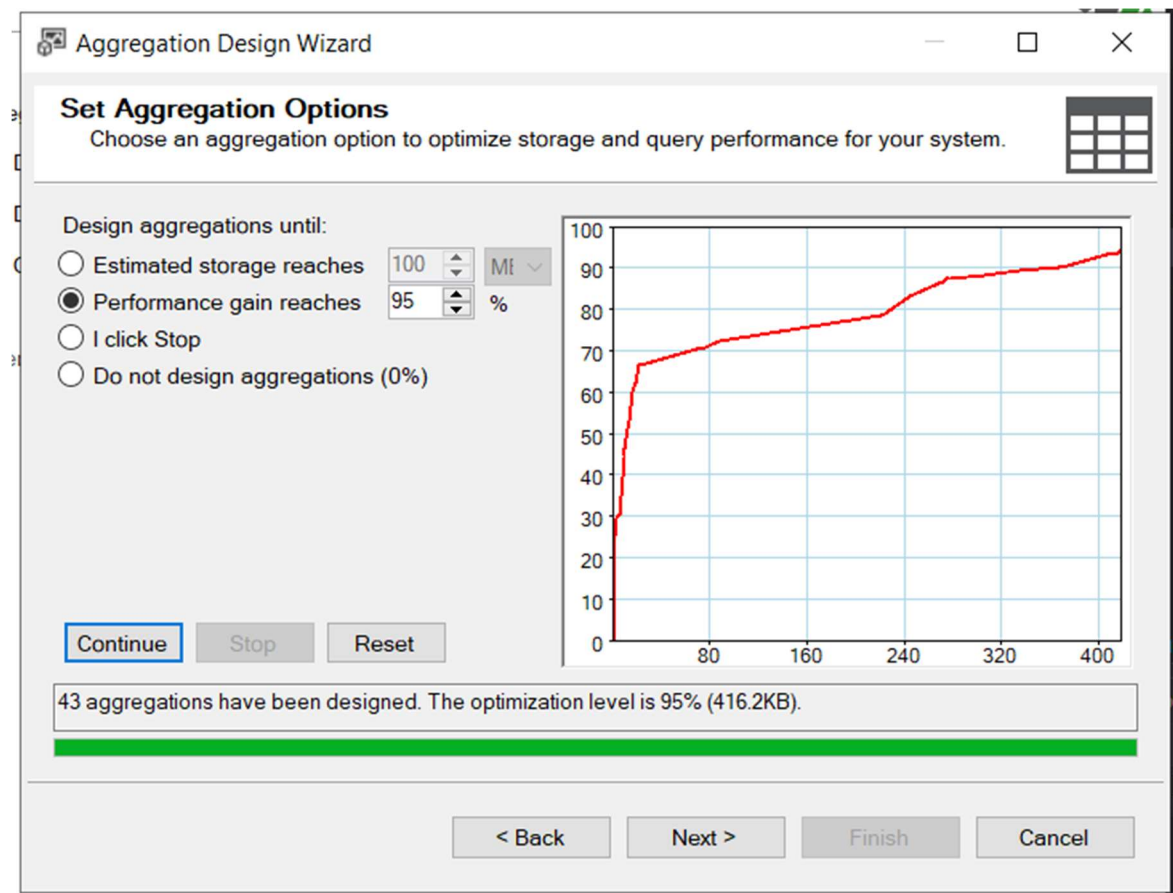
Kategorii nie trzeba rozwijać wszystkich – rozwinięcie jednej z nich automatycznie aktualizuje wykres. Na przykładzie powyżej, sprzedaż rowerów zajmowała ok. 80%, kategorię tę podzielono na podkategorie co pozwoliło na wykonanie bardziej przejrzystej wizualizacji danych. Tytuły rozwiniętych kategorii rozpoczynają się nazwą rozwiniętej kategorii „Bikes”

Zad. 3. Partycje

Zasilanie partycji za pomocą zapytań SQL

```
SELECT [Rasz].[Fact_Sales].[ProductID],[Rasz].[Fact_Sales].[CustomerID],  
       [Rasz].[Fact_Sales].[SalesPersonID], [Rasz].[Fact_Sales].[OrderDate],  
       [Rasz].[Fact_Sales].[ShipDate], [Rasz].[Fact_Sales].[OrderQty],  
       [Rasz].[Fact_Sales].[UnitPrice], [Rasz].[Fact_Sales].[UnitPriceDiscount],  
       [Rasz].[Fact_Sales].[LineTotal]  
FROM [Rasz].[Fact_Sales]  
JOIN [Rasz].[Dim_Time]  
ON [Rasz].[Fact_Sales].[OrderDate]=[Rasz].[Dim_Time].[Time] AND  
   [Rasz].[Dim_Time].[Year]=2011
```

Zapytanie powtórzono dla lat 2012, 2013, 2014



Agregację dla jednej z tabeli zoptymalizowano do 95% poziomu optymalizacji. Wynikowe partycje:

| Item | Partition Name ↑ | Source | Estimated Rows | Storage Mode | Aggregation Design |
|------|------------------|------------------------------------------------|----------------|--------------|--------------------|
| 1 | Fact Sales 2014 | SELECT [Rasz].[Fact_Sales].[ProductID],[Ras... | 0 | MOLAP | AggregationDesign |
| 2 | Fact Sales 2011 | SELECT [Rasz].[Fact_Sales].[ProductID],[Ras... | 5716 | MOLAP | AggregationDesign |
| 3 | Fact Sales 2012 | SELECT [Rasz].[Fact_Sales].[ProductID],[Ras... | 0 | MOLAP | AggregationDesign |
| 4 | Fact Sales 2013 | SELECT [Rasz].[Fact_Sales].[ProductID],[Ras... | 0 | MOLAP | AggregationDesign |

Ten sposób zasilania partycji jest korzystny w przypadku, kiedy nasze dane do podziału znajdują się w jednej tabeli. Zapytanie T-SQL pozwala na uniknięcie modyfikacji bazy danych.

Zasilanie partycji za pomocą osobnych tabel

```
CREATE TABLE [Rasz].[Fact_Sales_2011] (  
    [ProductID] INT NOT NULL,  
    [CustomerID] INT NOT NULL,  
    [SalesPersonID] INT NULL,  
    [OrderDate] INT NOT NULL,  
    [ShipDate] INT NULL,  
    [OrderQty] SMALLINT NOT NULL,  
    [UnitPrice] MONEY NOT NULL,  
    [UnitPriceDiscount] MONEY NOT NULL,  
    [LineTotal] NUMERIC(38, 6) NOT NULL  
);  
  
INSERT INTO [Rasz].[Fact_Sales_2011]  
([ProductID],[CustomerID],[SalesPersonID],[OrderDate], [ShipDate], [OrderQty],  
[UnitPrice], [UnitPriceDiscount], [LineTotal])  
(  
    SELECT S.*  
    FROM [Rasz].[Fact_Sales] S  
    JOIN [Rasz].[Dim_Time] T on S.OrderDate = T.Time and T.Year = 2011  
)
```

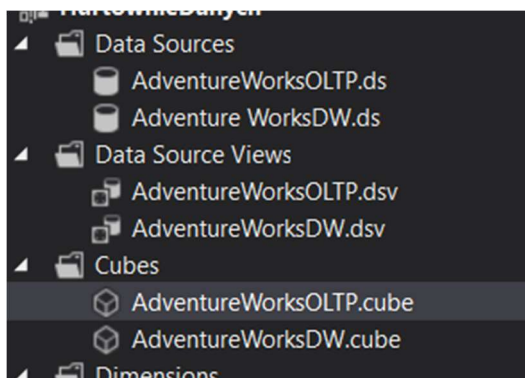
Tabele utworzono dla lat 2011, 2012, 2013, 2014. Bez większych trudów na ich podstawie utworzono partycje:

| Item | Partition Name | Source | Estimated Rows | Storage Mode | Aggregation Design |
|------|------------------------|---------------------------------------------|----------------|--------------|---------------------|
| 1 | Fact Sales 2011 UQ ... | Fact_Sales_2011 (Rasz) [AdventureWorksOL... | 0 | MOLAP | |
| 2 | Fact Sales 2012 | Fact_Sales_2012 (Rasz) [AdventureWorksOL... | 0 | MOLAP | |
| 3 | Fact Sales 2013 | Fact_Sales_2013 (Rasz) [AdventureWorksOL... | 0 | MOLAP | |
| 4 | Fact Sales 2014 | Fact_Sales_2014 (Rasz) [AdventureWorksOL... | 37339 | MOLAP | AggregationDesign 1 |

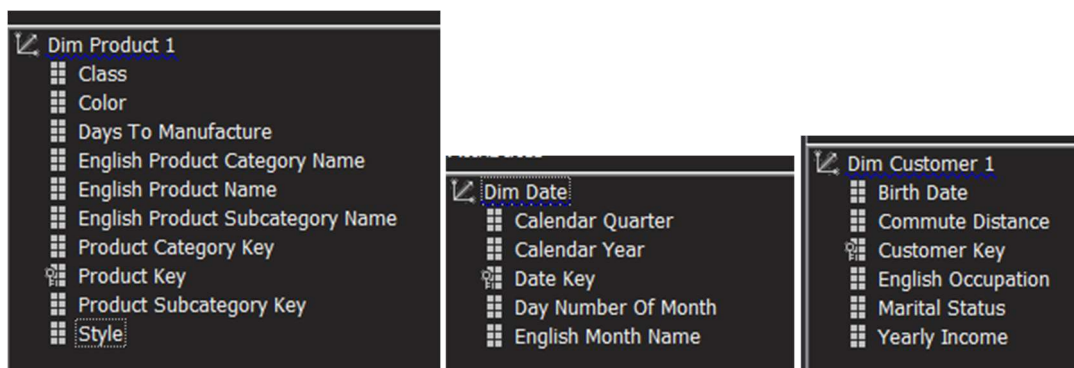
Taki sposób tworzenia partycji ma sens w przypadku danych podzielonych już na osobne tabele (np. osobna tabela dla użytkowników, lata itp.). W powyższym przypadku, konieczne było zmodyfikowanie bazy danych poprzez dodanie nowych tabel, co powieli liczbę danych oraz utrudnia proces.

Zad. 4.

Utworzono nowe obiekty dla bazy danych AdventureWorksDW:



Wybrano wymiary podobne do poprzednich zadań:



Wymiar Customer:

| Atrybut | Uzasadnienie |
|--------------------|------------------------------------------------------------------------------------|
| Commute Distance | Długość dojazdu do pracy – liczba kupowanych pojazdów może być zależna od długości |
| Customer Key | Klucz główny |
| English Occupation | Zawód – może wpływać na kupowane produkty |
| Marital Status | Stan cywilny – może dać ciekawe wyniki i wnioski |
| Yearly Income | Roczny dochód – może być powiązany z klasą kupowanych pojazdów |

Wymiar Product:

| Atrybut | Uzasadnienie |
|--------------------------|-------------------------------------------------------------------------|
| Color | Kolor – może pokazać preferencje klientów |
| Days To Manufacture | Czas wytwarzania – prawdopodobnie zależny od końcowej ceny |
| English Product Name | Nazwa |
| English Category Name | Nazwa Kategorii – może pokazać które produkty sprzedają się najczęściej |
| English Subcategory Name | Nazwa podkategorii – dalszy podział produktów |
| Product Key | Klucz Główny |
| Style | Styl – może pokazać preferencje klientów |
| Class | Klasa produktu – jak wyżej |

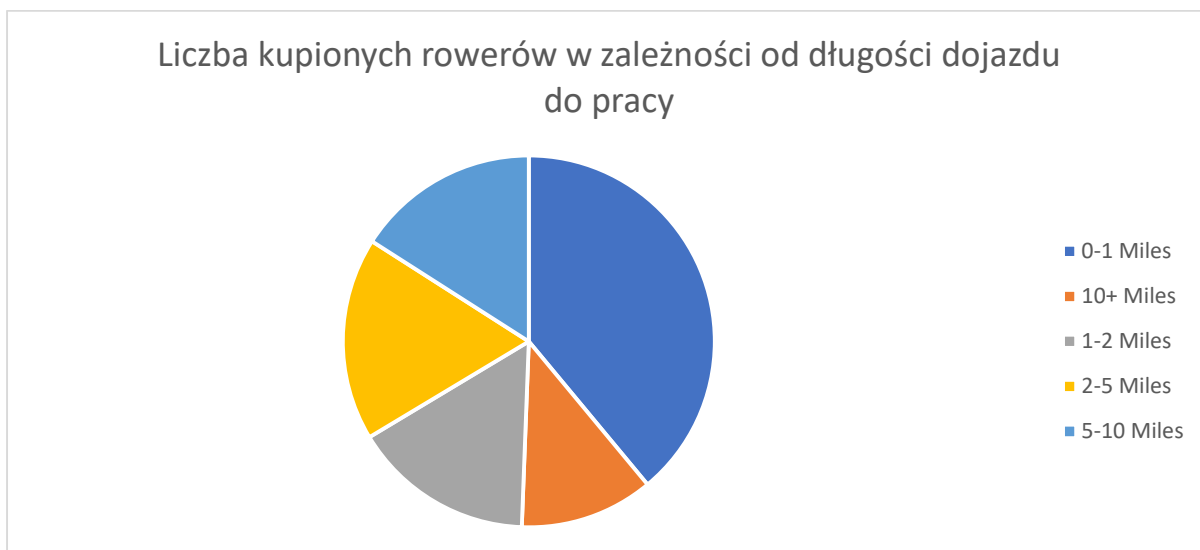
Wymiar Date:

| Atrybut | Uzasadnienie |
|---------------------|-------------------|
| CalendarYear | Do grupowania dat |
| CalendarQuarter | |
| EnglishMonthName | |
| Day Number Of Month | |

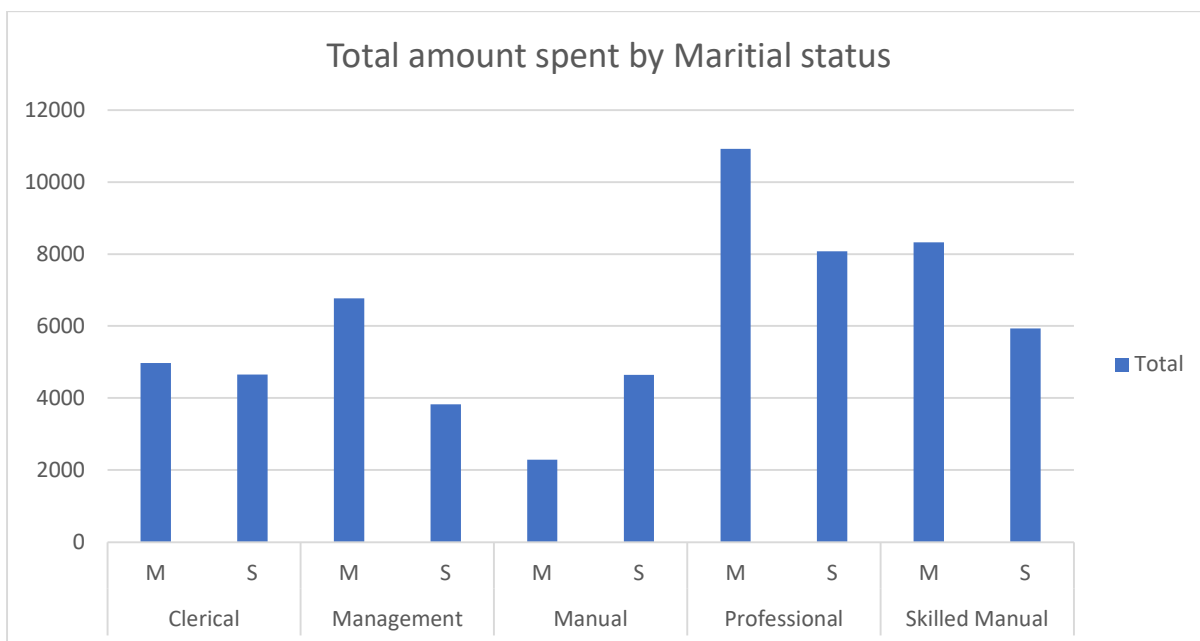
Miary wybrano z myślą o przeprowadzeniu badań na cenach produktów oraz ich obniżkach:

| |
|-------------------------|
| AdventureWorksDW |
| [Fact Internet Sales] |
| Order Quantity |
| Unit Price |
| Unit Price Discount Pct |
| Discount Amount |
| Product Standard Cost |
| Total Product Cost |
| Tax Amt |

Przykładowe wykresy:



Co ciekawe, najwięcej rowerów kupują klienci o najkrótszym dojeździe do pracy.



Widać tendencję do większych zakupów przez klientów w związku małżeńskim



Wszystkie sprzedaże w bazie danych AdventureWorksDW nie zawierają promocji. Zapytanie do bazy daje podobny rezultat:

```
select SUM(FactInternetSales.UnitPriceDiscountPct) from FactInternetSales;
```

108 %

Results Messages

| | (No column name) |
|---|------------------|
| 1 | 0 |