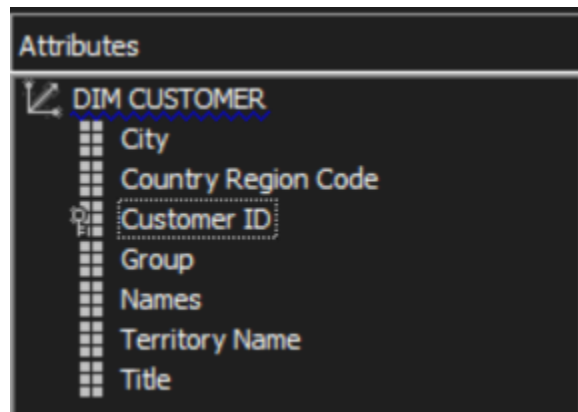


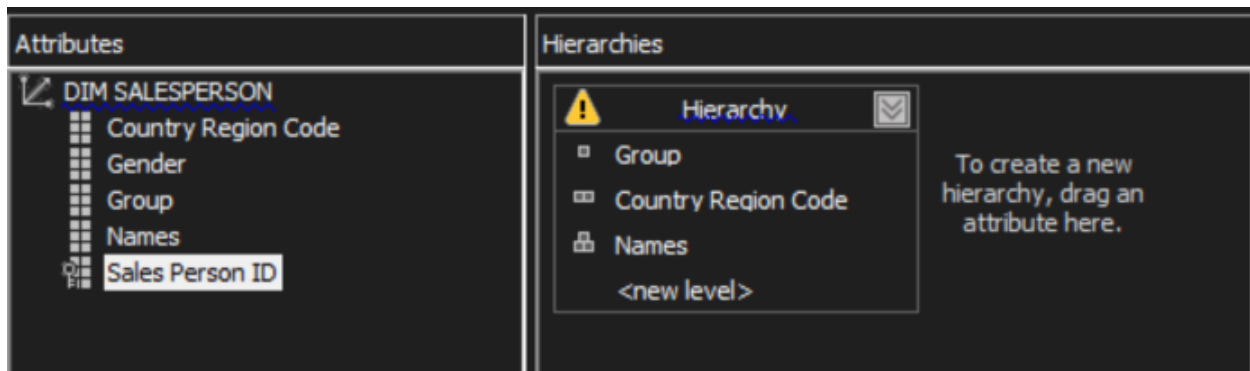
# Hurtownie Danych – lista 6

## Zad. 1. Modyfikacja wymiarów i tabeli faktów

- a) Bazując na kostce utworzonej przy realizacji listy 4, należy: a) zmodyfikować definicję wymiarów tak, aby:
- w wymiarach CUSTOMER i SALESPERSON nie można było korzystać z atrybutów FirstName oraz LastName. W zamian dodać atrybut Names



- w wymiarze SALESPERSON pojawiła się hierarchia Group – CountryRegionCode – Names



- w wymiarze CUSTOMER pojawiła się hierarchia Group – CountryRegionCode – Names

Attributes	Hierarchies
DIM CUSTOMER <ul style="list-style-type: none"> <li>City</li> <li>Country Region Code</li> <li>Customer ID</li> <li>Group</li> <li>Names</li> <li>Territory Name</li> <li>Title</li> </ul>	Hierarchy <ul style="list-style-type: none"> <li>Group</li> <li>Country Region Code</li> <li>Names</li> <li>&lt;new level&gt;</li> </ul> <p>To create a new hierarchy, drag an attribute here.</p>

- w wymiarze PRODUCT pojawiła się hierarchia CategoryName – SubCategoryName – Name

Attributes	Hierarchies
DIM PRODUCT <ul style="list-style-type: none"> <li>Category Name</li> <li>Color</li> <li>Is Purchased</li> <li>List Price</li> <li>Name</li> <li>Product ID</li> <li>Sub Category Name</li> </ul>	Hierarchy <ul style="list-style-type: none"> <li>Category Name</li> <li>Sub Category Name</li> <li>Name</li> <li>&lt;new level&gt;</li> </ul> <p>To create a new hierarchy, drag an attribute here.</p>

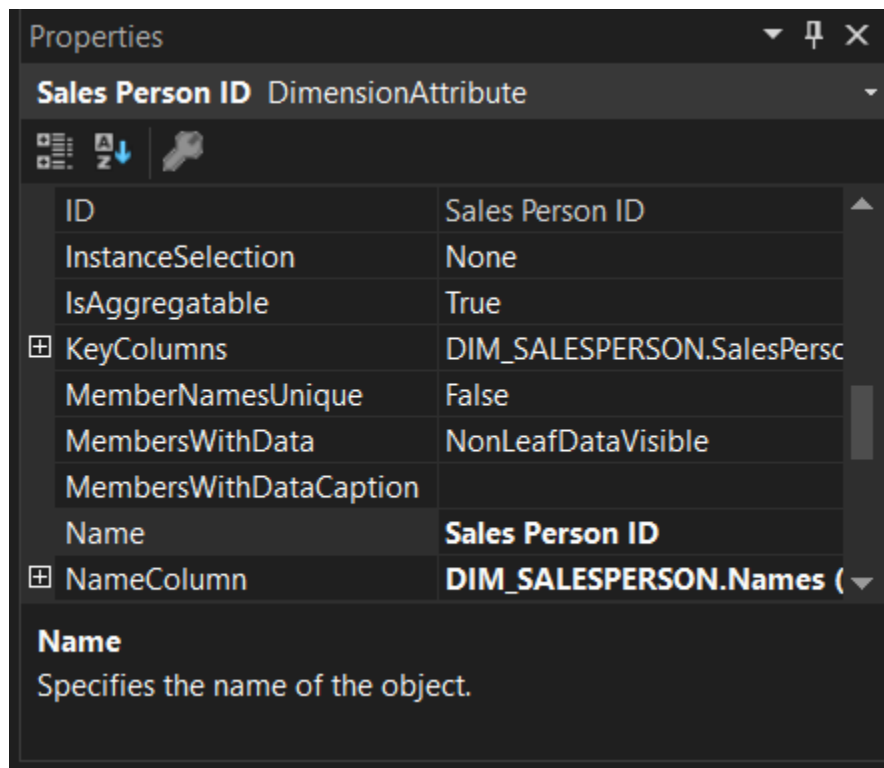
- w wymiarze TIME pojawiła się hierarchia Rok – Kwartał – Miesiąc – Dzień miesiąca

Attributes	Hierarchies
DIM TIME <ul style="list-style-type: none"> <li>Data</li> <li>Dzień Miesiaca</li> <li>Dzień Tygodnia Slow</li> <li>Kwartał</li> <li>Miesiąc</li> <li>Miesiąc Slow</li> <li>Rok</li> </ul>	Hierarchy <ul style="list-style-type: none"> <li>Rok</li> <li>Kwartał</li> <li>Miesiąc</li> <li>Dzień Miesiaca</li> <li>&lt;new level&gt;</li> </ul> <p>To create a new hierarchy, drag an attribute here.</p>

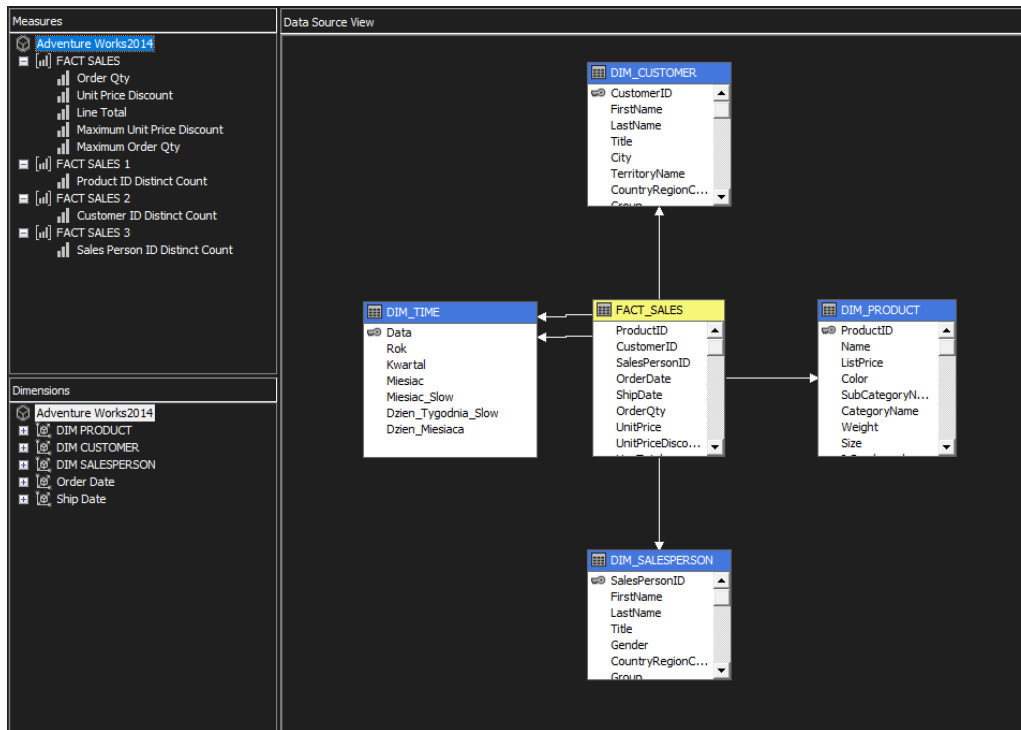
- b) Dla każdego atrybutu kluczowego wymiaru, którego wartościami są liczby całkowite, zmodyfikować właściwości (Properties). Zmodyfikować parametr NameColumn, tak aby nazwy kolejnych elementów wymiaru nie były liczbami. (Przykładowo dla wymiaru dotyczącego Produktu można wykorzystać atrybut Name)

Przypisujemy:

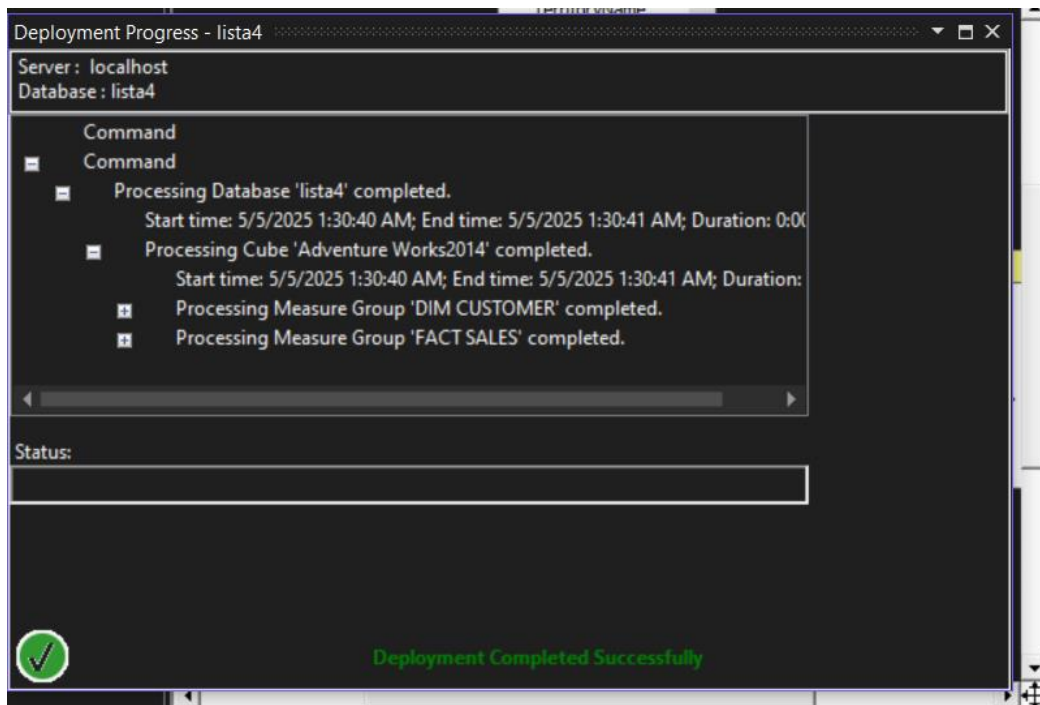
- DIM\_Salesperson -> Names
- DIM\_Product -> Name
- DIM\_Customer -> Names



- c) utworzyć nowe miary, które będą odzwierciedlać:
- Liczbę różnych klientów (aggregatedFunction: distinct count)
  - Liczbę różnych produktów
  - Maksymalną wartość rabatu (aggregatedFunction: max)
  - Maksymalną liczbę zamówionych produktów
  - Liczbę różnych sprzedawców realizujących zamówienia



d) wdrożyć i przetworzyć kostkę.



## Zadanie 2. Przegląd danych i tworzenie zestawień

Przy użyciu zakładki Browser:

- a) Sprawdzić, czy dane zapisane w kostce zgadzają się z danymi zapisanymi w tabelach, przeciągając za pomocą myszy:
  - a. atrybuty wymiarów w region wierszy
  - b. miary w część centralną widoku.

Browser:

Color	Line Total	Product ID Distinct Count
Black	38247022.5599971	79
Red	21623269.739998	31
Yellow	18669506.2100008	34

SQL Server Management Studio:

```

SELECT Czech.DIM_PRODUCT.Color, SUM(Czech.FACT_SALES.LineTotal) AS LineTotal, COUNT(DISTINCT Czech.FACT_SALES.ProductID) AS ProductIDDistinctCount
FROM Czech.FACT_SALES JOIN Czech.DIM_PRODUCT ON Czech.FACT_SALES.ProductID = Czech.DIM_PRODUCT.ProductID
WHERE Czech.DIM_PRODUCT.Color='Black' OR Czech.DIM_PRODUCT.Color='Red' OR Czech.DIM_PRODUCT.Color='Yellow'
GROUP BY Czech.DIM_PRODUCT.Color;
  
```

	Color	Line Total	ProductIDDistinctCount
1	Black	38247022.56	79
2	Red	21623269.74	31
3	Yellow	18669506.21	34

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

The screenshot shows the MDX query editor interface. On the left is a 'Metadata' tree with a search bar and a list of dimensions including 'Order Date', 'Color', and 'Product ID'. The main area displays a query table with columns: Dimension, Hierarchy, Operator, Filter Expression, and Parameters. The query table contains the following data:

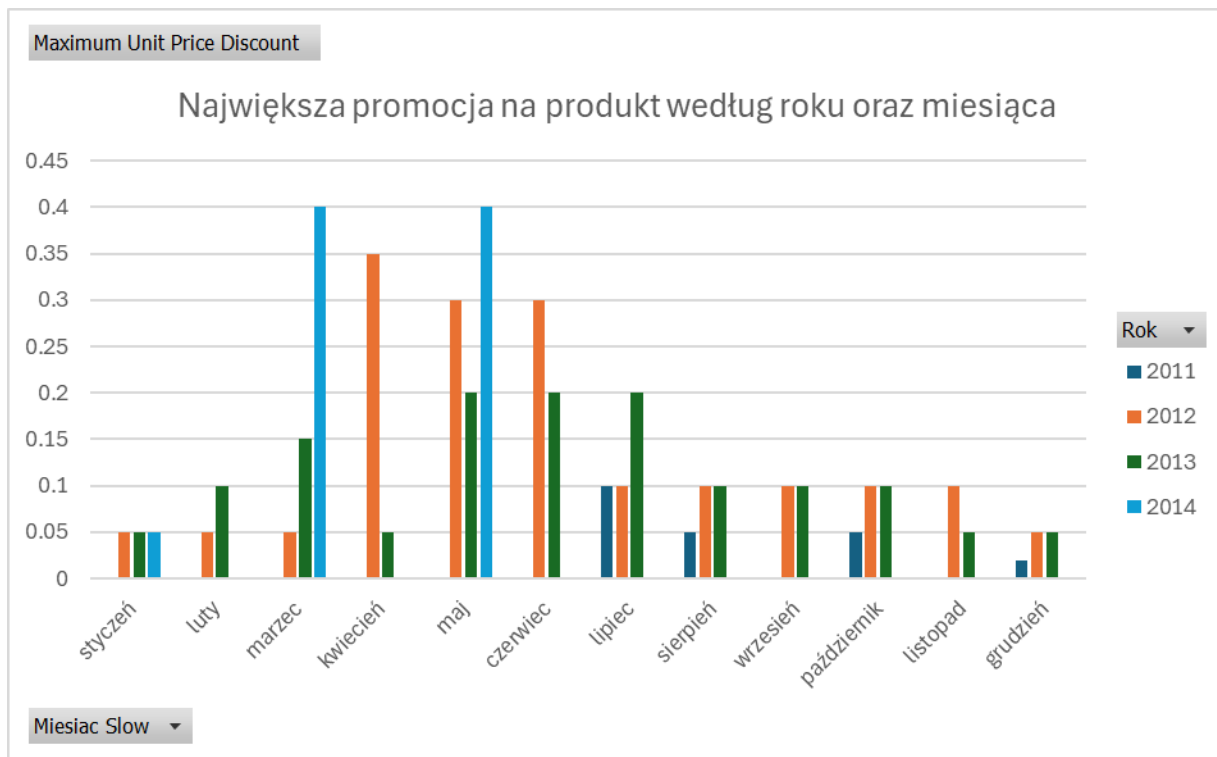
Dimension	Hierarchy	Operator	Filter Expression	Parameters
DIM PRODUCT	Color	Equal	{ Black, Yellow, Red }	<input checked="" type="checkbox"/> <input type="checkbox"/>
Order Date	Order Date.Rok	Range (Inclusive)	2011 : 2012	<input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				

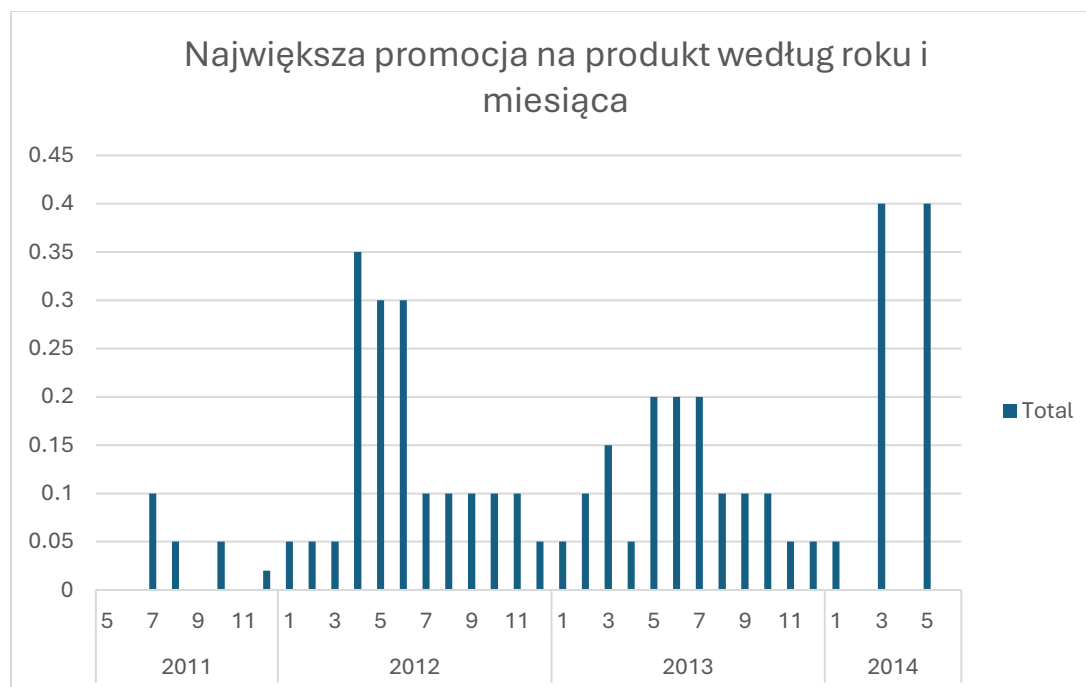
Below the query table is a result table with columns: Rok, Color, Line Total, and Product ID Distinct Count. The result table contains the following data:

Rok	Color	Line Total	Product ID Distinct Count
2011	Black	372467...	79
2011	Red	601941...	31
2011	Yellow	(null)	34
2012	Black	139178...	79
2012	Red	115332...	31
2012	Yellow	188574...	34

- c) Przygotować przykładowe tabele i wykresy przestawne oraz zinterpretować uzyskane wyniki (proszę zapisać wnioski!)

*Max UnitPriceDiscount według roku i miesiąca*





#### Wnioski:

- Największe zniżki można było zaobserwować od marca do czerwca (szczególnie w 2012 oraz 2014 roku). Świadczy to prawdopodobnie o kampaniach promocyjnych firmy.
- Najmniejsze zniżki obserwujemy w grudniu i styczniu. Przeczy to spodziewanym kampaniom zniżkowym organizowanym na święta wielkanocne. Należy jednak pamiętać, że duża część sprzedaży firmy AdventureWorks to sprzedaże sklepom, a nie osobom fizycznym.

*MaximumOrderQty vs CountryRegionCode i CategoryName*

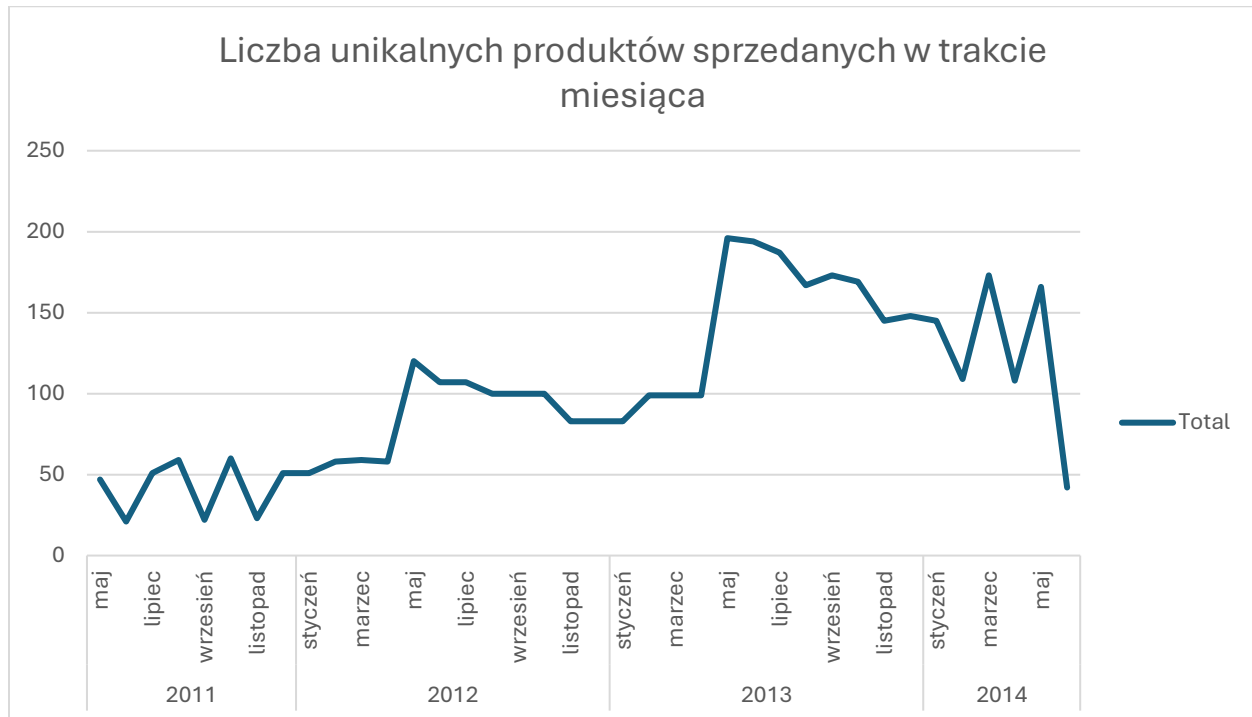
Maximum Order Qty	Column Labels				Grand Total
Row Labels	Accessories	Bikes	Clothing	Components	
<b>Europe</b>	<b>29</b>	<b>21</b>	<b>36</b>	<b>15</b>	<b>36</b>
DE	29	16	36	10	36
FR	28	19	25	13	28
GB	24	21	27	15	27
<b>North America</b>	<b>26</b>	<b>30</b>	<b>44</b>	<b>23</b>	<b>44</b>
CA	21	22	32	18	32
US	26	30	44	23	44
<b>Pacific</b>	<b>13</b>	<b>13</b>	<b>22</b>	<b>12</b>	<b>22</b>
AU	13	13	22	12	22
<b>Grand Total</b>	<b>29</b>	<b>30</b>	<b>44</b>	<b>23</b>	<b>44</b>

- Największe maksymalne zamówienia występują w kategoriach **Clothing** oraz **Bikes**:
  - USA** wyróżnia się pod tym względem, bo to tam występują najwyższe wartości zamówień, zarówno w przypadku odzieży, jak i rowerów.
- W **North America (głównie w USA i Kanadzie)** maksymalne wartości zamówień są wyraźnie wyższe niż w innych regionach:
  - W USA zarówno rowery, jak i odzież mają najwyższe limity zamówienia, wynoszące odpowiednio 30 i 44 sztuki.
  - Kanada również ma duże wartości zamówień, ale mniejsze niż USA.
- W **Europie** (głównie w **Niemczech**) największe zamówienia są na średnim poziomie:
  - Niemcy** mają najwyższe wartości zamówień — po 36 sztuk dla **odzieży** i 29 dla **akcesoriów**.
  - Francja** oraz **Wielka Brytania** mają nieco mniejsze wartości maksymalnych zamówień.
- Australia** ma najmniejsze wartości zamówień:
  - Maksymalne zamówienia to tylko 13 sztuk **akcesoriów** i **rowerów** oraz 22 sztuki odzieży.
- Podsumowanie regionalne:
  - North America (USA i Kanada)** ma największe maksymalne zamówienia, szczególnie w kategorii odzieży i rowerów. **Jest to rynek z największym potencjałem na hurtowe zamówienia.**
  - Europa** ma średnie wartości, z **Niemcami** na czołowej pozycji.



- c. **Pacyfik (Australia)** to region, gdzie maksymalne zamówienia są stosunkowo małe, co może wynikać z mniejszego rynku lub mniejszej liczby dużych zamówień.

*DistinctProductIDCount vs Rok i Miesiąc*



#### Wnioski:

- Spośród 266 produktów sprzedanych w trakcie funkcjonowania firmy, najwięcej w jednym miesiącu sprzedało się 196 (maj 2013 roku).
- Firma zaczęła od około 50 sprzedających się miesięcznie różnych produktów w 2011 roku. W 2012 roku rozwinęła swoją ofertę tak, że sprzedawało się około 100 różnych produktów, by w 2013 roku znowu podwoić ofertę i sprzedawać prawie 200 różnych produktów miesięcznie.
- Najwięcej różnych produktów sprzedaje się w maju (być może jest to spowodowane zmienną pogodą w danych rejonach lub tym, że ludzie zaczynają znowu kupować produkty przystosowane do letnich warunków).

## Zadanie 3. Miary kalkulowane

W zakładce Calculations dodać dwie miary kalkulowane (ang. calculated members):

- średnią liczbę zamówionych towarów na zamówienie

Name: AvgOrderQty

Parent Properties

Parent hierarchy: Measures

Parent member:

Change

Expression

`[Measures].[OrderQty] / [Measures].[OrderCount]`

No issues found

Ln: 1 Ch: 48 SPC CRLF

Additional Properties

- średnią ważoną liczbę towarów na zamówienie. Jako wagę należy wybrać cenę danego produktu.

Name: WeightedQtyAvg

Parent Properties

Parent hierarchy: Measures

Parent member:

Change

Expression

`[Measures].[OrderQtyWeightedSum]/[Measures].[LineTotal]`

No issues found

Ln: 1 Ch: 56 SPC CRLF

Adventure Works2014

Metadata

Search Model

<All>

Adventure Works2014

- Measures
- KPIs
- DIM CUSTOMER
- DIM PRODUCT
- DIM SALESPERSON
- Order Date
- Ship Date

Dimension	Hierarchy
<Select dimension>	
AvgOrderQty	WeightedQtyAvg
2.26607977447514	909.794216061618
OrderCount	Order Qty
121317	274914

Zgodnie z oczekiwaniami  $\text{AvgOrderQty} = \text{OrderQty} / \text{OrderCount} = 2.266$  produktu na zamówienie.

WeightedQtyAvg = 909.794 co znaczy, że średnio na zamówienie wydawano ponad \$900.

Możemy także obliczyć średnią cenę kupowanego produktu jako  $\text{WeightedQtyAvg} / \text{AvgOrderQty}$  (wychodzi około \$401.50)

## Zadanie 4. Partycje

Podzielić zawartość kostki na partycje (zakładka Partitions). Każda partycja powinna odzwierciedlać jeden rok. Istnieją dwa podstawowe sposoby podziału partycjonowania kostek:

- dane do zasilania poszczególnych partycji znajdują się w osobnych tabelach

Item	Partition Name	Source	Estimated Rows	Storage Mode	Aggregation Design
1	FACT SALES 1	FACT SALES	121317	MOLAP	AggregationDesign 3
2	FACT SALES 2011	FACT SALES_2011_VIEW	576	MOLAP	AggregationDesign
3	FACT SALES 2012	FACT SALES_2012_VIEW	21689	MOLAP	AggregationDesign 1
4	FACT SALES	SELECT [Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID],[Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID],[Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID]	0	MOLAP	AggregationDesign

Uzasadnienie:

- można **ładować i przetwarzać tylko potrzebny rok**
- szybsze niż przetwarzanie całej tabeli dla każdego z lat (łącznie n razy)
- dane do zasilania poszczególnych partycji znajdują się w tej samej tabeli, zaś każda z partycji ma przypisanie zapytanie SQL, którego wynik służy do jej zasilenia.

Item	Partition Name	Source	Estimated Rows	Storage Mode	Aggregation Design
1	FACT SALES 2014	SELECT [Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID],[Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID],[Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID]	121317	MOLAP	AggregationDesign 2
2	FACT SALES 2013	SELECT [Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID],[Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID],[Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID]	121317	MOLAP	AggregationDesign 1
3	FACT SALES 2012	SELECT [Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID],[Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID],[Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID]	121317	MOLAP	AggregationDesign
4	FACT SALES 2011	SELECT [Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID],[Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID],[Czech].[FACT SALES].[ProductID],[Czech].[FACT SALES].[CustomerID]	0	MOLAP	AggregationDesign

Uzasadnienie:

- Bardziej elastyczne, nie trzeba zapewniać takich samych pól i relacji przed tworzeniem partycji.
- Idealne gdy dane są w jednej tabeli.
- Nie trzeba dodawać tabeli do bazy danych.

Proszę przygotować partycje na dwa sposoby i znaleźć uzasadnienie dla każdej opcji.

## Wnioski:

### Zadanie 1. Modyfikacja wymiarów i faktów

a)

- Wymiana atrybutów **FirstName** oraz **LastName** na **Names** w wymiarach **CUSTOMER** i **SALESPERSON** pozwala na lepszą organizację danych i unika potencjalnych problemów związanych z wieloma wartościami tych atrybutów. Dzięki temu można łatwiej zarządzać danymi i analizować je w kontekście pełnych nazw.
- Wprowadzenie hierarchii w wymiarach **SALESPERSON** i **CUSTOMER** (Group – CountryRegionCode – Names) sprawia, że struktura danych staje się bardziej zorganizowana i pozwala na lepsze analizy geograficzne oraz demograficzne.
- Hierarchia w wymiarze **PRODUCT** (CategoryName – SubCategoryName – Name) sprawia, że analiza produktów staje się bardziej ziarnista, umożliwiając łatwiejsze porównanie sprzedaży produktów w różnych kategoriach i podkategoriach.
- Hierarchia w wymiarze **TIME** (Rok – Kwartał – Miesiąc – Dzień miesiąca) umożliwia analizę danych w kontekście czasu, zarówno na poziomie rocznym, kwartalnym, jak i miesięcznym, co jest pomocne w analizach trendów sprzedaży.

b)

- Przypisanie odpowiednich atrybutów kluczowych, takich jak **Names** dla **DIM\_Salesperson**, **Name** dla **DIM\_Product** i **DIM\_Customer**, sprawia, że dane są bardziej czytelne i łatwiejsze w analizie. Zamiast liczb, prezentowane są bardziej zrozumiałe opisy, co zwiększa przejrzystość wyników analiz.

c) Nowe miary:

- **Liczba różnych klientów:** Pomaga to w śledzeniu liczby unikalnych klientów, co jest niezbędne do analizowania lojalności i segmentacji klientów.
- **Liczba różnych produktów:** Umożliwia ocenę różnorodności oferty firmy i jej zmiany w czasie.
- **Maksymalna wartość rabatu:** Pozwala na ocenę najwyższych rabatów oferowanych klientom, co może być istotne w kontekście kampanii promocyjnych.
- **Maksymalna liczba zamówionych produktów:** Wskazuje na najwyższe zamówienia złożone przez klientów, co może pomóc w analizie dużych transakcji i identyfikacji trendów w zamówieniach.
- **Liczba różnych sprzedawców realizujących zamówienia:** Informacja o liczbie aktywnych sprzedawców pozwala na ocenę wydajności zespołu sprzedaży i jego zasięgu.

## Zadanie 2. Tworzenie zestawień

- Dzięki możliwości wykorzystania zakładki **Browser** w SSAS, łatwo można porównać dane w kostce z danymi w tabelach źródłowych, co pozwala upewnić się, że agregacje i miary są poprawnie obliczane. Jest to kluczowy krok w procesie weryfikacji poprawności danych.
- Testowanie operatorów i filtrów w zakładce **Browser** umożliwia bardziej zaawansowaną analitykę, gdzie użytkownicy mogą dynamicznie dostosowywać widok danych i badać różne scenariusze, takie jak selekcja danych po określonych wartościach lub grupach.

## Zadanie 3: Miary kalkulowane

### 1. Średnia liczba zamówionych towarów na zamówienie:

- Wnioski: Średnia liczba produktów na zamówienie wynosi około 2,27, co oznacza, że na każde zamówienie średnio przypada ponad dwa produkty. To może świadczyć o tym, że klienci składają zamówienia na większe ilości towarów.

### 2. Średnia ważona liczba towarów na zamówienie:

- Wnioski: Średnia ważona wynosi 909,794, co oznacza, że średnia wartość zamówienia jest znacznie wyższa (ponad 900 dolarów). Z kolei średnia cena produktu na zamówienie to około 401,50 USD, co wskazuje na to, że firma sprzedaje w głównej mierze droższe produkty

## Zadanie 4: Partycje

- **Pierwsza opcja (osobne tabele):**

- **Uzasadnienie:** Ta opcja jest szybsza, ponieważ umożliwia ładowanie i przetwarzanie tylko danych z jednego roku. Jest to efektywne, gdy liczba danych w każdym roku jest duża i przetwarzanie całej tabeli zajęłoby zbyt dużo czasu.

- **Druga opcja (zapytania SQL w jednej tabeli):**

- **Uzasadnienie:** Ta opcja jest bardziej elastyczna, ponieważ nie wymaga tworzenia nowych tabel, a dane w jednej tabeli mogą być przetwarzane dynamicznie. Idealna, gdy dane są w jednej tabeli i nie chcemy tworzyć dodatkowych struktur w bazie danych. Jest to również bardziej elastyczne w przypadku zmiany struktury danych w przyszłości, ponieważ nie musimy dodawać nowych tabel, wystarczy zmodyfikować zapytania.