

HURTOWNIE DANYCH

Laboratorium 6

Maciej Kopiński 254578

Zad. 1. Modyfikacja wymiarów i tabeli faktów

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
 - w wymiarze PRODUCT pojawiła się hierarchia CategoryName – SubCategoryName – Name
 - w wymiarze TIME pojawiła się hierarchia Rok – Kwartał – Miesiąc – Dzień miesiąca
- 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).
- 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 przeprocесować kostkę.

Zad. 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:
 - atrybuty wymiarów w region wierszy
 - miary w część centralną widoku
- b) Przetestować możliwości przeglądarki (Browser) – operator wyboru danych (Operator), wyrażenia filtrujące dane (Filter Expression) itp.
- c) Przygotować przykładowe tabele i wykresy przestawne (Excel) oraz zinterpretować uzyskane wyniki (wnioski!)

Zad. 3. 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
- 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.

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

Zad. 4.

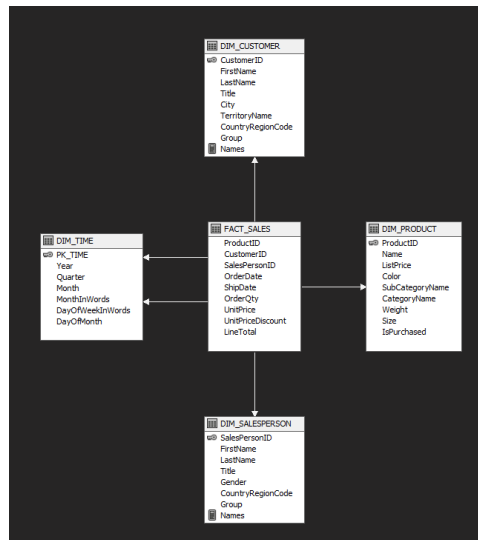
Korzystając z bazy danych AdventureWorksDW:

- a) Utworzyć nowe źródło danych
- b) Utworzyć nowy widok źródła danych
- c) Utworzyć nową kostkę wybierając jako tabelę faktów FactInternetSales, wymiary według własnego uznania – uzasadnić wybór
- d) Zbadać, jak zmienia się czas przetwarzania kostki w zależności od liczby zdefiniowanych wymiarów.
- e) Poćwiczyć edycję wymiarów i korzystanie z przeglądarki kostki danych (Browser) oraz przygotować ciekawe zestawienia – tabele i wykresy przestawne w MS Excel.

Rozwiązania:

Zad.1

- a) Kostka z atrybutami Names zamiast FirstName i LastName



Hierarchie:

Attributes	Hierarchies
<ul style="list-style-type: none">DIM CUSTOMERCityCountry Region CodeCustomer IDGroupNamesTerritory NameTitle	<ul style="list-style-type: none">HierarchyGroupCountry Region CodeNames<new level>

Attributes	Hierarchies
<ul style="list-style-type: none">DIM TIMEDay Of MonthDay Of Week In WordsMonthMonth In WordsPK TIMEQuarterYear	<ul style="list-style-type: none">HierarchyYearQuarterMonthDay Of Month<new level>

Attributes	Hierarchies
<ul style="list-style-type: none">DIM SALESPERSONCountry Region CodeGenderGroupNamesSales Person IDTitle	<ul style="list-style-type: none">HierarchyGroupCountry Region CodeNames<new level>

Attributes	Hierarchies
<ul style="list-style-type: none">DIM PRODUCTCategory NameColorIs PurchasedList PriceNameProduct IDSizeSub Category NameWeight	<ul style="list-style-type: none">HierarchyCategory NameSub Category NameName<new level>

b)

Name	Product ID
NameColumn	DIM_PRODUCT.Name (WChar)

Name	PK TIME
NameColumn	DIM_TIME.DayOfWeekInWords (WChar)

Name	Sales Person ID
NameColumn	DIM SALESPERSON.Names (WChar)

Name	Customer ID
NameColumn	DIM_CUSTOMER.Names (WChar)

c)


Adventure Works2019	
[]	FACT SALES
[]	Order Qty
[]	Unit Price
[]	Unit Price Discount
[]	Line Total
[]	FACT SALES Count
[]	Maximum Unit Price Discount
[]	Maximum Order Qty
[]	FACT SALES 1
[]	Customer ID Distinct Count
[]	FACT SALES 2
[]	Product ID Distinct Count
[]	FACT SALES 3
[]	Sales Person ID Distinct Count

d)

Process Progress	
Command	
Processing Cube 'Adventure Works2019' completed.	
Start time: 12.04.2022 21:50:08; End time: 12.04.2022 21:50:12; Duration: 0:00:04	
[]	Processing Measure Group 'FACT SALES' completed.
[]	Processing Measure Group 'FACT SALES 1' completed.
[]	Processing Measure Group 'FACT SALES 2' completed.
[]	Processing Measure Group 'FACT SALES 3' completed.
Status:	
Process succeeded.	
Stop	Reprocess
View Details...	Copy
Close	Help


Zad.2

a)

Dimension	Hierarchy	Operator	Filter Expression	Para
DIM CUSTOMER	 City	Equal		<input type="checkbox"/>
<Select dimension>				<input type="checkbox"/>

Line Total	Maximum Order Qty	FACT SALES Count	Maximum Unit Price Discount	Order Qty	Unit Price	Unit Price Discount
1098463...	44	121317	0,4	274914	5642374...	342,850000000004

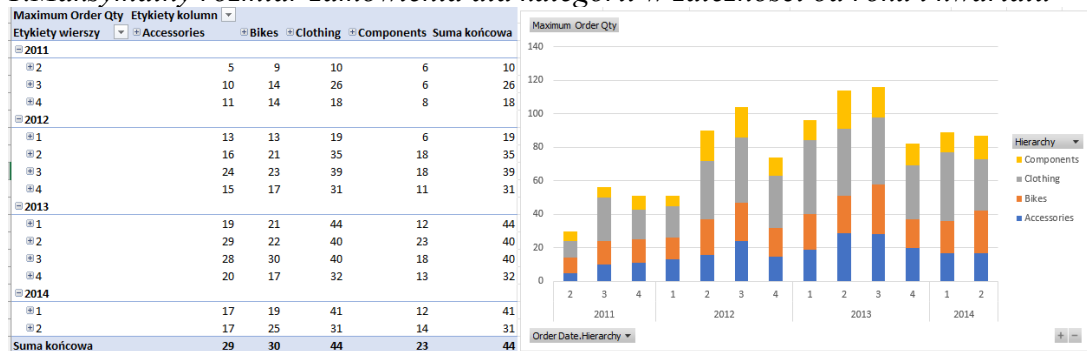
b)

Dimension	Hierarchy	Operator	Filter Expression	Parameter
DIM CUSTOMER	 City	Equal	{ Abingdon }	<input type="checkbox"/>
<Select dimension>				<input type="checkbox"/>

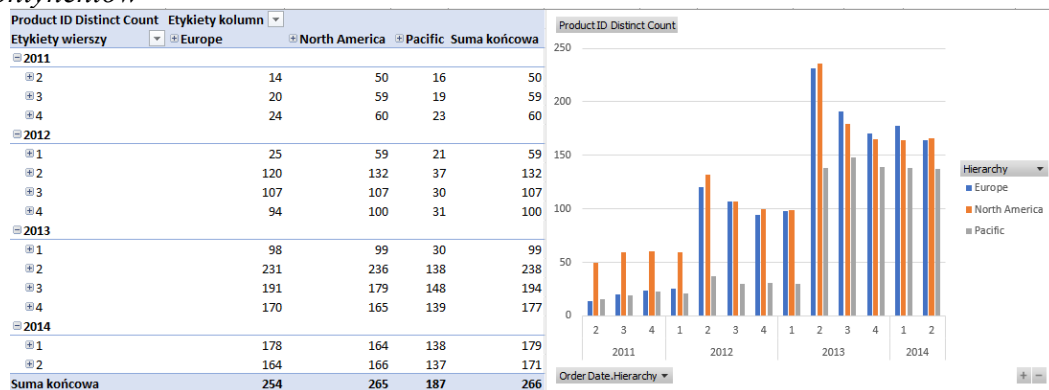
Line Total	Maximum Order Qty	FACT SALES Count	Maximum Unit Price Discount	Order Qty	Unit Price	Unit Price Discount
59,334	7	2	0	11	10,788	0

c)

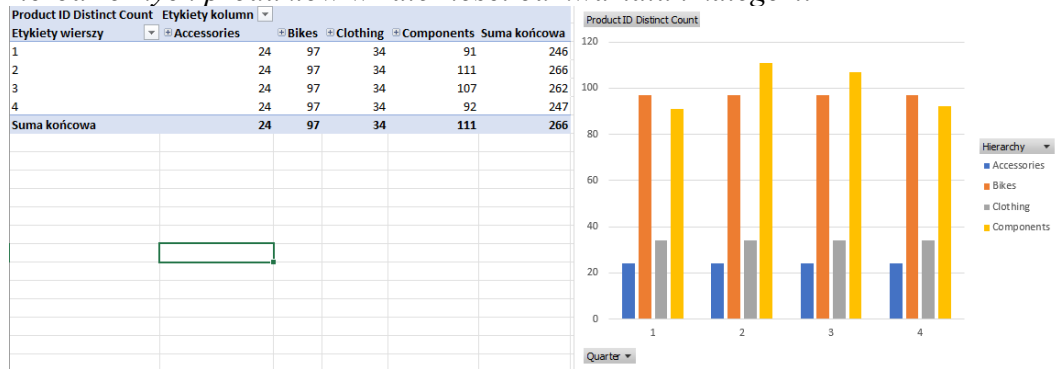
1. Maksymalny rozmiar zamówienia dla kategorii w zależności od roku i kwartału



2. Liczba różnych produktów w zależności od roku i kwartału dla poszczególnych kontynentów



3. Liczba różnych produktów w zależności od kwartału i kategorii



Zad.3

Partycje:

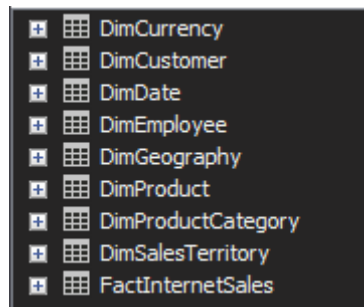
Item	Partition Name	Source	Estimated Rows	Storage Mode	Aggregation Design
1	FACT SALES 2011	SELECT [Kopinski].[FACT_SALES].[ProductID],[Kopinski]...	0	MOLAP	
2	FACT SALES 2012	SELECT [Kopinski].[FACT_SALES].[ProductID],[Kopinski]...	21689	MOLAP	AggregationDesign_2012
3	FACT SALES 2013	SELECT [Kopinski].[FACT_SALES].[ProductID],[Kopinski]...	56573	MOLAP	AggregationDesign_2013
4	FACT SALES 2014	SELECT [Kopinski].[FACT_SALES].[ProductID],[Kopinski]...	37339	MOLAP	AggregationDesign_2014

Zad.4

a) oraz b)

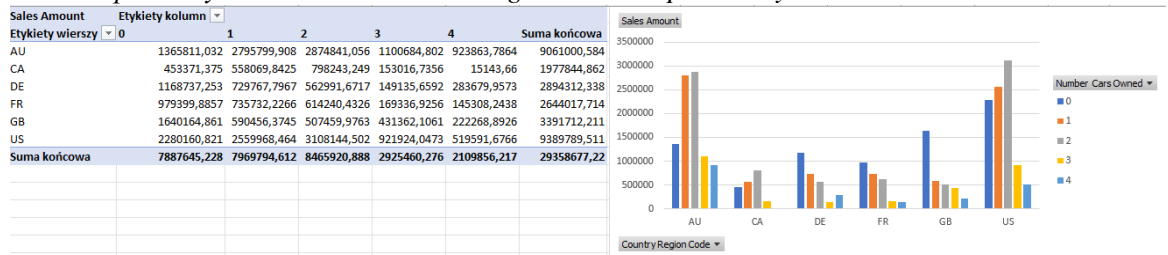


c)

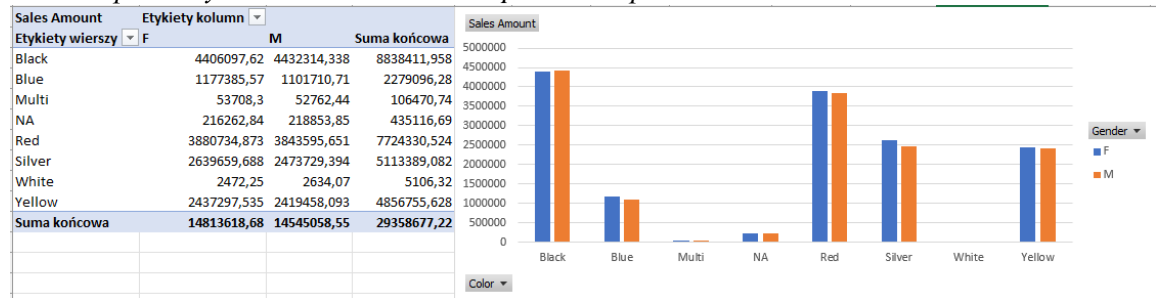


e)

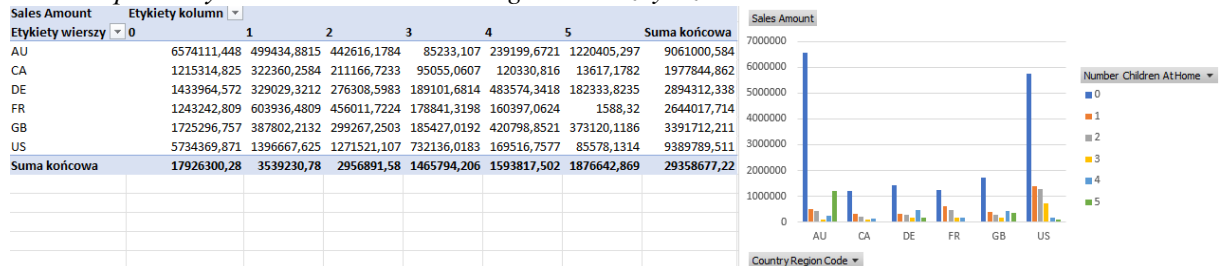
1. Suma sprzedaży w zależności od kodu regionu i ilości posiadanych samochodów



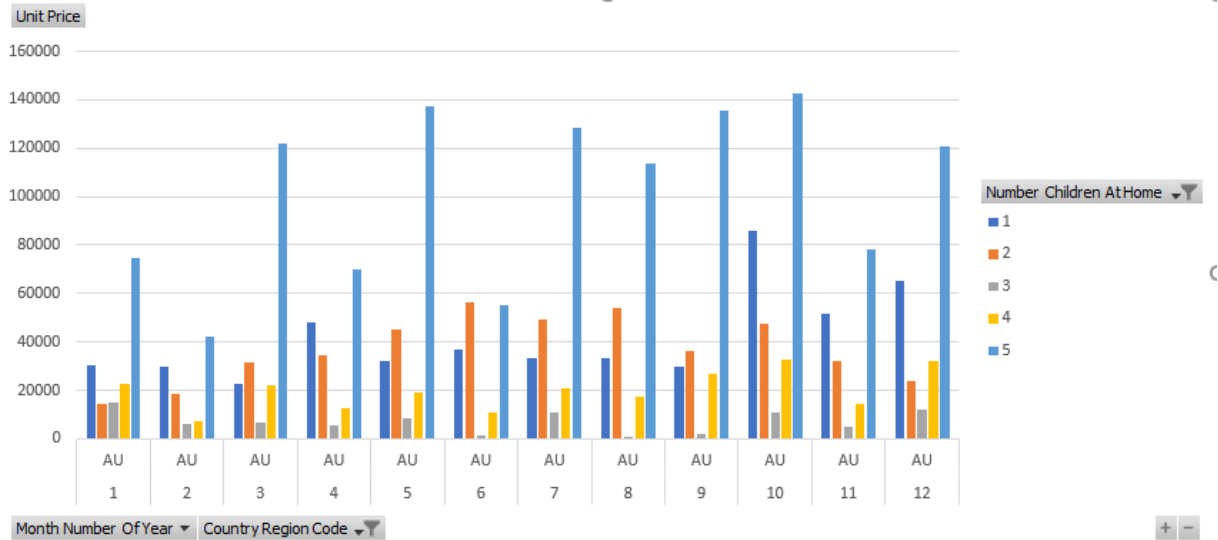
2. Suma sprzedaży w zależności od koloru produktu i płci



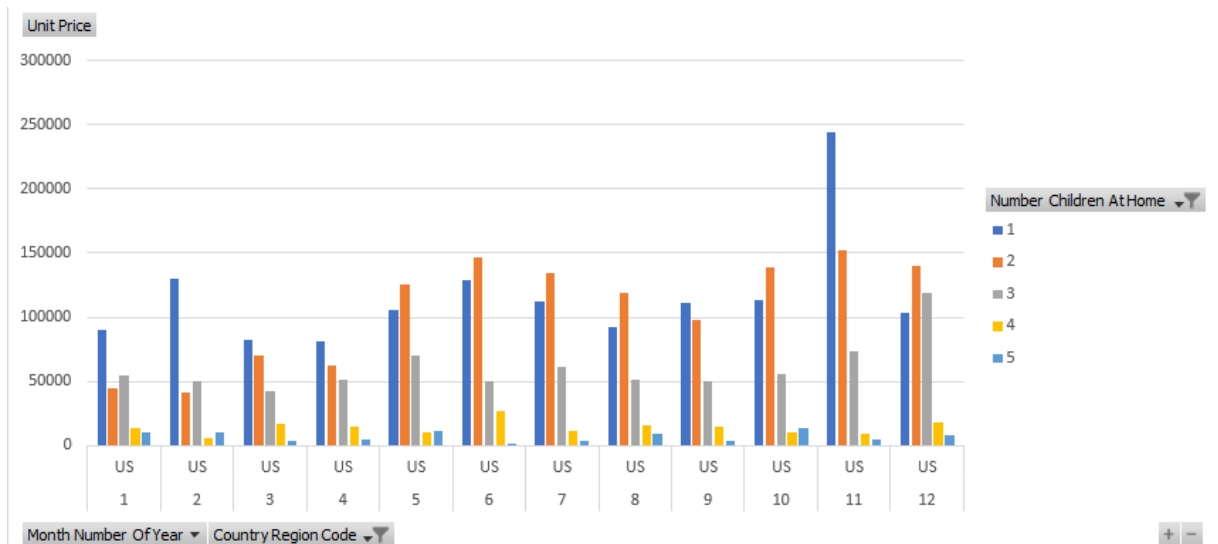
3. Suma sprzedaży w zależności od kodu regionu i liczby dzieci



4. Suma sprzedaży dla osób z dziećmi według miesięcy w Australii



Dla Stanów Zjednoczonych



Wnioski:

Wykres 2.c.1:

Klienci kupują w jednym zamówieniu najwięcej produktów z kategorii ubrania.

Wykres 2.c.2:

Sklep z roku na rok oferuje coraz więcej różnych produktów.

Wykres 2.c.3:

Najwięcej różnych produktów jest kupowane z kategorii rowery i komponenty.

Wykres 4.e.1:

Im więcej klient posiada samochodów, tym mniej kupuje.

Wykres 4.e.2:

Płeć klienta nie wpływa na kolory produktów, które kupuje.

Wykres 4.e.3:

Im większa liczba dzieci, tym niższe ilości produktów w zamówieniu. Wyjątkiem jest Australia, gdzie wciąż najwięcej kupują osoby bezdzietne, natomiast w przypadku posiadania 5 dzieci kupują więcej niż w przypadku posiadania mniejszej ilości dzieci.

Projekt typu Analysis Services pozwala na wygodne wybieranie, przetwarzanie i prezentowanie danych. Pozwala on na modyfikowanie przedstawionych danych bez konieczności modyfikowania ich w bazie poprzez funkcję Named Calculations czy tworzenie hierarchii pozwalającej na późniejsze wygodniejsze doprecyzowanie danych przykładowo na wykresie. Nie wymaga pisania zapytań do samej bazy w celu utworzenia tabel przestawnych prezentujących potrzebne nam dane.