







































Hurtownie danych Laboratorium Czw 11:15

Lista 6

Kajetan Pynka 254495

Zad 1

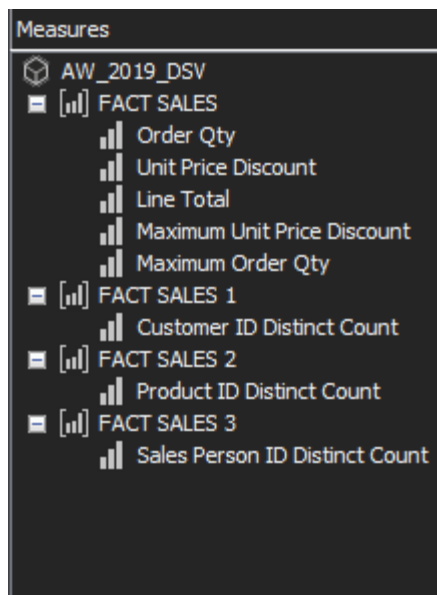
a)

Attributes  DIM CUSTOMER  City  Country Region Code  Customer ID  Group  Names  Territory Name  Title	Hierarchies  Hierarchy  <ul style="list-style-type: none"> ▪ Group ▪ Country Region Code ▪ Names <new level>
Attributes  DIM SALESPERSON  Country Region Code  Gender  Group  Names  Sales Person ID  Title	Hierarchies  Hierarchy  <ul style="list-style-type: none"> ▪ Group ▪ Country Region Code ▪ Names <new level>
Attributes  DIM TIME  Dzien Miesiaca  Dzien Tygodnia  Kwartal  Miesiac  Miesiac Slownie  PK TIME  Rok	Hierarchies  Hierarchy  <ul style="list-style-type: none"> ▪ Rok ▪ Kwartal ▪ Miesiac ▪ Dzien Miesiaca <new level>
Attributes  DIM PRODUCT  Category Name  Color  List Price  Name  Product ID  Sub Category Name	Hierarchies  Hierarchy  <ul style="list-style-type: none"> ▪ Category Name ▪ Sub Category Name ▪ Name <new level>

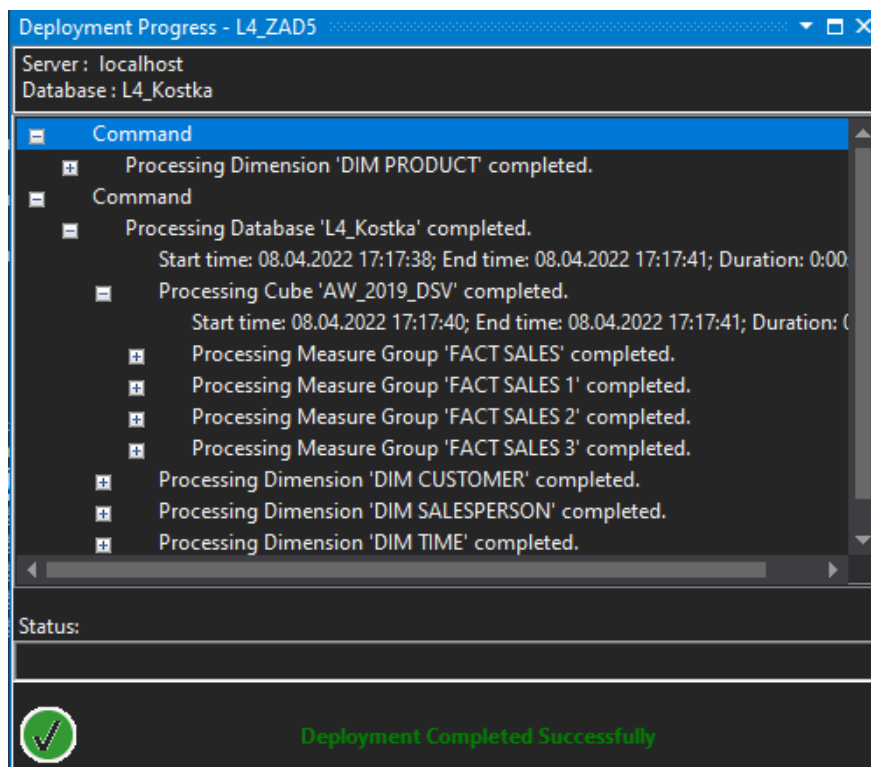
b)

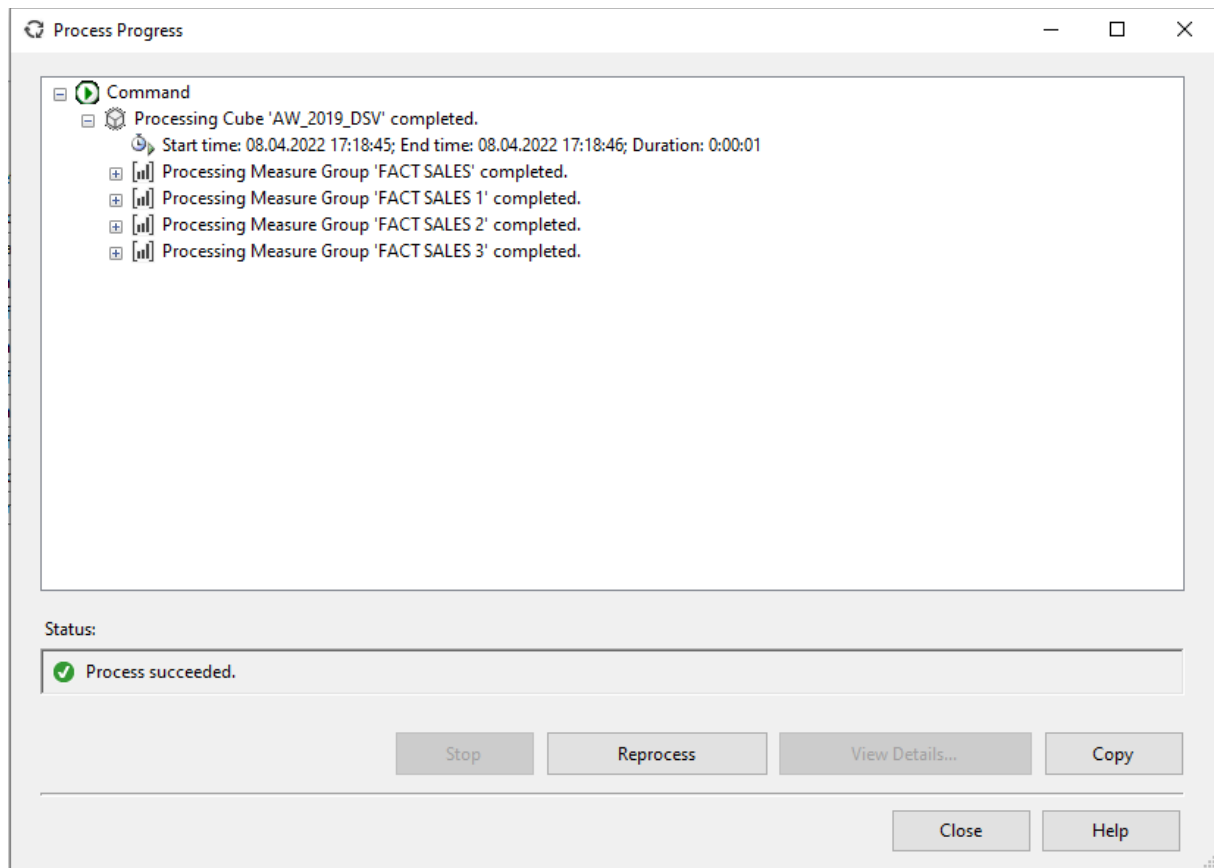
⊞ KeyColumns	DIM_PRODUCT.ProductID (Integer)
⊞ NameColumn	DIM_PRODUCT.Name (WChar)
⊞ KeyColumns	DIM_CUSTOMER.CustomerID (Integer)
⊞ NameColumn	DIM_CUSTOMER.Names (WChar)
⊞ KeyColumns	DIM SALESPERSON.SalesPersonID (Integer)
⊞ NameColumn	DIM SALESPERSON.Names (WChar)
⊞ KeyColumns	DIM_TIME.PK_TIME (Integer)
⊞ NameColumn	DIM_TIME.Dzien_miesiaca (WChar)

c)



d)





Zad 2

a)

Dimension	Hierarchy	Operator
DIM CUSTOMER	City	Equal
DIM CUSTOMER	Country Region Code	Equal
DIM CUSTOMER	Customer ID	Equal
DIM CUSTOMER	Group	Equal
DIM CUSTOMER	Names	Equal
DIM CUSTOMER	Territory Name	Equal
DIM CUSTOMER	Title	Equal
<Select dimension>		

Customer ID Distinct Count
19119

Dimension	Hierarchy	Operator
DIM PRODUCT	Category Name	Equal
DIM PRODUCT	Color	Equal
DIM PRODUCT	List Price	Equal
DIM PRODUCT	Name	Equal
DIM PRODUCT	Product ID	Equal
DIM PRODUCT	Sub Category Name	Equal
<Select dimension>		

Product ID Distinct Count
266

Dimension	Hierarchy	Operator
DIM SALESPERSON	Country Region Code	Equal
DIM SALESPERSON	Gender	Equal
DIM SALESPERSON	Group	Equal
DIM SALESPERSON	Names	Equal
DIM SALESPERSON	Sales Person ID	Equal
DIM SALESPERSON	Title	Equal
<Select dimension>		

Sales Person ID Distinct Count
18

b)

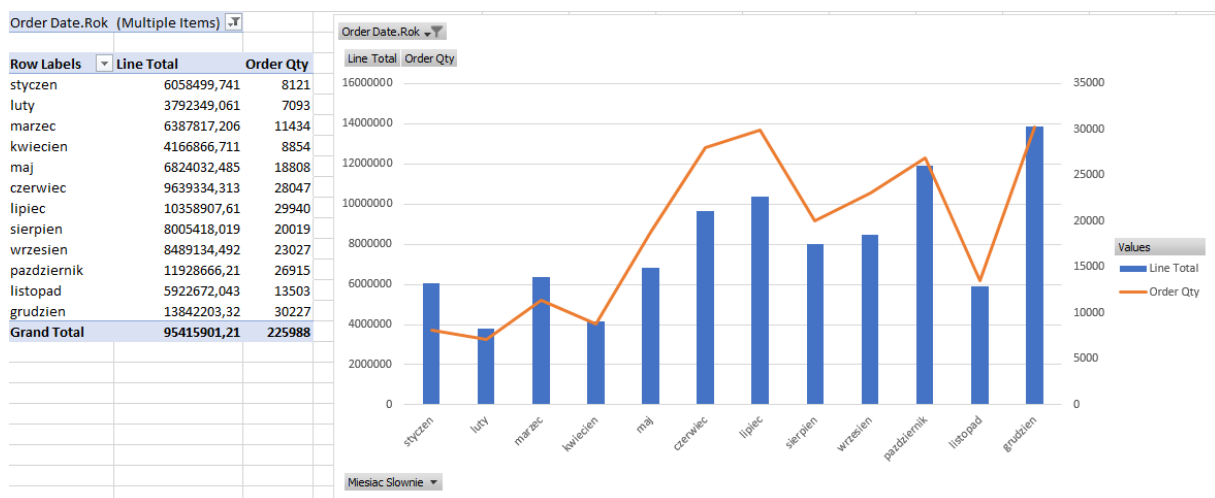
Dimension	Hierarchy	Operator	Filter Expression
DIM SALESPERSON	Sales Person ID	Equal	
DIM SALESPERSON	Names	Equal	
DIM SALESPERSON	Gender	Equal	{ F }
DIM SALESPERSON	Country Region Code	Equal	
DIM SALESPERSON	Title	Equal	
DIM SALESPERSON	Group	Equal	{ North America }
<Select dimension>			

Sales Person ID Distinct Count	Customer ID Distinct Count
3	219

Dimension	Hierarchy	Operator	Filter Expression
Order Date	Order Date.PK TIME	Equal	
Order Date	Order Date.Rok	Not Equal	{ 2014 }
Order Date	Order Date.Miesiac Słownie	Equal	
Order Date	Order Date.Miesiac	Range (Inclusive)	6 : 8
Order Date	Order Date.Kwartal	Equal	
Order Date	Order Date.Dzien Tygodnia	Equal	
Order Date	Order Date.Dzien Miesiaca	Equal	
<Select dimension>			

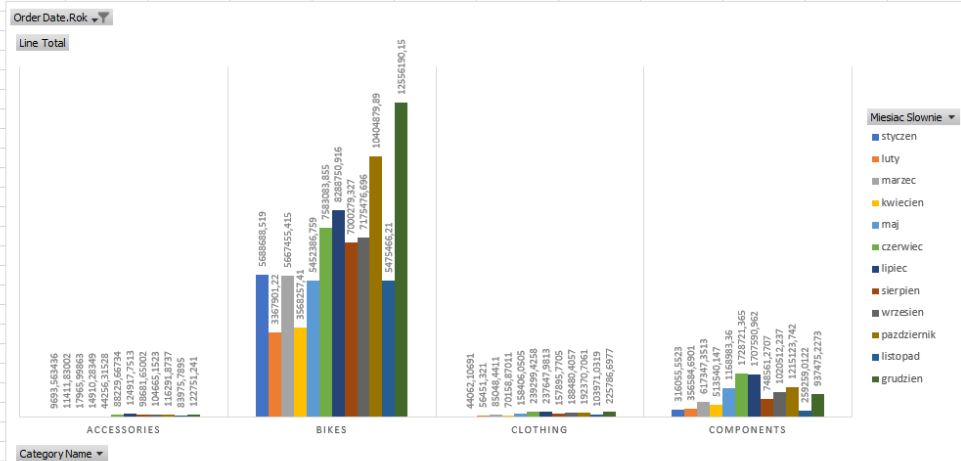
Product ID Distinct Count	Order Qty	Line Total
266	78006	28003659,9425666

c)

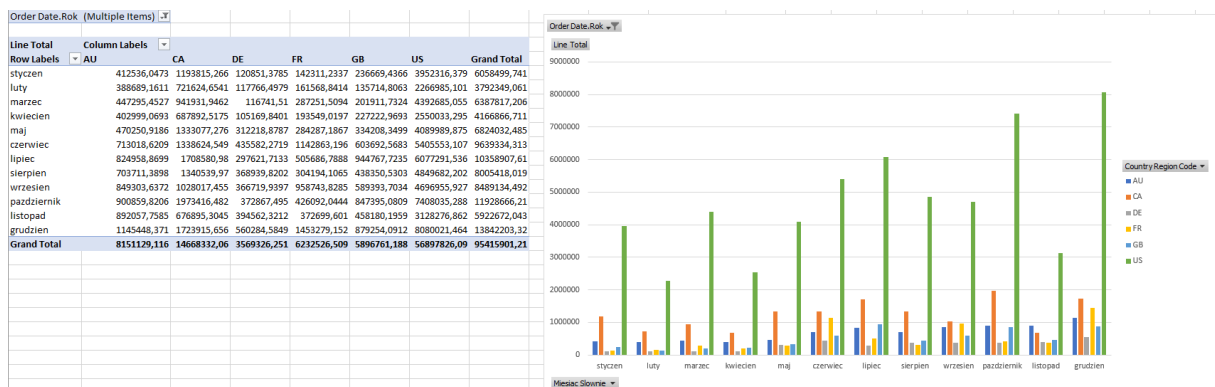


Wniosek: Dane uwzględniają jedynie lata 2011-2013 (pominąłem rok 2014, ponieważ ostatnie transakcje kończą się na czerwcu przez co zaburza to trend dla drugiej połowy roku). Na podstawie zebranych danych możemy stwierdzić, że początek roku (od stycznia do kwietnia włącznie) jest najmniej przychodowy (w porównaniu z innymi miesiącami). Jedynie listopad wydaje się ponownie spadać do poziomu pierwszego kwartału.

Order Date,Rok	(Multiple Items)												Y
Line Total	Column Labels												
Row Labels	styczen	luty	marzec	kwiecień	maj	czerwiec	lipiec	sierpień	wrzesień	październik	listopad	grudzień	Grand Total
Accessories	9693,563436	11411,83002	17965,99863	14910,28349	44256,31528	88229,66734	124917,7513	98681,65002	104665,1523	116291,8737	83975,7895	122751,241	837751,1161
Bikes	568868,519	3367901,22	5667455,415	3568257,41	5452386,759	7583083,855	8288750,916	7000279,327	7175476,696	10404879,89	5475466,21	12556190,15	82228816,37
Clothing	44062,10691	56451,321	85048,4411	70158,87011	158406,0505	239299,4258	237647,9813	157895,7705	188480,4057	192370,7061	103971,0319	225786,6977	1759578,809
Components	316055,5523	356584,6901	617347,3513	513540,147	1168983,36	1728721,365	1707590,962	748561,2707	1020512,237	1215123,742	259259,0122	937475,2273	10589754,92
Grand Total	6058499,741	3792349,061	6387817,206	4166866,711	6824032,485	9639334,313	10358907,61	8005418,019	8489134,492	11928666,21	5922672,043	13842203,32	95415901,21



Wniosek: Analizując miesięczne przychody ze względu na kategorie możemy dostrzec, że dla każdej kategorii produktów ta sama tendencja jest zachowana (styczeń-kwiecień + listopad są najgorsze). Jeśli chodzi o najbardziej przychodową kategorię to widać różnice: dla rowerów jest to grudzień (prezenty świąteczne?), a przykładowo dla części czerwiec+lipiec (naprawy w sezonie rowerowym?).



Wniosek: Analizując miesięczne przychody ze względu na kraje pochodzenia klientów możemy zauważyć silną korelację między słupkami Stanów Zjednoczonych a słupkami rowerów (z poprzedniego zrzutu ekranu). Oznacza to, że jednocześnie najbardziej dochodowe są rowery (ze wszystkich kategorii) a najwięcej pieniędzy wydają na nie ludzie ze Stanów Zjednoczonych. Dla pozostałych regionów ogólnie potwierdza się, że najgorszym jest okres styczeń-kwiecień.

Zad 3

a)

The screenshot displays the SQL Server Data Tools (SSDT) interface. On the left, a data flow diagram shows an 'OLE DB Source' connected to a 'Conditional Split' task. The 'Conditional Split' task has two output paths. To the right, the 'Columns' pane is empty. Below it, a table lists the conditions for the split:

Order	Output Name	Condition
1	FACT_SALES_2011	LEFT((DT_WSTR,8)OrderDate,4) == "2011" LEFT((DT_WST...
2	FACT_SALES_2012	LEFT((DT_WSTR,8)OrderDate,4) == "2012" LEFT((DT_WST...
3	FACT_SALES_2013	LEFT((DT_WSTR,8)OrderDate,4) == "2013" LEFT((DT_WST...
4	FACT_SALES_2014	LEFT((DT_WSTR,8)OrderDate,4) == "2014" LEFT((DT_WST...

Below the table, the 'Partition Wizard' dialog is open, showing the 'Specify Source Information' step. The 'Measure group' is 'FACT SALES 2011'. The 'Partition Source' is 'Data Source View.AW_2019_DSV_PARTI'. The 'Available tables' list includes:

- ☐ FACT_SALES_2011
- ☒ FACT_SALES_2013
- ☐ FACT_SALES_2012
- ☐ FACT_SALES_2014

b)

Item	Partition Name	Source	Estimated Rows	Storage Mode
1	FACT SALES 2011	SELECT [Pynka].[FACT_SALES].[ProductID],[Pynka].[FACT_...	0	MOLAP
2	FACT SALES 2012	SELECT [Pynka].[FACT_SALES].[ProductID],[Pynka].[FACT_...	0	MOLAP
3	FACT SALES 2013	SELECT [Pynka].[FACT_SALES].[ProductID],[Pynka].[FACT_...	0	MOLAP
4	FACT SALES 2014	<div> <div>Partition Source - FACT SALES 2014</div> <div> <div>Binding type: Query Binding</div> <div>Data source: DS_AW2019</div> <div> SELECT [Pynka].[FACT_SALES].[ProductID],[Pynka].[FACT_SALES].[CustomerID],[Pynka].[FACT_SALES].[SalesPersonID],[Pynka].[FACT_SALES].[OrderDate],[Pynka].[FACT_SALES].[ShipDate],[Pynka].[FACT_SALES].[OrderQty],[Pynka].[FACT_SALES].[UnitPrice],[Pynka].[FACT_SALES].[UnitPriceDiscount],[Pynka].[FACT_SALES].[LineTotal] FROM [Pynka].[FACT_SALES] WHERE LEFT([Pynka].[FACT_SALES].[OrderDate], 4)=2014 OR LEFT([Pynka].[FACT_SALES].[ShipDate], 4)=2014; </div> </div> </div>		

Zad 4

a)

Data Source Wizard

Select how to define the connection
You can select from a number of ways in which your data source will define its connection string.

☐ Create a data source based on another object

☒ Create a data source based on an existing or new connection

Data connections:

- 1KVKMNB.AdventureWorks2019.root
- 1KVKMNB.AdventureWorksDW2019.root

Data connection properties:

Property	Value
Data Source	DESKTOP-1KVKMNB
Initial Catalog	AdventureWorksDW2019
Persist Security...	True
Provider	SQLNCLI11.1
User ID	root

New... Delete

< Back Next > Finish >>| Cancel

b)

Select Tables and Views

Select objects from the relational database to be included in the data source view.



Available objects:

Name	Type
DatabaseLog (dbo)	Table
AdventureWorksDWBuildVersion (d...	Table
NewFactCurrencyRate (dbo)	Table
ProspectiveBuyer (dbo)	Table
sysdiagrams (dbo)	Table
<input type="checkbox"/> vAssocSeqLineItems (dbo)	View
<input type="checkbox"/> vAssocSeqOrders (dbo)	View
<input type="checkbox"/> vDMPrep (dbo)	View
<input type="checkbox"/> vTargetMail (dbo)	View
<input checked="" type="checkbox"/> vTimeSeries (dbo)	View

Filter:

☐ Show system objects

Included objects:

Name	Type
DimProductCategory (dbo)	Table
DimProductSubcategory (dbo)	Table
DimPromotion (dbo)	Table
DimReseller (dbo)	Table
DimSalesReason (dbo)	Table
DimSalesTerritory (dbo)	Table
DimScenario (dbo)	Table
FactAdditionalInternationalProductD...	Table
FactCallCenter (dbo)	Table
FactCurrencyRate (dbo)	Table
FactFinance (dbo)	Table
FactInternetSales (dbo)	Table
FactInternetSalesReason (dbo)	Table
FactProductInventory (dbo)	Table

> < >> <<

Add Related Tables

c)

Select Measure Group Tables

Select a data source view or diagram and then select the tables that will be used for measure groups.



Data source view:

AW_DW_2019_DSV

Measure group tables:

☐ DimScenario

☐ FactAdditionalInternationalProductDescription

☐ FactCallCenter

☐ FactCurrencyRate

☐ FactFinance

☒ FactInternetSales

☐ FactInternetSalesReason

☐ FactProductInventory

☐ FactResellerSales

☐ FactSalesQuota

☐ FactSurveyResponse

Suggest

< Back Next > Finish >> Cancel

Select New Dimensions

Select new dimensions to be created, based on available tables.



Dimension

- ☐ Dim Promotion
- ☒ Dim Date
 - ☒ DimDate
- ☒ Dim Sales Territory
 - ☒ DimSalesTerritory
- ☒ Dim Customer
 - ☒ DimCustomer
 - ☒ DimGeography
- ☒ Dim Currency
 - ☒ DimCurrency
- ☒ Dim Product
 - ☒ DimProduct
 - ☒ DimProductSubcategory
 - ☒ DimProductCategory
- ☐ Fact Internet Sales
 - ☐ FactInternetSales

< Back Next > Finish >> Cancel

d)

Na moim sprzęcie nie zaobserwowałem znaczącej różnicy między mniejszą a większą liczbą wybranych wymiarów (tzn. nawet wybierając wszystkie wymiary kostka procesuje się w 1s lub mniej). Zakładam, że posiadając dostatecznie dużo danych i wymiarów zauważyłbym wzrost czasu w przypadku zwiększenia liczby wymiarów.

e)

Attributes	
Dim Product	
Class	
Color	
Days To Manufacture	
English Product Category Name	
English Product Name	
English Product Subcategory Name	
List Price	
Product Category Key	
Product Key	
Product Subcategory Key	
Size	
Status	
Style	
Weight	

AW_DW_2019_Cube

Metadata

Search Model

<All>

AW_DW_2019_Cube

Measures

KPIs

Dim Currency

Dim Customer

Customer Key

English Education

Gender

Geography Key

Yearly Income

Dim Customer - Sales Territory

Dim Product

Dim Product.Class

Dim Product.Color

Dim Product.Days To Manufacture

Dim Product.English Product Category

Dim Product.English Product Name

Dim Product.English Product Subcategory

Dim Product.List Price

Dim Product.Product Category

Dim Product.Product Key

Dim Product.Product Subcategory

Dim Product.Size

Dim Product.Status

Dim Product.Style

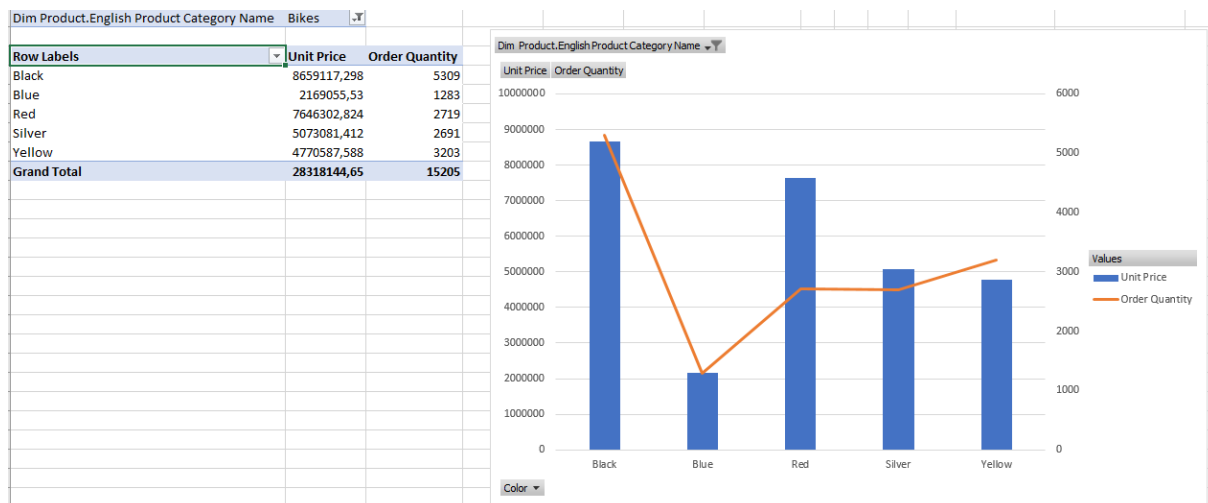
Dim Product.Weight

Dim Sales Territory

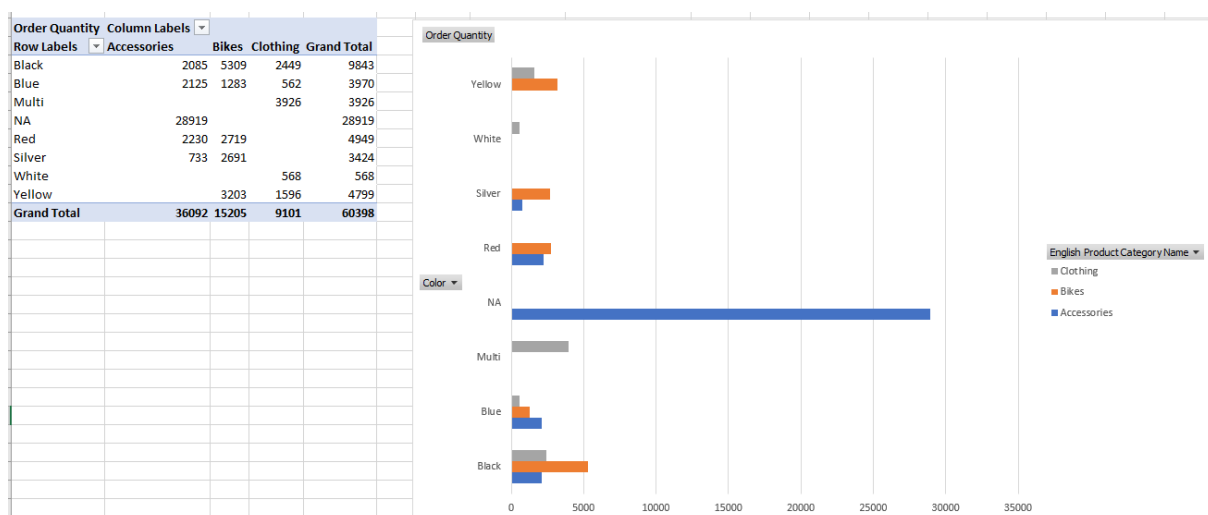
Due Date

Dimension	Hierarchy	Operator	Filter Expression
Dim Product	Dim Product.English Product Category...	Equal	{ Bikes }
Dim Customer	Yearly Income	Range (Inclusive)	20000 : 80000
<Select dimension>			

Color	Yearly Income	Order Quantity	Unit Price
Black	20000	482	604607,...
Black	30000	655	818819,...
Black	40000	728	1171536,...
Black	50000	133	200628,...
Black	60000	587	911611,...
Black	70000	862	1485392,...
Black	80000	419	801090,...
Blue	20000	111	132245,25
Blue	30000	143	194056,...
Blue	40000	186	279353,...
Blue	50000	42	56124,96
Blue	60000	180	260396,...
Blue	70000	204	419560,...
Blue	80000	104	220754,82
Red	20000	221	561709,...
Red	30000	330	880498,...
Red	40000	392	1078260,...
Red	50000	83	247662,...
Red	60000	436	1302287,...
Red	70000	331	959111,...
Red	80000	179	510307,...
Silver	20000	166	281884,...
Silver	30000	249	385223,...
Silver	40000	455	733398,...
Silver	50000	75	103877,...
Silver	60000	351	590390,...
Silver	70000	430	913028,...
Silver	80000	207	468527,...
Yellow	20000	264	346487,...

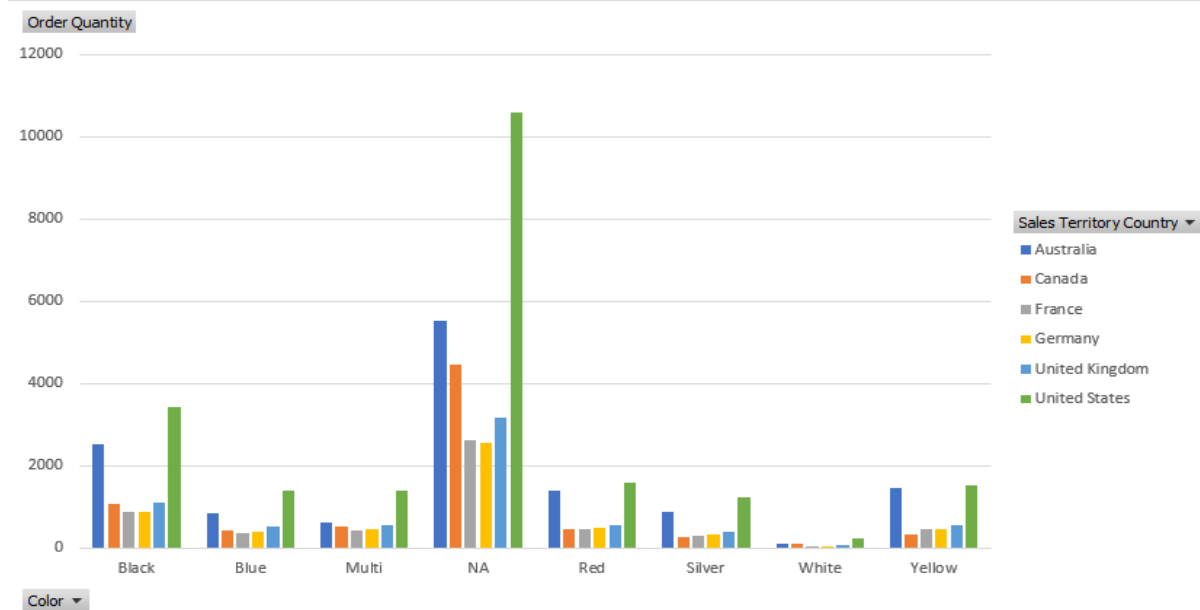


Wniosek: Najdroższe są rowery w kolorach czarnym oraz czerwonym, natomiast zdecydowanie najwięcej klienci zamawiali czarnych rowerów.

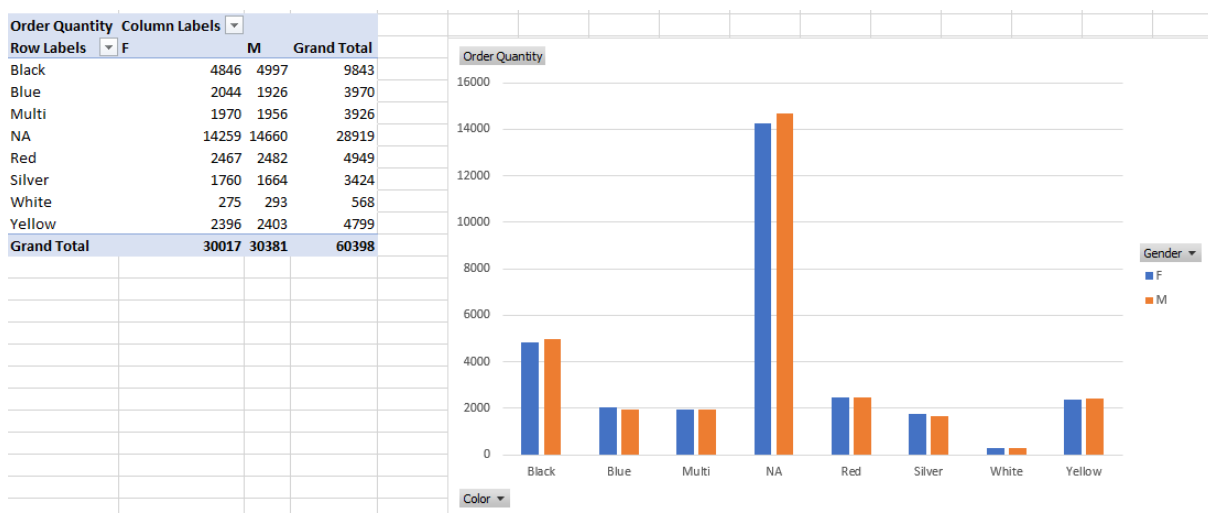


Wniosek: Patrząc ogólniej, dokonuję analizy liczby zamówionych przez klientów sztuk produktów ze względu na kolor oraz kategorię produktu. Okazuje się, że czarny kolor przeważa jedynie w przypadku rowerów, dla ubrań najpopularniejszy jest wariant „Multi”, natomiast dla akcesoriów zdecydowanie przeważają produkty bez podanego koloru (a wśród tych z określonym kolorem czarny jest tuż za czerwonym i niebieskim).

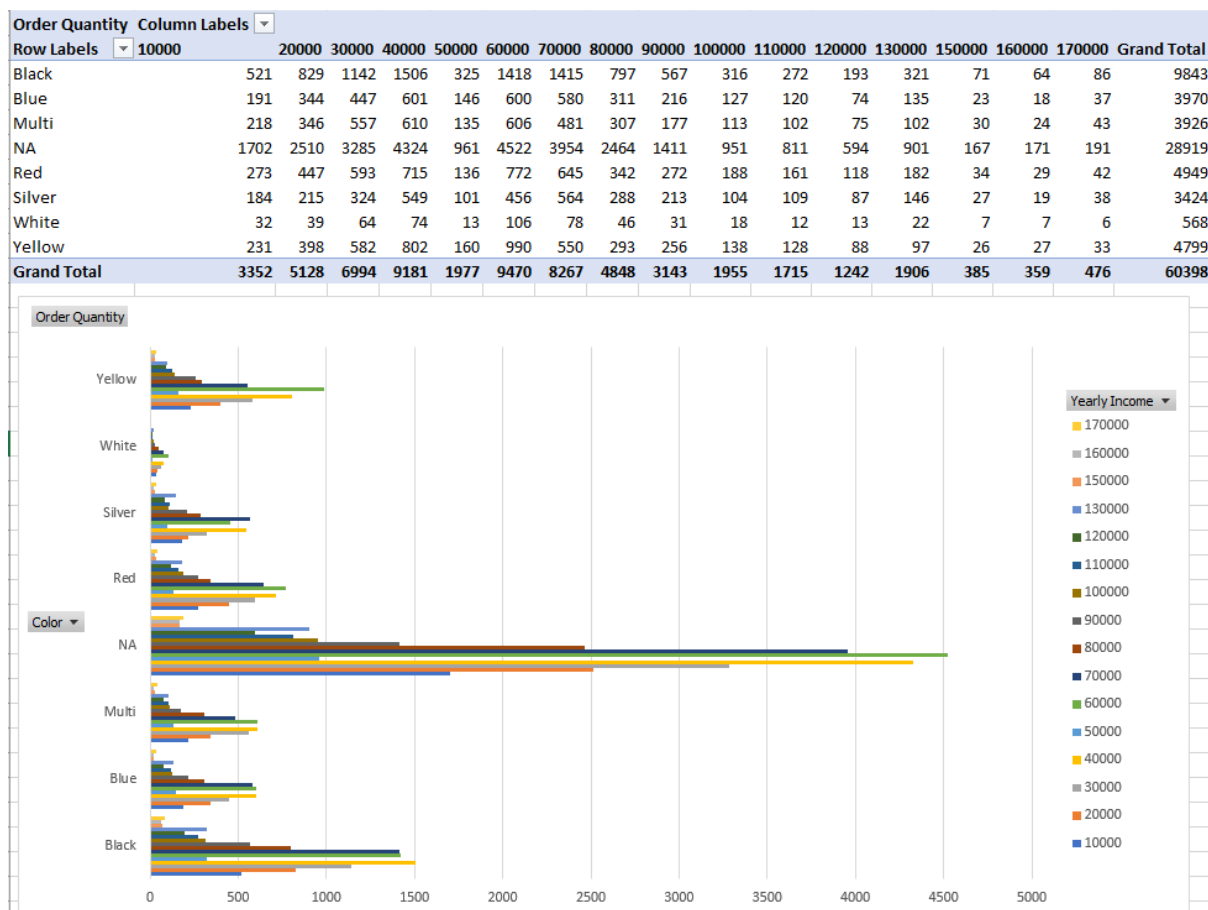
Order Quantity	Column Labels						
Row Labels	Australia	Canada	France	Germany	United Kingdom	United States	Grand Total
Black	2515	1061	882	874	1094	3417	9843
Blue	851	439	376	409	507	1388	3970
Multi	605	509	412	461	562	1377	3926
NA	5528	4450	2626	2556	3177	10582	28919
Red	1391	470	472	496	545	1575	4949
Silver	877	263	307	331	404	1242	3424
White	112	89	32	35	63	237	568
Yellow	1466	339	451	463	554	1526	4799
Grand Total	13345	7620	5558	5625	6906	21344	60398



Wniosek: Niezależnie od kraju pochodzenia klienta, nadal przeważający jest kolor czarny (jeśli pominiemy masę zakupionych akcesoriów bez określonego koloru).



Wniosek: Niezależnie od płci klienta nadal dominującym kolorem ze względu na liczbę sprzedanych sztuk jest czarny.



Wniosek: Najwięcej sztuk produktów zamawiają klienci z rocznymi przychodami 40000 lub 60000 (zakładam ich lokalnej waluty). Niezależnie jednak od rocznych przychodów, kolor czarny wydaje się dominować w przypadku liczby kupowanych sztuk.

Ogólnie można stwierdzić, że ze względu na dane demograficzne jednoznacznie najczęściej kupowane są produkty w kolorze czarnym (niezależnie od płci, przychodów czy kraju pochodzenia klienta) lub bez określonego koloru, jeżeli bierzemy je pod uwagę (całość takowych stanowią akcesoria). Patrząc na kategorie produktów z asortymentu sklepowego to kolor czarny dominuje jedynie dla rowerów.

Wnioski:

- Jako że nie zawsze potrzebujemy wszystkich rekordów z tablicy faktów do analizy, to warto je podzielić na mniejsze „porcje” w celu zwiększenia wydajności procesu analizy danych. Możemy tego dokonać tworząc partycje, albo ze względu na już istniejący podział tabel faktów, albo niejako filtrując odpowiednie rekordy z jednej tablicy np. po roku.
- Narzędzia takie jak SSIS i SSAS w ramach Visual Studio pozwalają w bardzo przejrzysty i szybki sposób przygotować dane do analizy oraz dopasować wymiary / miary pod konkretne zagadnienia, które pragniemy przeanalizować.