

## Data Warehouse Lab 5

### Task 1

1. i. Which products are sold the most on [Year, Semester, Quarter, Month, Day of week, Season, Hour, time of day (like morning, afternoon, evening, night)]

Tables and Attributes: Product (Name), SalesOrderDetail (LineTotal), SalesOrderHeader (DueDate), SalesOrderHeader (Status)

ii. Which product types [Categories/Subcategories/ProductionLine/Color/Size/etc.] generate the most profit

Tables and Attributes: ProductCategory (Name), ProductSubcategory (Name), Product (ProductLine), Product (Color), Product (Size), SalesOrderDetail (LineTotal), SalesOrderHeader (Status)

iii. What is the average discount given in a certain sales territory [Name/Group/Continent] for products [still in sales/ discontinued]

Tables and Attributes: SalesTerritory (Name), SalesTerritory (Group), SalesOrderDetail (UnitPriceDiscount), Product (FinishedGoodsFlag)

iv. Which products have the highest increase in cost in last [Year/Month/Week/etc.]

Tables and Attributes: Product (Name), SalesOrderDetail (UnitPrice), SalesOrderHeader (DueDate)

v. Do most popular (occur on most transactions) products generate more profit than less popular products.

Tables and Attributes: Product (Name), SalesOrderDetail (OrderQty), SalesOrderDetail (LineTotal)

### 2. Data Profiling of the tables related to Products and SalesTerritory

	Table	Attribute name	Type of data	Unique values	Null values	Quality assessment
1.	Product	Name	nvarchar(50)	504	0	All clear
2.	Product	Color	nvarchar(15)	0	248	Null values
3.	Product	Size	nvarchar(5)	0	293	Null values
4.	Product	ProductLine	nchar(2)	0	226	Null values
5.	Product	FinishedGoods Flag	bit	0	0	All clear

6.	SalesTerritory	Name	nvarchar(50)	10	0	All clear
7.	SalesTerritory	Group	nvarchar(50)	1	0	All clear

b. Prepare quality checks for identified data anomalies – checks should be handled by proper SQL queries

1. SELECT \*  
FROM Production.Product  
WHERE ProductColor is NULL
2. SELECT \*  
FROM Production.Product  
WHERE ProductSize is NULL
3. SELECT \*  
FROM Production.Product  
WHERE ProductLine is NULL

## Task 2

1. PRODUCT table has been imported with useful attributes and data

	ProductID	ProductName	ProductColor	ProductSize	ProductLine	ProductFinishedGoodsFlag
1	1	Adjustable Race	NULL	NULL	NULL	0
2	2	Bearing Ball	NULL	NULL	NULL	0
3	3	BB Ball Bearing	NULL	NULL	NULL	0
4	4	Headset Ball Bearings	NULL	NULL	NULL	0
5	316	Blade	NULL	NULL	NULL	0
6	317	LL Crankarm	Black	NULL	NULL	0
7	318	ML Crankarm	Black	NULL	NULL	0
8	319	HL Crankarm	Black	NULL	NULL	0
9	320	Chainring Bolts	Silver	NULL	NULL	0
10	321	Chainring Nut	Silver	NULL	NULL	0
11	322	Chainring	Black	NULL	NULL	0
12	323	Crown Race	NULL	NULL	NULL	0
13	324	Chain Stays	NULL	NULL	NULL	0
14	325	Decal 1	NULL	NULL	NULL	0
15	326	Decal 2	NULL	NULL	NULL	0
16	327	Down Tube	NULL	NULL	NULL	0
17	328	Mountain End Caps	NULL	NULL	NULL	0

2. SALES TERRITORY table has been imported to the new database

	TerritoryID	TerritoryName	TerritoryGroup
1	1	Northwest	North America
2	2	Northeast	North America
3	3	Central	North America
4	4	Southwest	North America
5	5	Southeast	North America
6	6	Canada	North America
7	7	France	Europe
8	8	Germany	Europe
9	9	Australia	Pacific
10	10	United Kingdom	Europe

3. PRODUCT CATEGORY table has been imported to the new database

	ProductCategoryID	ProductCategoryName
1	1	Bikes
2	2	Components
3	3	Clothing
4	4	Accessories

4. PRODUCT SUBCATEGORY table has been imported to the new database

	ProductSubcategoryID	ProductSubcategoryName	ProductCategoryID
1	1	Mountain Bikes	1
2	2	Road Bikes	1
3	3	Touring Bikes	1
4	4	Handlebars	2
5	5	Bottom Brackets	2
6	6	Brakes	2
7	7	Chains	2
8	8	Cranksets	2
9	9	Deraillleurs	2
10	10	Forks	2
11	11	Headsets	2
12	12	Mountain Frames	2
13	13	Pedals	2

5. SALES ORDER DETAIL table has been imported to the new database

	SalesOrderID	SalesOrderDetailID	LineTotal	UnitPriceDiscount	UnitPrice	OrderQty
1	54409	52401	21.490000	0.00	21.49	1
2	54409	52402	3.990000	0.00	3.99	1
3	54409	52403	2.290000	0.00	2.29	1
4	54410	52404	4.990000	0.00	4.99	1
5	54410	52405	34.990000	0.00	34.99	1
6	54411	52406	4.990000	0.00	4.99	1
7	54411	52407	34.990000	0.00	34.99	1
8	54412	52408	35.000000	0.00	35.00	1
9	54412	52409	4.990000	0.00	4.99	1
10	54413	52410	4.990000	0.00	4.99	1
11	54413	52411	35.000000	0.00	35.00	1
12	54413	52412	8.990000	0.00	8.99	1
13	54414	52413	35.000000	0.00	35.00	1
14	54414	52414	2.290000	0.00	2.29	1
15	54415	52415	9.990000	0.00	9.99	1

6. SALES ORDER HEADER table has been imported

	SalesOrderID	DueDate	Status
1	43659	2011-06-12 00:00:00.000	5
2	43660	2011-06-12 00:00:00.000	5
3	43661	2011-06-12 00:00:00.000	5
4	43662	2011-06-12 00:00:00.000	5
5	43663	2011-06-12 00:00:00.000	5
6	43664	2011-06-12 00:00:00.000	5
7	43665	2011-06-12 00:00:00.000	5
8	43666	2011-06-12 00:00:00.000	5
9	43667	2011-06-12 00:00:00.000	5
10	43668	2011-06-12 00:00:00.000	5
11	43669	2011-06-12 00:00:00.000	5
12	43670	2011-06-12 00:00:00.000	5

### Task 3

In the TimeDIM table year, quarter, day attributes data type should be stored as an int.

In OrdersFact table Quantity attribute should be stored in Int data type.

In the ProductDIM table ListPrice, Weight, StandardCost, DaysToManufacture should be stored as an Int datatype and IsPurchased should be stored as Bit. No attributes about discounts.

#### Task 4

2. a. Prepare SQL query (using SELECT INTO) to create ProductDIM and move the data.

```
SELECT a.ProductID, a.ProductName, a.ProductColor, a.ProductSize, a.ProductLine,
a.ProductFinishedGoodsFlag, b.ProductSubcategoryName, c.ProductCategoryName
INTO ProductDIM
FROM dbo.Product AS a
    LEFT JOIN dbo.ProductSubcategory AS b
        ON b.ProductSubcategoryID = a.ProductSubcategoryID
    LEFT JOIN dbo.ProductCategory AS c
        ON c.ProductCategoryID = b.ProductCategoryID;
GO
```

b. Prepare SQL statements (using INSERT INTO) to create SalesLocationDIM and move the data.

```
SELECT *
INTO SalesLocationDIM
FROM dbo.SalesTerritory AS a;
GO
```

c. Prepare SQL queries/statements to create TimeDIM

```
SELECT a.SalesOrderID AS 'TimeDIM_ID' , CAST(a.DueDate AS DATE) AS 'Date',
a.DueDate AS 'FullDate', Datepart(year, a.DueDate) AS 'Year', Datepart(quarter, a.DueDate)
AS 'Quarter', Datepart(month, a.DueDate) AS 'Month', Datepart(day, a.DueDate) AS 'Day'
INTO TimeDIM
FROM dbo.SalesOrderHeader AS a;
GO
```

d. Prepare SQL statements to create OrdersFact table and move the data.

```
SELECT a.*, b.Status, b.TerritoryID
INTO OrdersFact
FROM dbo.SalesOrderDetail AS a
    LEFT JOIN dbo.SalesOrderHeader AS b
        ON b.SalesOrderID = a.SalesOrderID
GO
```

3. Please verify data quality and propose cleaning transformations

!All the queries are in Text File and SQL Files will be attached

a. Use the quality checks identified and defined earlier

ProductDIM NULL Values has been removed, now it decrease from 504 to 211 rows

	ProductID	ProductName	ProductColor	ProductSize	ProductLine	ProductFinishedGoodsFlag	ProductSubcategoryName	ProductCategoryName
1	680	HL Road Frame - Black, 58	Black	58	R	1	Road Frames	Components
2	706	HL Road Frame - Red, 58	Red	58	R	1	Road Frames	Components
3	709	Mountain Bike Socks, M	White	M	M	1	Socks	Clothing
4	710	Mountain Bike Socks, L	White	L	M	1	Socks	Clothing
5	713	Long-Sleeve Logo Jersey, S	Multi	S	S	1	Jerseys	Clothing
6	714	Long-Sleeve Logo Jersey, M	Multi	M	S	1	Jerseys	Clothing
7	715	Long-Sleeve Logo Jersey, L	Multi	L	S	1	Jerseys	Clothing
8	716	Long-Sleeve Logo Jersey, XL	Multi	XL	S	1	Jerseys	Clothing
9	717	HL Road Frame - Red, 62	Red	62	R	1	Road Frames	Components
10	718	HL Road Frame - Red, 44	Red	44	R	1	Road Frames	Components
11	719	HL Road Frame - Red, 48	Red	48	R	1	Road Frames	Components
12	720	HL Road Frame - Red, 52	Red	52	R	1	Road Frames	Components
13	721	HL Road Frame - Red, 56	Red	56	R	1	Road Frames	Components
14	722	LL Road Frame - Black, 58	Black	58	R	1	Road Frames	Components
15	723	LL Road Frame - Black, 60	Black	60	R	1	Road Frames	Components

Other tables do not have NULL values and duplicate values.

b. Think about extending available information

Number of month has been changed into month name as “June”

	TimeDIM_ID	Date	FullDate	Year	Quarter	Month	Day
1	43659	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
2	43660	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
3	43661	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
4	43662	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
5	43663	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
6	43664	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
7	43665	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
8	43666	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
9	43667	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
10	43668	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
11	43669	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
12	43670	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
13	43671	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
14	43672	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
15	43673	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12
16	43674	2011-06-12	2011-06-12 00:00:00.000	2011	2	June	12

“1” Product Finished Goods Flag values have been changed to verbose names, like “Salable/Not salable”

	ProductID	ProductName	ProductColor	ProductSize	ProductLine	ProductFinishedGoodsFlag	ProductSubcategoryName	ProductCategoryName
1	680	HL Road Frame - Black, 58	Black	58	R	Salable	Road Frames	Components
2	706	HL Road Frame - Red, 58	Red	58	R	Salable	Road Frames	Components
3	709	Mountain Bike Socks, M	White	M	M	Salable	Socks	Clothing
4	710	Mountain Bike Socks, L	White	L	M	Salable	Socks	Clothing
5	713	Long-Sleeve Logo Jersey, S	Multi	S	S	Salable	Jerseys	Clothing
6	714	Long-Sleeve Logo Jersey, M	Multi	M	S	Salable	Jerseys	Clothing
7	715	Long-Sleeve Logo Jersey, L	Multi	L	S	Salable	Jerseys	Clothing
8	716	Long-Sleeve Logo Jersey, XL	Multi	XL	S	Salable	Jerseys	Clothing
9	717	HL Road Frame - Red, 62	Red	62	R	Salable	Road Frames	Components
10	718	HL Road Frame - Red, 44	Red	44	R	Salable	Road Frames	Components
11	719	HL Road Frame - Red, 48	Red	48	R	Salable	Road Frames	Components
12	720	HL Road Frame - Red, 52	Red	52	R	Salable	Road Frames	Components
13	721	HL Road Frame - Red, 56	Red	56	R	Salable	Road Frames	Components
14	722	LL Road Frame - Black, 58	Black	58	R	Salable	Road Frames	Components
15	723	LL Road Frame - Black, 60	Black	60	R	Salable	Road Frames	Components
16	724	LL Road Frame - Black, 62	Black	62	R	Salable	Road Frames	Components
17	725	LL Road Frame - Red, 44	Red	44	R	Salable	Road Frames	Components
18	726	LL Road Frame - Red, 48	Red	48	R	Salable	Road Frames	Components
19	727	LL Road Frame - Red, 52	Red	52	R	Salable	Road Frames	Components
20	728	LL Road Frame - Red, 58	Red	58	R	Salable	Road Frames	Components

“5” Status values in the OrdersFact table have been changed to verbose names, like “Shipped”

	TimeDIM_ID	SalesOrderDetailID	LineTotal	UnitPriceDiscount	UnitPrice	OrderQty	ProductID	Status	TerritoryID
1	43667	77	17.100000	0.00	5.70	3	710	Shipped	3
2	43667	78	2039.994000	0.00	2039.994	1	773	Shipped	3
3	43667	79	2024.994000	0.00	2024.994	1	778	Shipped	3
4	43667	80	2024.994000	0.00	2024.994	1	775	Shipped	3
5	43669	110	714.704300	0.00	714.7043	1	747	Shipped	1
6	43684	243	1749.588000	0.00	874.794	2	754	Shipped	5
7	43684	244	838.917800	0.00	419.4589	2	762	Shipped	5
8	43684	245	419.458900	0.00	419.4589	1	760	Shipped	5
9	43684	246	1749.588000	0.00	874.794	2	758	Shipped	5
10	43684	247	419.458900	0.00	419.4589	1	770	Shipped	5
11	43684	248	419.458900	0.00	419.4589	1	763	Shipped	5
12	43686	253	2624.382000	0.00	874.794	3	758	Shipped	1
13	43686	254	419.458900	0.00	419.4589	1	762	Shipped	1
14	43686	255	419.458900	0.00	419.4589	1	770	Shipped	1

## Task 5

1. Product's sales over time and location, including:

a. Revenue for different (years, quarters, months, days).

Year	LineTotal
2011	12,390,235.63
2	747,358.66
June	747,358.66
12	501,902.79
13	13,931.52
14	15,012.18
15	7,156.54
16	15,012.18
17	14,313.08
18	7,855.64
19	7,855.64
20	20,909.78

b. Revenue for different sales locations.

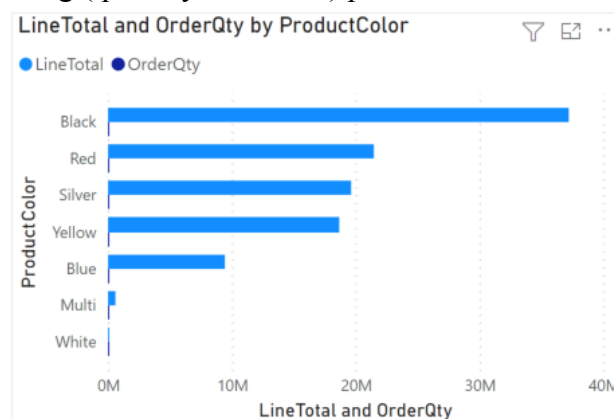
LineTotal by TerritoryName and TerritoryGroup

TerritoryGroup ● Europe ● North America ● Pacific

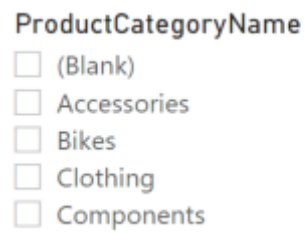


2. Product's sales, excluding product categories and including:

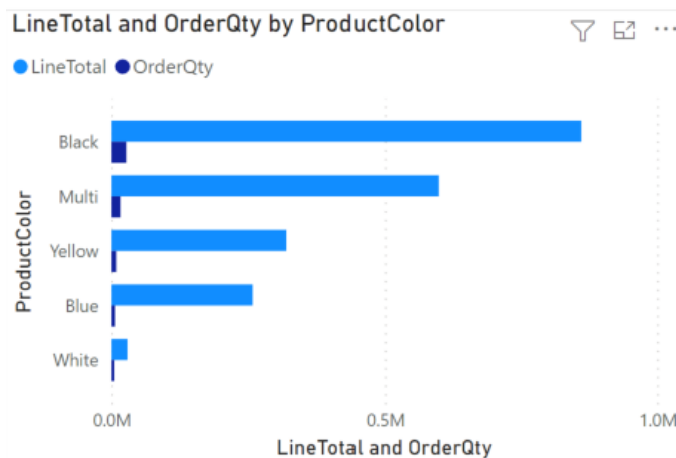
c. Worst and bestselling (quantity and value) product's color.



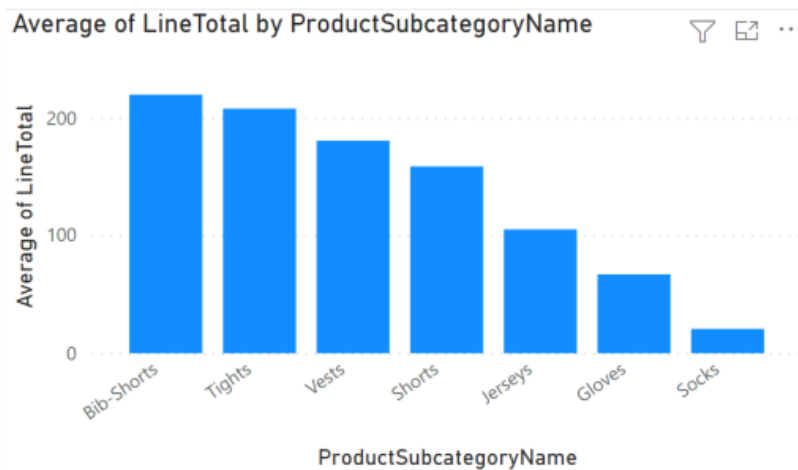
3. Product's category sales, including:
- e. For a selected category (use slicer)



- i. worst and bestselling (quantity and value) product's color, subcategory, etc.



- ii. average price depending on different subcategories



- iii. revenue (quantity and turnover) from product subcategories, products for different years (quarters, months, days, days of week).

Year	Bib-Shorts	Gloves	Jerseys	Shorts	Socks	Tights	Vests	Total
2011			29,524.55		3,690.39			33,214.94
2012	108,395.99	95,614.00	113,762.30	52,824.97	2,883.00	130,724.89		504,205.14
2013	58,694.68	113,818.13	385,047.35	222,373.13	13,965.52	72,131.73	157,503.51	1,023,534.06
2014	467.95	34,079.86	223,925.19	138,402.40	9,206.22	292.46	101,984.85	508,358.94
<b>Total</b>	<b>167,558.62</b>	<b>243,511.98</b>	<b>752,259.39</b>	<b>413,600.51</b>	<b>29,745.13</b>	<b>203,149.08</b>	<b>259,488.37</b>	<b>2,069,313.08</b>