UTM
UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING
Faculty of Engineering

GROUP PROJECT

MCSD1123 – 01

BIG DATA MANAGEMENT

Lecturer's Name: Dr. Nor Haizan Mohamed Radzi

Student's Name:

Ahmed Hashim Taha Salim (MCS211041)

Divyasini A.Konasegar (MCS222002)

Ibrahim Sunusi Alkali (MCS222003)

TABLE OF CONTENTS

1.0 Executive Summary

1.1 Selected Dataset

The Adventure works dataset is obtained from Kaggle, an open-source platform which provides a large number of datasets and problems for users to work on, and it also provides tools for users to submit and evaluate their models. Adventure Works | Kaggle

The Adventure Works dataset consists of 18,000 customers, both women and men are involved in purchasing the company's products. Each of the customers' data recorded consists of their personal information including customer Key, first Name, last Name, email address, and annual income.

The dataset also involves products and its subcategories that includes different brands of bikes that are categorized into four groups: bikes, components, clothing, and accessories. Since the company manufactures its own components, accessories, and clothing. The primary selling products of AW are several different brands of bikes including mountain, road, and touring bikes.

It also has more than 56,000 sales records from years 2015, 2016, and 2017 with their returned orders and the territories where it has shipped from and into.

1.2 Business Background

Adventure works cycles, is a company on which the adventure works sample databases are based, is a large, multinational manufacturing company. The company manufactures and sells metal and composite bicycles to ten territories, including north-American, European and Asian commercial markets. The company is focusing on four different categories like Bikes, Components, Clothing and Accessories. While its base operation is located in Bothell, several regional sales teams are located throughout their market base.

In early 2001, adventure works cycles bought a small manufacturing plant located in Mexico. And most of the products were modelled by outside vendors. In such instances, AW becomes the reseller. The company started manufactures several critical subcategories for the AW cycles products. These subcategories are shipped to those vendor's location for final product assembly. In late 2001, the company stock started to store different assembled products and ship these products to customers throughout the United States, Canada, France, Germany, Australia, and the United Kingdom

1.3 The Schema

After data collection, we have created the Adventure Works Database Schema by combining different attributes of the data and created ten different tables. we have combined different types of products and its subcategories with the products main four categories, and we've collected the sales of these products in three-year operations and the sales return in this period and combined it with sales territories such as regions and countries. We also obtained the customers with their personal data such as names, Email and their annual income.

We had to make sure that the date type of all the columns is in the right format, for example DATE type had to be changed into "YYYY-MM-DD" format, as well as Annual Income containing US dollar Symbol "$", we had to remove it so it matches the integer data type and gives us access to make statistical queries or order by ascending and descending. We also had to assure that the primary keys must be unique.

For more understanding the relationships of the tables can be summarized in Diagram: 1 and in Table: 1 sequentially.

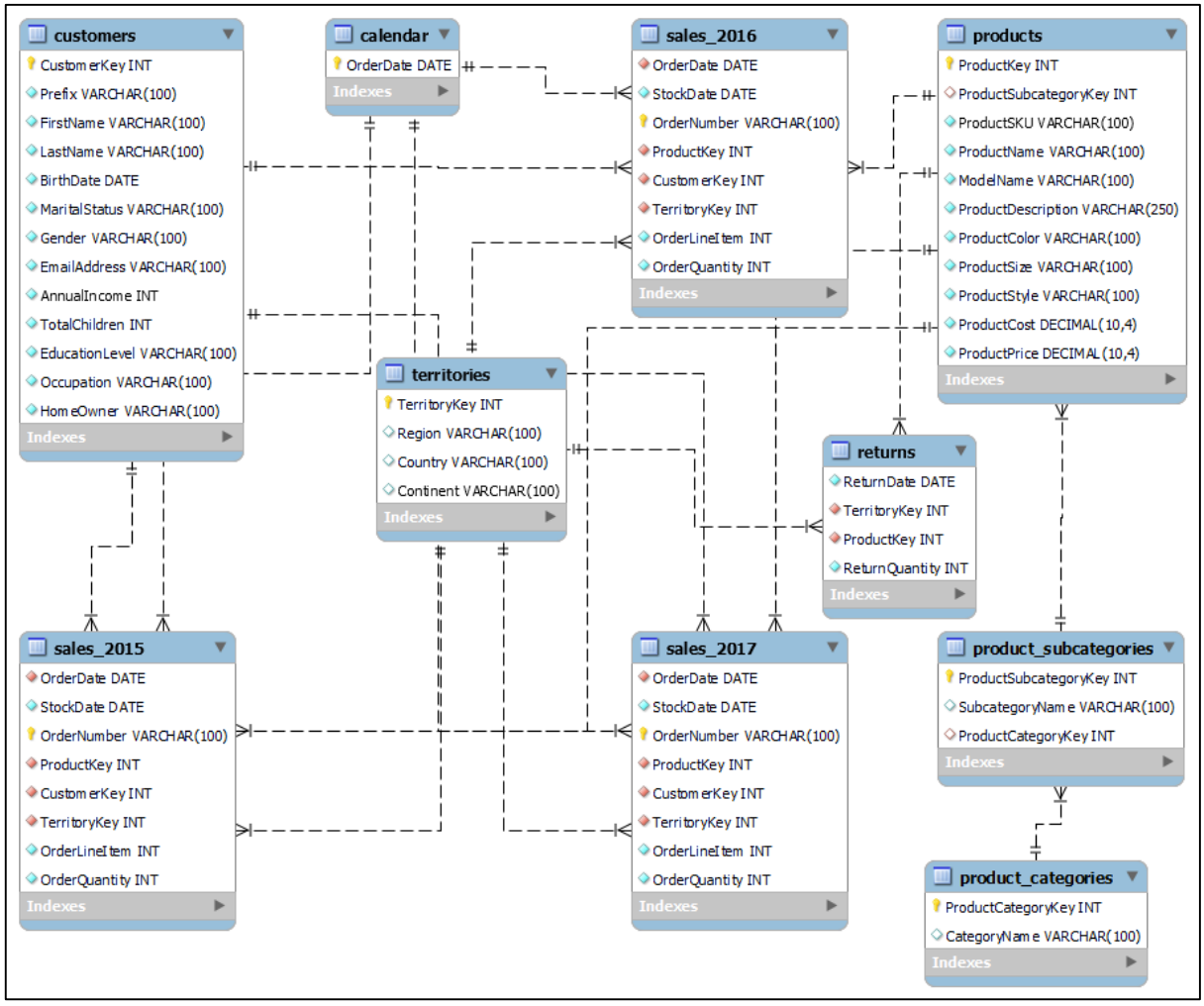


Diagram 1: Reverse engineer a database

Diagram 1 shows the relational schema diagram of the Adventure Works database, which elucidates that the selected database consists of ten relations tables namely, customers, products, products_subcategories, product_categories, calendar, territories, sales_2015, sales_2016, sales_2017 and returns.

In each of these sales tables, we had to change the order numbers in the OrderNumber table due to duplicates found in the dataset, and as it is a primary key it must involves unique values. These sales table have combined most of the other tables primary keys as a foreign key's in it, as well as sharing one-to-many relations from them.

No.	Schema	Contains objects related to	Examples
1	Products	Product Keys **(PK)** Product subcategories **(FK)** Product names, model names Product price	Products table Subcategories table Categories table
2	Sales return	Return quantity Product key and Territory key	Return table Products table Territories table
3	Sales in 2015 Sales in 2016 Sales in 2017	Order Numbers **(PK)** Customer Keys **(PK)** Territory Keys **(FK)** Product Keys **(FK)** Order Date **(FK)** Stock Date and Order Quantity Customers and sales related data	Sales_2015 table Sales_2016 table Sales_2017 table Customer table Products table Territories table Calendar table
4	Customers	Customer Key **(PK)** Customers personal data (First Name, Email Address Annual Income. Etc.)	Sales in 2016 Customer table Products table Territories table Calendar table

Table 1: Relationships between tables

2.0 Statistical queries

Statistical queries were performed for a better understanding of the selected dataset. We have created and performed statistical queries to examine the dataset which is able to produce count, mean, median, and average values. For instance, average sales for the most purchased products in 2015, 2016, and 2017, Median value of annual income of our customers and mean value of product sales from years 2015, 2016, and 2017.

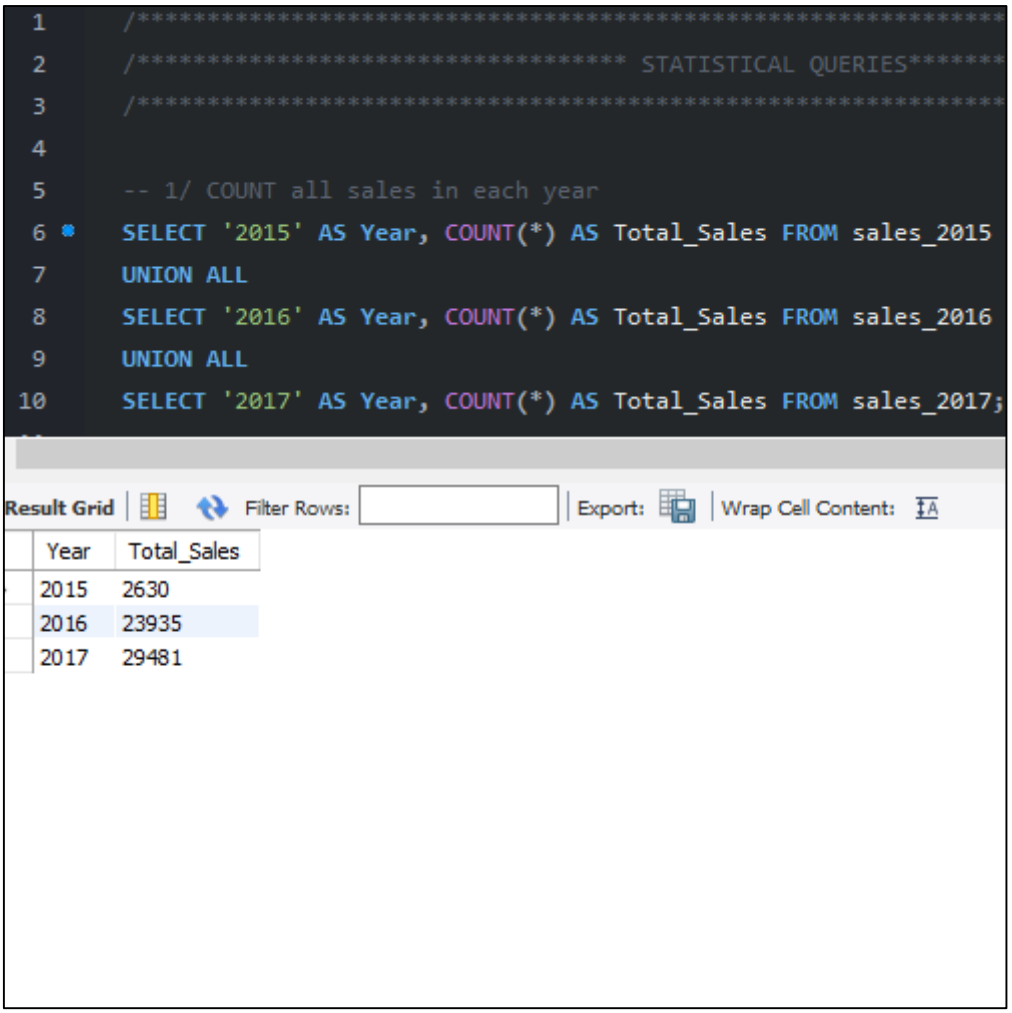


Figure 1: All sales counts in each year

Figure 1 using a statistical metric ‘COUNTS’ and shows the total sales counts in each year.

```
13 •  SELECT '2015' AS Year, SUM(ReturnQuantity) AS Total_Returns FROM returns
14    WHERE ReturnDate BETWEEN '2015-01-01' AND '2015-12-31'
15    UNION ALL
16    SELECT '2016' AS Year, SUM(ReturnQuantity) AS Total_Returns FROM returns
17    WHERE ReturnDate BETWEEN '2016-01-01' AND '2016-12-31'
18    UNION ALL
19    SELECT '2017' AS Year, SUM(ReturnQuantity) AS Total_Returns FROM returns
20    WHERE ReturnDate BETWEEN '2017-01-01' AND '2017-12-31';
```

esult Grid | Filter Rows: | Export: | Wrap Cell Content:

Year	Total_Returns
2015	86
2016	770
2017	972

Figure 2: SUM of return quantities in each year

The above Figure 2 shows a statistical metric ‘SUM’ and is used to calculate the sum of all return quantities in each year.

```
21
22    -- 3/ Calculate the average age of all customers
23 •  SELECT AVG(EXTRACT( YEAR FROM DATE('2023-01-17')) - (EXTRACT(YEAR FROM BirthDate))) AS average_age
24    FROM customers;
25
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

average_age
61.4156

Figure 3: AVG age of all customers

Figure 3 illustrate the average age of all customers, where it extracts the current date and substrate it with the Birth Date column, whereas in DATE format.

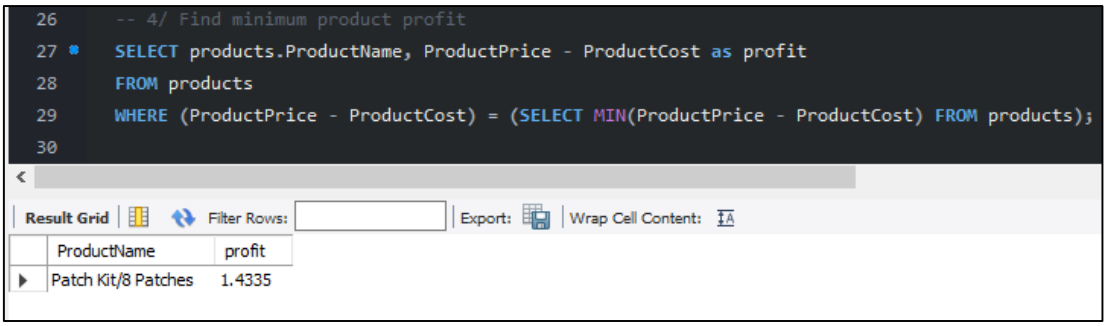


Figure 4: minimum product profit

Figure 4 shows the minimum product profit by subtracting the product price by its initial cost.

3.0 Scenario for Adventure Works

The management of Adventure Works would like to have a complete analysis during the three-years operation. An analysis of how well a firm's products is selling, and the success of their sales forces by using product's initial costs analysis. As well as using customer segmentation to further understand the customers and do better marketing or promotions to the them, by analysing customer's personal data like annual incomes and their occupation and mark their preferred products.

The company would like to know the top sold products in each year. In addition, this analysis may aim at discovering the territory and regions with which most of the sales were made. Moreover, in order to increase revenue, we would like to discover the weakness of the company, for instance; products that were returned by the customers as a result of some damages, faults or any kind of complaint lodged by the customers. So that the management would tackle and handle such lapses in the future. Setting role-specific targets, such as a revenue target and a sales productivity goal, is a typical example of a sales analytics activity.

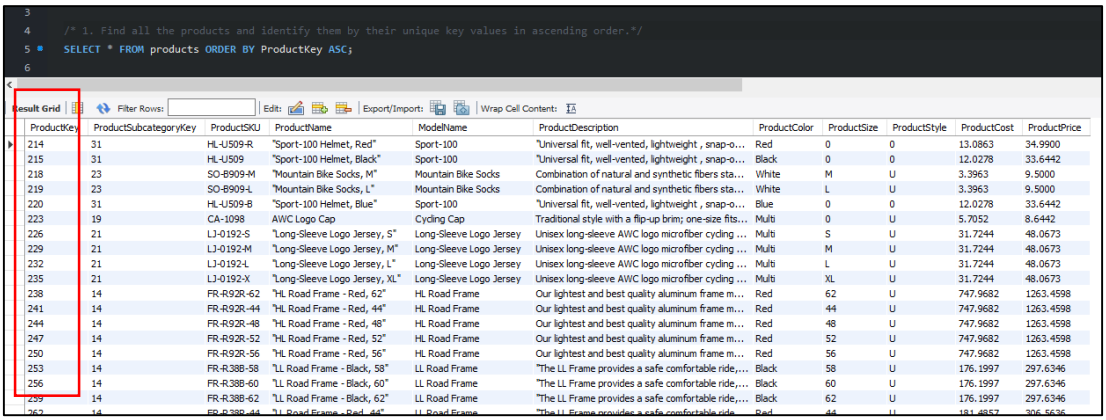
Based on the available dataset and the database we created earlier, we would have to undergo some queries as follows:

3.1 Products queries

3.1.1 Find all the products and identify them by their unique key values in ascending order

Query - `select *from products order by ProductKey asc;`

Output:



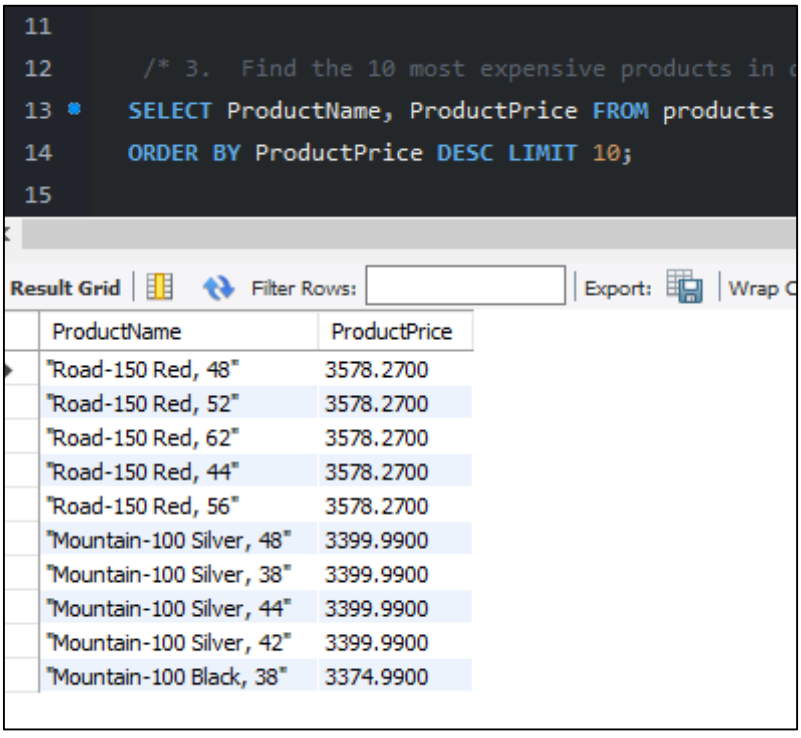
```
/* 1. Find all the products and identify them by their unique key values in ascending order.*/
SELECT * FROM products ORDER BY ProductKey ASC;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

ProductKey	ProductSubcategoryKey	ProductSKU	ProductName	ModelName	ProductDescription	ProductColor	ProductSize	ProductStyle	ProductCost	ProductPrice
214	31	HL-U509-R	"Sport-100 Helmet, Red"	Sport-100	"Universal fit, well-vented, lightweight , snap-o...	Red	0	0	13.0863	34.9900
215	31	HL-U509	"Sport-100 Helmet, Black"	Sport-100	"Universal fit, well-vented, lightweight , snap-o...	Black	0	0	12.0278	33.6442
218	23	SO-B909-M	"Mountain Bike Socks, M"	Mountain Bike Socks	Combination of natural and synthetic fibers sta...	White	M	U	3.3963	9.5000
219	23	SO-B909-L	"Mountain Bike Socks, L"	Mountain Bike Socks	Combination of natural and synthetic fibers sta...	White	L	U	3.3963	9.5000
220	31	HL-U509-B	"Sport-100 Helmet, Blue"	Sport-100	"Universal fit, well-vented, lightweight , snap-o...	Blue	0	0	12.0278	33.6442
223	19	CA-1098	AWC Logo Cap	Cycling Cap	Traditional style with a flip-up brim; one-size fits...	Multi	0	U	5.7052	8.6442
226	21	LJ-0192-S	"Long-Sleeve Logo Jersey, S"	Long-Sleeve Logo Jersey	Unisex long-sleeve AWC logo microfiber cycling ...	Multi	S	U	31.7244	48.0673
229	21	LJ-0192-M	"Long-Sleeve Logo Jersey, M"	Long-Sleeve Logo Jersey	Unisex long-sleeve AWC logo microfiber cycling ...	Multi	M	U	31.7244	48.0673
232	21	LJ-0192-L	"Long-Sleeve Logo Jersey, L"	Long-Sleeve Logo Jersey	Unisex long-sleeve AWC logo microfiber cycling ...	Multi	L	U	31.7244	48.0673
235	21	LJ-0192-X	"Long-Sleeve Logo Jersey, XL"	Long-Sleeve Logo Jersey	Unisex long-sleeve AWC logo microfiber cycling ...	Multi	XL	U	31.7244	48.0673
238	14	FR-R92R-62	"HL Road Frame - Red, 62"	HL Road Frame	Our lightest and best quality aluminum frame m...	Red	62	U	747.9682	1263.4598
241	14	FR-R92R-44	"HL Road Frame - Red, 44"	HL Road Frame	Our lightest and best quality aluminum frame m...	Red	44	U	747.9682	1263.4598
244	14	FR-R92R-48	"HL Road Frame - Red, 48"	HL Road Frame	Our lightest and best quality aluminum frame m...	Red	48	U	747.9682	1263.4598
247	14	FR-R92R-52	"HL Road Frame - Red, 52"	HL Road Frame	Our lightest and best quality aluminum frame m...	Red	52	U	747.9682	1263.4598
250	14	FR-R92R-56	"HL Road Frame - Red, 56"	HL Road Frame	Our lightest and best quality aluminum frame m...	Red	56	U	747.9682	1263.4598
253	14	FR-R38B-58	"LL Road Frame - Black, 58"	LL Road Frame	"The LL Frame provides a safe comfortable ride,...	Black	58	U	176.1997	297.6346
256	14	FR-R38B-60	"LL Road Frame - Black, 60"	LL Road Frame	"The LL Frame provides a safe comfortable ride,...	Black	60	U	176.1997	297.6346
259	14	FR-R38B-62	"LL Road Frame - Black, 62"	LL Road Frame	"The LL Frame provides a safe comfortable ride,...	Black	62	U	176.1997	297.6346

Figure 5: products with ProductKey in ascending order by

3.1.2 Find the 10 most expensive products in descending order.



```
/* 3.  Find the 10 most expensive products in
SELECT ProductName, ProductPrice FROM products
ORDER BY ProductPrice DESC LIMIT 10;
```

Result Grid | Filter Rows: | Export: | Wrap

ProductName	ProductPrice
"Road-150 Red, 48"	3578.2700
"Road-150 Red, 52"	3578.2700
"Road-150 Red, 62"	3578.2700
"Road-150 Red, 44"	3578.2700
"Road-150 Red, 56"	3578.2700
"Mountain-100 Silver, 48"	3399.9900
"Mountain-100 Silver, 38"	3399.9900
"Mountain-100 Silver, 44"	3399.9900
"Mountain-100 Silver, 42"	3399.9900
"Mountain-100 Black, 38"	3374.9900

Figure 6: most 10 expensive products

3.1.3 Find all the products profit and identify them by their names in ascending order

Query -

```
SELECT ProductName, ProductCost, ProductPrice,
ProductPrice-ProductCost AS Profit
FROM products ORDER BY profit DESC;
```

Output:

```
7    /* 2. Find all the products profit and identify th
8 *  SELECT ProductName, ProductCost, ProductPrice,
9    ProductPrice-ProductCost AS Profit
10   FROM products ORDER BY profit DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell

ProductName	ProductCost	ProductPrice	Profit
"Mountain-100 Silver, 38"	1912.1544	3399.9900	1487.8356
"Mountain-100 Silver, 42"	1912.1544	3399.9900	1487.8356
"Mountain-100 Silver, 44"	1912.1544	3399.9900	1487.8356
"Mountain-100 Silver, 48"	1912.1544	3399.9900	1487.8356
"Mountain-100 Black, 38"	1898.0944	3374.9900	1476.8956
"Mountain-100 Black, 42"	1898.0944	3374.9900	1476.8956
"Mountain-100 Black, 44"	1898.0944	3374.9900	1476.8956
"Mountain-100 Black, 48"	1898.0944	3374.9900	1476.8956
"Road-150 Red, 62"	2171.2942	3578.2700	1406.9758
"Road-150 Red, 44"	2171.2942	3578.2700	1406.9758
"Road-150 Red, 48"	2171.2942	3578.2700	1406.9758
"Road-150 Red, 52"	2171.2942	3578.2700	1406.9758
"Road-150 Red, 56"	2171.2942	3578.2700	1406.9758
"Mountain-200 Silver, 38"	1117.8559	2071.4196	953.5637
"Mountain-200 Silver, 42"	1117.8559	2071.4196	953.5637
"Mountain-200 Silver, 46"	1117.8559	2071.4196	953.5637
"Mountain-200 Black, 38"	1105.8100	2049.0982	943.2882
"Mountain-200 Black, 42"	1105.8100	2049.0982	943.2882
"Mountain-200 Black, 46"	1105.8100	2049.0982	943.2882
"Road-250 Red, 44"	1518.7864	2443.3500	924.5636
"Road-250 Red, 48"	1518.7864	2443.3500	924.5636
"Road-250 Red, 52"	1518.7864	2443.3500	924.5636
"Touring-1000 Yellow, 46"	1481.9379	2384.0700	902.1321
"Touring-1000 Yellow, 50"	1481.9379	2384.0700	902.1321

Figure 7: Total products and profit

3.1.4 Find the 10 cheapest products in ascending order

```
15
16      /* 4.Find the 10 cheapest products in ascendin
17  *   SELECT ProductName, ProductPrice FROM products
18      ORDER BY ProductPrice ASC LIMIT 10;
19
```

Result Grid | Filter Rows: | Export: | Wrap C

ProductName	ProductPrice
Patch Kit/8 Patches	2.2900
Road Tire Tube	3.9900
Water Bottle - 30 oz.	4.9900
Touring Tire Tube	4.9900
Mountain Tire Tube	4.9900
Bike Wash - Dissolver	7.9500
AWC Logo Cap	8.6442
"Racing Socks, M"	8.9900
Road Bottle Cage	8.9900
"Racing Socks, L"	8.9900

Figure 8: most 10 cheapest products

3.1.5 Find the average price from products and products greater than the average

```
21  *   SELECT  ProductName, ProductPrice FROM products
22      HAVING ProductPrice >  (SELECT
23       AVG(ProductPrice) FROM products)
24      ORDER BY ProductPrice ASC;
```

Result Grid | Filter Rows: | Export: | Wrap C

ProductName	ProductPrice
"Touring-3000 Blue, 44"	742.3500
"Touring-3000 Blue, 50"	742.3500
"Touring-3000 Blue, 62"	742.3500
"Touring-3000 Yellow, 44"	742.3500
"Touring-3000 Yellow, 50"	742.3500
"Touring-3000 Yellow, 54"	742.3500
"Touring-3000 Yellow, 58"	742.3500
"Touring-3000 Yellow, 62"	742.3500
"Touring-3000 Blue, 58"	742.3500
"Touring-3000 Blue, 54"	742.3500
"Mountain-400-W Silver, ...	769.4900

Figure 9: product price is greater than the average

3.1.6 List all products whose size is medium, red in color and the product cost less than 800

```
SELECT ProductKey, ProductName,
ProductSize,
ProductColor, ProductCost
FROM products
WHERE ProductSize > 20
AND ProductColor='red'
AND ProductCost < 800;
```

Filter Rows: Edit: Export/Impo

ctKey	ProductName	ProductSize	ProductColor	ProductCost
	"HL Road Frame - Red, 62"	62	Red	747.9682
	"HL Road Frame - Red, 44"	44	Red	747.9682
	"HL Road Frame - Red, 48"	48	Red	747.9682
	"HL Road Frame - Red, 52"	52	Red	747.9682
	"HL Road Frame - Red, 56"	56	Red	747.9682
	"LL Road Frame - Red, 44"	44	Red	181.4857
	"LL Road Frame - Red, 48"	48	Red	181.4857
	"LL Road Frame - Red, 52"	52	Red	181.4857
	"LL Road Frame - Red, 58"	58	Red	181.4857
	"LL Road Frame - Red, 60"	60	Red	181.4857
	"LL Road Frame - Red, 62"	62	Red	181.4857
	"ML Road Frame - Red, 44"	44	Red	352.1394
	"ML Road Frame - Red, 48"	48	Red	352.1394
	"ML Road Frame - Red, 52"	52	Red	352.1394
	"ML Road Frame - Red, 58"	58	Red	352.1394
	"ML Road Frame - Red, 60"	60	Red	352.1394
	"Road-650 Red, 58"	58	Red	413.1463
	"Road-650 Red, 60"	60	Red	413.1463

Figure 10: Red products with price less than 800

3.1.7 List all products based on subcategory names

```
35      /* 7. List all products based on subcategories.*/
36 •    SELECT ProductKey, ProductName, subcategoryName
37      FROM products
38      JOIN product_subcategories
39      ON products.ProductSubcategoryKey
40      =product_subcategories.ProductSubcategoryKey;
```

esult Grid | Filter Rows: | Export: | Wrap Cell C

ProductKey	ProductName	subcategoryName
344	"Mountain-100 Silver, 38"	Mountain Bikes
345	"Mountain-100 Silver, 42"	Mountain Bikes
346	"Mountain-100 Silver, 44"	Mountain Bikes
347	"Mountain-100 Silver, 48"	Mountain Bikes
348	"Mountain-100 Black, 38"	Mountain Bikes
349	"Mountain-100 Black, 42"	Mountain Bikes
350	"Mountain-100 Black, 44"	Mountain Bikes
351	"Mountain-100 Black, 48"	Mountain Bikes
352	"Mountain-200 Silver, 38"	Mountain Bikes
354	"Mountain-200 Silver, 42"	Mountain Bikes
356	"Mountain-200 Silver, 46"	Mountain Bikes
358	"Mountain-200 Black, 38"	Mountain Bikes
360	"Mountain-200 Black, 42"	Mountain Bikes
362	"Mountain-200 Black, 46"	Mountain Bikes
364	"Mountain-300 Black, 38"	Mountain Bikes
365	"Mountain-300 Black, 40"	Mountain Bikes
366	"Mountain-300 Black, 44"	Mountain Bikes
367	"Mountain-300 Black, 48"	Mountain Bikes
587	"Mountain-400-W Silver,...	Mountain Bikes
588	"Mountain-400-W Silver,...	Mountain Bikes
589	"Mountain-400-W Silver,...	Mountain Bikes
590	"Mountain-400-W Silver,...	Mountain Bikes

Figure 11: products and subcategory names

3.2 Customers queries

3.2.1 List all customers who owns house by gender by descending order of Annual Income

```
43      /* 8.  List all customers who owns house by g
44  •   SELECT gender, FirstName, LastName,
45      AnnualIncome, HomeOwner
46      FROM customers
47      WHERE HomeOwner = 'Y'
48      ORDER BY AnnualIncome DESC;
49
```

esult Grid | Filter Rows: | Export: | Wra

gender	FirstName	LastName	AnnualIncome	HomeOwner
M	TERRY	NARA	170000	Y
M	LUIS	ROBERTS	170000	Y
M	NATHAN	WRIGHT	170000	Y
M	DEVIN	JOHNSON	170000	Y
M	ALEX	BROOKS	170000	Y
M	DAMIEN	CHANDER	170000	Y
M	JAMES	CARTER	170000	Y
M	VINCENT	SUN	170000	Y
M	NELSON	ORTEGA	170000	Y
M	CAMERON	ANDERSON	170000	Y
M	DEVIN	MARTIN	170000	Y
M	WESLEY	LIANG	170000	Y
M	HECTOR	ALONSO	170000	Y
M	SHANNON	NAVARRO	170000	Y
F	SHANNON	LIU	170000	Y
M	LEONARD	GOEL	170000	Y
M	HUNTER	GRIFFIN	170000	Y
F	VANESSA	PATTERSON	170000	Y
F	JOY	GOMEZ	170000	Y
M	DUSTIN	DENG	170000	Y
F	LORI	DOMINGUEZ	170000	Y

stomers 24 ×

Figure 12: customers name with high income

3.2.2 Find married customers that own a house and their occupation by ascending order of birth date

```
51 •  SELECT FirstName,BirthDate, MaritalStatus,
52    EducationLevel, Occupation
53    FROM customers
54    WHERE MaritalStatus= 'm'
55    AND HomeOwner ='Y'
56    ORDER BY BirthDate ASC;
57
```

esult Grid | Filter Rows: | Export: | Wrap Cell Content

FirstName	BirthDate	MaritalStatus	EducationLevel	Occupation
FRANKLIN	1911-12-07	M	Partial High School	Manual
DALE	1912-08-14	M	Partial High School	Manual
BRIANNA	1916-10-11	M	Bachelors	Skilled Manual
HALEY	1916-12-07	M	High School	Manual
MARIAH	1917-10-15	M	Partial High School	Manual
SHARON	1920-07-13	M	Bachelors	Clerical
DAKOTA	1922-02-04	M	Bachelors	Clerical
CAROLINE	1922-02-13	M	Graduate Degree	Clerical
DAVID	1922-06-03	M	Bachelors	Professional
LUIS	1922-09-20	M	Partial College	Manual
CHLOE	1923-06-10	M	Partial High School	Skilled Manual
MATTHEW	1923-08-03	M	Graduate Degree	Clerical
GRACE	1924-04-23	M	Bachelors	Skilled Manual
ELIZABETH	1925-03-09	M	Bachelors	Skilled Manual
MIGUEL	1926-02-13	M	Bachelors	Clerical
ALEXAND...	1926-07-08	M	Bachelors	Clerical
CHASE	1926-07-28	M	Bachelors	Clerical
JAMES	1927-03-09	M	Partial College	Skilled Manual
JESSE	1927-07-27	M	High School	Manual
ROBERT	1927-08-06	M	Bachelors	Clerical

Figure 13: Married customers by ascending birthdate

3.2.3 Find customers that are single and whose annual income is greater than 50,000 in ascending order

```
59 •  SELECT FirstName, LastName,
60    MaritalStatus, AnnualIncome
61    FROM customers
62    WHERE MaritalStatus='s' AND
63    AnnualIncome > 50000
64    ORDER BY AnnualIncome ASC;
65
```

Result Grid | Filter Rows: | Export:

FirstName	LastName	MaritalStatus	AnnualIncome
EUGENE	HUANG	S	60000
JACK	ADAMS	S	60000
ROBERT	ZHANG	S	60000
CARLOS	RICHARDSON	S	60000
ROBIN	VERHOFF	S	60000
JONATHAN	FOSTER	S	60000
KATHERINE	GRAY	S	60000
MORGAN	SCOTT	S	60000
MIGUEL	PARKER	S	60000
CALEB	CARTER	S	60000
ISABELLA	BAILEY	S	60000
AUSTIN	RODRIGUEZ	S	60000
DEVIN	WHITE	S	60000
ARIANNA	HUGHES	S	60000
DANIEL	RODRIGUEZ	S	60000
ABIGAIL	WALKER	S	60000

Figure 14: customers with Annual income greater than 50,000

3.2.4 Among the female customers who are married, find the ones that have houses and their annual income is greater than average income.

```
57 •  SELECT CustomerKey, FirstName,
58    LastName, MaritalStatus,gender, HomeOwner, AnnualIncome
59    FROM customers
70    WHERE MaritalStatus = 'M'
71    AND gender = 'F'
72    AND HomeOwner='Y'
73    AND AnnualIncome > (select avg(AnnualIncome)from customers)
74    ORDER BY AnnualIncome;
```

esult Grid | Filter Rows: | Edit: | Export/Import:

CustomerKey	FirstName	LastName	MaritalStatus	gender	HomeOwner	AnnualIncome
18675	ABIGAIL	MARTIN	M	F	Y	60000
11036	JENNIFER	RUSSELL	M	F	Y	60000
11041	AMANDA	CARTER	M	F	Y	60000
16798	LATASHA	MARTIN	M	F	Y	60000
16799	STEPHANIE	GONZALES	M	F	Y	60000
16807	VIRGINIA	SURI	M	F	Y	60000
16812	ALEXA	RICHARDSON	M	F	Y	60000
18684	MAKAYLA	WATSON	M	F	Y	60000
29207	KATELYN	JAMES	M	F	Y	60000
18687	OLIVIA	SMITH	M	F	Y	60000
11111	MEREDITH	GUTIERREZ	M	F	Y	60000
11112	CRYSTAL	WANG	M	F	Y	60000
29254	ERIN	SANCHEZ	M	F	Y	60000
19537	MICHELLE	BROOKS	M	F	Y	60000
29287	SUZANNE	LIN	M	F	Y	60000
29299	BETHANY	KUMAR	M	F	Y	60000
29314	MEREDITH	MALHOTRA	M	F	Y	60000
11187	JENNIFER	COOPER	M	F	Y	60000
11194	JACQUELINE	PRICE	M	F	Y	60000
11195	MEGAN	HENDERSON	M	F	Y	60000
26577	CINDY	EDWARDS	M	F	Y	60000

Figure 15: Married females with annual income greater than average income

3.2.5 List all the customers that their annual income is less than 20,000 and bought items in 2015

```
77 *  SELECT  FirstName, LastName,
78    AnnualIncome, ProductName,
79    YEAR(OrderDate) AS Year
80    FROM sales_2015
81    JOIN products ON sales_2015.ProductKey = products.ProductKey
82    JOIN customers ON sales_2015.CustomerKey = customers.CustomerKey
83    HAVING AnnualIncome < 20000;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

FirstName	LastName	AnnualIncome	ProductName	Year
SABRINA	BLANCO	10000	"Road-150 Red, 56"	2015
GEOFFREY	RODRIGUEZ	10000	"Road-150 Red, 52"	2015
SHAWN	LUO	10000	"Mountain-100 Black, 48"	2015
ANDREA	COLLINS	10000	"Mountain-100 Silver, 38"	2015
SABRINA	RUBIO	10000	"Road-150 Red, 48"	2015
DEVIN	SANDERS	10000	"Mountain-100 Silver, 44"	2015
PRESTON	PRASAD	10000	"Road-150 Red, 62"	2015
DEVIN	TAYLOR	10000	"Mountain-100 Silver, 38"	2015
FRANCISCO	GONZALEZ	10000	"Mountain-100 Silver, 44"	2015
CASSANDRA	SARA	10000	"Road-150 Red, 44"	2015
MEREDITH	VANCE	10000	"Road-150 Red, 52"	2015
GILBERT	LIANG	10000	"Mountain-100 Black, 42"	2015
KRISTINE	SANDBERG	10000	"Mountain-100 Silver, 48"	2015
MARGARET	LU	10000	"Road-150 Red, 56"	2015
TRACY	ANDERSEN	10000	"Road-150 Red, 62"	2015
MAYRA	RAMAN	10000	"Road-150 Red, 52"	2015
ARIANNA	WARD	10000	"Road-150 Red, 52"	2015
KRISTY	RUBIO	10000	"Mountain-100 Black, 38"	2015
RACHEL	POWELL	10000	"Road-150 Red, 56"	2015
CLAYTON	SHARMA	10000	"Mountain-100 Silver, 42"	2015
JENNIFER	GREEN	10000	"Mountain-200 Silver, 42"	2015

Figure 16: customers with low income and bought products in 2015

3.3 Sales and returns in 2015,2016 and 2017

3.3.1 List all sales from 2015 in ascending order by order Number, product key and customer Key and in day/month/year format

```
87 •  SELECT OrderNumber,products.ProductKey,
88    customers.CustomerKey,DAY(OrderDate) AS Day,
89    MONTH(OrderDate) AS MONTH, YEAR(OrderDate) AS Year,
90    sales_2015.OrderQuantity * products.ProductPrice AS Sales
91    from sales_2015 join products
92    ON sales_2015.ProductKey = products.ProductKey
93    JOIN customers ON
94    sales_2015.CustomerKey = customers.CustomerKey
95    ORDER BY Sales ASC;
```

esult Grid | Filter Rows: | Export: | Wrap Cell Content:

OrderNumber	ProductKey	CustomerKey	Day	MONTH	Year	Sales
SO2015-1	332	14657	1	1	2015	699.0982
SO2015-676	324	25920	20	4	2015	699.0982
SO2015-1665	342	15287	28	8	2015	699.0982
SO2015-671	334	14757	20	4	2015	699.0982
SO2015-67	342	14656	13	1	2015	699.0982
SO2015-665	326	19439	19	4	2015	699.0982
SO2015-659	328	18148	18	4	2015	699.0982
SO2015-626	332	19450	14	4	2015	699.0982
SO2015-614	330	14880	12	4	2015	699.0982
SO2015-612	338	19443	12	4	2015	699.0982
SO2015-600	326	25919	10	4	2015	699.0982
SO2015-2116	328	15398	1	11	2015	699.0982
SO2015-1676	332	20807	30	8	2015	699.0982
SO2015-584	338	14754	7	4	2015	699.0982
SO2015-581	340	18143	7	4	2015	699.0982
SO2015-2252	330	15396	24	11	2015	699.0982

Figure 17: sales from 2015 in ascending order

3.3.2 List all sales from 2016 in descending

```
98 *  SELECT OrderNumber,products.ProductKey,
99    customers.CustomerKey,DAY(OrderDate) AS Day,
00    MONTH(OrderDate) AS MONTH, YEAR(OrderDate) AS Year,
01    sales_2016.OrderQuantity * products.ProductPrice AS Sales
02    from sales_2016 join products
03    ON sales_2016.ProductKey = products.ProductKey
04    JOIN customers ON
05    sales_2016.CustomerKey = customers.CustomerKey
06    ORDER BY Sales DESC;
```

esult Grid | Filter Rows: | Export: | Wrap Cell Content:

OrderNumber	ProductKey	CustomerKey	Day	MONTH	Year	Sales
SO2016-889	369	24171	12	4	2016	2443.3500
SO2016-902	369	17833	13	4	2016	2443.3500
SO2016-911	369	13896	14	4	2016	2443.3500
SO2016-917	368	24162	15	4	2016	2443.3500
SO2016-922	369	18000	15	4	2016	2443.3500
SO2016-930	370	16404	16	4	2016	2443.3500
SO2016-939	369	16412	17	4	2016	2443.3500

Figure 18: sales from 2016 in descending order

3.3.3 List all sales from 2017 in descending

```
9 *   SELECT OrderNumber,products.ProductKey,
0     customers.CustomerKey,DAY(OrderDate) AS Day,
1     MONTH(OrderDate) AS MONTH, YEAR(OrderDate) AS Year,
2     sales_2017.OrderQuantity * products.ProductPrice AS Sales
3     from sales_2017 join products
4     ON sales_2017.ProductKey = products.ProductKey
5     JOIN customers ON
6     sales_2017.CustomerKey = customers.CustomerKey
7     ORDER BY Sales DESC;
```

sult Grid | Filter Rows: | Export: | Wrap Cell Content:

OrderNumber	ProductKey	CustomerKey	Day	MONTH	Year	Sales
SO2017-9457	575	11237	4	3	2017	2384.0700
SO2017-945	563	24683	7	1	2017	2384.0700
SO2017-9433	573	26694	4	3	2017	2384.0700
SO2017-7714	563	11917	20	2	2017	2384.0700
SO2017-8905	564	26511	1	3	2017	2384.0700
SO2017-943	573	27866	7	1	2017	2384.0700
SO2017-7718	573	26247	20	2	2017	2384.0700

Figure 19: sales from 2017 in descending order

3.3.4 List all the customers that purchased the more than one product in the year that has higher sales 2017

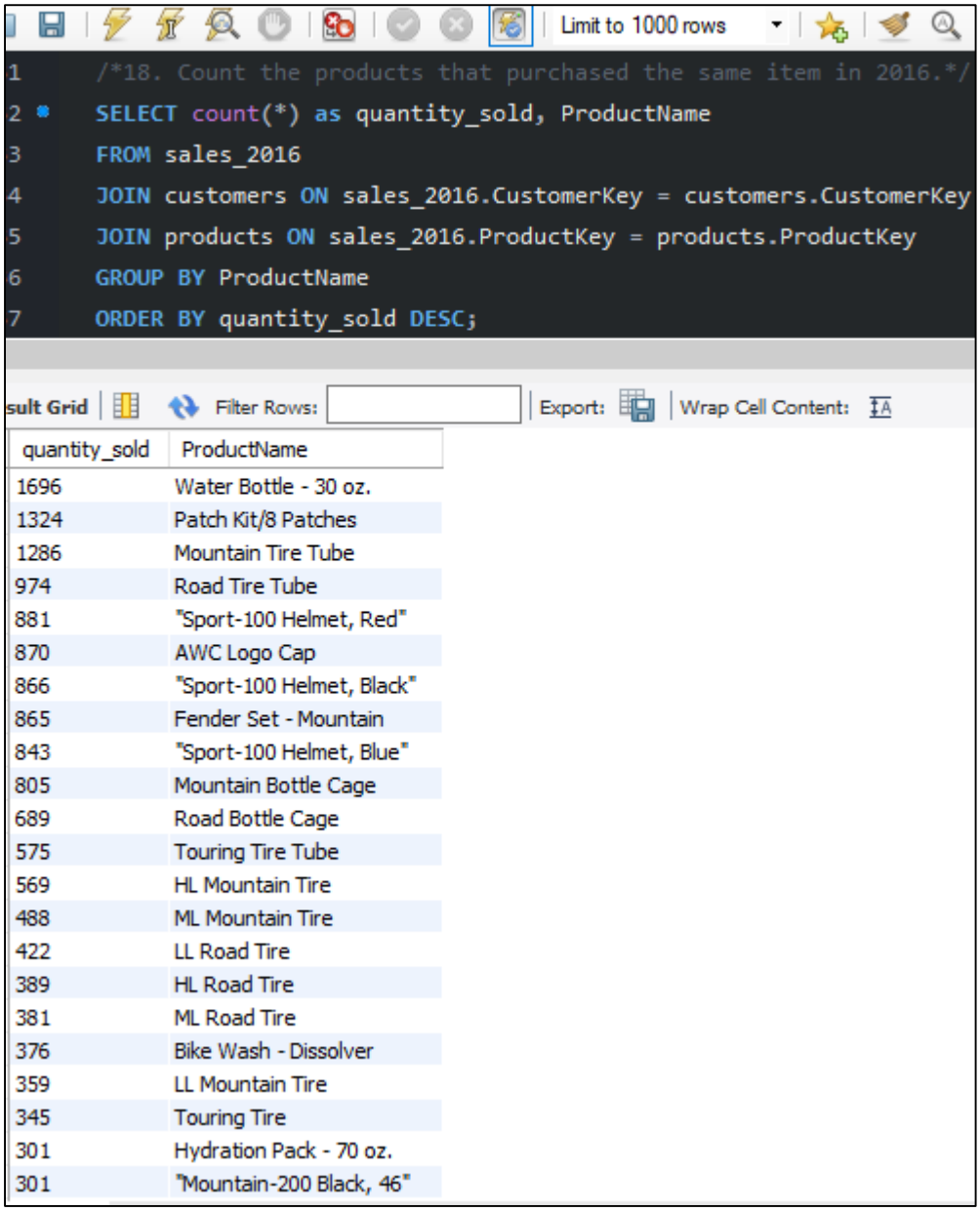
```
SELECT customers.CustomerKey,
FirstName, LastName,
ProductName, OrderQuantity, OrderDate
FROM sales_2017
JOIN customers ON
sales_2017.CustomerKey = customers.CustomerKey
JOIN products ON
sales_2017.ProductKey = products.ProductKey
WHERE OrderQuantity > (SELECT AVG(OrderQuantity)
FROM sales_2017);
```

ult Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch r

CustomerKey	FirstName	LastName	ProductName	OrderQuantity	OrderDate
23791	AARON	GRIFFIN	Road Tire Tube	2	2017-01-01
11530	ANDREW	MARTINEZ	ML Mountain Tire	2	2017-01-01
26502	TAYLOR	BROWN	Patch Kit/8 Patches	2	2017-03-07
26502	TAYLOR	BROWN	LL Mountain Tire	2	2017-03-07
13199	TIMOTHY	STEWART	Patch Kit/8 Patches	2	2017-03-07
13199	TIMOTHY	STEWART	Mountain Tire Tube	2	2017-03-07
16163	WILLIE	CHANDE	Mountain Tire Tube	3	2017-03-07
16163	WILLIE	CHANDE	LL Mountain Tire	2	2017-03-07
21218	WYATT	RODRIGUEZ	Road Tire Tube	2	2017-03-07
14611	ZOE	JAMES	Mountain Bottle Cage	3	2017-03-07
14611	ZOE	JAMES	Water Bottle - 30 oz.	2	2017-03-07
23450	ALEX	ROGERS	"Half-Finger Gloves, M"	3	2017-03-08
23450	ALEX	ROGERS	Water Bottle - 30 oz.	2	2017-03-08
14537	CLAYTON	CAI	Patch Kit/8 Patches	2	2017-01-08
23450	ALEX	ROGERS	Road Bottle Cage	2	2017-03-08
21553	ALLEN	SURI	AWC Logo Cap	2	2017-03-08

Figure 20: customers with more than one order quantity in 2017

3.3.5 Count the products that purchased the same item in 2016.



```
/*18. Count the products that purchased the same item in 2016.*/
SELECT count(*) as quantity_sold, ProductName
FROM sales_2016
JOIN customers ON sales_2016.CustomerKey = customers.CustomerKey
JOIN products ON sales_2016.ProductKey = products.ProductKey
GROUP BY ProductName
ORDER BY quantity_sold DESC;
```

quantity_sold	ProductName
1696	Water Bottle - 30 oz.
1324	Patch Kit/8 Patches
1286	Mountain Tire Tube
974	Road Tire Tube
881	"Sport-100 Helmet, Red"
870	AWC Logo Cap
866	"Sport-100 Helmet, Black"
865	Fender Set - Mountain
843	"Sport-100 Helmet, Blue"
805	Mountain Bottle Cage
689	Road Bottle Cage
575	Touring Tire Tube
569	HL Mountain Tire
488	ML Mountain Tire
422	LL Road Tire
389	HL Road Tire
381	ML Road Tire
376	Bike Wash - Dissolver
359	LL Mountain Tire
345	Touring Tire
301	Hydration Pack - 70 oz.
301	"Mountain-200 Black, 46"

Figure 21: counts of products purchased in 2016

3.3.6 List all products that have been returned based on continent, country and region and order by the return date

```
SELECT products.ProductKey,
ProductName,ReturnDate,
Continent, Country, Region
FROM returns
JOIN products ON
returns.ProductKey = products.ProductKey
JOIN territories ON
returns.TerritoryKey = territories.TerritoryKey
ORDER BY ReturnDate;
```

sult Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

ProductKey	ProductName	ReturnDate	Continent	Country	Region
312	"Road-150 Red, 48"	2015-01-18	Pacific	Australia	Australia
310	"Road-150 Red, 62"	2015-01-18	Europe	United Kingdom	United Kingdom
346	"Mountain-100 Silver, 44"	2015-01-21	Europe	Germany	Germany
311	"Road-150 Red, 44"	2015-01-22	North America	United States	Southwest
312	"Road-150 Red, 48"	2015-02-02	North America	Canada	Canada
312	"Road-150 Red, 48"	2015-02-15	North America	United States	Northwest
311	"Road-150 Red, 44"	2015-02-19	Pacific	Australia	Australia
314	"Road-150 Red, 56"	2015-02-24	Europe	Germany	Germany
350	"Mountain-100 Black, 44"	2015-03-08	Europe	Germany	Germany
350	"Mountain-100 Black, 44"	2015-03-13	Pacific	Australia	Australia
346	"Mountain-100 Silver, 44"	2015-03-14	North America	United States	Southwest
340	"Road-650 Black, 48"	2015-03-15	Pacific	Australia	Australia
311	"Road-150 Red, 44"	2015-03-22	North America	United States	Southwest
312	"Road-150 Red, 48"	2015-03-26	Europe	United Kingdom	United Kingdom
312	"Road-150 Red, 48"	2015-03-28	Europe	France	France
314	"Road-150 Red, 56"	2015-03-28	Pacific	Australia	Australia
311	"Road-150 Red, 44"	2015-03-29	Pacific	Australia	Australia
311	"Road-150 Red, 44"	2015-04-01	Europe	Germany	Germany
311	"Road-150 Red, 44"	2015-04-07	Pacific	Australia	Australia
351	"Mountain-100 Black, 48"	2015-04-07	Pacific	Australia	Australia

ult 55 ×

Figure 22: returned products by date and territory

3.3.7 Count the returned products group by region

```
SELECT count(*) AS Total_Return, Region
FROM returns
JOIN territories ON
returns.TerritoryKey = territories.TerritoryKey
GROUP BY region;
```

Result Grid | Filter Rows: | Export: | Wrap

Total_Return	Region
269	Northwest
354	Southwest
1	Southeast
234	Canada
186	France
161	Germany
400	Australia
204	United Kingdom

Figure 23: Total returns in each region

3.3.8 Find out the profit of the top 5 products for 2017

```
/*21. Find out the profit of the top 5 products for 2017.*/
SELECT products.ProductKey, ProductName,ProductCost,
ProductPrice, ProductPrice - ProductCost AS Profit, OrderDate
FROM sales_2017
JOIN products ON sales_2017.ProductKey = products.ProductKey
LIMIT 5;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch

ProductKey	ProductName	ProductCost	ProductPrice	Profit	OrderDate
529	Road Tire Tube	1.4923	3.9900	2.4977	2017-01-01
536	ML Mountain Tire	11.2163	29.9900	18.7737	2017-01-01
479	Road Bottle Cage	3.3623	8.9900	5.6277	2017-01-02
215	"Sport-100 Helmet, Black"	12.0278	33.6442	21.6164	2017-01-08
480	Patch Kit/8 Patches	0.8565	2.2900	1.4335	2017-03-07

Figure 24: top 5 products with profit

3.3.9 Find the average returns in each year

```
224     /* 21. Find the average returns in each year. */
225 *   SELECT '2017' AS Year, AVG(ReturnQuantity) AS Average_returns FROM returns
226     WHERE ReturnDate BETWEEN '2017-01-01' AND '2017-12-31'
227     UNION ALL
228     SELECT '2016' AS Year, AVG(ReturnQuantity) AS Average_returns FROM returns
229     WHERE ReturnDate BETWEEN '2016-01-01' AND '2016-12-31'
230     UNION ALL
231     SELECT '2015' AS Year, AVG(ReturnQuantity) AS Average_returns FROM returns
232     WHERE ReturnDate BETWEEN '2015-01-01' AND '2015-12-31';
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Year	Average_returns
2017	1.0125
2016	1.0079
2015	1.0118

Figure 25: average returns for each year

3.3.10 Find total quantities ordered in each year and at all times within each region

```
69      FROM cte2015
70      JOIN cte2016
71      ON cte2015.Region = cte2016.Region and cte2015.TerritoryKey = cte2016.TerritoryKey
72      JOIN cte2017
73      ON cte2016.Region = cte2017.Region and cte2016.TerritoryKey = cte2017.TerritoryKey
74      JOIN cte_all_times
75      ON cte2017.Region = cte_all_times.Region and cte2017.TerritoryKey = cte_all_times.TerritoryKey
76      GROUP BY cte2015.Region, cte2015.TerritoryKey
77      ORDER BY total_quantities2015 DESC;
```

esult Grid | Filter Rows: | Export: | Wrap Cell Content:

Region	TerritoryKey	total_quantities2015	total_quantities2016	total_quantities2017	total_quantities_all_times
Australia	9	848	7616	9487	17951
Southwest	4	514	7287	9390	17191
Northwest	1	334	5390	6789	12513
United Kingdom	10	255	4346	5093	9694
France	7	229	3397	4236	7862
Germany	8	226	3338	4386	7950
Canada	6	223	4811	5860	10894
Southeast	5	1	13	35	49

Figure 26: Total quantities in each year and overall in each region