CIS 467 final group project (due by Wednesday, December 15 at 11:59 PM).
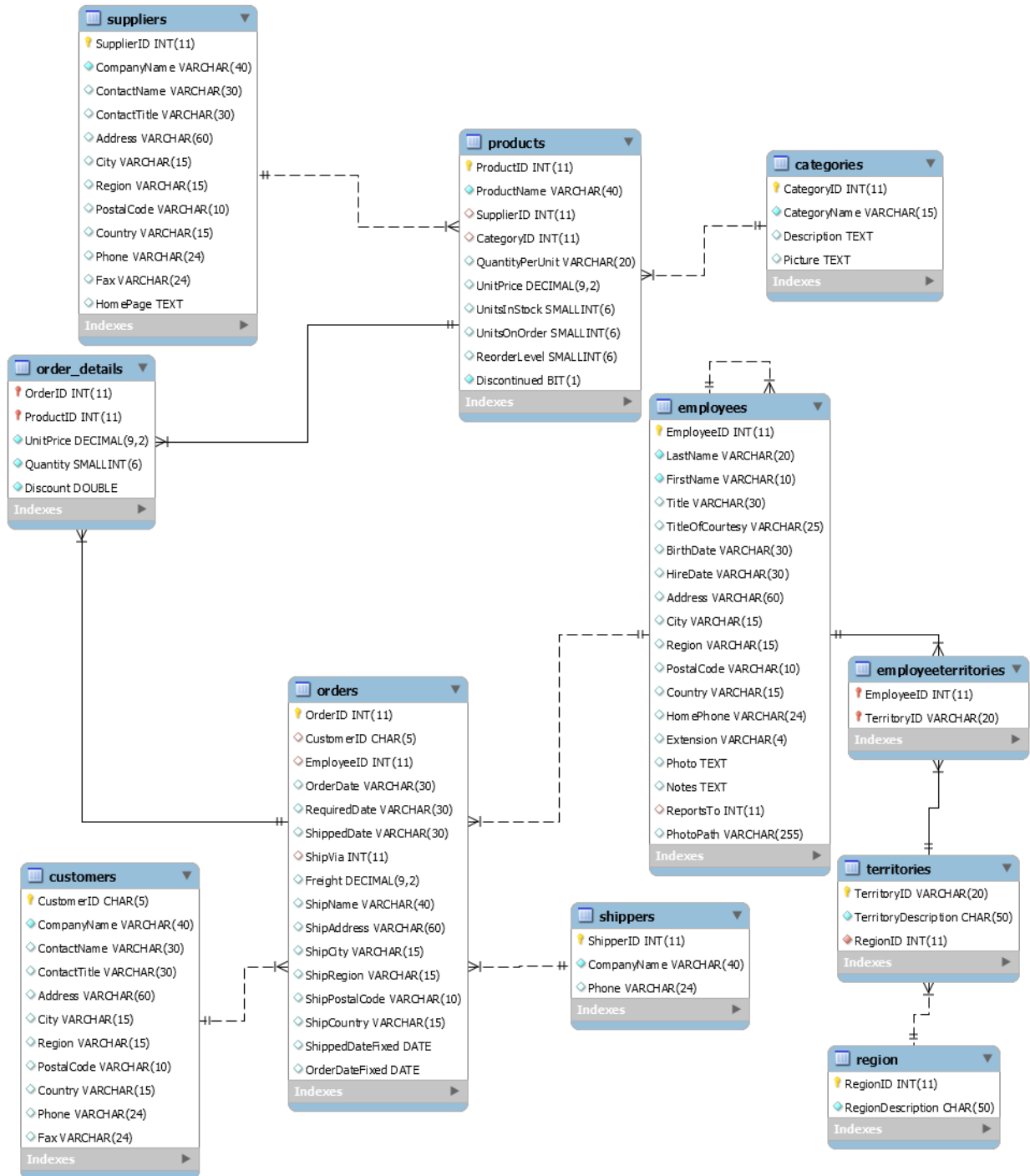
This is a group project (total 400 points). The groups have been created on Blackboard. Please make only one submission per group and put all your team members' full names into this Word document. Please also submit a Tableau Workbook file .twb into the Final Team Project folder on Blackboard together with the Excel file of your Data Warehouse which you uploaded to Tableau and used for visualizations. If you cannot attach the Tableau Workbook .twb file and the Excel file for your Data Warehouse to the Final Team Project folder on Blackboard, please email (one email per group) the Tableau Workbook .twb file and the Excel file for your Data Warehouse to me at mlysyako@simon.rochester.edu indicating your class section and your team name from Blackboard and all members in the email.

YOU WILL HAVE 5 ATTEMPTS. Please submit once if possible, and use other attempts if the submitted files in the first attempt were wrong or corrupted or if you find major mistakes in the first attempt. The last attempt will be graded (as long as it is before the deadline).

The script file final_project_database_cis467.sql creates a database which contains the 11 tables (the database schema is below), with transactional data related to some company operations.

Please check early that you can create the final project database on your machine.

Please review Panopto video for week 7 for a description of this dataset.

suppliers
- SupplierID INT(11)
- CompanyName VARCHAR(40)
- ContactName VARCHAR(30)
- ContactTitle VARCHAR(30)
- Address VARCHAR(60)
- City VARCHAR(15)
- Region VARCHAR(15)
- PostalCode VARCHAR(10)
- Country VARCHAR(15)
- Phone VARCHAR(24)
- Fax VARCHAR(24)
- HomePage TEXT
- Indexes

products
- ProductID INT(11)
- ProductName VARCHAR(40)
- SupplierID INT(11)
- CategoryID INT(11)
- QuantityPerUnit VARCHAR(20)
- UnitPrice DECIMAL(9,2)
- UnitsInStock SMALLINT(6)
- UnitsOnOrder SMALLINT(6)
- ReorderLevel SMALLINT(6)
- Discontinued BIT(1)
- Indexes

categories
- CategoryID INT(11)
- CategoryName VARCHAR(15)
- Description TEXT
- Picture TEXT
- Indexes

order_details
- OrderID INT(11)
- ProductID INT(11)
- UnitPrice DECIMAL(9,2)
- Quantity SMALLINT(6)
- Discount DOUBLE
- Indexes

employees
- EmployeeID INT(11)
- LastName VARCHAR(20)
- FirstName VARCHAR(10)
- Title VARCHAR(30)
- TitleOfCourtesy VARCHAR(25)
- BirthDate VARCHAR(30)
- HireDate VARCHAR(30)
- Address VARCHAR(60)
- City VARCHAR(15)
- Region VARCHAR(15)
- PostalCode VARCHAR(10)
- Country VARCHAR(15)
- HomePhone VARCHAR(24)
- Extension VARCHAR(4)
- Photo TEXT
- Notes TEXT
- ReportsTo INT(11)
- PhotoPath VARCHAR(255)
- Indexes

employeeterritories
- EmployeeID INT(11)
- TerritoryID VARCHAR(20)
- Indexes

orders
- OrderID INT(11)
- CustomerID CHAR(5)
- EmployeeID INT(11)
- OrderDate VARCHAR(30)
- RequiredDate VARCHAR(30)
- ShippedDate VARCHAR(30)
- ShipVia INT(11)
- Freight DECIMAL(9,2)
- ShipName VARCHAR(40)
- ShipAddress VARCHAR(60)
- ShipCity VARCHAR(15)
- ShipRegion VARCHAR(15)
- ShipPostalCode VARCHAR(10)
- ShipCountry VARCHAR(15)
- ShippedDateFixed DATE
- OrderDateFixed DATE
- Indexes

territories
- TerritoryID VARCHAR(20)
- TerritoryDescription CHAR(50)
- RegionID INT(11)
- Indexes

customers
- CustomerID CHAR(5)
- CompanyName VARCHAR(40)
- ContactName VARCHAR(30)
- ContactTitle VARCHAR(30)
- Address VARCHAR(60)
- City VARCHAR(15)
- Region VARCHAR(15)
- PostalCode VARCHAR(10)
- Country VARCHAR(15)
- Phone VARCHAR(24)
- Fax VARCHAR(24)
- Indexes

shippers
- ShipperID INT(11)
- CompanyName VARCHAR(40)
- Phone VARCHAR(24)
- Indexes

region
- RegionID INT(11)
- RegionDescription CHAR(50)
- Indexes

Please put all of your work into **this single Word doc and also submit a Tableau Workbook file .twb and the Excel file of your Data Warehouse that you used for Tableau visualizations**. Please see instructions for Tableau below in the question 3.

1. **(160 points)** Design and create a data warehouse for the provided database. The decisions about which fields to include and how to aggregate the data are left to you. You do not need to include every single data point from the 11 tables given. Use your judgement as to what will be interesting/useful for the organization. But please make sure that you pull (combine) data from **at least six tables** and compute relevant aggregate statistics. Please compute relevant aggregate statistics for each table that you join. In your queries later in part 2, you may join your Data Warehouse with other tables to answer useful questions. Please see many examples from class lectures and you may adapt those codes for your purpose (for this dataset).

Description of our Datawarehouse:
Through this warehouse, we are gathering information from product, suppliers, categories, order, order details, employees and customers tables from the schema.
Using the aggregate functions and JOIN ON function, we calulated extended price, total product sales, stock-order ratio(to reflect the inventory status) as crucial metrics to measure products and orders. By counting the order number, employee number, and shipper number we can easily return the quantities of orders, employees, and shippers for further analysis. We also included the shipping day and delayed day of the orders, which can help the company to inspect their parternership with shipping companies to increase shipping efficiency and customer satisfaction. We also calculated orders number per supplier and per customer, which can be used to identify our primary customers and suppliers.

```
#create data warehouse#
CREATE TABLE DW_cis467 AS
SELECT
Products.ProductID, Products.ProductName, Orders.OrderID,
categories.CategoryName, Products.Discontinued,
Order_Details.UnitPrice, Order_Details.Quantity, Order_Details.Discount,
Orders.CustomerID, Customers.CompanyName AS CustomerName, Orders.ShipCountry,
STR_TO_DATE(Orders.ShippedDate,"%m/%d/%Y") AS ShippedDate,
STR_TO_DATE(Orders.OrderDate,"%m/%d/%Y") AS OrderDate,
STR_TO_DATE(Orders.RequiredDate,"%m/%d/%Y") AS RequiredDate,
suppliers.CompanyName AS SupplierName,
suppliers.Country AS SupplierCountry,
CONCAT(Employees.FirstName, ' ', Employees.LastName) AS Salesperson,
Employees.Title,
s.OrdersCount AS OrderPerSupplier,
c.OrdersCount AS OrderPerCustomer,
d.ProductSales AS SalesPerCategory,
shippers.CompanyName AS ShipperName,

-- compute with aggregate function --
ROUND((Order_Details.UnitPrice*Quantity*(1-Discount)/100)*100,2) AS
ExtendedPrice,
ROUND(Sum((Order_Details.UnitPrice*Quantity*(1-Discount)/100)*100),2) AS
ProductSales,
ROUND(Products.UnitsOnOrder/Products.UnitsInstock,2) AS OrderToStock_Ratio,
```

```sql
    DATEDIFF(STR_TO_DATE(Orders.ShippedDate,"%m/%d/%Y"),
    STR_TO_DATE(Orders.OrderDate,"%m/%d/%Y")) AS ShipDay,
    DATEDIFF(STR_TO_DATE(Orders.ShippedDate,"%m/%d/%Y"),
    STR_TO_DATE(Orders.RequiredDate,"%m/%d/%Y")) AS DelayedDay,
    COUNT(Orders.OrderID) AS OrderNumber,
    COUNT(Employees.EmployeeID) AS EmployeeNumber,
    COUNT(shippers.ShipperID) AS ShipperNumber

-- join tables:
categories/products/employees/customer/order/order.details/suppliers --
FROM    suppliers JOIN
        (categories JOIN
                (Products  JOIN
                    (
                            (Employees  JOIN
                                (Customers  JOIN (shippers JOIN Orders
ON shippers.ShipperID = orders.ShipVia)
                    ON Customers.CustomerID = Orders.CustomerID)
                            ON Employees.EmployeeID = Orders.EmployeeID)
                    JOIN Order_Details ON Orders.OrderID =
Order_Details.OrderID)
                ON Products.ProductID = Order_Details.ProductID)
        ON categories.CategoryID = Products.CategoryID)
        ON suppliers.SupplierID = Products.SupplierID

-- use subqueries to caculate orders' number per supplier --
    JOIN (SELECT COUNT(Order_details.OrderID) AS OrdersCount,
suppliers.CompanyName
    FROM Suppliers JOIN (Order_details JOIN Products ON Products.ProductID =
Order_details.ProductID)
    ON  Suppliers.SupplierID = Products.SupplierID
    GROUP BY Suppliers.CompanyName) S
    ON S.CompanyName = suppliers.CompanyName

-- use subqueries to caculate orders' number per customer --
    JOIN (SELECT COUNT(Orders.OrderID) AS OrdersCount, customers.CompanyName
AS CustomerName
    FROM customers JOIN Orders ON customers.CustomerID = Orders.CustomerID
    GROUP BY CompanyName) C
    ON C.CustomerName = Customers.CompanyName

-- use subqueries to return values under each categories --
    JOIN (SELECT ROUND(Sum((Order_Details.UnitPrice*Quantity*(1-
Discount)/100)*100),2) AS ProductSales, CategoryName
    FROM Order_details JOIN (products JOIN categories ON products.CategoryID
= categories.CategoryID) ON Order_details.ProductID = products.ProductID
    GROUP BY CategoryName) D
    ON D.CategoryName = categories.CategoryName

GROUP BY ProductName, OrderID
ORDER BY ProductID;
```

2. **(140 points)** Create **eight** SQL queries **on your data warehouse** (not on the original dataset) that answer interesting questions. At least **6** queries should be more complex queries. For example, more complex queries could include Joins, a Group By, UNION elements or a subquery or use some aggregate functions and summary calculations (see examples in the class lectures' slides).

```
# Section 1 Queries for logistic study
-- Descrption for query 1: To know which shippers'shipment cycle is longer
than average and the average time they are taking.
-- United Package has the longest average shipping days, which is 8.9938.
SELECT ShipperName, AVG(ShipDay) AS Average_ShipDay, ShipCountry
FROM dw_cis467
GROUP BY ShipperName, ShipCountry
HAVING AVG(ShipDay) > (SELECT AVG(ShipDay) FROM dw_cis467)
ORDER BY ShipDay DESC;
```

ShipperName	Average_ShipDay
United Package	8.9938

```sql
-- Descrption for query 2: To find the orders that were delivered after the
required date and how many days were they delayed
-- Then determine which shippers are responsible for the delayed orders,
which shippers have the most delayed orders, and the highest average delayed
days.
-- United Package has the highest Delayed Days of 23 days.
SELECT OrderID, ShipperName, DelayedDay
FROM dw_cis467
WHERE DelayedDay > 0
GROUP BY OrderID
ORDER BY DelayedDay DESC;
```

OrderID	ShipperName	DelayedDay
▶ 10777	United Package	23
10726	Speedy Express	18
10423	Federal Shipping	18
10970	Speedy Express	17
10515	Speedy Express	16
10827	United Package	11
10663	United Package	9
10660	Speedy Express	9
10828	Speedy Express	8
10593	United Package	7
10924	United Package	7
10545	United Package	7
10427	United Package	7
10380	Federal Shipping	7
10451	Federal Shipping	7
10309	Speedy Express	6
10709	Federal Shipping	6
10705	United Package	6
10960	Speedy Express	6
10927	Speedy Express	6
10847	Federal Shipping	5
10727	Speedy Express	4
10483	United Package	4
10596	Speedy Express	4

```sql
-- design a new index to measure shippers' performance - 'delayindex', the
higher the more delays days they had
-- Speedy Express has the highest Delay Index meaning the company should
examinate the partnership with them and maybe switch to a new partner in
order to increase customer satistiifcation.

SELECT ShipperName, AVG(DelayedDay) AS AverageDelay, COUNT(OrderID) AS
DelayOrderNumber,
ROUND(AVG(DelayedDay) * COUNT(OrderID),0) AS DelayIndex
FROM dw_cis467
WHERE DelayedDay > 0
GROUP BY ShipperName
ORDER BY AVG(DelayedDay) * COUNT(OrderID) DESC;
```

	ShipperName	AverageDelay	DelayOrderNumber	DelayIndex
▶	Speedy Express	7.9677	31	247
	United Package	4.6944	36	169
	Federal Shipping	6.0800	25	152

Section 2 Queries for customer study--
-- **Description for query 3:** The TOP25 MVCs (most valued customer)by demographic information based on purchase frequency and total sales.
-- Most valued customer list (based on total sales and purchase frequency)
-- Returned by such queries are our top 25 most valued customer list with their total sales and shopping frequency defined by the number of orders they placed during their entire shopping cycle as well as their locations.

```sql
SELECT CustomerID, ROUND(SUM(ProductSales),2) AS Sales_customer, ShipCountry,
-- define frequency as the order placed in 100 days --
COUNT(OrderID)/DATEDIFF(MAX(OrderDate), MIN(OrderDate))*100 AS Frequency
FROM DW_cis467
GROUP BY CustomerID
ORDER BY Sales_customer*Frequency DESC
LIMIT 25;
```

	CustomerID	Sales_customer	ShipCountry	Frequency
▶	SAVEA	104361.96	USA	20.3509
	ERNSH	104874.99	Austria	15.5251
	QUICK	110277.32	Germany	13.9384
	RATTC	51097.79	USA	10.8729
	HUN	49979.9	Ireland	9.1362
	MEREP	28872.2	Canada	8.4656
	BERGS	24927.57	Sweden	9.1388
	KOENE	30908.38	Germany	7.0144
	FOLKO	29567.55	Sweden	7.0093
	FRANK	26656.55	Germany	7.7544
	QUEEN	25717.5	Brazil	7.7519
	WHITC	27363.6	USA	6.2598
	BONAP	21963.25	France	7.7601
	HANAR	32841.37	Brazil	4.8632
	HILAA	22768.76	Venezuela	6.9124
	BOTTM	20801.6	Canada	7.1429
	SUPRD	24088.78	Belgium	5.9908
	LINOD	16476.56	Venezuela	7.4468
	LEHMS	19261.41	Germany	6.1905
	GREAL	18507.46	USA	6.1281
	PICCO	23128.86	Austria	4.3396
	VAFFE	15843.92	Denmark	6.3265
	SEVES	16215.33	UK	5.9091
	WARTH	15648.7	Finland	5.8917
	BLONP	18534.08	France	4.8507

Section 3 Queries for customer study --
-- Description for query 4: To find orders which have a large/null order-stock ratio (insufficient availability inventory)and suppliers who are responsible for these orders. Also the products' categories with low stock.
-- Returned results are a list of products that faced with insufficient inventory or were sold out when ordered with their categories and their suppliers.
-- The company should examine the corresponding suppliers for future partnership and use such data as a record for underperformance of those suppliers.

```sql
SELECT ProductName, CategoryName, SupplierName,
CASE WHEN OrderToStock_Ratio IS NULL THEN "Sold Out"
WHEN OrderToStock_Ratio > 1 THEN "Not Sufficient Inventory"
WHEN OrderToStock_Ratio < 1 AND OrderToStock_Ratio > 0 THEN "Product In Stock"
ELSE "Not Ordered"
END AS InventoryStatus
FROM dw_cis467
WHERE OrderToStock_Ratio IS NULL OR OrderToStock_Ratio > 1
GROUP BY ProductName
ORDER BY OrderToStock_Ratio DESC;
```

ProductName	CategoryName	SupplierName	InventoryStatus
Louisiana Hot Spiced Okra	Condiments	New Orleans Cajun Delights	Not Sufficient Inventory
Rogede sild	Seafood	Lyngbysild	Not Sufficient Inventory
Sir Rodney's Scones	Confections	Specialty Biscuits, Ltd.	Not Sufficient Inventory
Maxilaku	Confections	Karkki Oy	Not Sufficient Inventory
Aniseed Syrup	Condiments	Exotic Liquids	Not Sufficient Inventory
Longlife Tofu	Produce	Tokyo Traders	Not Sufficient Inventory
Chocolade	Confections	Zaanse Snoepfabriek	Not Sufficient Inventory
Gravad lax	Seafood	Svensk Sj??a AB	Not Sufficient Inventory
Mascarpone Fabioli	Dairy Products	Formaggi Fortini s.r.l.	Not Sufficient Inventory
Wimmers gute Semmelkn?el	Grains/Cereals	Plutzer Lebensmittelgro??kte AG	Not Sufficient Inventory
Chang	Beverages	Exotic Liquids	Not Sufficient Inventory
Scottish Longbreads	Confections	Specialty Biscuits, Ltd.	Not Sufficient Inventory
Queso Cabrales	Dairy Products	Cooperativa de Quesos 'Las Ca...	Not Sufficient Inventory
Chef Anton's Gumbo Mix	Condiments	New Orleans Cajun Delights	Sold Out
Alice Mutton	Meat/Poultry	Pavlova, Ltd.	Sold Out
Th?inger Rostbratwurst	Meat/Poultry	Plutzer Lebensmittelgro??kte AG	Sold Out
rnzola Telino	Dairy Products	Formaggi Fortini s.r.l.	Sold Out
Perth Pasties	Meat/Poultry	G'day, Mate	Sold Out

Section 4 Queries for employee study --
-- Description for query 5: Sales people who performed the best regarding total product sales, and information about their title to know whether sales people with higher title tend to perform better or not.
-- Also, our output derives information on the loyal customers who made large purchases with one salesperson.
-- People with higher title do not necessarily perform better. Sales representatives generally have better performances. We also returned a list of loyal customers' ID for each salesperson for future CRM and remarketing purposes.

```sql
SELECT dw_cis467.Salesperson, dw_cis467.Title, dw_cis467.CustomerID AS
Loyal_Customer,
ROUND(SUM(ProductSales),2) AS Sales_employee,
SUM(OrderNumber) AS Order_total
FROM dw_cis467
JOIN (SELECT MAX(TotalSales) AS Max_Sales, Salesperson, CustomerID FROM
(SELECT CustomerID, Salesperson, ROUND(SUM(ProductSales),2) AS TotalSales
FROM dw_cis467
GROUP BY Salesperson, CustomerID
ORDER BY SUM(ProductSales) desc) Em
GROUP BY Salesperson
) emstudy ON emstudy.CustomerID = dw_cis467.CustomerID
GROUP BY Salesperson
ORDER BY Sales_employee DESC , Order_total DESC;
```

Salesperson	Title	Loyal_Customer	Sales_employee	Order_total
Andrew Fuller	Vice President, Sales	QUICK	170262.18	112
Nancy Davolio	Sales Representative	QUICK	132593.09	183
Janet Leverling	Sales Representative	SAVEA	120276.33	108
Robert King	Sales Representative	SAVEA	108475.99	68
Margaret Peacock	Sales Representative	RATTC	107353.7	138
Laura Callahan	Inside Sales Coordinator	SAVEA	84545.69	103
Steven Buchanan	Sales Manager	SAVEA	67362.88	68
Michael Suyama	Sales Representative	SAVEA	58342.64	73
Anne Dodsworth	Sales Representative	ERNSH	55040.08	42

```sql
-- To see the total sales under different titles
-- Sales Representative has the highest total sales and total orders and
outperform among all the titles
SELECT dw_cis467.Title,
ROUND(SUM(ProductSales),2) AS Sales_employee,
SUM(OrderNumber) AS Order_total
FROM dw_cis467
JOIN (SELECT MAX(TotalSales) AS Max_Sales, Salesperson, CustomerID FROM
(SELECT CustomerID, Salesperson, ROUND(SUM(ProductSales),2) AS TotalSales
FROM dw_cis467
GROUP BY Salesperson, CustomerID
ORDER BY SUM(ProductSales) desc) Em
GROUP BY Salesperson
) emstudy ON emstudy.CustomerID = dw_cis467.CustomerID
GROUP BY Title
ORDER BY Sales_employee DESC , Order_total DESC;
```

Title	Sales_employee	Order_total
Sales Representative	582081.83	612
Vice President, Sales	170262.18	112
Inside Sales Coordinator	84545.69	103
Sales Manager	67362.88	68

Section 5 Queries for Supplier study --
-- **Description for query 6**: To find out geographically distribution of suppliers with the most varied products/categories and whose products are most ordered
-- Supplier named Plutzer located in Germany has the highest category variety, product variety and total order and total product sales.

```
SELECT SupplierName, SupplierCountry, ProductSales,
SUM(OrderNumber) AS Orders_Total,
COUNT(DISTINCT CategoryName) AS CategoryVariety,
COUNT(DISTINCT ProductName) AS ProductVariety
FROM dw_cis467
GROUP BY SupplierName
ORDER BY Orders_Total DESC, ProductVariety DESC, CategoryVariety DESC;
```

SupplierName	SupplierCountry	ProductSales	Orders_Total	CategoryVariety	ProductVariety
Plutzer Lebensmittelgro??kte AG	Germany	665	179	5	5
Pavlova, Ltd.	Australia	570	163	5	5
Specialty Biscuits, Ltd.	UK	340	126	1	4
Norske Meierier	Norway	40	105	1	3
Gai p?urage	France	440	105	1	2
Formaggi Fortini s.r.l.	Italy	250	104	1	3
G'day, Mate	Australia	2120	98	3	3
Exotic Liquids	UK	810	94	2	3
New England Seafood Cannery	USA	183.35	88	1	2
Leka Trading	Singapore	441.6	82	3	3
Pasta Buttini s.r.l.	Italy	390	73	1	2
For?s d?ables	Canada	936.7	72	2	2
New Orleans Cajun Delights	USA	1299.84	70	1	4
Karkki Oy	Finland	240	70	2	3
Ma Maison	Canada	202.64	69	1	2
Mayumi's	Japan	60	68	3	3
Bigfoot Breweries	USA	144	65	1	3
Heli Süßwaren GmbH & Co. KG	Germany	373.5	59	1	3
Grandma Kelly's Homestead	USA	1750	54	2	3
Aux joyeux eccl?iastiques	France	283.5	54	1	2
Cooperativa de Quesos 'Las Ca...	Spain	1162.8	52	1	2
Tokyo Traders	Japan	595.2	51	3	3
Svensk Sj??a AB	Sweden	421.2	51	1	3
Refrescos Americanas LTDA	Brazil	101.25	51	1	1

-- suppliers geographical distribution
-- 3 Suppliers in France contribute most product sales and total orders as compared to other countries.

```
SELECT SupplierCountry, ProductSales,
SUM(OrderNumber) AS Orders_Total,
COUNT(DISTINCT SupplierName) AS SupplierNumber
FROM dw_cis467
GROUP BY SupplierCountry
ORDER BY ProductSales DESC, Orders_Total DESC;
```

	SupplierCountry	ProductSales	Orders_Total	SupplierNumber
▶	France	2040	177	3
	Italy	583.8	177	2
	Norway	576	105	1
	Germany	396	270	3
	Netherlands	380	27	1
	Singapore	276	82	1
	Spain	210	52	1
	Sweden	201.6	85	2
	Denmark	190	41	1
	Finland	180	70	1
	USA	147	277	4
	UK	131.4	220	2
	Japan	60	119	2
	Canada	35.4	141	2
	Australia	33.75	261	2
	Brazil	22.5	51	1

#Section 6 Queries for Product study

-- **Description for query 7**: To find out which products have the highest sales, order amount, sales per order and which categories are they in.
-- To also know whether these products are discontinued.
-- Ce de Blaye has the highest sale and orders amount. It is not discontinued. However, there are three products whose total sales and orders are rather significant but were discontinued. The company should find replacements for such products in order to maintain a high-level sales.

```sql
SELECT ProductName, CategoryName, ProductSales, SUM(OrderNumber) AS Orders_Total,
ROUND(ProductSales/ SUM(OrderNumber),2) AS SalesPerOrder,
CASE
WHEN Discontinued = 0 THEN "Not Discontinued"
WHEN Discontinued = 1 THEN "Discontinued"
END AS DiscontinueStatus
FROM dw_cis467
GROUP BY ProductName
ORDER BY ProductSales*Orders_Total, SalesPerOrder DESC;
```

ProductName	CategoryName	ProductSales	Orders_Total	SalesPerOrder	DiscontinueStatus
C?e de Blaye	Beverages	8263.36	24	344.31	Not Discontinued
Mishi Kobe Niku	Meat/Poultry	1396.8	5	279.36	Discontinued
Schoggi Schokolade	Confections	1755	9	195	Not Discontinued
Th?inger Rostbratwurst	Meat/Poultry	4642.12	32	145.07	Discontinued
Carnarvon Tigers	Seafood	3125	27	115.74	Not Discontinued
Alice Mutton	Meat/Poultry	3900	37	105.41	Discontinued
Chocolade	Confections	606.9	6	101.15	Not Discontinued
Northwoods Cranberry Sauce	Condiments	1200	13	92.31	Not Discontinued
Louisiana Hot Spiced Okra	Condiments	693.6	8	86.7	Not Discontinued
Grandma's Boysenberry Spread	Condiments	1000	12	83.33	Not Discontinued
Sirop d'?able	Condiments	1598.85	24	66.62	Not Discontinued
Genen Shouyu	Condiments	368.12	6	61.35	Not Discontinued
Gustaf's Kn?kebr?	Grains/Cereals	735	14	52.5	Not Discontinued
Vegie-spread	Condiments	842.88	17	49.58	Not Discontinued
Camembert Pierrot	Dairy Products	2427.6	51	47.6	Not Discontinued
Laughing Lumberjack Lager	Beverages	476	10	47.6	Not Discontinued
Chef Anton's Cajun Seasoning	Condiments	935	20	46.75	Not Discontinued
Chef Anton's Gumbo Mix	Condiments	427	10	42.7	Discontinued
Raclette Courdavault	Dairy Products	2090	54	38.7	Not Discontinued
Uncle Bob's Organic Dried Pears	Produce	1050	29	36.21	Not Discontinued
Gravad lax	Seafood	208	6	34.67	Not Discontinued
Mozzarella di Giovanni	Dairy Products	1218	38	32.05	Not Discontinued
P??chinois	Meat/Poultry	960	33	29.09	Not Discontinued
Sir Rodney's Marmalade	Confections	405	16	25.31	Not Discontinued
Tunnbr?	Grains/Cereals	450	20	22.5	Not Discontinued

-- **Description for query 8**: To find not discontinued products with the highest discounted frequency and to know the average of discounts. --
-- Ce de Blaye generates the most sales among all the discounted products orders, and among all the orders involving this product, 16.67% of the orders are discounted. The average discount is around 11% off. The company should look into this frequency and discount rate and conduct experiments to test whether this is the optimal choice for other products.

```sql
SELECT dw_cis467.ProductName, COUNT(DISTINCT Discount) AS
NonDiscounted_Number,
Quantity, ProductSales,
pr.Orders_Total, ROUND(AVG(Discount), 2) AS Average_Discount,
COUNT(DISTINCT dw_cis467.Discount)/pr.Orders_total AS Discount_Proportion
FROM dw_cis467
JOIN (SELECT SUM(OrderNumber) AS Orders_total, ProductName FROM dw_cis467
GROUP BY ProductName) pr
ON pr.ProductName = dw_cis467.ProductName
WHERE Discount != 0 AND dw_cis467.Discontinued = 0
GROUP BY ProductName
ORDER BY ProductSales DESC;
```

ProductName	NonDiscounted_Number	Quantity	ProductSales	Orders_Total	Average_Discount	Discount_Proportion
C?e de Blaye	4	49	8263.36	24	0.11	0.1667
Ikura	5	100	2635	33	0.13	0.1515
Carnarvon Tigers	5	40	2000	27	0.17	0.1852
Camembert Pierrot	6	70	1428	51	0.14	0.1176
Raclette Courdavault	5	30	1402.5	54	0.13	0.0926
Sir Rodney's Marmalade	5	21	1360.8	16	0.11	0.3125
Flotemysost	5	60	1032	42	0.15	0.1190
Wimmers gute Semmelkn?el	5	35	837.9	30	0.12	0.1667
Ipoh Coffee	4	20	782	28	0.18	0.1429
Tarte au sucre	5	21	776.47	48	0.15	0.1042
Boston Crab Meat	5	50	736	41	0.13	0.1220
rnzola Telino	5	70	700	51	0.12	0.0980
Gnocchi di nonna Alice	5	20	646	50	0.16	0.1000
Mozzarella di Giovanni	4	24	600.48	38	0.13	0.1053
Vegie-spread	4	16	505.44	17	0.18	0.2353
Escarts de Bourgne	5	40	503.5	18	0.16	0.2778
Inlagd Sill	5	30	456	31	0.19	0.1613
Maxilaku	3	30	456	21	0.11	0.1429
Mascarpone Fabioli	4	15	456	15	0.14	0.2667
Teatime Chocolate Biscuits	5	50	437	37	0.14	0.1351
Konbu	5	84	428.4	40	0.13	0.1250
Gravad lax	2	18	421.2	6	0.15	0.3333
Steeleye Stout	5	25	405	36	0.15	0.1389
Schoggi Schokolade	1	15	394.88	9	0.25	0.1111
Grandma's Boysenberry S...	3	20	375	12	0.09	0.2500

3. (**100 points**) Create **five** Tableau individual visualizations (graphs) **on your data warehouse** with valuable information to present findings to senior management of the company. Save each visualization as a png file (as we will practice in the lab 5) and paste each individual visualization png file **into this Word** document with the full explanation of what the visualizations show, how they are useful to a company and how company management could make decisions based on what you show. Finally, combine those **five** visualizations into one **Dashboard** (as we will practice in the lab 5), and save this Dashboard as a png file and **paste the Dashboard into this Word** document.

ProductSales-ShipCountry
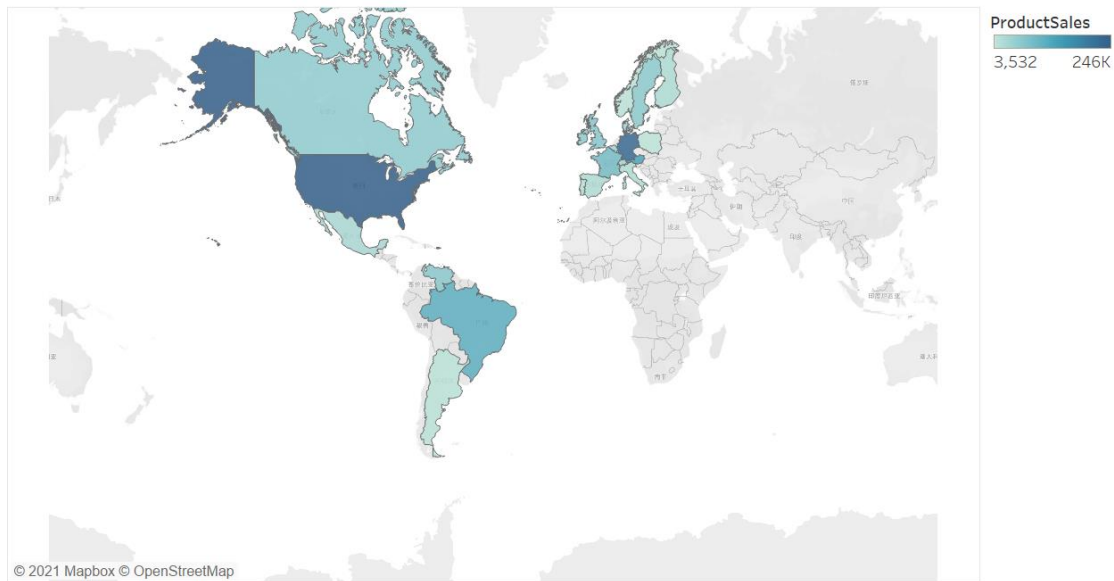
Map based on Longitude (generated) and Latitude (generated). Color shows sum of ProductSales. Details are shown for ShipCountry.

Sheet1: Ship countries' product sales

Explanation: This graph shows the sum of product sales in different ship countries

Why useful: The management team can base different countries' product sales to know briefly about the global sales situation (e.g.: Norway has the lowest total sales which is $5,735 and the USA has the largest total sales which is $245,585). Therefore, the management team can allocate resources such as human talents and promotion budgets to further develop the country which has better market potential.

SalesPerson-ProductSales and Category

| Salesperson	Beverages	Dairy Products	Meat/Poultry	CategoryName Confections	Seafood	Condiments	Grains/Cereals	Produce
Margaret Peacock I.	50,308	33,550	30,867	27,769	27,316	23,315	22,580	17,187
Nancy Davolio	46,599	36,023	15,038	28,569	24,144	13,562	8,466	19,706
Janet Leverling	44,757	32,321	20,503	33,622	25,032	13,382	21,235	11,961
Andrew Fuller	40,248	23,813	29,874	21,456	15,748	14,851	11,173	9,376
Robert King	27,964	27,622	21,177	14,519	7,147	8,851	6,536	10,753
Anne Dodsworth	19,643	21,101	8,677	8,053	8,149	10,126	1,245	315
Laura Callahan	17,898	21,101	16,395	21,700	12,042	14,638	11,072	12,017
Steven Buchanan	11,001	21,938	11,488	4,810	5,744	2,675	4,028	7,109
Michael Suyama	9,450	17,039	9,004	6,860	5,941	4,648	9,411	11,561

Sum of ProductSales broken down by CategoryName vs. Salesperson. Color shows sum of ProductSales. The marks are labeled by sum of ProductSales.

Sheet 2: Sales persons' total product sales among different categories.

Explanation: The graph shows 9 sales persons' product sales in 8 product categories.

Why useful: Based on the graph, the manager and human resources team can briefly identify the overall performance and the sales in various categories. Management team can use the graph as a key performance indicator to evaluate employee performances, give them some incentives. Furthermore, human resource team can allocate employees who are prominent to be a specialist in selling one category.

## ShipperName-DelayedDays and ShipDays
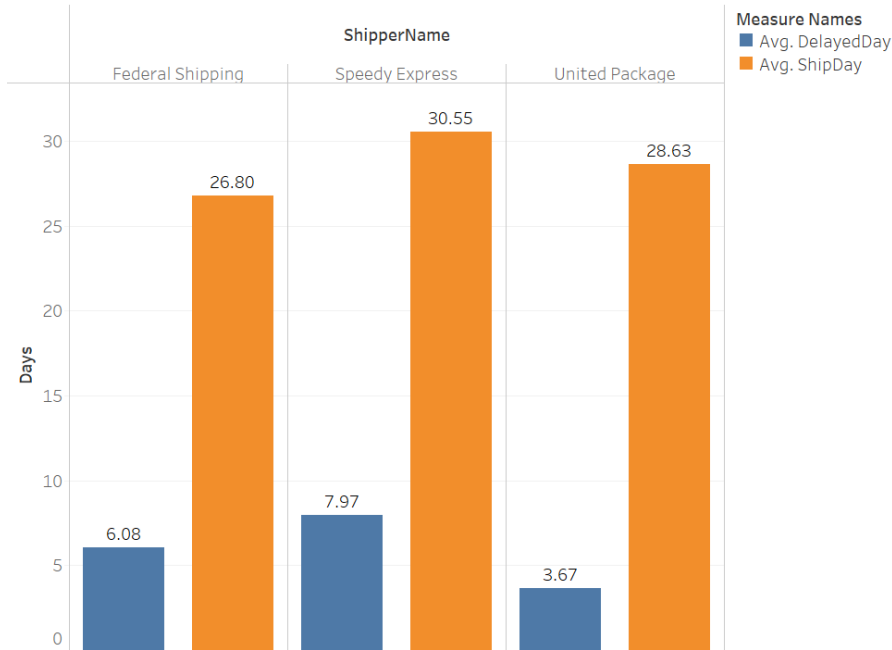
Avg. DelayedDay and Avg. ShipDay for each ShipperName. Color shows details about Avg. DelayedDay and Avg. ShipDay. The data is filtered on DelayedDay and ShipDay. The DelayedDay filter ranges from 0 to 23 and keeps Null values. The ShipDay filter ranges from 0 to 37 and keeps Null values.

Sheet 3: Shippers' average delayed days and ship days.

Explanation: Using shipped date -required date, we can get delayed day (if delayed day <0, the package arrive earlier than the required date). By filtering the delayed day >0, we can calculate the average delayed day for three shippers. Using order date- shipped date, we can get ship day. The graph shows the average of delayed day and ship day for the three shippers.

Why useful: Management team could see the shipping speed as well as the delayed situation for the three shippers (Speedy Express has both the highest average delayed day and ship day) to evaluate the performances.

OrderDate-Category
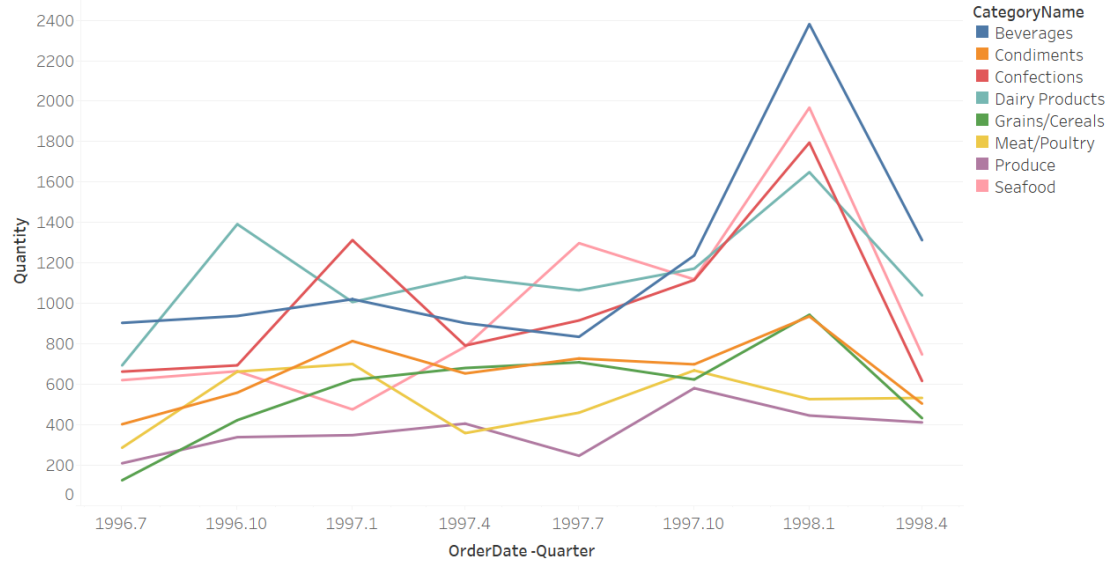
The trend of sum of Quantity for OrderDate Quarter. Color shows details about CategoryName.

Sheet 4: Sales quantity for product categories

Explanation: The graph shows the total quantity of different product categories changes quarterly from 1996.7 to 1998.4.

Why useful: Besides looking at the changes of quantity sold, the graph also shows a sharp increase in sales quantity from 1997.10 to 1998.1, followed by a decrease afterwards. Management team could recall what business strategies they have utilized during 1997.10-1998.1 and cooperate with business consulting teams to identify the sharp decrease from 1998.1 to 1998.4 to come up with business plans that make sure the business has a steady growth.
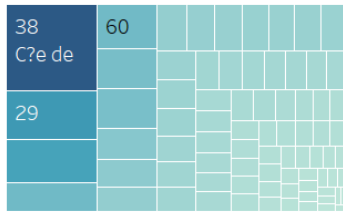
Sales-Product
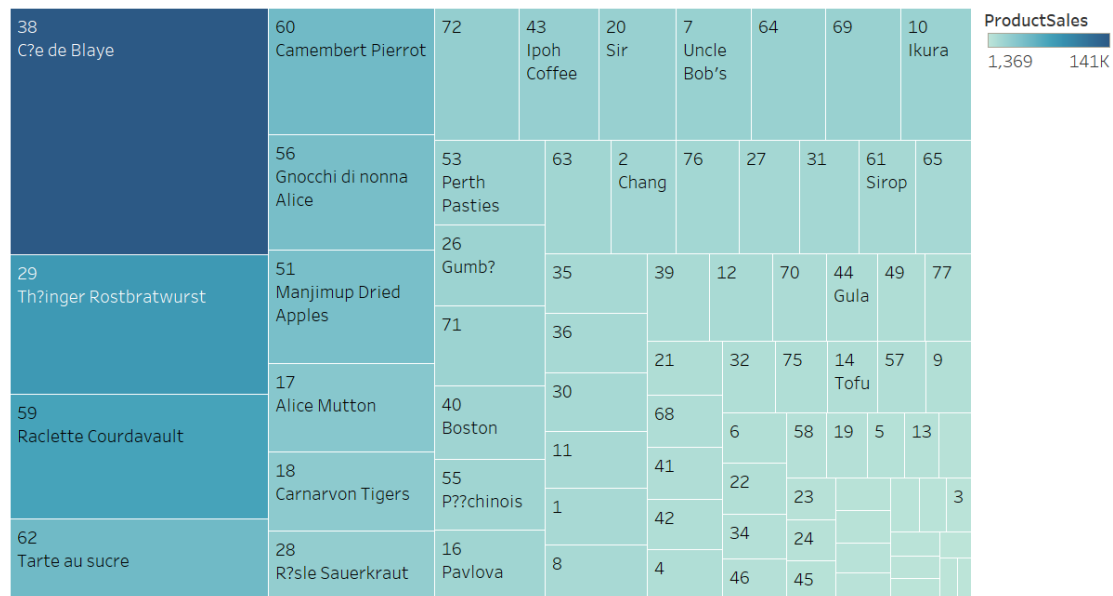
ProductID and ProductName. Color shows sum of ProductSales. Size shows sum of ProductSales. The marks are labeled by ProductID and ProductName.

Sheet 5: Total sales for different products.

Explanation: The graph shows the ranking of the sales for all products.

Why useful: Management team can be based on the graph to identify which products are popular among customers. Therefore, it can help managers to make decisions to adjust the supply chain to increase production on those who are popular to make more sales.
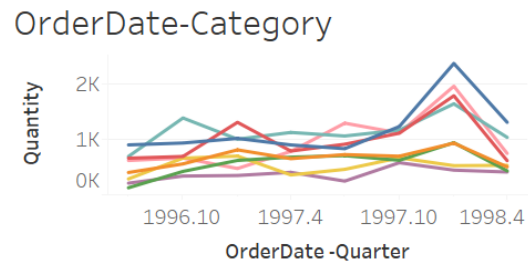
Sales-Product

ProductSales-ShipCountry
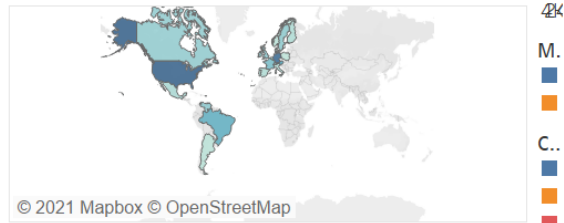
© 2021 Mapbox © OpenStreetMap

SalesPerson-ProductSales and Category

Salesperson	Beverages ⌐	Dairy Products	Meat/Poultry	Confections	Seafood
Margaret Peacock	50,308	33,550	30,867	27,769	2
Nancy Davolio	46,599	36,023	15,038	28,569	24
Janet Leverling	44,757	32,321	20,503	33,622	25
Andrew Fuller	40,248	23,813	29,874	21,456	1
Robert King	27,964	27,622	21,177	14,519	7

ShipperName-DelayedDays and ShipDays

OrderDate-Category

Please also save the whole Tableau project as a Tableau Workbook file .twb (In Tableau use File - Save as) and submit to the Final Team Project folder on Blackboard together with this Word document and together with the Excel file of your Data Warehouse which you uploaded to Tableau and used for visualizations. If you cannot attach the Tableau Workbook .twb file and the Excel file for your Data Warehouse to the Final Team Project folder on Blackboard, please email the Tableau Workbook .twb file and the Excel file for your Data Warehouse to me at mlysyako@simon.rochester.edu indicating your class section and your team name from Blackboard and all members in the email.

General grading criteria: Your completed work will be evaluated using the criteria below. I encourage you to use your creativity and other business skills (communication, presentation, critical thinking) in addition to the data management concepts and the SQL and Tableau skills that we have covered in CIS467.

High score	Score between high and good	Good/medium score	Low score
All required parts of the final project are complete and technically correct. Queries are useful/interesting and provide valuable information for senior management to act upon. Not just random queries. Tableau visualizations provide interesting useful information based on which senior management of the company can make important decisions.	All required parts of the final project are complete and technically correct (with possibly a few minor errors). Queries are useful/interesting and provide valuable information for senior management to act upon. Not just random queries (with possibly a few minor errors). Tableau visualizations provide interesting useful information based on which senior management of the company can make important decisions (with possibly a few minor errors).	Some required parts of the final project are missing and/or there are more significant errors. Some queries appear random and do not answer any useful/interesting questions. Tableau visualizations are very simple but may still provide interesting useful information based on which senior management of the company can make important decisions.	The final project has large portions missing and/or major conceptual errors. Most/all queries (if any) appear random and do not answer any useful/interesting questions. Tableau visualizations are very simple and **do not** provide interesting useful information based on which senior management of the company can make important decisions.