

Team: B-A2

Team Members: Jiajian Chen, Yiwei Gong, Yuwei Duan, Yushan Zhu, Xinyang Ru

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).

Submit a screenshot of the first 25 rows of your data warehouse (paste into this Word document) and the SQL code that you used to create it. Please copy and paste your SQL code into this Word document. If your PC does not show 25 rows of data, please submit what you have (i.e., rows you can see on a screenshot) with a comment that you cannot show 25 rows of data. Please add a description of what your Data Warehouse will be tracking for a company.

Description of the Data Warehouse:

This Data Warehouse centers around orders mainly tracking their shipments, client sources, item structures and subtotal structures individually.

```
CREATE OR REPLACE VIEW dw_v2 AS
SELECT Orders.OrderID, STR_TO_DATE(Orders.OrderDate, "%m/%d/%Y") AS OrderDate, Orders.EmployeeID,
       STR_TO_DATE(Orders.RequiredDate, "%m/%d/%Y") AS RequiredDate,
       STR_TO_DATE(Orders.ShippedDate, "%m/%d/%Y") AS ShippedDate,
       IF(STR_TO_DATE(Orders.ShippedDate, "%m/%d/%Y") IS NULL, NULL,
          IF (STR_TO_DATE(Orders.RequiredDate, "%m/%d/%Y")-STR_TO_DATE(Orders.ShippedDate, "%m/%d/%Y")<0, 'delayed', 'on_time')) as 'delayed_order',
       Shippers.CompanyName AS ShipperName, Orders.ShipCountry, Orders.CustomerID, Customers.CompanyName AS CustomerName, Customers.Country AS CustomerCountry,
       COUNT(DISTINCT Suppliers.SupplierID) AS "#Number_Supplier",
       IFNULL(SUM(CASE WHEN Categories.CategoryID = 1 THEN Order_Details.Quantity END), 0) AS "Quantity_Beverage",
       IFNULL(SUM(CASE WHEN Categories.CategoryID = 2 THEN Order_Details.Quantity END), 0) AS "Quantity_Condiments",
       IFNULL(SUM(CASE WHEN Categories.CategoryID = 3 THEN Order_Details.Quantity END), 0) AS "Quantity_Confections",
```

```

    IFNULL(SUM(CASE WHEN Categories.CategoryID = 4 THEN Order_Details.Quantity END),
0) AS "Quantity_Dairy_Products",
    IFNULL(SUM(CASE WHEN Categories.CategoryID = 5 THEN Order_Details.Quantity END),
0) AS "Quantity_Grains/Cereals",
    IFNULL(SUM(CASE WHEN Categories.CategoryID = 6 THEN Order_Details.Quantity END),
0) AS "Quantity_Meat/Poultry",
    IFNULL(SUM(CASE WHEN Categories.CategoryID = 7 THEN Order_Details.Quantity END),
0) AS "Quantity_Produce",
    IFNULL(SUM(CASE WHEN Categories.CategoryID = 8 THEN Order_Details.Quantity END),
0) AS "Quantity_Seafood",
    IFNULL(ROUND(SUM(CASE WHEN Categories.CategoryID = 1 THEN ROUND((Order_Details.Un
itPrice* Order_Details.Quantity*(1-Order_Details.Discount)/100)*100, 2) END), 2), 0) AS
"SubtotalBeverages",
    IFNULL(ROUND(SUM(CASE WHEN Categories.CategoryID = 2 THEN ROUND((Order_Details.Un
itPrice* Order_Details.Quantity*(1-Order_Details.Discount)/100)*100, 2) END), 2), 0) A
S "SubtotalCondiments",
    IFNULL(ROUND(SUM(CASE WHEN Categories.CategoryID = 3 THEN ROUND((Order_Details.Un
itPrice* Order_Details.Quantity*(1-Order_Details.Discount)/100)*100, 2) END), 2), 0) AS
"SubtotalConfections",
    IFNULL(ROUND(SUM(CASE WHEN Categories.CategoryID = 4 THEN ROUND((Order_Details.Un
itPrice* Order_Details.Quantity*(1-Order_Details.Discount)/100)*100, 2) END), 2), 0) AS
"SubtotalDairy_Products",
    IFNULL(ROUND(SUM(CASE WHEN Categories.CategoryID = 5 THEN ROUND((Order_Details.Un
itPrice* Order_Details.Quantity*(1-Order_Details.Discount)/100)*100, 2) END), 2), 0) AS
"SubtotalGrains/Cereals",
    IFNULL(ROUND(SUM(CASE WHEN Categories.CategoryID = 6 THEN ROUND((Order_Details.Un
itPrice* Order_Details.Quantity*(1-Order_Details.Discount)/100)*100, 2) END), 2), 0) AS
"SubtotalMeat/Poultry",
    IFNULL(ROUND(SUM(CASE WHEN Categories.CategoryID = 7 THEN ROUND((Order_Details.Un
itPrice* Order_Details.Quantity*(1-Order_Details.Discount)/100)*100, 2) END), 2), 0) AS
"SubtotalProduce",
    IFNULL(ROUND(SUM(CASE WHEN Categories.CategoryID = 8 THEN ROUND((Order_Details.Un
itPrice* Order_Details.Quantity*(1-Order_Details.Discount)/100)*100, 2) END), 2), 0) AS
"SubtotalSeafood",
    ROUND(SUM(ROUND((Order_Details.UnitPrice* Order_Details.Quantity*(1-Order_Detail
s.Discount)/100)*100, 2)), 2) AS OrderAmount,
    ROUND(SUM(Order_Details.UnitPrice* Order_Details.Quantity), 2) AS BookAmount,
    ROUND(1 - ROUND(SUM(ROUND((Order_Details.UnitPrice* Order_Details.Quantity*(1-Ord
er_Details.Discount)/100)*100, 2)), 2)/ROUND(SUM(Order_Details.UnitPrice* Order_Detail
s.Quantity), 2), 2) AS OrderDiscount,
    Orders.Freight
FROM Categories JOIN
    (Suppliers JOIN
        (Shippers JOIN

```

```
(Products JOIN
  ((Employees JOIN
    (Customers JOIN Orders
      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 Shippers.ShipperID = Orders.ShipVia)
      ON Suppliers.SupplierID = Products.SupplierID)
      ON Categories.CategoryID = Products.CategoryID

GROUP BY OrderID
ORDER BY OrderID;
```

OrderID	OrderDate	EmployeeID	RequiredDate	ShippedDate	delayed_order	ShipperName	ShipCountry	CustomerID	CustomerName	CustomerCount...	#Number_Suppli...	Quantity_Beverage	Quantity_L...
10248	1996-07-04	5	1996-08-01	1996-07-16	on_time	Federal Shipping	France	VINET	Vins et alcools Chevalier	France	3	0	0
10249	1996-07-05	6	1996-08-16	1996-07-10	on_time	Speedy Express	Germany	TOMSP	Toms Spezialitäten	Germany	2	0	0
10250	1996-07-08	4	1996-08-05	1996-07-12	on_time	United Package	Brazil	HANAR	Hanari Carnes	Brazil	3	0	15
10251	1996-07-08	3	1996-08-05	1996-07-15	on_time	Speedy Express	France	VICTE	Victuailles en stock	France	3	0	20
10252	1996-07-09	4	1996-08-06	1996-07-11	on_time	United Package	Belgium	SUPRD	Suprêmes d'ômes	Belgium	3	0	0
10253	1996-07-10	3	1996-07-24	1996-07-16	on_time	United Package	Brazil	HANAR	Hanari Carnes	Brazil	3	42	0
10254	1996-07-11	5	1996-08-08	1996-07-23	on_time	United Package	Switzerland	CHOPS	Chop-suey Chinese	Switzerland	3	15	0
10255	1996-07-12	9	1996-08-09	1996-07-15	on_time	Federal Shipping	Switzerland	RICSU	Richter Supermarkt	Switzerland	4	20	0
10256	1996-07-15	3	1996-08-12	1996-07-17	on_time	United Package	Brazil	WELLI	Wellington Importadora	Brazil	2	0	12
10257	1996-07-16	4	1996-08-13	1996-07-22	on_time	Federal Shipping	Venezuela	HILAA	HILARION-Abastos	Venezuela	3	6	15
10258	1996-07-17	1	1996-08-14	1996-07-23	on_time	Speedy Express	Austria	ERNSH	Ernst Handel	Austria	3	50	65
10259	1996-07-18	4	1996-08-15	1996-07-25	on_time	Federal Shipping	Mexico	CENTC	Centro comercial Moct...	Mexico	2	0	0
10260	1996-07-19	4	1996-08-16	1996-07-29	on_time	Speedy Express	Germany	OTTIK	Ottiles K?laden	Germany	4	21	0
10261	1996-07-19	4	1996-08-16	1996-07-30	on_time	United Package	Brazil	QUEDA	Que Del?ia	Brazil	2	20	0
10262	1996-07-22	8	1996-08-19	1996-07-25	on_time	Federal Shipping	USA	RATTIC	Rattlesnake Canyon Gr...	USA	3	0	12
10263	1996-07-23	9	1996-08-20	1996-07-31	on_time	Federal Shipping	Austria	ERNSH	Ernst Handel	Austria	4	28	0
10264	1996-07-24	6	1996-08-21	1996-08-23	delayed	Federal Shipping	Sweden	FOLKO	Folk och I?HB	Sweden	2	35	0
10265	1996-07-25	2	1996-08-22	1996-08-12	on_time	Speedy Express	France	BLONP	Blondesd?sl p?e et fils	France	1	20	0
10266	1996-07-26	3	1996-09-06	1996-07-31	on_time	Federal Shipping	Finland	WARTH	Wartian Herku	Finland	1	0	0
10267	1996-07-29	4	1996-08-26	1996-08-06	on_time	Speedy Express	Germany	FRANK	Frankenversand	Germany	3	15	0
10268	1996-07-30	8	1996-08-27	1996-08-02	on_time	Federal Shipping	Venezuela	GROSR	GROSELLA-Restaurante	Venezuela	2	0	0
10269	1996-07-31	5	1996-08-14	1996-08-09	on_time	Speedy Express	USA	WHITC	White Clover Markets	USA	2	0	0
10270	1996-08-01	1	1996-08-29	1996-08-02	on_time	Speedy Express	Finland	WARTH	Wartian Herku	Finland	2	25	0
10271	1996-08-01	6	1996-08-29	1996-08-30	delayed	United Package	USA	SPLIR	Split Rail Beer & Ale	USA	1	0	0
10272	1996-08-02	6	1996-08-30	1996-08-06	on_time	United Package	USA	RATTIC	Rattlesnake Canyon Gr...	USA	2	0	0
10273	1996-08-05	2	1996-08-22	1996-08-12	on_time	Federal Shipping	Germany	CHICK	CHICK Stew	Germany	5	22	0

. Quantity_Beverage	Quantity_Condiment	Quantity_Confectio...	Quantity_Dairy_Produ...	Quantity_Grains/Cere...	Quantity_Meat/Poul...	Quantity_Produce	Quantity_Seafood	Quantity_Sugar
0	0	0	17	10	0	0	0	0
0	0	0	0	0	0	49	0	0
0	15	0	0	0	0	35	10	0
0	20	0	0	21	0	0	0	0
0	0	40	65	0	0	0	0	0
42	0	40	20	0	0	0	0	0
15	0	0	0	0	21	21	0	0
20	0	35	30	0	0	0	25	0
0	12	0	0	0	15	0	0	0
6	15	25	0	0	0	0	0	0
50	65	0	6	0	0	0	0	0
0	0	10	0	0	0	0	1	0
21	0	15	0	50	0	0	0	16
20	0	20	0	0	0	0	0	0
0	12	0	0	2	0	15	0	0
28	0	60	0	0	0	36	60	0
35	0	0	0	0	0	0	25	0
20	0	0	0	0	30	0	0	0
0	0	0	12	0	0	0	0	0
15	0	0	70	0	0	0	50	0
0	0	0	4	0	10	0	0	0
0	0	0	80	0	0	0	0	0
25	0	0	0	0	0	0	30	0
0	0	0	24	0	0	0	0	0
0	0	6	64	0	0	0	0	0
33	0	0	35	0	0	0	84	0

	SubtotalBeverag...	SubtotalCondime...	SubtotalConfectio...	SubtotalDairy_Produ...	SubtotalGrains/Cere...	SubtotalMeat/Poul...	SubtotalProdu...	SubtotalSeafo...	OrderAmount	BookAmount	OrderDiscou...	Freight
0	0	0	342	98	0	0	440	440.00	0	32.38		
0	0	0	0	0	1863.4	0	1863.4	1863.40	0	11.61		
0	214.2	0	0	0	1261.4	77	1552.6	1813.00	0.14	65.83		
0	336	0	0	318.06	0	0	654.06	670.80	0.02	41.34		
0	0	2462.4	1135.5	0	0	0	3597.9	3730.00	0.04	51.30		
604.8	0	640	200	0	0	0	1444.8	1444.80	0	58.17		
45.9	0	0	0	0	342.72	168	0	556.62	625.20	0.11	22.98	
304	0	486.5	1320	0	0	0	380	2490.5	2490.50	0	148.33	
0	124.8	0	0	393	0	0	517.8	517.80	0	13.97		
86.4	156	877.5	0	0	0	0	0	1119.9	1119.90	0	81.91	
608	884	0	122.88	0	0	0	0	1614.88	2018.60	0.2	140.51	
0	0	80	0	0	0	0	20.8	100.8	100.80	0	3.25	
189	0	443.25	0	780	0	0	92.4	1504.65	1746.20	0.14	55.09	
288	0	160	0	0	0	0	0	448	448.00	0	3.05	
0	163.2	0	0	60.8	0	360	0	584	624.80	0.07	48.29	
100.8	0	625.5	0	0	0	216	931.5	1873.8	2464.80	0.24	146.06	
532	0	0	0	0	0	0	163.62	695.62	724.50	0.04	3.67	
240	0	0	0	0	936	0	0	1176	1176.00	0	55.28	
0	0	0	346.56	0	0	0	0	346.56	364.80	0.05	25.73	
183.6	0	0	2618	0	0	0	735	3536.6	4031.00	0.12	208.58	
0	0	0	111.2	0	990	0	0	1101.2	1101.20	0	66.29	
0	0	0	642.2	0	0	0	0	642.2	676.00	0.05	4.56	
920	0	0	0	0	0	0	456	1376	1376.00	0	136.54	
0	0	0	48	0	0	0	0	48	48.00	0	4.54	
0	0	388.8	1067.2	0	0	0	0	1456	1456.00	0	98.03	
451.44	0	0	182.5	0	0	0	1403.34	2037.28	2142.40	0.05	76.07	

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).

Submit a copy of each query SQL code (paste into this Word document), and the screenshot of each query results (or the first 25 rows if it is longer or how many rows you can get on your PC) and full description of the question your SQL code was addressing and what you found in the results. The question that each query answers should be useful for a company to make decisions and act upon.

-- Query 1:

-- From the perspective of order value, which salesman should we give higher compensation?

```
SELECT YEAR(OrderDate), CONCAT.FirstName, ' ', LastName AS Salesperson, ROUND(SUM(OrderAmount), 2) AS orderValuebySalesperson
FROM dw_v2 d
JOIN Employees e
ON d.EmployeeID = e.EmployeeID
GROUP BY YEAR(OrderDate), Salesperson
ORDER BY YEAR(OrderDate), orderValuebySalesperson DESC;
```

-- Finding: Margaret Peacock is best salesman for 1996 and 1997 but she was a little bit left behind for 1998. The management team could probably communicate with her to ask why.

YEAR(OrderDate)	Salesperson	orderValuebySalesperson
▶ 1996	Margaret Peacock	49945.12
1996	Nancy Davolio	35764.51
1996	Laura Callahan	22240.12
1996	Andrew Fuller	21757.06
1996	Steven Buchanan	18383.92
1996	Janet Leverling	18223.96
1996	Michael Suyama	16642.6
1996	Robert King	15232.16
1996	Anne Dodsworth	9894.52
1997	Margaret Peacock	128809.82
1997	Janet Leverling	108026.12
1997	Nancy Davolio	93148.1
1997	Andrew Fuller	70444.14
1997	Robert King	60471.19
1997	Laura Callahan	56032.61
1997	Michael Suyama	43126.38
1997	Steven Buchanan	30716.48
1997	Anne Dodsworth	26310.39

-- Query 2:

-- Which country has order numbers over average?

```
SELECT ShipCountry, COUNT(OrderID) AS numberOforders
FROM dw_v2
GROUP BY ShipCountry
HAVING COUNT(OrderID) > (SELECT COUNT(*)/COUNT(DISTINCT ShipCountry) FROM dw_v2)
ORDER BY numberOforders DESC;
```

-- Finding: seven countries have order numbers over average, they are Germany, USA, Brazil, France, UK, Venezuela and Austria. We should emphasize more on those countries.

Result Grid Filter Rows: Search

ShipCountry	numberOforders
► Germany	122
USA	122
Brazil	83
France	77
UK	56
Venezuela	46
Austria	40

-- Query 3:

-- Order delay rates by country

```
SELECT ShipCountry,
       COUNT(IF (delayed_order = "delayed", 1, null))/COUNT(ShippedDate) AS delayRates
FROM dw_v2
GROUP BY ShipCountry
ORDER BY delayRates DESC;
```

-- Finding: orders shipped to Ireland have the highest delay rates; orders shipped to Switzerland, Mexico, Canada, Denmark, Poland and Norway never delay. We should investigate why the delay rate in Ireland is the highest to improve our service.

Result Grid Filter Rows: Search Export:

ShipCountry	delayRates
Ireland	0.1075
Sweden	0.0811
Portugal	0.0769
Italy	0.0741
UK	0.0714
Argentina	0.0714
USA	0.0588
Belgium	0.0526
Venezuela	0.0465
Finland	0.0455
Spain	0.0435
France	0.0400
Brazil	0.0370
Germany	0.0333
Austria	0.0263
Switzerland	0.0000
Mexico	0.0000
Canada	0.0000
Denmark	0.0000

Result 4

-- Query 4:

-- one dollar cost in freight will generate how much revenue in each order? used to judge which shipper is more valuable to us

```
SELECT ShipCountry, ShipperName, Round(SUM(OrderAmount)/SUM(Freight), 3) AS RevenueGeneratedPerDollarFreightCost, Round(SUM(OrderAmount), 2) AS total_revenue_by_shipper
FROM dw_v2
GROUP BY ShipCountry, ShipperName
ORDER BY ShipCountry DESC, RevenueGeneratedPerDollarFreightCost DESC;
```

-- Finding: The value of the shipper varies in different destination country,

-- for example, for destinations to the USA we probably could use more Speedy Express since for every dollar freight, it could generate more revenue for us.

-- But at the same time, we should investigate more since we are not sure whether the company tends to give smaller weight but higher value orders to Speedy Express or not, this might introduce some bias.

ShipCountry	ShipperName	RevenueGeneratedPerDollarFreightCost	total_revenue_by_shipper
Venezuela	Speedy Express	24.682	20272.15
Venezuela	United Package	20.469	23599.55
Venezuela	Federal Shipping	17.004	12938.92
USA	Speedy Express	23.418	53737.11
USA	United Package	17.442	100790.78
USA	Federal Shipping	15.981	91056.73
UK	United Package	21.874	26985.86
UK	Federal Shipping	19.193	18899.45
UK	Speedy Express	17.783	13086.01
Switzerland	Speedy Express	24.347	6099.56
Switzerland	United Package	24.286	11967.8
Switzerland	Federal Shipping	21.793	13625.28
Sweden	Federal Shipping	26.255	13704.85
Sweden	Speedy Express	15.339	14858.91
Sweden	United Package	14.844	25931.36
Spain	Federal Shipping	50.919	6544.1
Spain	United Package	15.696	8602

-- Query 5:

-- From the perspective of order value, find customers with higher spending per order than average, their total_spending in our company and divide them into 5 groups with rank

```
CREATE TABLE top20 AS
```

```

SELECT CustomerName, CustomerCountry, ROUND(SUM(OrderAmount)/COUNT(DISTINCT(OrderID)), 2) AS amount_per_order,
       ROUND(SUM(OrderAmount), 2) AS total_spending,
       NTILE(5) OVER(ORDER BY ROUND(SUM(OrderAmount), 2)DESC) AS customerRanking
FROM dw_v2
GROUP BY CustomerName, CustomerCountry
HAVING amount_per_order > (SELECT SUM(OrderAmount)/COUNT(DISTINCT(OrderID)) FROM dw_v2)
ORDER BY customerRanking;

```

-- Finding: Maintaining the relationship with those top-quality buyers, such as in rank 1 and rank 2. Finding ways to encourage customers in rank 3/4/5 to spend more to increase our order value.

Result Grid					
	CustomerName	CustomerCountry	amount_per_order	total_spending	customerRanking
▶	QUICK-Stop	Germany	3938.48	110277.32	1
◀	Ernst Handel	Austria	3495.83	104874.99	1
◀	Save-a-lot Markets	USA	3366.51	104361.96	1
◀	Rattlesnake Canyon Grocery	USA	2838.77	51097.79	1
◀	Hungry Owl All-Night Grocers	Ireland	2630.52	49979.9	1
◀	Hanari Carnes	Brazil	2345.81	32841.37	2
◀	K?iglich Essen	Germany	2207.74	30908.38	2
◀	Folk och f?HB	Sweden	1556.19	29567.55	2
◀	M?e Paillarde	Canada	2220.94	28872.2	2
◀	White Clover Markets	USA	1954.54	27363.6	2
◀	Frankenversand	Germany	1777.1	26656.55	3
◀	Queen Cozinha	Brazil	1978.27	25717.5	3
◀	Supr?es d?ices	Belgium	2007.4	24088.78	3
◀	Piccolo und mehr	Austria	2312.89	23128.86	3
◀	Richter Supermarkt	Switzerland	1934.38	19343.76	4
◀	Blondesdssl p?e et fils	France	1684.92	18534.08	4
◀	Great Lakes Food Market	USA	1682.5	18507.46	4
◀	Simons bistro	Denmark	2402.44	16817.1	4

-- Query 6:

-- Analyse year-month level order value in our company to observe the seasonality and the top 3 sales countries in this month

```
WITH monthly_revenue AS (
```

```
    SELECT
```

```
        OrderDate, YEAR(OrderDate) AS orderYear, MONTH(OrderDate) AS orderMonth, ShipCountry,
```

```
        ROUND(SUM(OrderAmount), 2) AS monthly_revenue,
```

```
        RANK() OVER (
```

```
            PARTITION BY YEAR(OrderDate), MONTH(OrderDate)
```

```

        ORDER BY ROUND(SUM(OrderAmount), 2) DESC
    ) revenueRank
FROM dw_v2
GROUP BY YEAR(OrderDate), MONTH(OrderDate), ShipCountry
)
SELECT orderYear, orderMonth, ShipCountry, monthly_revenue, revenueRank
FROM monthly_revenue
WHERE revenueRank <=3
ORDER BY YEAR(OrderDate), MONTH(OrderDate) ASC;

```

- Findings: Germany and USA have most times on the top3, so they are very big market

Result Grid		Filter Rows:		Search	Export:
orderYear	orderMonth	ShipCountry	monthly_revenue	revenueRank	
▶ 1996	7	Germany	6904.65	1	
1996	7	Brazil	3963.2	2	
1996	7	Belgium	3597.9	3	
▶ 1996	8	Germany	9518.82	1	
1996	8	Brazil	4781.52	2	
1996	8	USA	3391.6	3	
1996	9	USA	9718.6	1	
1996	9	Ireland	4407	2	
1996	9	Belgium	2708.8	3	
1996	10	USA	10788.65	1	
1996	10	Germany	7553.04	2	
1996	10	Canada	5140.88	3	
1996	11	Austria	15681.78	1	
1996	11	France	10011.26	2	
1996	11	Germany	6888.24	3	
1996	12	Brazil	10240.18	1	
1996	12	USA	8924.15	2	
1996	12	Austria	6430.88	3	

-- Query 7:

-- Analyze the proportion of category in all orders by order value and by different years

```
SELECT      YEAR(OrderDate),
```

```

        ROUND(sum(SubtotalBeverages)/sum(OrderAmount), 2) as total_Beverages_percentage,
        ROUND(sum(SubtotalCondiments)/sum(OrderAmount), 2) as total_Condiments_percentage,
        ROUND(sum(SubtotalConfections)/sum(OrderAmount), 2) as total_Confections_percentage,
        ROUND(sum(SubtotalDairy_Products)/sum(OrderAmount), 2) total_Dairy_Products_percentage,
        ROUND(sum(`SubtotalGrains/Cereals`)/sum(OrderAmount), 2) as total_Grains_percentage,
        ROUND(sum(`SubtotalMeat/Poultry`)/sum(OrderAmount), 2) as total_Meat_percentage,
        ROUND(sum(SubtotalProduce)/sum(OrderAmount), 2) AS total_Produce_percentage,
        ROUND(sum(SubtotalSeafood)/sum(OrderAmount), 2) as total_Seafood_percentage
FROM dw_v2
GROUP BY YEAR(OrderDate);

```

-- Findings: There is no significant variation of orderAmount within each category

-- However, we still noticed that the Beverage category represents only 17% of the revenue in 1997, lower than 23% in 1996 and 26% in 1998.

-- It could be investigated more about this problem.

YEAR(OrderDate)	total_Beverages_percentage	total_Condiments_percentage	total_Confections_percentage	total_Dairy_Products_percentage
1996	0.23	0.09	0.14	0.2
1997	0.17	0.09	0.13	0.19
1998	0.26	0.07	0.12	0.18

total_Dairy_Products_percentage	total_Grains_percentage	total_Meat_percentage	total_Produce_percentage	total_Seafood_percentage
0.2	0.05	0.14	0.07	0.09
0.19	0.09	0.13	0.09	0.11
0.18	0.07	0.12	0.07	0.1

-- Query 8:

-- Analyze the delay of the shipment of orders after the order date by month to observe the seasonal impact

```

SELECT DATE_FORMAT(OrderDate, "%b") AS Order_month,
        ROUND(AVG(DATEDIFF(ShippedDate, OrderDate)), 2) AS shipment_delay

```

```

FROM dw_v2
GROUP BY Order_month
ORDER BY shipment_delay DESC;

-- Findings: orders in Sep and Jan are most delayed, probably because people are on vacation during these months

-- Probably hire more part-time employees for shipping orders in Jan since Jan is a peak-season for the company.

```

Order_month	shipment_delay
Sep	9.70
Jan	9.40
Mar	9.08
Dec	8.94
Jun	8.83
Nov	8.58
Jul	8.44
May	8.42
Feb	7.95
Oct	7.81
Apr	7.43
Aug	7.31

- Query 9: (additional one, for visualization)

```

CREATE TABLE revenuegenerate AS

SELECT ShipperName, Round(SUM(OrderAmount)/SUM(Freight), 3) AS RevenueGeneratedPerDollarFreightCost
FROM dw_v2
GROUP BY ShipperName
ORDER BY RevenueGeneratedPerDollarFreightCost DESC;

```

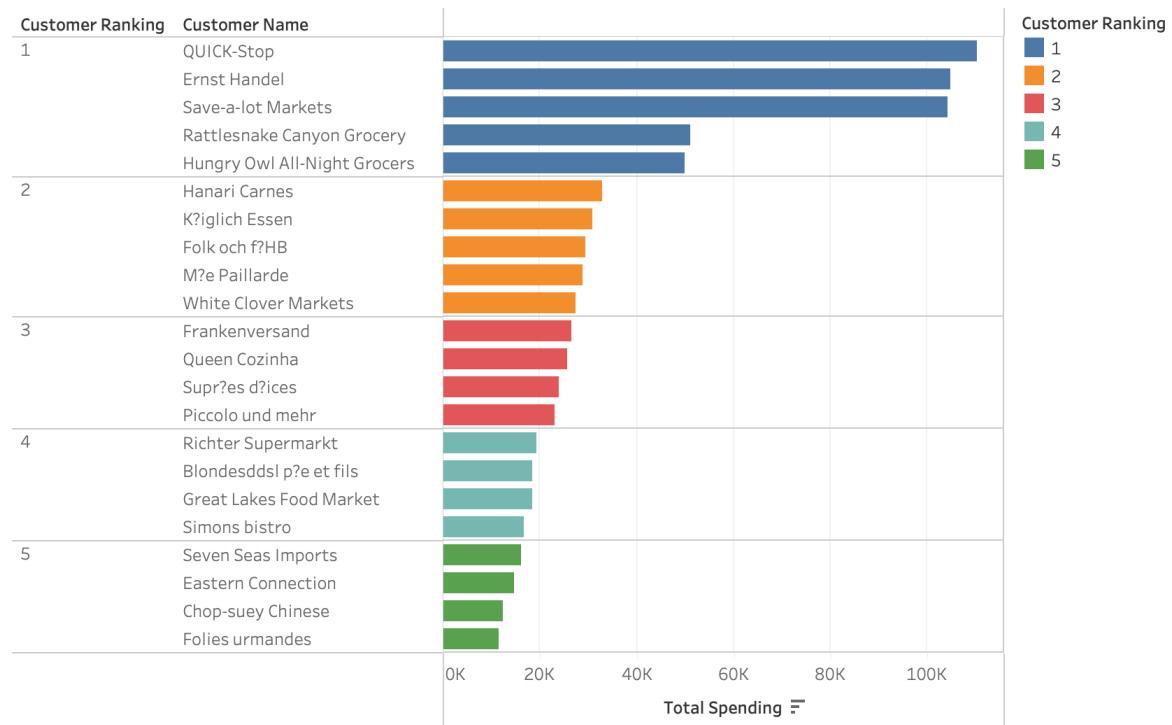
-- Findings: SpeedyExpress is the most efficient one and the average revenue per freight from the other two shippers are roughly the same.

-- Pay more attention to the partnership with SpeedyExpress, and at the same time find ways to lower the other two shipment company's freight cost.

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 exp

loration 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**.

Consumer Division by Revenue

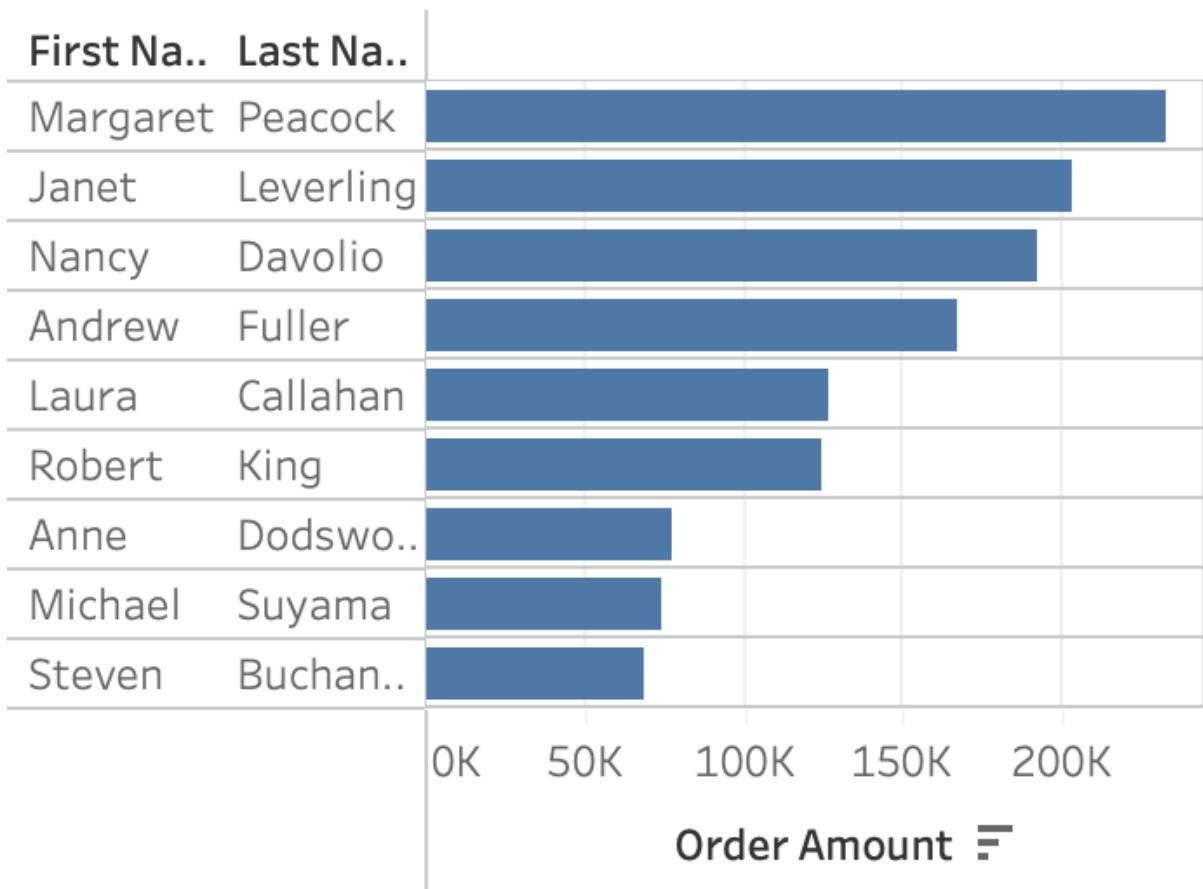


Sum of Total Spending for each Customer Name broken down by Customer Ranking. Color shows details about Customer Ranking. The data is filtered on sum of Order Amount and Customer Country. The sum of Order Amount filter includes everything. The Customer Country filter keeps 21 of 21 members.

We have divided consumers into five groups by revenue. This graph shows the revenue of each customer and their group number.

With this graph, our company can easily find out which customers have spent more money. Our further research can focus on their other factors like location, brand and preference. This can help us to find out the similarity in each consumer group in order to increase our profit.

Employee Revenue

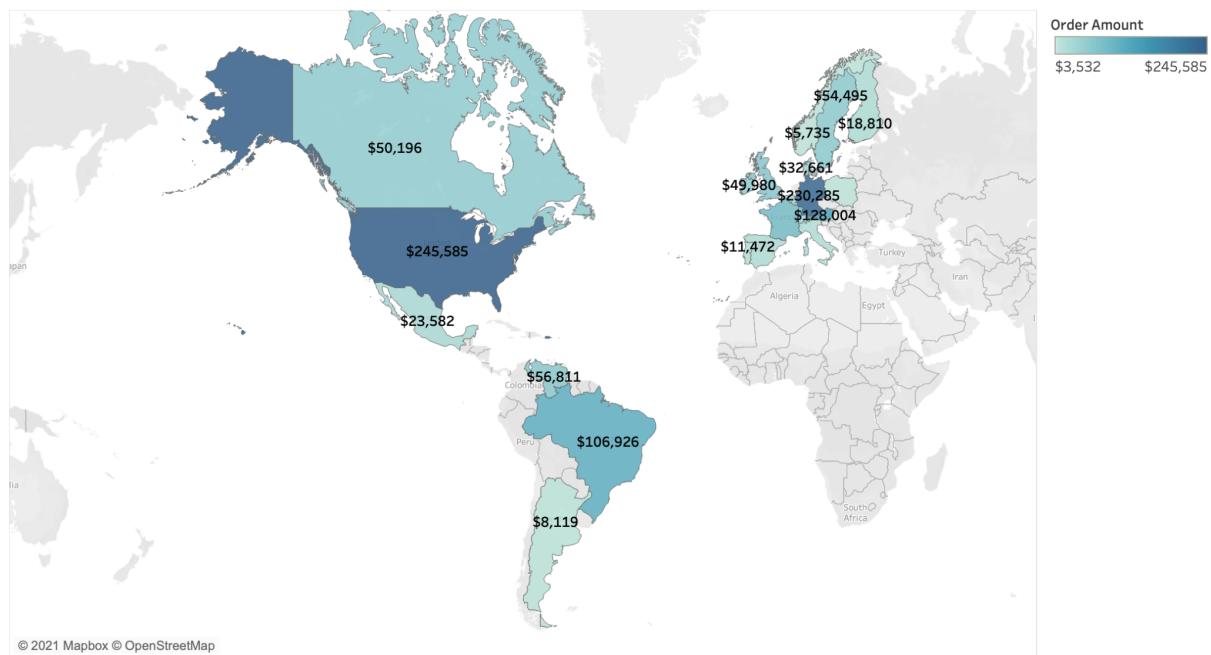


Sum of Order Amount for each Last Name broken down by First Name.

In this graph, we calculate the order value made by each salesperson (here all the salesperson joined the company before 1996 so the time span for all employees is the same). We find that Margaret Peacock is the best salesperson, Janet Leverling is the second and Nancy Davolio is the third. So we can conclude that Margaret Peacock brought most revenue to the company.

This order value generated by the salesperson is important for the company because the management team can see who brings the most benefits to the company, who did the best and who did the worst. With this information, the management team can design new systems to award the best person and encourage the worst person which could make the competition among salespeople more fierce and generate more revenue to the company.

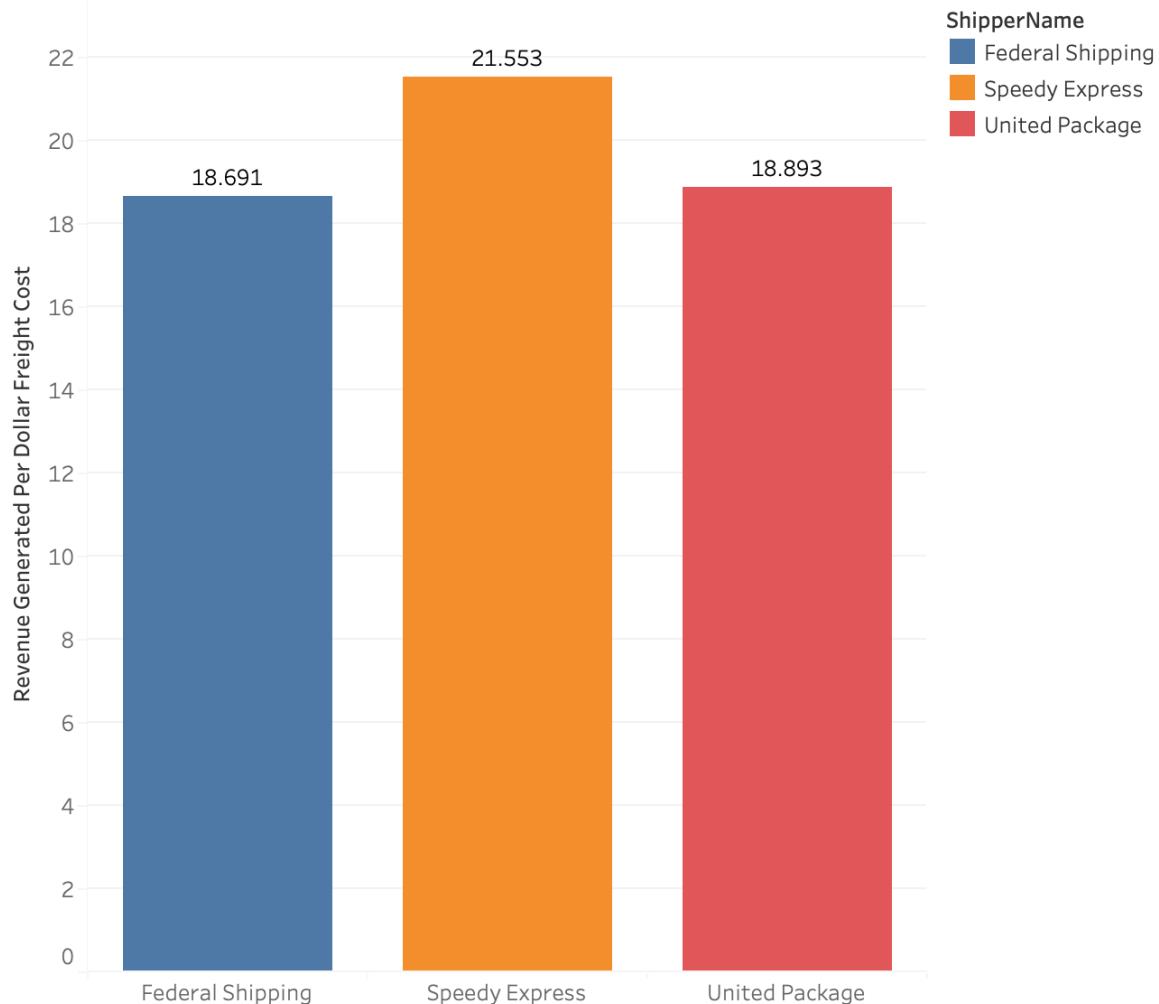
Revenue Distribution



We wanted to also analyze the revenue by shipping country to have a business overview of our revenue break-down by geological dimension. We summed the revenue by ShipCountry.

With the map, from the color scale, it could be easily shown that the United-States (245k\$), Germany (230k\$), and Austria (128k\$) are TOP 3 revenue resources (higher total order value). There are some areas with lighter color, which are also our market but still with potential in the future. The management team could use this information to decide which country could be developed more.

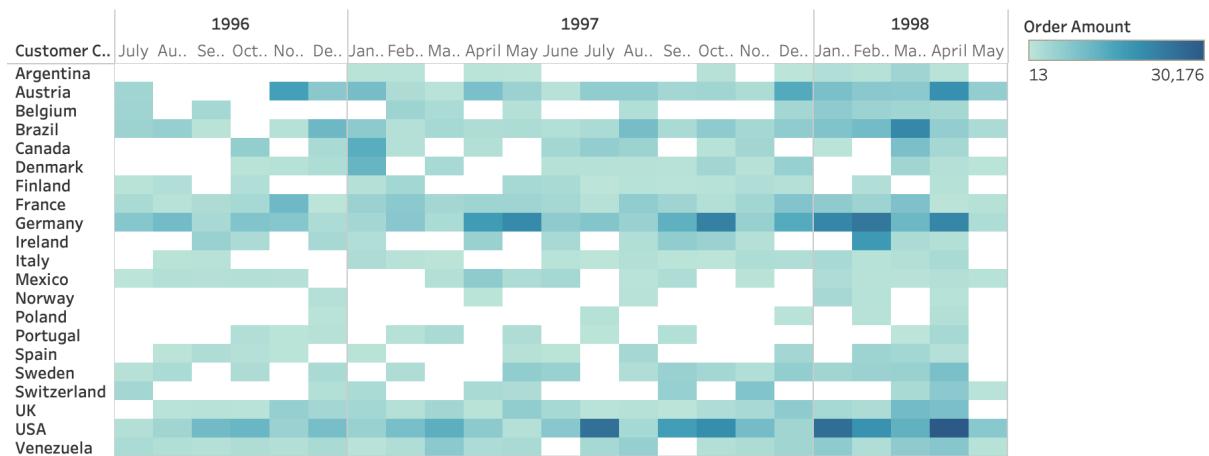
Shipping Analysis



Sum of Revenue Generated Per Dollar Freight Cost for each ShipperName (delayrate).
Color shows details about ShipperName (delayrate). The marks are labeled by sum of Revenue Generated Per Dollar Freight Cost.

We created this chart to show how much revenue each shipping company can generate per freight cost. Freight cost is very meaningful for a company to consider when calculating the total expenditure. So we are trying to figure out how much revenue we can generate per dollar of freight cost to see which shipper company we should use to ship our products. As you can see from this chart, Speedy Express provides the most benefits for us, and we should use them to send our products the most for the following shipments. United Package and Federal shipping companies do not show that much difference to using either of them for the following.

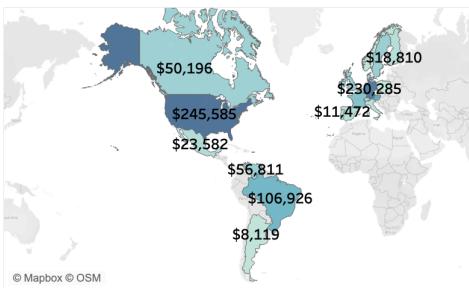
Monthly Order Revenue from each Country



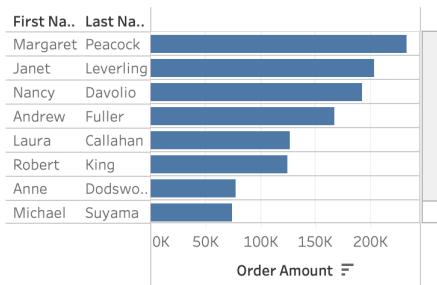
Sum of Order Amount (color) broken down by Order Date Year and Order Date Month vs. Customer Country.

The topic of this heat map is Monthly Order Revenue from each Country. By using this visualization rather than a line chart, we can better show the new markets (white cells) and the main markets (USA, Germany, Brazil, France, UK and Venezuela) at the same time. Company management could make decisions about resource allocation and market exploration based on this graph. For example, in July 1997, decision makers in the company could clearly notice that the USA market performed extraordinarily that year and something bad happened in the Mexico market so the company can learn from experience in the USA and avoid a similar situation in Mexico happening again.

Revenue Distribution



Employee Revenue



Order Amount
13 30,176

(Customer Division) Country

- Argentina
- Austria
- Belgium
- Brazil
- Canada
- Denmark
- Finland
- France
- Germany
- Ireland
- Italy
- Mexico
- Norway

Customer Ranking

- 1
- 2
- 3
- 4
- 5

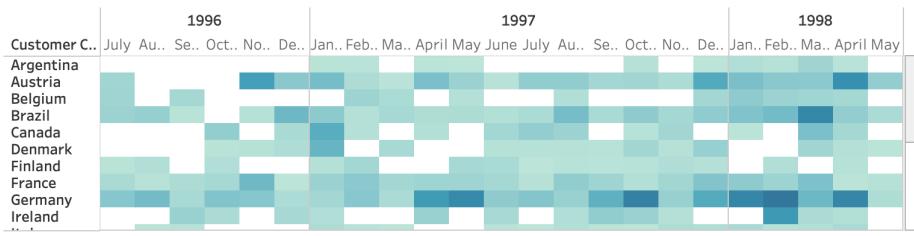
ShipperName

- Federal Shipping
- Speedy Express
- United Package

(Revenue Distribution) Country

- Argentina
- Austria
- Belgium
- Brazil
- Canada
- Denmark
- Finland
- France
- Germany
- Ireland
- Italy
- Mexico
- Norway

Year-month Level Revenue for Each Country in Company



Consumer Division by Revenue



Shipping Analysis

