Orion Star Sports & Outdoors

Orion Star Sports & Outdoors is a fictitious global sports and outdoors retailer with *traditional stores*, an *online store*, and a large *catalog business*.

The corporate headquarters is located in the United States with offices and stores in many countries throughout the world.

Orion Star has about 1,000 employees and 90,000 customers, processes approximately 150,000 orders annually, and purchases products from 64 suppliers.

Orion Star Data

As is the case with most organizations, Orion Star has a large amount of data about its ***customers***, ***suppliers***, ***products***, and ***employees***. Much of this information is stored in transactional systems in various formats.

This transactional information was extracted, transformed, and loaded into a data warehouse.

Data sets were created to meet the needs of specific departments such as Marketing, Revenue Management, or Human Resources.

Potential objectives of an analysis

- Customer relationship management
  - Encourage customer loyalty
  - Study the buying behavior of the customers
    - Occasional or loyal shopper
  - Where to concentrate any marketing efforts
    - Based on past purchases or upcoming birthdays, send out promotional coupons
- Human Resources
  - Which sales employees perform well
  - Select employee of the month and reward based on a function
  - Send notification to those employees who haven't completed mandatory training.
- Purchasing
  - Find out which suppliers offer the best selling products
- Supply Chain – omni-channel operations
  - Which order type (online, retail or catalog) is placed by which customer segment

Potential objectives of an analysis

- Customer relationship management
  - Encourage customer loyalty
  - Study the buying behavior of the customers
    - Occasional or loyal shopper
  - Where to concentrate any marketing efforts
    - Based on past purchases or upcoming birthdays, send out promotional coupons
- Human Resources
  - Which sales employees perform well
  - Select employee of the month and reward based on a function
  - Send notification to those employees who haven’t completed mandatory trainings.
- Purchasing
  - Find out which suppliers offer the best selling products
- Supply Chain – omni-channel operations
  - Which order type (online, retail or catalog) is placed by which customer segment

Potential objectives of an analysis

- Customer relationship management
  - Encourage customer loyalty
  - Study the buying behavior of the customers
    - Occasional or loyal shopper
  - Where to concentrate any marketing efforts
    - Based on past purchases or upcoming birthdays, send out promotional coupons
- Human Resources
  - Which sales employees perform well
  - Select employee of the month and reward based on a function
  - Send notification to those employees who haven't completed mandatory training.
- Purchasing
  - Find out which suppliers offer the best selling products
- Supply Chain – omni-channel operations
  - Which order type (online, retail or catalog) is placed by which customer segment

The Orion Star Company

We will use a subset of Orion Star data including *employees* in the United States and Australia *customers* from Australia, Canada, Germany, Israel, South Africa, the United States, and Turkey.

This is a relational database.

Employee Related Data

Several data sets deal with employees and their information.

- EMPLOYEE_ADDRESSES
- EMPLOYEE_DETAIL
- EMPLOYEE_DONATIONS
- EMPLOYEE_ORGANIZATION
- EMPLOYEE_PAYROLL
- EMPLOYEE_TRAINING
- SALESASSOCIATES
- employee_info.xlsx
- ACTIVEEMPLOYEES
- NEWEMPLOYEES

EMPLOYEE_ADDRESSES (n=424)

#	Variable	Type
6	City	Char
9	Country	Char
1	Employee_ID	Num
2	Employee_Name	Char
8	Postal_Code	Char
7	State	Char
3	Street_ID	Num
5	Street_Name	Char
4	Street_Number	Num

EMPLOYEE_DONATIONS (n=124)

#	Variable	Type	Len
1	Employee_ID	Num	8
7	Paid_By	Char	17
2	Qtr1	Num	8
3	Qtr2	Num	8
4	Qtr3	Num	8
5	Qtr4	Num	8
6	Recipients	Char	65

EMPLOYEE_TRAINING (n=696)

#	Variable	Type	Len
3	Course	Char	25
4	Course_Date	Num	8
6	Department	Char	40
1	Employee_ID	Num	8
5	Job_Title	Char	25
2	Name	Char	40

EMPLOYEE_PAYROLL (n=424)

#	Variable	Type	Len
4	Birth_Date	Num	8
8	Dependents	Num	8
2	Employee_Gender	Char	1
5	Employee_Hire_Date	Num	8
1	Employee_ID	Num	8
6	Employee_Term_Date	Num	8
7	Marital_Status	Char	1
3	Salary	Num	8

EMPLOYEE_DETAIL (n=424)

#	Variable	Type	Len
8	Birth_Date	Num	8
12	City	Char	12
14	Country	Char	2
4	Department	Char	40
6	Employee_Gender	Char	1
9	Employee_Hire_Date	Num	8
1	Employee_ID	Num	8
2	Employee_Name	Char	28
10	Employee_Term_Date	Num	8
3	Job_Title	Char	25
5	Manager_ID	Num	8
11	Marital_Status	Char	1
7	Salary	Num	8
13	State	Char	2

ACTIVEEMPLOYEES (n=308)

#	Variable	Type	Len
14	Birth_Date	Num	8
5	City	Char	30
8	Country	Char	2
10	Department	Char	40
17	Dependents	Num	8
1	Employee_ID	Num	8
12	Gender	Char	1
15	Hire_Date	Num	8
9	Job_Title	Char	25
11	Manager_ID	Num	8
16	Marital_Status	Char	1
2	Name	Char	40
7	Postal_Code	Char	10
13	Salary	Num	8
6	State	Char	2
4	Street_Name	Char	40
3	Street_Number	Num	8

SALESASSOCIATES (n=424)

#	Variable	Type	Len
5	Birth_Date	Num	8
6	Emp_Hire_Date	Num	8
7	Emp_Term_Date	Num	8
1	Employee_ID	Num	8
10	Employee_Name	Char	40
4	Gender	Char	1
2	Job_Title	Char	25
8	Manager_ID	Num	8
9	SSN	Char	16
3	Salary	Num	8

EMPLOYEE_INFO (n=923)

#	Variable	Type	Len
1	Employee_ID	Num	8
3	Phone_Number	Char	15
2	Phone_Type	Char	4

EMPLOYEE_ORGANIZATION (n=424)

#	Variable	Type	Len
3	Department	Char	40
1	Employee_ID	Num	8
2	Job_Title	Char	25
4	Manager_ID	Num	8

EMPLOYEE_ADDRESSES (n=424)

#	Variable	Type
6	City	Char
9	Country	Char
1	Employee_ID	Num
2	Employee_Name	Char
8	Postal_Code	Char
7	State	Char
3	Street_ID	Num
5	Street_Name	Char
4	Street_Number	Num

EMPLOYEE_DONATIONS (n=124)

#	Variable	Type	Len
1	Employee_ID	Num	8
7	Paid_By	Char	17
2	Qtr1	Num	8
3	Qtr2	Num	8
4	Qtr3	Num	8
5	Qtr4	Num	8
6	Recipients	Char	65

EMPLOYEE_TRAINING (n=696)

#	Variable	Type	Len
3	Course	Char	25
4	Course_Date	Num	8
6	Department	Char	40
1	Employee_ID	Num	8
5	Job_Title	Char	25
2	Name	Char	40

EMPLOYEE_PAYROLL (n=424)

#	Variable	Type	Len
4	Birth_Date	Num	8
8	Dependents	Num	8
2	Employee_Gender	Char	1
5	Employee_Hire_Date	Num	8
1	Employee_ID	Num	8
6	Employee_Term_Date	Num	8
7	Marital_Status	Char	1
3	Salary	Num	8

EMPLOYEE_DETAIL (n=424)

#	Variable	Type	Len
8	Birth_Date	Num	8
12	City	Char	12
14	Country	Char	2
4	Department	Char	40
6	Employee_Gender	Char	1
9	Employee_Hire_Date	Num	8
1	Employee_ID	Num	8
2	Employee_Name	Char	28
10	Employee_Term_Date	Num	8
3	Job_Title	Char	25
5	Manager_ID	Num	8
11	Marital_Status	Char	1
7	Salary	Num	8
13	State	Char	2

ACTIVEEMPLOYEES (n=308)

#	Variable	Type	Len
14	Birth_Date	Num	8
5	City	Char	30
8	Country	Char	2
10	Department	Char	40
17	Dependents	Num	8
1	Employee_ID	Num	8
12	Gender	Char	1
15	Hire_Date	Num	8
9	Job_Title	Char	25
11	Manager_ID	Num	8
16	Marital_Status	Char	1
2	Name	Char	40
7	Postal_Code	Char	10
13	Salary	Num	8
6	State	Char	2
4	Street_Name	Char	40
3	Street_Number	Num	8

SALESASSOCIATES (n=424)

#	Variable	Type	Len
5	Birth_Date	Num	8
6	Emp_Hire_Date	Num	8
7	Emp_Term_Date	Num	8
1	Employee_ID	Num	8
10	Employee_Name	Char	40
4	Gender	Char	1
2	Job_Title	Char	25
8	Manager_ID	Num	8
9	SSN	Char	16
3	Salary	Num	8

EMPLOYEE_INFO (n=923)

#	Variable	Type	Len
1	Employee_ID	Num	8
3	Phone_Number	Char	15
2	Phone_Type	Char	4

EMPLOYEE_ORGANIZATION (n=424)

#	Variable	Type	Len
3	Department	Char	40
1	Employee_ID	Num	8
2	Job_Title	Char	25
4	Manager_ID	Num	8

Employee_ID is the primary key for HR data

Customers Related Data

Customer data resides in one file.

- CUSTOMERS

CUSTOMERS (n=77)

#	Variable	Type	Len
11	Customer_Age	Num	8
8	Customer_Age_Group	Char	12
7	Customer_BirthDate	Num	8
2	Customer_Country	Char	2
5	Customer_FirstName	Char	20
3	Customer_Gender	Char	1
10	Customer_Group	Char	40
1	Customer_ID	Num	8
6	Customer_LastName	Char	30
4	Customer_Name	Char	40
9	Customer_Type	Char	40

Customers Related Data

Customer data resides in one file.

- CUSTOMERS

Customer_ID is the primary key for Marketing and Customer Relations data

CUSTOMERS (n=77)

#	Variable	Type	Len
11	Customer_Age	Num	8
8	Customer_Age_Group	Char	12
7	Customer_BirthDate	Num	8
2	Customer_Country	Char	2
5	Customer_FirstName	Char	20
3	Customer_Gender	Char	1
10	Customer_Group	Char	40
1	Customer_ID	Num	8
6	Customer_LastName	Char	30
4	Customer_Name	Char	40
9	Customer_Type	Char	40

Orders, Products and Suppliers Related Data

Few data sets deal with orders.

- ORDERS
- PURCHASED_PRODUCTS
- SHOE_VENDORS
- QUARTERLY_SALES

SHOE_VENDORS (n=361)

#	Variable	Type	Len
10	Category_Name	Char	45
9	Group_Name	Char	45
11	Line_Name	Char	45
12	Mfg_Suggested_Retail_Price	Num	8
2	Product_Category	Num	8
3	Product_Group	Num	8
4	Product_ID	Num	8
1	Product_Line	Num	8
5	Product_Name	Char	45
8	Supplier_Country	Char	2
6	Supplier_ID	Num	4
7	Supplier_Name	Char	30

ORDERS (n=617)

#	Variable	Type	Len
10	CostPrice_Per_Unit	Num	8
2	Customer_ID	Num	8
5	Delivery_Date	Num	8
11	Discount	Num	8
3	Employee_ID	Num	8
4	Order_Date	Num	8
1	Order_ID	Num	8
6	Order_Type	Num	8
7	Product_ID	Num	8
13	Profit	Num	8
8	Quantity	Num	8
12	Shipping	Num	8
9	Total_Retail_Price	Num	8

PURCHASED_PRODUCTS (n=1441)

#	Variable	Type	Len
3	Customer_Age_Group	Char	12
2	Customer_Country	Char	2
1	Customer_ID	Num	8
5	Order_Date	Num	8
4	Order_ID	Num	8
6	Product_ID	Num	8
7	Product_Name	Char	45
8	Total_Retail_Price	Num	8

QUARTERLY_PURCHASES (n=75)

#	Variable	Type	Len
3	Customer_BirthDate	Num	8
1	Customer_ID	Num	8
2	Customer_Name	Char	40
4	Qtr1Purchases	Num	8
5	Qtr2Purchases	Num	8
6	Qtr3Purchases	Num	8
7	Qtr4Purchases	Num	8

SHOE_VENDORS (n=361)

#	Variable	Type	Len
10	Category_Name	Char	45
9	Group_Name	Char	45
11	Line_Name	Char	45
12	Mfg_Suggested_Retail_Price	Num	8
2	Product_Category	Num	8
3	Product_Group	Num	8
4	Product_ID	Num	8
1	Product_Line	Num	8
5	Product_Name	Char	45
8	Supplier_Country	Char	2
6	Supplier_ID	Num	4
7	Supplier_Name	Char	30

ORDERS (n=617)

#	Variable	Type	Len
10	CostPrice_Per_Unit	Num	8
2	Customer_ID	Num	8
5	Delivery_Date	Num	8
11	Discount	Num	8
3	Employee_ID	Num	8
4	Order_Date	Num	8
1	Order_ID	Num	8
6	Order_Type	Num	8
7	Product_ID	Num	8
13	Profit	Num	8
8	Quantity	Num	8
12	Shipping	Num	8
9	Total_Retail_Price	Num	8

PURCHASED_PRODUCTS (n=1441)

#	Variable	Type	Len
3	Customer_Age_Group	Char	12
2	Customer_Country	Char	2
1	Customer_ID	Num	8
5	Order_Date	Num	8
4	Order_ID	Num	8
6	Product_ID	Num	8
7	Product_Name	Char	45
8	Total_Retail_Price	Num	8

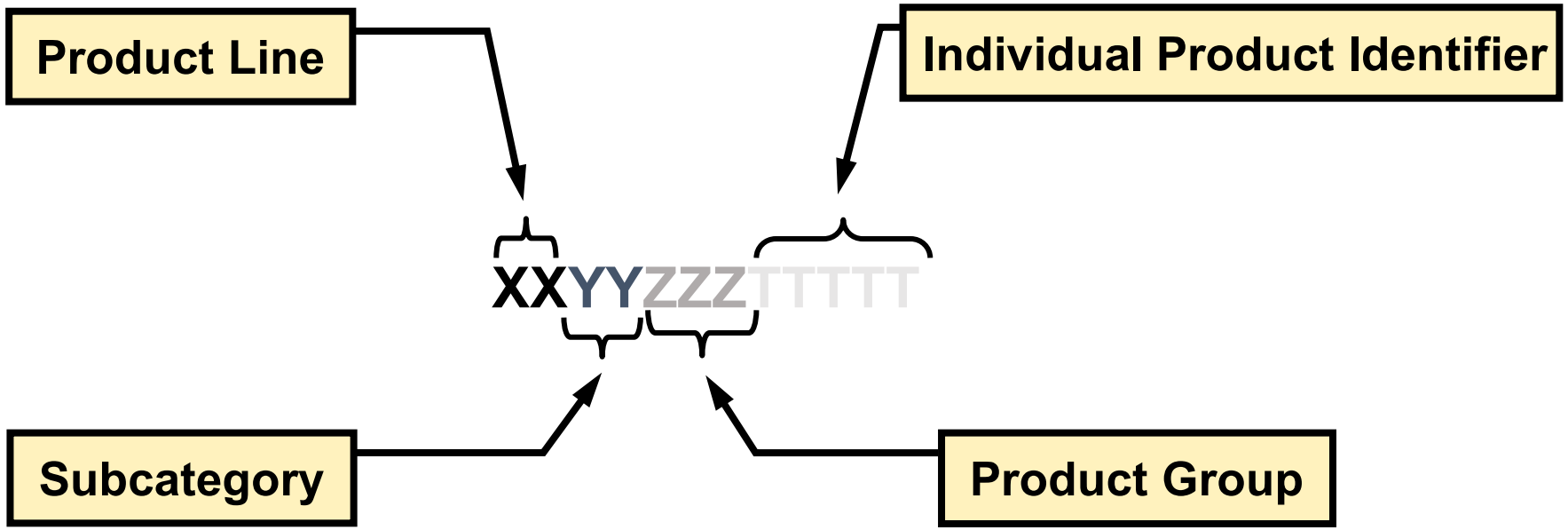
QUARTERLY_PURCHASES (n=75)

#	Variable	Type	Len
3	Customer_BirthDate	Num	8
1	Customer_ID	Num	8
2	Customer_Name	Char	40
4	Qtr1Purchases	Num	8
5	Qtr2Purchases	Num	8
6	Qtr3Purchases	Num	8
7	Qtr4Purchases	Num	8

Order_ID and Product_ID are two key variables for Sales data.

Orion Product ID Codes

Codes are numeric in the form **XXYYZZZTTTTT.**



Orion Product ID Codes

Variable	Label	Value
Line_Name	Product Line Name	Clothes & Shoes
		Children
		Sports

Variable	Label	Value
Category_Name	Category Name	Shoes
		Children Sports
		Clothes
		Golf
		Assorted Sports Articles

Variable	Label	Value
Group_Name	Group Name	Eclipse Shoes
		Tracker Shoes
		Shoes
		Eclipse, Kid's Shoes
		Orion
		Tracker Kid's Clothes
		LSF
		Golf
		Assorted Sports articles
		N.D. Gear, Kids
		Olssons, Kids
		Eclipse Clothing
		Osprey
		Osprey, Kids

Variable	Label	Value	Frequency Count
Product_Name	Product Name	4men Men's Air Golden Shoes	1
		Alexis Women's Classic Shoes	1
		Alexis Women's Classic Shoes Lar	1
		Allinall Star 2000 Women's Canva	1
		Allinall Star Men's Canvas Shoes	1
		Arctic Children's Shoes	1
		Armadillo Men's Running Shoes	1
		Armadillo Road Dmx Men's Running	1
		Armadillo Road Dmx Women's Runni	1
		Armadillo Women's Running Shoes	1
		Atmosphere Acma Men's Running Sh	1
		Atmosphere Deschutz Shoes	1
		Atmosphere Imara Women's Running	1
		Atmosphere Mic Plus.Men's Runnin	1
		Atmosphere Mickey Shoes	1
		Atmosphere Shatter Mid Shoes	1
		Atmosphere Terra Grande Shoes	1
		Atx 7500 Ch Children's Shoes	1
		Banto Light Men's Shoes	1
		Big Guy Men's Air Ace 4 Plus Lo	1
		Big Guy Men's Air Trainer Struc	1
		Big Guy Men's Air 120 Soccer Sho	1
		Big Guy Men's Air 45 Trainer Sho	1
		Big Guy Men's Air Align Shoes	1
		Big Guy Men's Air Aragon Shoes	1
		Big Guy Men's Air Baltoro Shoes	1
		Big Guy Men's Air Cortez Shoes	1
		Big Guy Men's Air Croos Trainer	1
		Big Guy Men's Air Deschutz Shoes	1
		Big Guy Men's Air Deschutz Viii	1
		All other values	331

Supplier

Name	Variable Number	Type
Category_Name	10	Character
Group_Name	9	Character
Line_Name	11	Character
Mfg_Suggested_Retail_Price	12	Numeric
Product_Category	2	Numeric
Product_Group	3	Numeric
Product_ID	4	Numeric
Product_Line	1	Numeric
Product_Name	5	Character
Supplier_Country	8	Character
Supplier_ID	6	Numeric
Supplier_Name	7	Character

Variable	Label	Value
Supplier_Country	Country	US
		CA
		GB
		SE
		ES
		NL
		DK

Variable	Label	Value
Supplier_Name	Supplier Name	Eclipse Inc
		3Top Sports
		Fuller Trading Co.
		Pro Sportswear Inc
		Greenline Sports Ltd
		Petterson AB
		Twain Inc
		Ultra Sporting Goods Inc
		Rubby Zapatos S.A.
		Van Dammeren International
		Sportico
		Triple Sportswear Inc
		Teamsports Inc
		CrystalClear Optics Inc
		Luna sastreria S.A.
		Roll-Over Inc
		Top Sports

The **ORION.Orders** table contains the key identifier columns **Employee_ID**, **Customer_ID**, **Order_ID**, and **Product_ID**, and these can be used to relate data from the different groups to each other.

ORDERS (n=617)

#	Variable	Type	Len
10	CostPrice_Per_Unit	Num	8
2	Customer_ID	Num	8
5	Delivery_Date	Num	8
11	Discount	Num	8
3	Employee_ID	Num	8
4	Order_Date	Num	8
1	Order_ID	Num	8
6	Order_Type	Num	8
7	Product_ID	Num	8
13	Profit	Num	8
8	Quantity	Num	8
12	Shipping	Num	8
9	Total_Retail_Price	Num	8

To identify the name of an employee who made any sale to a customer in South Africa in 2007, you need information from three different tables: **Employee_Name** from **ORION.EMPLOYEE_ADDRESSES, Order_Date** from **ORION.ORDERS, and Country** from **ORION.CUSTOMERS**

CUSTOMERS (n=77)

#	Variable	Type	Len
11	Customer_Age	Num	8
8	Customer_Age_Group	Char	12
7	Customer_BirthDate	Num	8
2	Customer_Country	Char	2
5	Customer_FirstName	Char	20
3	Customer_Gender	Char	1
10	Customer_Group	Char	40
1	Customer_ID	Num	8
6	Customer_LastName	Char	30
4	Customer_Name	Char	40
9	Customer_Type	Char	40

ORDERS (n=617)

#	Variable	Type	Len
10	CostPrice_Per_Unit	Num	8
2	Customer_ID	Num	8
5	Delivery_Date	Num	8
11	Discount	Num	8
3	Employee_ID	Num	8
4	Order_Date	Num	8
1	Order_ID	Num	8
6	Order_Type	Num	8
7	Product_ID	Num	8
13	Profit	Num	8
8	Quantity	Num	8
12	Shipping	Num	8
9	Total_Retail_Price	Num	8

EMPLOYEE_ADDRESSES (n=424)

#	Variable	Type
6	City	Char
9	Country	Char
1	Employee_ID	Num
2	Employee_Name	Char
8	Postal_Code	Char
7	State	Char
3	Street_ID	Num
5	Street_Name	Char
4	Street_Number	Num

COUNTRY_REGION_LOOKUP Table

country_code	country_name	region
ZA	South Africa	Africa
IL	Israel	Asia/Pacific
TR	Turkey	Asia/Pacific
AU	Australia	Asia/Pacific
DE	Germany	Europe
CA	Canada	North America
US	United States	North America

How to discover the table contents and common variables?

- Browse Tasks > Data > Data Set Attributes
- Click on Add and locate ORION datasets such as:
  EMPLOYEE_ADDRESSES,
  EMPLOYEE_DETAIL,EMPLOYEE_DONATIONS,
  EMPLOYEE_ORGANIZATION, EMPLOYEE_PAYROLL,
  EMPLOYEE_TRAINING, SALESASSOCIATES,
  ACTIVEEMPLOYEES, NEWEMPLOYEES, CUSTOMERS, ORDERS,
  PURCHASED_PRODUCTS, SHOE_VENDORS, QUARTERLY_SALES,
  COUNTRY_REGION_LOOKUP ETC.
- Hold ctrl key and select all dataset you want to analyze.
- Click Finish (It might take a few minutes time to run)
- Observe that datasets are added on Process Flow and are linked to the Data Set Attributes task.
- Go to Data Set Attributes tab and scroll down to see variables in each table.

How to discover the table contents and common variables?

Alternatively run the following code:

```
PROC SQL;
CREATE TABLE OrionDataDict AS
SELECT v.name, v.memname, v.type, v.label, v.format
FROM DICTIONARY.COLUMNS AS v
WHERE v.memname IN ("EMPLOYEE_ADDRESSES", "EMPLOYEE_DETAIL","EMPLOYEE_DONATIONS",
                    "EMPLOYEE_ORGANIZATION", "EMPLOYEE_PAYROLL",
                    "EMPLOYEE_TRAINING",
                    "SALESASSOCIATES", "ACTIVEEMPLOYEES", "NEWEMPLOYEES",
                    "CUSTOMERS",
                    "ORDERS", "PURCHASED_PRODUCTS", "SHOE_VENDORS",
                    "QUARTERLY_SALES",
                    "COUNTRY_REGION_LOOKUP")
ORDER BY v.name, v.memname;
/*Order by name first if you want to see the common variables among data tables,*/
/*Order by memname first if you want to see all the variables in each data table.*/
QUIT;
```