

HW 5 Data Warehouse Analysis

Jackson Lee

1. Use the information_schema to find out how many rows there are in each table in the adventureworks data warehouse. Show the table name and its row count.

```

7 SELECT TABLE_SCHEMA, TABLE_NAME, TABLE_ROWS FROM TABLES
8 WHERE TABLE_SCHEMA = 'aw';

```

100% 1:6

Result Grid Filter Rows: Search Export:

TABLE_SCHEMA	TABLE_NAME	TABLE_ROWS
aw	DimAccount	99
aw	DimCurrency	0
aw	DimCustomer	18304
aw	DimDepartmentGroup	7
aw	DimEmployee	296
aw	DimGeography	655
aw	DimOrganization	14
aw	DimProduct	158
aw	DimProductCategory	4
aw	DimProductSubcategory	37
aw	DimPromotion	16
aw	DimReseller	701
aw	DimSalesReason	10
aw	DimSalesTerritory	11
aw	DimScenario	3
aw	DimTime	1158
aw	FactCurrencyRate	0
aw	FactFinance	38480
aw	FactInternetSales	59800

2. Use the information_schema to list out each table in the adventureworks data warehouse and its primary key.

```

10 SELECT TABLE_SCHEMA, TABLE_NAME, COLUMN_NAME, COLUMN_KEY
11 FROM COLUMNS
12 WHERE TABLE_SCHEMA = 'aw' AND COLUMN_KEY = 'PRI';
13

```

100% 1:22

Result Grid Filter Rows: Search Export:

TABLE_SCHEMA	TABLE_NAME	COLUMN_NAME	COLUMN_KEY
aw	DimAccount	AccountKey	PRI
aw	DimCurrency	CurrencyKey	PRI
aw	DimCustomer	CustomerKey	PRI
aw	DimDepartmentGroup	DepartmentGroupKey	PRI
aw	DimEmployee	EmployeeKey	PRI
aw	DimGeography	GeographyKey	PRI
aw	DimOrganization	OrganizationKey	PRI
aw	DimProduct	ProductKey	PRI
aw	DimProductCategory	ProductCategoryKey	PRI
aw	DimProductSubcategory	ProductSubcategoryKey	PRI
aw	DimPromotion	PromotionKey	PRI
aw	DimReseller	ResellerKey	PRI
aw	DimSalesReason	SalesReasonKey	PRI
aw	DimSalesTerritory	SalesTerritoryKey	PRI
aw	DimScenario	ScenarioKey	PRI
aw	DimTime	TimeKey	PRI
aw	FactInternetSales	SalesOrderNumber	PRI
aw	FactInternetSales	SalesOrderLineNumber	PRI

3. What standard table naming convention did the AdventureWorksDW database designers use to differentiate dimension tables from fact tables in this star schema data warehouse
 1. To name the dimension tables they start each table with “Dim”. To differentiate the fact tables they start each table name with “Fact”.
4. What do you think is the purpose of the recursive relationship on DimEmployee?
 1. The purpose of the recursive relationship is to allow for employee hierarchy that exists in the AdventureWorks business. At their business each employee can have a boss and/or a subordinate beneath them so a self referencing relationship is needed.
5. What are the three types of models of bikes sold by AdventureWorks?
 1. AdventureWorks sells Mountain, Road and Touring bikes.
6. Of these three, which type of bike model had the highest sales (in dollar volume) in 2003?

```

37 SELECT DTIM.CalendarYear, DPRO.ProductSubcategoryKey, SUM(FSAL.UnitPrice) AS TotalSales
38 FROM FactInternetSales AS FSAL JOIN DimProduct AS DPRO JOIN DimTime AS DTIM ON
39 FSAL.ProductKey=DPRO.ProductKey AND FSAL.OrderDateKey=DTIM.TimeKey
40 WHERE DTIM.CalendarYear='2003' AND
41 (DPRO.ProductSubcategoryKey=1 OR DPRO.ProductSubcategoryKey=2 OR DPRO.ProductSubcategoryKey=3)
42 GROUP BY DTIM.CalendarYear, DPRO.ProductSubcategoryKey
43 ORDER BY TotalSales DESC;
44

```

100% 69:41

Result Grid Filter Rows: Search Export:

CalendarYear	ProductSubcategoryKey	TotalSales
2003	1	3989373.00
2003	2	3951673.00
2003	3	1417351.00

1. AdventureWorks Mountain Bike had the highest sales in 2003
7. List five of the other non-bike products sold by AdventureWorks. (Pick any five.)

```

22 SELECT ProductKey, EnglishProductName, Modelname FROM DimProduct
23 WHERE ProductKey = 214 OR ProductKey = 225 OR ProductKey = 228 OR ProductKey = 473 OR ProductKey = 489;

```

100% 26:20

Result Grid Filter Rows: Search Edit: Export/Import:

ProductKey	EnglishProductName	Modelname
214	Sport-100 Helmet, Red	Sport-100
225	AWC Logo Cap	Cycling Cap
228	Long-Sleeve Logo Jersey, S	Long-Sleeve Logo Jersey
473	Classic Vest, L	Classic Vest
489	Short-Sleeve Classic Jersey, M	Short-Sleeve Classic Jersey

8. Compare and rank the total counts of the bikes sold by AdventureWorks for each of the years 2001 – 2004 by color. What was the most popular color of bikes sold in each of these 4 years? Provide your SQL query, and your answer set along with your answer to the question. You can assume that one row in the fact table equals one sale.
HINT: Since the fact table contains sales for all kinds of products, you should include only fact rows where the sale is for a bike. One easy way to do this is a WHERE clause selecting

only rows where EnglishProductSubcategoryName contains the string “bikes”. There are other ways to determine this as well.

```

34 SELECT DTIM.CalendarYear, DPRO.COLOR, SUM(FSAL.OrderQuantity) as TotalSold
35 FROM FactInternetSales AS FSAL JOIN DimProduct AS DPRO JOIN DimTime AS DTIM ON FSAL.ProductKey=DPRO.ProductKey
36 AND FSAL.OrderDateKey=DTIM.TimeKey
37 WHERE (DTIM.CalendarYear='2001' OR DTIM.CalendarYear='2002' OR DTIM.CalendarYear='2003' OR DTIM.CalendarYear='2004')
38 AND (DPRO.ProductSubcategoryKey=1 OR DPRO.ProductSubcategoryKey=2 OR DPRO.ProductSubcategoryKey=3)
39 GROUP BY DTIM.CalendarYear, DPRO.COLOR
40 ORDER BY DTIM.CalendarYear, TotalSold DESC;
41

```

100% 57:31

Result Grid Filter Rows: Search Export:

CalendarYear	COLOR	TotalSold
2001	Red	775
2001	Black	154
2001	Silver	84
2002	Red	1380
2002	Black	868
2002	Silver	283
2002	Yellow	146
2003	Black	2321
2003	Yellow	1268
2003	Silver	1119
2003	Red	501
2003	Blue	501
2004	Black	1966
2004	Yellow	1789
2004	Silver	1205
2004	Blue	782
2004	Red	63

1. In 2001 and 2002 the most popular color was red. In 2003 and 2004, black was the most popular.
9. 9. AdventureWorks tracks the level of education of each of their customers. Which model of bike is the highest seller among customers with graduate degrees? You can assume that one row in the fact table equals one sale.

```

44 SELECT DPRO.ProductSubcategoryKey, DCUS.EnglishEducation, SUM(FSAL.OrderQuantity) AS TotalSold
45 FROM FactInternetSales AS FSAL JOIN DimProduct AS DPRO JOIN DimCustomer AS DCUS
46 ON FSAL.ProductKey=DPRO.ProductKey AND FSAL.CustomerKey = DCUS.CustomerKey
47 WHERE DCUS.EnglishEducation = 'Graduate Degree' AND
48 (DPRO.ProductSubcategoryKey=1 OR DPRO.ProductSubcategoryKey=2 OR DPRO.ProductSubcategoryKey=3)
49 GROUP BY DPRO.ProductSubcategoryKey
50 ORDER BY TotalSold DESC;
51

```

100% 41:39

Result Grid Filter Rows: Search Export:

ProductSubcategoryKey	EnglishEducation	TotalSold
2	Graduate Degree	1334
1	Graduate Degree	953
3	Graduate Degree	433

Road Bikes are the highest selling type of bike among customers with graduate degrees.

10. 10. For the year 2004, which State/Province yielded the highest margin for AdventureWorks? (HINT: use the customer's State/Province.) Provide your SQL query, and your answer set along with your answer to the question.

California yielded the highest margin for AdventureWorks

```

55 SELECT DTIM.CalendarYear, DGEO.StateProvinceName, SUM(FSAL.UnitPrice - DPRO.StandardCost) AS Margin
56 FROM FactInternetSales as FSAL JOIN DimProduct AS DPRO JOIN DimTime AS DTIM JOIN DimGeography AS DGEO JOIN DimCustomer AS DCUS
57 ON FSAL.ProductKey=DPRO.ProductKey
58 AND FSAL.OrderDateKey=DTIM.TimeKey
59 AND FSAL.CustomerKey = DCUS.CustomerKey
60 AND DCUS.GeographyKey = DGEO.GeographyKey
61 WHERE DTIM.CalendarYear = '2004'
62 GROUP BY DTIM.CalendarYear, DGEO.StateProvinceName
63 ORDER BY Margin DESC;

```

CalendarYear	StateProvinceName	Margin	
2004	California	847226.00	
2004	England	499735.00	
2004	New South Wales	464461.00	
2004	Washington	373392.00	
2004	British Columbia	288089.00	
2004	Victoria	247257.00	
2004	Queensland	230767.00	
2004	Oregon	170851.00	
2004	Saarland	115289.00	
2004	Hessen	103598.00	
2004	Nordrhein-Westfalen	93326.00	
2004	Seine (Paris)	72431.00	
2004	Hamburg	68368.00	
2004	South Australia	67413.00	
2004	Bayern	59083.00	
2004	Seine Saint Denis	55086.00	
2004	Nord	52012.00	
2004	Yveline	46509.00	
2004	Hauts de Seine	39594.00	
2004	Essonne	38227.00	
2004	Tasmania	20852.00	
2004	Seine et Marne	16647.00	
2004	Moselle	15486.00	
2004	Loiret	10585.00	
2004	Brandenburg	8631.00	
2004	Garonne (Haute)	8372.00	
2004	Val d'Oise	7904.00	
2004	Charente-Maritime	5078.00	
2004	Somme	4658.00	
2004	Val de Marne	4615.00	
2004	Alberta	2448.00	
2004	Pas de Calais	2380.00	
2004	Loir et Cher	2363.00	
2004	Florida	2064.00	
2004	South Carolina	1105.00	
2004	New York	1097.00	
2004	Wyoming	443.00	
2004	Texas	438.00	
2004	Georgia	378.00	
2004	Ohio	107.00	
2004	Illinois	88.00	
2004	Kentucky	61.00	
2004	Minnesota	35.00	
2004	Mississippi	33.00	
2004	Virginia	25.00	
2004	Alabama	22.00	