

## Hw5

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.

**Query:**

```
use information_schema;
```

```
select TABLE_Name, TABLE_ROWS
      from TABLES
      where TABLE_SCHEMA like 'aw';
```

**Answer Set:**

TABLE_NAME	TABLE_ROWS
DimAccount	99
DimCurrency	0
DimCustomer	18356
DimDepartmentGroup	7
DimEmployee	296
DimGeography	655
DimOrganization	14
DimProduct	158
DimProductCategory	4
DimProductSubcategory	37
DimPromotion	16
DimReseller	701
DimSalesReason	10
DimSalesTerritory	11
DimScenario	3
DimTime	1158
FactCurrencyRate	0
FactFinance	38480
FactInternetSales	59800

2. Use the *information\_schema* to list out each table in the *adventureworks data warehouse* and its primary key.

**Query:**

```
use information_schema;
```

```
select TABLE_NAME, COLUMN_NAME, COLUMN_KEY
      from COLUMNS
      where TABLE_SCHEMA like 'aw'
      and COLUMN_KEY like 'PRI';
```

**Answer Set:**

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

**For all dimensional tables start with the prefix “Dim”, and every fact table starts with the prefix “Fact”.**

4. What do you think is the purpose of the recursive relationship on DimEmployee?

**I think the purpose of the recursive relationship on DimEmployee, is a way of tracking who manages who. This is due to the facts there is foreign key called ParentEmployeeKey for each employee. We can make a sanity check of my guess by noting the recursive relationship is a one to many which would work as a manager can have many employees.**

5. What are the three types of models of bikes sold by AdventureWorks?

**Query:**

*use aw;*

```
select EnglishProductSubcategoryName
      from DimProductSubcategory
      where EnglishProductSubcategoryName like '%bikes%';
```

**Answer Set:**

	EnglishProductSubcategoryName
▶	Mountain Bikes
	Road Bikes
	Touring Bikes

**Answer:**

*I found this using a sql query, that the three models are mountain, road, and touring bikes.*

6. 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.

**Query:**

```

select YEAR(DT.FullDateAlternateKey) as Year, count(FI.ProductKey) as Sales, totalSales.Sales as
TotalSales, DP.Color, count(DP.Color) as colorCount
    From FactInternetSales FI, DimProduct DP, DimProductSubcategory DPS, DimTime DT,
        (select YEAR(DT.FullDateAlternateKey) as Year, count(FI.ProductKey) as Sales
        From FactInternetSales FI, DimProduct DP, DimProductSubcategory
        DPS, DimTime DT
        WHERE DP.ProductKey = FI.ProductKey
        AND DP.ProductSubcategoryKey = DPS.ProductSubcategoryKey
        AND FI.OrderDateKey = DT.TimeKey
        AND DPS.EnglishProductSubcategoryName like '%bikes%'
        Group by YEAR(DT.FullDateAlternateKey) ) AS totalSales
    WHERE DP.ProductKey = FI.ProductKey
    AND DP.ProductSubcategoryKey = DPS.ProductSubcategoryKey
    AND FI.OrderDateKey = DT.TimeKey
    AND YEAR(DT.FullDateAlternateKey) = totalSales.Year
    AND DPS.EnglishProductSubcategoryName like '%bikes%'
    Group by YEAR(DT.FullDateAlternateKey), DP.Color, totalSales.Sales
    Order By totalSales.Sales desc;

```

**Answer Set:**

Year	Sales	TotalSales	Color	colorCount
2004	782	5805	Blue	782
2004	1789	5805	Yellow	1789
2004	63	5805	Red	63
2004	1966	5805	Black	1966
2004	1205	5805	Silver	1205
2003	1119	5710	Silver	1119
2003	501	5710	Blue	501
2003	1268	5710	Yellow	1268
2003	2321	5710	Black	2321
2003	501	5710	Red	501
2002	283	2677	Silver	283
2002	146	2677	Yellow	146
2002	868	2677	Black	868
2002	1380	2677	Red	1380
2001	775	1013	Red	775
2001	84	1013	Silver	84
2001	154	1013	Black	154

**Answer:** As you can see in total sales each year AdventureWorks increased in sales. The most popular color of bikes sold in 2001 is Red. The most popular color of bikes sold in 2002 is Red. The most popular color of bikes sold in 2003 is Black. The most popular color of bikes sold in 2004 is Black.

7. List and compare the total sales quantities of bikes sold (all model types) by customer gender by year and month. In which year and month were bike sales to females the highest?

**Query:**

```
select YEAR(DT.FullDateAlternateKey) as Year, month(DT.FullDateAlternateKey) as
Month, DC.Gender, count(FI.ProductKey) as Sales
    From FactInternetSales FI, DimProduct DP, DimProductSubcategory
DPS, DimTime DT, DimCustomer DC
    WHERE DP.ProductKey = FI.ProductKey
    AND DP.ProductSubcategoryKey = DPS.ProductSubcategoryKey
    AND FI.OrderDateKey = DT.TimeKey
    AND FI.CustomerKey = DC.CustomerKey
    AND DPS.EnglishProductSubcategoryName like '%bikes%'
    Group by DC.Gender, YEAR(DT.FullDateAlternateKey),
month(DT.FullDateAlternateKey)
    Order By Year(DT.FullDateAlternateKey), Month(DT.FullDateAlternateKey) desc;
```

**Answer Set:**

Year	Month	Gender	Sales
2001	12	M	120
2001	12	F	115
2001	11	F	81
2001	11	M	87
2001	10	F	74
2001	10	M	87
2001	9	M	77
2001	9	F	69
2001	8	M	62
2001	8	F	94
2001	7	F	67
2001	7	M	79
2002	12	M	160
2002	12	F	170
2002	11	F	99
2002	11	M	94
2002	10	F	117
2002	10	M	112
2002	9	F	110
2002	9	M	88
2002	8	M	129
2002	8	F	152
2002	7	M	130
2002	7	F	123
2002	6	M	108
2002	6	F	106
2002	5	M	96
2002	5	F	118
2002	4	M	100
2002	4	F	107
2002	3	M	105
2002	3	F	94
2002	2	F	80
2002	2	M	91
2002	1	F	98
2002	1	M	90
2003	12	M	528
2003	12	F	530
2003	11	M	330
2003	11	F	368
2003	10	M	281
2003	10	F	341
2003	9	F	284
2003	9	M	301
2003	8	F	241
2003	8	M	254
2003	7	M	252
2003	7	F	261
2003	6	F	174
2003	6	M	147
2003	5	M	172
2003	5	F	163
2003	4	M	153
2003	4	F	141
2003	3	F	129
2003	3	M	143
2003	2	F	141
2003	2	M	131
2003	1	F	110
2003	1	M	134
2004	6	M	597
2004	6	F	589
2004	5	F	583
2004	5	M	554
2004	4	M	462
2004	4	F	513
2004	3	M	453
2004	3	F	428
2004	2	F	410
2004	2	M	437
2004	1	F	400
2004	1	M	379

**Answer:**

In 2004, month 6 (June) Bike Sales For Females was highest with 589 sales.

8. *For the year 2003, which model of bike yielded the highest margin for AdventureWorks? Provide your SQL query, and your answer set along with your answer to the question.*

**Query:**

```
select DPS.EnglishProductSubcategoryName as Model, Year(DT.FullDateAlternateKey) as Year,
sum(FI.UnitPrice - FI.ProductStandardCost) as Margin
  From FactInternetSales FI, DimProduct DP, DimProductSubcategory DPS,DimTime DT
 WHERE DP.ProductKey = FI.ProductKey
 AND DP.ProductSubcategoryKey = DPS.ProductSubcategoryKey
 AND FI.OrderDateKey = DT.TimeKey
 AND DPS.EnglishProductSubcategoryName like '%bikes%'
 AND YEAR(DT.FullDateAlternateKey) = '2003'
 Group by DPS.EnglishProductSubcategoryName,Year(DT.FullDateAlternateKey)
 Order by Margin desc
 Limit 1;
```

**Answer Set:**

	Model	Year	Margin
►	Mountain Bikes	2003	1820800.00