

Homework 4

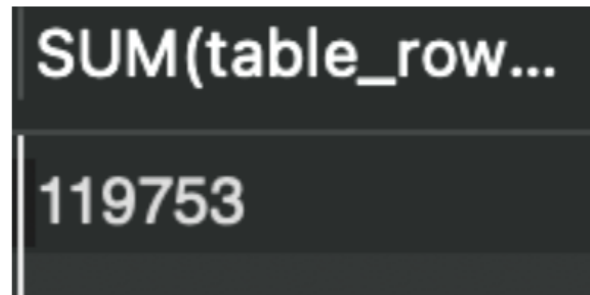
Adrian Ornelas Ruvalcaba, November-13-2022 , Partner: Jonathan Goins

Question 1

Question 1.1

```
use information_schema;

SELECT SUM(table_rows)
  FROM information_schema.tables
 WHERE table_schema = 'aw';
```

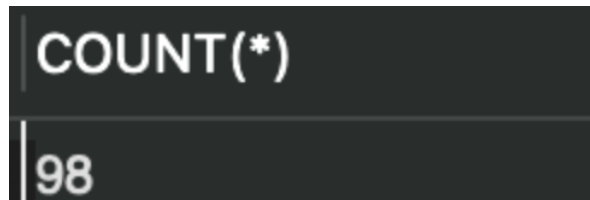


A terminal window with a dark background. The top line shows the command 'SUM(table_row...' and the bottom line shows the result '119753'.

Question 1.2

```
use information_schema;

SELECT COUNT(*)
  FROM information_schema.tables;
```



A terminal window with a dark background. The top line shows the command 'COUNT(*)' and the bottom line shows the result '98'.

Question 1.3

Using a manual count actually counts how many rows are present in each table while schema uses an approximation while taking into account recent changes like delete & insert, resulting in a bad approximation.

Question 1.4

The SELECT COUNT (*) is less effective because it goes through each row manually counting the number of rows. While schema doesn't enter the rows at all but rather takes an estimate. So it is less effective to use information_schema but is less computationally heavy.

Question 2

```
use information_schema;

SELECT DISTINCT table_name, column_name
  FROM information_schema.COLUMNS
 WHERE table_schema = 'aw'
    AND column_key = 'PRI';
```

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

Question 3

All tables are named using PascalCase for readability. Dimension tables are named such that they start with “dim” followed by the rest of the table name. The names of fact tables are preceded by “Fact”.

Question 4

The purpose of the recession relation in the columns of VacationHours on Phone is used to represent that the phone numbers are available if vacation hours are valid.

Question 5

```
use aw;

SELECT EnglishProductSubcategoryName
FROM DimProductSubcategory
WHERE ProductCategoryKey = 1;
```

	EnglishProductSubcategoryName
►	Mountain Bikes
	Road Bikes
	Touring Bikes

Result: The three types of bikes are: Mountain, Road, & Touring Bikes.

Question 6

```
SELECT DimProductSubcategory.EnglishProductSubcategoryName,
       ProfitVolume.DollarVolumeOfUnit
FROM (SELECT SUM(ProductsUnitPrice.unitprice) AS DollarVolumeOfUnit,
             DimProduct.ProductSubcategoryKey AS ProductSubcategoryKey
FROM (SELECT ProductKey, UnitPrice
FROM FactInternetSales
RIGHT JOIN DimTime ON FactInternetSales.OrderDateKey
WHERE FullDateAlternateKey BETWEEN '2004-01-01'AND'2004-12-31')
      AS ProductsUnitPrice
JOIN DimProduct ON ProductsUnitPrice.ProductKey = DimProduct.ProductKey
WHERE DimProduct.ProductSubcategoryKey = 1 ||
      DimProduct.ProductSubcategoryKey = 2 ||
      DimProduct.ProductSubcategoryKey = 3
GROUP BY DimProduct.ProductSubcategoryKey) AS ProfitVolume
JOIN DimProductSubcategory ON
DimProductSubcategory.ProductSubcategoryKey = ProfitVolume.ProductSubcategoryKey;
```

Volume	BikeType
2428349976.00	Mountain Bikes
3542742872.00	Road Bikes
938077520.00	Touring Bikes

Result: It looks like Touring Bikes were the least sold.

Question 7

```
SELECT EnglishProductSubcategoryName
FROM DimProductSubcategory
WHERE ProductCategoryKey != 1;
```

EnglishProductSubcategoryName
► Handlebars
Bottom Brackets
Brakes
Chains
Cranksets
Derailleurs
Forks
Headsets
Mountain Frames
Pedals
Road Frames
Saddles
Touring Frames
Wheels
Bib-Shorts
Caps
Gloves
Jerseys
Shorts
Socks
Tights
Vests
Bike Racks
Bike Stands
Bottles and Cages
Cleaners
Fenders
Helmets
Hydration Packs
Lights
Locks
Panniers
Pumps
Tires and Tubes

Result: The 6 non-bike items are: Forks, Socks, Saddles, Jerseys, Headsets, & Panniers.

Question 8

```
-- Option 2
SELECT COUNT(*), DimProduct.Color
FROM FactInternetSales
JOIN DimProduct on DimProduct.ProductKey = FactInternetSales.ProductKey
JOIN DimProductSubcategory on
    DimProductSubcategory.ProductSubcategoryKey = DimProduct.ProductSubcategoryKey
RIGHT JOIN DimTime on FactInternetSales.OrderDateKey
WHERE CalendarYear = 2002
    AND DimProductSubcategory.EnglishProductSubcategoryName LIKE '%Bikes%'
Group by DimProduct.Color
ORDER BY DimProduct.Color ASC;
```

```
-- Option 2
SELECT COUNT(*), DimProduct.Color
FROM FactInternetSales
JOIN DimProduct on DimProduct.ProductKey = FactInternetSales.ProductKey
JOIN DimProductSubcategory on
    DimProductSubcategory.ProductSubcategoryKey = DimProduct.ProductSubcategoryKey
RIGHT JOIN DimTime on FactInternetSales.OrderDateKey
WHERE CalendarYear BETWEEN '2001' AND '2004'
    AND DimProductSubcategory.EnglishProductSubcategoryName LIKE '%Bikes%'
Group by DimProduct.Color
ORDER BY DimProduct.Color ASC;
```

COUNT(*)	Color
1295396	Black
313052	Blue
663436	Red
656604	Silver
781532	Yellow

Result: Black was the most popular color.

Question 9

```
SELECT DimCustomer.Gender, sum(Sumie.OrderQuantity), Sumie.DayNumberOfMonth, Sumie.CalendarYear
FROM (SELECT FactInternetSales.OrderQuantity, FactInternetSales.CustomerKey, DimTime.DayNumberOfMonth, DimTime.CalendarYear
      FROM FactInternetSales
      JOIN DimProduct ON FactInternetSales.ProductKey = DimProduct.ProductKey
      JOIN DimTime ON FactInternetSales.OrderDateKey = DimTime.TimeKey
      WHERE DimProduct.ProductSubcategoryKey in (1,2,3)
      ) AS Sumie
```

```
JOIN DimCustomer ON Sumie.CustomerKey = DimCustomer.CustomerKey
GROUP BY DimCustomer.Gender, Sumie.DayNumberOfMonth, Sumie.CalendarYear
ORDER BY sum(Sumie.OrderQuantity) DESC;
```

M	45	34	10002
M	45	25	10002
M	44	14	10002
M	44	10	10002
M	44	16	10002
M	43	10	10002
M	43	22	10002
F	43	25	10002
M	42	10	10002
M	42	22	10002
M	42	25	10002
M	41	11	10002
F	41	4	10002
M	41	30	10002
M	40	9	10002
M	40	25	10002
M	40	26	10002
M	39	4	10002
F	39	5	10002
F	39	30	10002
F	39	30	10002
F	39	8	10002
F	39	7	10002
F	39	5	10002
F	39	5	10002
M	39	1	10002
M	39	1	10002
M	39	10	10002
M	39	10	10002
M	39	16	10002
F	39	16	10002
F	39	26	10002
M	39	26	10002
M	38	8	10001
M	38	8	10001
F	38	30	10001
M	38	10	10001
M	38	10	10001
M	38	6	10001
M	38	11	10001
M	38	21	10001
M	38	21	10001

Result: The Month of March of the Year 2003 was the highest sum for Females.

Question 10

```
SELECT StateProvinceName, SUM(SalesAmount - TotalProductCost) AS MarginByState
FROM FactInternetSales
JOIN DimCustomer ON DimCustomer.CustomerKey = FactInternetSales.CustomerKey
JOIN DimGeography ON DimGeography.GeographyKey = DimCustomer.GeographyKey
JOIN DimTime ON DimTime.TimeKey = FactInternetSales.OrderDateKey
WHERE DimTime.CalendarYear = '2004'
GROUP BY StateProvinceName
ORDER BY MarginByState DESC;
```

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

Result: The State with the highest Profit Margin for the AW database is California.