Assignment 2 Jushira Thelakkat(jt39634)

# PART A – Coffee Sales

The following questions are based on the Coffee store sales data. Answer the following queries and you must use SQL to extract data and not eyeball some data to answer the questions.

1. **Just for starters - SQL questions:**
   1. **In each state, find the area codes with sales more than 10% of the average sales of all area codes within that state for the year 2013.**

**Solution:**

WITH Q1 AS (

SELECT S.StateName, A.AreaID, SUM(F.ActSales) AS TotSales

FROM AreaCode A, FactCoffee F, States S

WHERE A.StateID=S.StateID AND extract(year from F.FactDate) =2013 AND

F.AreaID=A.AreaID

GROUP BY S.StateName, A.AreaID),

Q2 AS (

SELECT StateName, AVG(Q1.TotSales) AS AvgSales

FROM Q1

GROUP BY StateName)

SELECT Q1.AreaID, Q2.StateName, ROUND(Q2.AvgSales,2) AS AvgSales, Q1.TotSales

FROM Q1, Q2

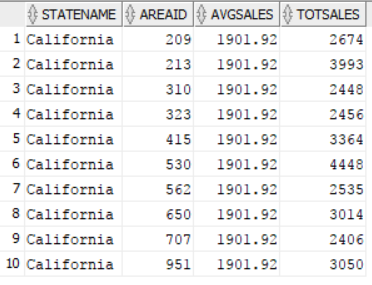
WHERE Q1.StateName = Q2.StateName

AND Q1.TotSales > 1.1\*Q2.AvgSales

GROUP BY Q1.AreaID,Q2.StateName, ROUND(Q2.AvgSales,2), Q1.TotSales

ORDER BY Q2.StateName

FETCH FIRST 10 ROWS ONLY;



**Insights:**  In each state, these are the area codes where total sales are more than 10% average sales. (Only first ten rows have been fetched)

* 1. **Find the products with profit margins as percentage of sales (profits/sales) of at least 15%. Display the results in descending order of total actual sales. Round the percentage to two digits using ROUND(….,2) function.**

**Solution:**

SELECT P.ProdName, SUM(F.ActSales) as TotSales, ROUND(100\*(SUM(F.ActProfit)/SUM(F.ActSales)),2) as Perc

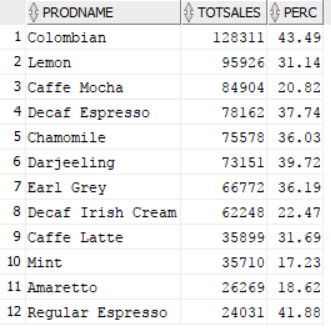
FROM ProdCoffee P, FactCoffee F

WHERE P.ProductID = F.ProductID

GROUP BY P.ProdName

HAVING ROUND(100\*(SUM(F.ActProfit)/SUM(F.ActSales)),2) > 15

ORDER BY TotSales DESC;



**Insights:** The 12 products above are the ones withprofit margins (profits/sales) above 15%.

* 1. **Find AreaIDs where the total profits from leaves in 2012 are 1.2 times greater than that from beans.**

**Solution:**

WITH Q1 AS(

SELECT A.AreaID, P.ProdLine, SUM(F.ActProfit) as TotProfit

FROM AreaCode A, ProdCoffee P, FactCoffee F

WHERE A.AreaID = F.AreaID AND P.ProductID = F.ProductID AND EXTRACT(year from F.FactDate) = 2012

GROUP BY A.AreaID, P.ProdLine)

SELECT \*

FROM Q1

PIVOT (

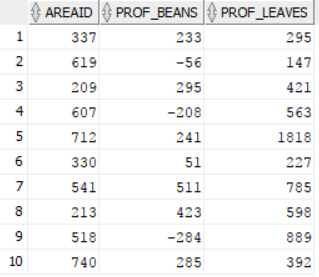
SUM(TotProfit)

FOR ProdLine IN ('Beans' as Prof\_Beans, 'Leaves' as Prof\_Leaves)

)

WHERE Prof\_Leaves > 1.2\*Prof\_Beans

FETCH FIRST 10 ROWS ONLY;



**Insights:** The above are the first ten cases where the actual profits of leaves are greater than 1.2 times the profits of beans in 2012.

1. **DECLINING PROFITS:**
   1. **Which are the top 5 area codes with declining profits and how much did the profits decline for these 5 area codes?**

**Solution:**

CREATE TABLE Temp1 AS

SELECT \* FROM (

SELECT AreaID, StateName, P\_2012, P\_2013, P\_2012-P\_2013 as ProfDec

FROM(

SELECT \* FROM(

SELECT A.AreaID, S.StateName, EXTRACT(year from F.FactDate) as ProfYear, SUM(F.ActProfit) As AreaProf

FROM AreaCode A, FactCoffee F, States S

WHERE A.AreaID = F.AreaID and S.StateID = A.StateID

GROUP BY A.AreaID, S.StateName, EXTRACT(year from F.FactDate)

)

PIVOT (

Sum(AreaProf)

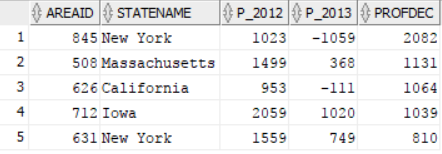
FOR ProfYear IN (2012 as P\_2012, 2013 as P\_2013)

))

WHERE 100\*(P\_2013-P\_2012)/abs(P\_2012) <> 0

ORDER BY ProfDec DESC)

WHERE ROWNUM <= 5;



**Insights:** Creating a temp table which I will be using in next part of this question.

Profits in 2013 are lesser than profits in 2012 (Decline in profits) in these five areas (845,508,626,712 and 631). Here I have taken decline to be (profits of 2012 – profits of 2013) divided by absolute value of the profits in 2012 and ordered by descending to know which have the highest declines.

* 1. **Among the five profit-declining area codes, are the profits consistently declining for all products? If not, identify the products for which they had significantly higher profit decline.**

**Solution:**

SELECT AreaID, StateName, ProdName, P\_2012 - P\_2013 as ProfDec

FROM (

SELECT \* FROM (

SELECT T.AreaID, T.StateName, P.ProdName, EXTRACT(year from F.FactDate) as Year, SUM(F.ActProfit) as ProdProfit

FROM Temp1 T, ProdCoffee P, FactCoffee F

WHERE T.AreaID = F.AreaID AND P.ProductID = F.ProductID

GROUP BY T.AreaID, T.StateName, P.ProdName, EXTRACT(year from F.FactDate)

)

PIVOT (

SUM(ProdProfit)

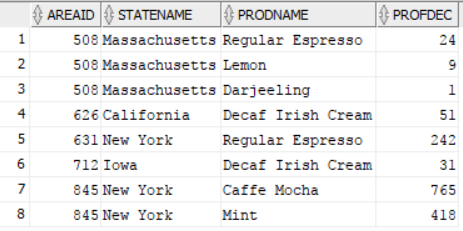
FOR Year IN (2012 as P\_2012, 2013 as P\_2013)

)

)

WHERE P\_2012 - P\_2013 > 0

ORDER BY AreaID, ProfDec DESC;



**Insights:** Using the table I created in previous part, I just checked the products for each of those area ids I got in previous part. We see that in Massachusetts (508), Regular Espresso had significantly higher profit decline (P\_2013-P\_2012) than the other products. For California (626), Decaf Irish Cream was the only product with a profit decline. For New York (631), Regular Espresso was the only product with a profit decline. For Iowa (712), Decaf Irish Cream was the only product with a profit decline. For New York (845), Caffe Mocha had a significantly higher profit decline than other products

1. **BUDGETED Numbers:** 
   1. **All the budgeted numbers are expected targets for 2012 and 2013. Identify the top 5 states for the year 2012 that have substantially higher actual numbers relative to budgeted numbers for profits and sales.**

**Solution:**

WITH Q1 AS(

SELECT S.StateName, SUM(F.ActProfit) - SUM(F.BudProfit) as PROFITDIFF, SUM(F.ActSales) - SUM(F.BudSales) as SALESDIFF

FROM States S, FactCoffee F, AreaCode A

WHERE S.StateID = A.StateID AND A.AreaID = F.AreaID AND EXTRACT(YEAR FROM FACTDATE)=2012

GROUP BY S.StateName

HAVING SUM(F.ActProfit) - SUM(BudProfit) > 0 OR SUM(F.ActSales) - SUM(F.BudSales) > 0

ORDER BY SALESDIFF DESC)

SELECT \* FROM Q1

FETCH FIRST 5 ROWS ONLY;



**Insights:** When I used AND in my HAVING line of code to only get an empty result. When I checked separately for profits and sales, there were no states that had actual profits more than the budgeted profits. I then changed it to OR and these states above are the ones that have higher actual sales than the budgeted sales.

* 1. **Identify area codes within these 5 states that beat budgeted sales and profits significantly (You need to define what significant means here).**

**Solution:**

WITH Q1 AS(

SELECT S.StateName, SUM(F.ActProfit) - SUM(F.BudProfit) as PROFITDIFF, SUM(F.ActSales) - SUM(F.BudSales) as SALESDIFF

FROM States S, FactCoffee F, AreaCode A

WHERE S.StateID = A.StateID AND A.AreaID = F.AreaID AND EXTRACT(YEAR FROM FACTDATE)=2012

GROUP BY S.StateName

HAVING SUM(F.ActProfit) - SUM(BudProfit) > 0 OR SUM(F.ActSales) - SUM(F.BudSales) > 0

ORDER BY SALESDIFF DESC)

SELECT Q1.StateName, A.AreaID,ROUND(100\*(SUM(F.ActProfit) - SUM(F.BudProfit))/SUM(F.BudProfit), 2) as PDP,

ROUND(100\*(SUM(F.ActSales) - SUM(F.BudSales))/SUM(F.BudSales), 2) as SDP

FROM Q1 , AreaCode A, FactCoffee F, States S

WHERE Q1.StateName = S.StateName

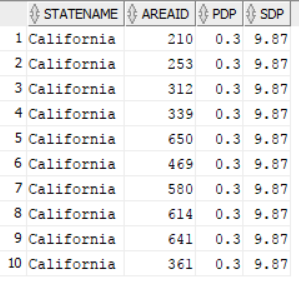
GROUP BY Q1.StateName, A.AreaID

HAVING 100\*(SUM(F.ActProfit) - SUM(F.BudProfit))/SUM(F.BudProfit) > 20

OR 100\*(SUM(F.ActSales) - SUM(F.BudSales))/SUM(F.BudSales) > 5

ORDER BY Q1.StateName, ROUND(100\*(SUM(F.ActSales) - SUM(F.BudSales))) DESC

FETCH FIRST 10 ROWS ONLY;



**Insights:** Here I have specified that if the percent of profit difference is above 20 or percentage of sales difference ((actual-budget)/budget) is greater than 5, then the difference is significant

(I came to these numbers with trial and error)

1. **PRODUCT related:** 
   1. **In each market, which products have the greatest increase in profits?**

**Solution:**

WITH Q1 AS (

SELECT S.StateMkt, P.ProdName, SUM(F.ActProfit) P\_2012

FROM FactCoffee F, States S, ProdCoffee P, AreaCode A

WHERE P.ProductID = F.ProductID and S.StateID = A.StateID and A.AreaID = F.AreaID and extract(YEAR FROM factdate) = 2012

GROUP BY S.StateMkt, P.ProdName),

Q2 AS (

SELECT S.StateMkt, P.ProdName, SUM(F.ActProfit) P\_2013

FROM FactCoffee F, States S, ProdCoffee P, AreaCode A

WHERE P.ProductID = F.ProductID and S.StateID = A.StateID and A.AreaID = F.AreaID and extract(YEAR FROM factdate) = 2013

GROUP BY S.StateMkt, P.ProdName)

SELECT \* FROM (

SELECT Q1.STATEMKT, ROW\_NUMBER() OVER (PARTITION BY Q1.STATEMKT ORDER BY (Q2.P\_2013-Q1.P\_2012)/Q1.P\_2012 DESC) AS RANKID,

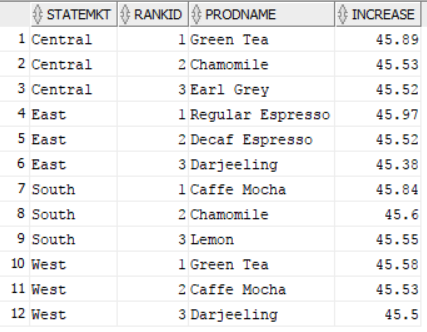
Q1.PRODNAME, ROUND(100\*(Q2.P\_2013-Q1.P\_2012)/Q1.P\_2012,2) INCREASE

FROM Q1, Q2

WHERE Q1.StateMkt = Q2.StateMkt AND Q1.PRODNAME = Q2.PRODNAME

ORDER BY Q1.STATEMKT) X

WHERE X.RANKID <=3;



**Insights:** After writing subqueries using WITH to find total profits for each market and product in 2012 and in 2013 separately, we subtract these profits to see which products in each market have the highest increase in profits.

* 1. **In each market, which product types have greatest increase in sales?**

**Solution:**

WITH Q1 AS (

SELECT S.StateMkt, P.ProdType, SUM(F.ActSales) S\_2012

FROM FactCoffee F, States S, ProdCoffee P, AreaCode A

WHERE P.ProductID = F.ProductID and S.StateID = A.StateID and A.AreaID = F.AreaID and extract(YEAR FROM factdate) = 2012

GROUP BY S.StateMkt, P.ProdType),

Q2 AS (

SELECT S.StateMkt, P.ProdType, SUM(F.ActSales) S\_2013

FROM FactCoffee F, States S, ProdCoffee P, AreaCode A

WHERE P.ProductID = F.ProductID and S.StateID = A.StateID and A.AreaID = F.AreaID and extract(YEAR FROM factdate) = 2013

GROUP BY S.StateMkt, P.ProdType)

SELECT \* FROM (

SELECT Q1.STATEMKT, ROW\_NUMBER() OVER (PARTITION BY Q1.STATEMKT ORDER BY (Q2.S\_2013-Q1.S\_2012)/Q1.S\_2012 DESC) AS RANKID,

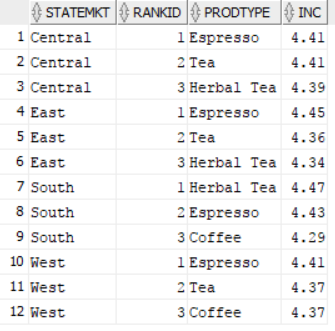
Q1.PRODTYPE, ROUND(100\*(Q2.S\_2013-Q1.S\_2012)/Q1.S\_2012,2) INC

FROM Q1, Q2

WHERE Q1.StateMkt = Q2.StateMkt AND Q1.PRODTYPE = Q2.PRODTYPE

ORDER BY Q1.STATEMKT) X

WHERE X.RANKID <=3;



**Insights:** After writing subqueries using WITH to find total sales for each market and product type in 2012 and in 2013 separately, we subtract these profits to see which product types in each market have the highest increase in sales.

* 1. **Have all products within the product types show similar behavior, or some products within a product type have greatest increase in sales?**

**Solution:**

WITH Q1 AS (

SELECT P.PRODTYPE, P.ProdNAME, SUM(F.ActSales) S1\_2012

FROM FactCoffee F, PRODCOFFEE P

WHERE P.ProductID = F.ProductID and extract(YEAR FROM factdate) = 2012

GROUP BY P.ProdType, P.PRODNAME),

Q2 AS (

SELECT P.PRODTYPE, P.ProdNAME, SUM(F.ActSales) S1\_2013

FROM FactCoffee F, PRODCOFFEE P

WHERE P.ProductID = F.ProductID and extract(YEAR FROM factdate) = 2013

GROUP BY P.ProdType, P.PRODNAME)

SELECT \* FROM (

SELECT Q1.PRODTYPE, ROW\_NUMBER() OVER (PARTITION BY Q1.PRODTYPE ORDER BY (Q2.S1\_2013-Q1.S1\_2012)/Q1.S1\_2012 DESC) AS RANKID,

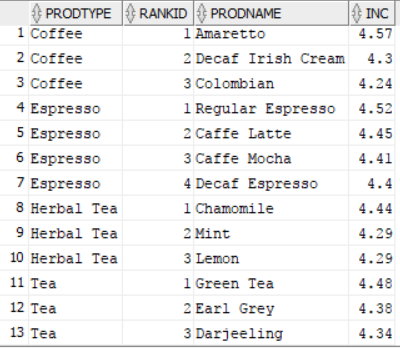
Q1.PRODNAME, ROUND(100\*(Q2.S1\_2013-Q1.S1\_2012)/Q1.S1\_2012,2) INC

FROM Q1, Q2

WHERE Q1.PRODTYPE = Q2.PRODTYPE AND Q1.PRODNAME = Q2.PRODNAME

ORDER BY Q1.PRODTYPE) X

WHERE X.RANKID <=5;



**Insights:** Within each of the product types, we see that the increase in sales percentage is showing similar trend. Everything in is the range of 4%-4.5%

We could probably check w.r.t. market

WITH Q1 AS (

SELECT S.STATEMKT, P.PRODTYPE, P.ProdNAME, SUM(F.ActSales) S2\_2012

FROM STATES S, FactCoffee F, PRODCOFFEE P, AREACODE A

WHERE S.STATEID=A.STATEID AND A.AREAID= F.AREAID AND P.ProductID = F.ProductID and extract(YEAR FROM factdate) = 2012

GROUP BY S.STATEMKT, P.ProdType, P.PRODNAME),

Q2 AS (

SELECT S.STATEMKT, P.PRODTYPE, P.ProdNAME, SUM(F.ActSales) S2\_2013

FROM STATES S, FactCoffee F, PRODCOFFEE P, AREACODE A

WHERE S.STATEID=A.STATEID AND A.AREAID= F.AREAID AND P.ProductID = F.ProductID and extract(YEAR FROM factdate) = 2013

GROUP BY S.STATEMKT, P.ProdType, P.PRODNAME),

Q3 AS (

SELECT Q1.STATEMKT, Q1.PRODTYPE, ROW\_NUMBER() OVER (PARTITION BY Q1.STATEMKT ORDER BY (Q2.S2\_2013-Q1.S2\_2012)/Q1.S2\_2012 DESC) AS RANKID,

Q1.PRODNAME, ROUND(100\*(Q2.S2\_2013-Q1.S2\_2012)/Q1.S2\_2012,2) INC

FROM Q1, Q2

WHERE Q1.STATEMKT=Q2.STATEMKT AND Q1.PRODTYPE = Q2.PRODTYPE AND Q1.PRODNAME = Q2.PRODNAME

ORDER BY Q1.STATEMKT)

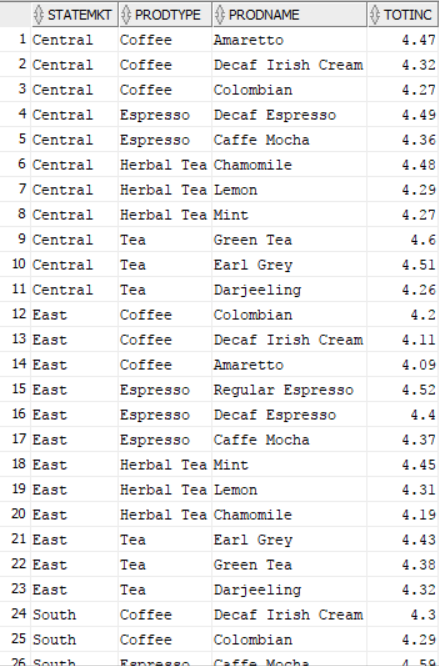
---WHERE X.RANKID <=10;

SELECT Q3.STATEMKT, Q3.PRODTYPE , Q3.PRODNAME, SUM(Q3.INC) AS TOTINC

FROM Q3

GROUP BY Q3.STATEMKT, Q3.PRODTYPE, Q3.PRODNAME

ORDER BY Q3.STATEMKT, Q3.PRODTYPE, TOTINC DESC;



**Insights:** We do not see any big variations. Products show similar behavior.

1. **MARKETING EXPENSES (LOWEST):**
   1. **Which top 5 states have the lowest market expenses as a percentage of their sales?**

**Solution:**

WITH Q1 AS(

SELECT S.STATENAME, ROUND(100\*(SUM(F.ACTMARKCOST)/SUM(F.ACTSALES)), 2) AS PERCSALES

FROM STATES S, FACTCOFFEE F, AREACODE A

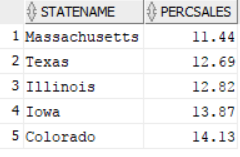
WHERE S.STATEID = A.STATEID AND F.AREAID = A.AREAID

GROUP BY S.STATENAME

ORDER BY PERCSALES)

SELECT \* FROM Q1

WHERE ROWNUM <= 5;



**Insights:** These are the five states that have the least market expenses as a percentage of their sales.

* 1. **Do the above 5 states also have the highest profits as a percentage of sales?**

**Solution:**

WITH Q1 AS(

SELECT S.STATENAME, ROUND(100\*(SUM(F.ACTPROFIT)/SUM(F.ACTSALES)), 2) AS PROFBYSALES

FROM STATES S, FACTCOFFEE F, AREACODE A

WHERE S.STATEID = A.STATEID AND F.AREAID = A.AREAID

GROUP BY S.STATENAME

ORDER BY PROFBYSALES DESC)

SELECT \* FROM Q1

WHERE ROWNUM <= 5;

**With temp table**--

CREATE TABLE Temp2 AS

SELECT \* FROM (

SELECT S.StateName, ROUND(100\*(SUM(F.ActProfit)/SUM(F.ActSales)), 2) AS ProfPercSales

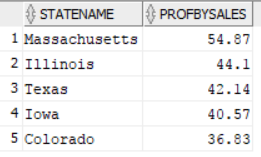
FROM States S, FactCoffee F, AreaCode A

WHERE S.StateID = A.StateID AND F.AreaID = A.AreaID

GROUP BY S.StateName

ORDER BY ProfPercSales DESC)

WHERE ROWNUM <= 5;



**Insights:** The table shows the top 5 states with the highest profits as a percentage of sales. The states are the same as the 5 with the lowest market expenses as a percentage of sales, but they are ranked differently. Illinois has higher profits as a percentage of sales than Texas but also higher market expenses as a percent of sales.

* 1. **Are there any particular product(s) within these markets with the least marketing expenses?**

**Solution:**

SELECT StateName, ProdName, MktExp FROM(

SELECT T.StateName, P.ProdName, SUM(F.ActMarkCost) AS MktExp, RANK() OVER (PARTITION BY T.StateName

ORDER BY SUM(F.ActMarkCost)) AS StateRank

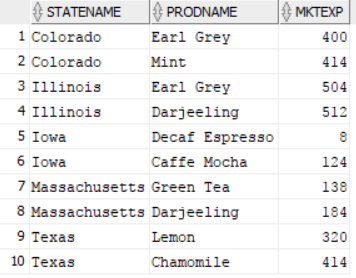
FROM Temp2 T, ProdCoffee P, FactCoffee F, States S, AreaCode A

WHERE T.StateName = S.StateName AND S.StateID = A.StateID AND A.AreaID = F.AreaID AND P.ProductID = F.ProductID

GROUP BY T.StateName, P.ProdName

ORDER BY StateName, MktExp)

WHERE StateRank <= 2;



**Insights:** These are the two products with least marketing expenses in each state. In Iowa, we see that Decaf Espresso has very low expenses in comparison with the other products in that state.

1. **MARKETING EXPENSES (highest):**
   1. **Which 5 states have the highest marketing expenses as a percentage of sales?**

**Are these marketing expenses justified? (Note: you need to think how you will justify high marketing expenses)?**

**Solution:**

SELECT \* FROM

(SELECT S.STATENAME, ROUND(100\*(SUM(F.ACTMARKCOST)/SUM(F.ACTSALES)), 2) AS COSTBYSALES, ROUND(100\*(SUM(F.ACTPROFIT)/SUM(F.ACTSALES)),2) AS PROFBYSALES,

ROUND(100\*(SUM(F.ACTMARKCOST)/SUM(F.ACTPROFIT)),2) AS COSTBYPROF

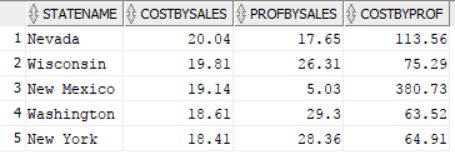
FROM STATES S, FACTCOFFEE F, AREACODE A

WHERE S.STATEID = A.STATEID AND F.AREAID = A.AREAID

GROUP BY S.STATENAME

ORDER BY COSTBYSALES DESC)

WHERE ROWNUM<=5;



WITH Q1 AS(

SELECT S.STATENAME, ROUND(100\*(SUM(F.ACTMARKCOST)/SUM(F.ACTSALES)), 2) AS COSTBYSALES, ROUND(100\*(SUM(F.ACTPROFIT)/SUM(F.ACTSALES)),2) AS PROFBYSALES,

ROUND(100\*(SUM(F.ACTMARKCOST)/SUM(F.ACTPROFIT)),2) AS COSTBYPROF

FROM STATES S, FACTCOFFEE F, AREACODE A

WHERE S.STATEID = A.STATEID AND F.AREAID = A.AREAID

GROUP BY S.STATENAME

ORDER BY COSTBYSALES DESC)

SELECT AVG(Q1.COSTBYPROF) AS AVGCOSTBYPROF, AVG(Q1.PROFBYSALES) AS AVGPROFBYSALES

FROM Q1

WHERE ROWNUM <= 5;



**Insights:** Nevada, Wisconsin, New Mexico, Washington and New York are the states with highest expenses as a percent of sales (CostBySales). To understand if these high expenses are justifiable, I checked two other ratios for comparison. The Profit by Sales ratio should have been in line with the Cost by Sales ratio but it does not seem to be happening. New Mexico’s cost is so high and profit/sales so low. Also, the Cost by Profit ratio for New Mexico tells us that it is very clearly spending much more than it is earning. So these expenses are not justifiable.

* 1. **In each of these 5 states, do any area codes spend too much on marketing expenses relative to others?**

**Solution:**

WITH Q1 AS(

SELECT \* FROM(

SELECT S.STATENAME, ROUND(100\*(SUM(F.ACTMARKCOST)/SUM(F.ACTSALES)), 2) AS COSTBYSALES, ROUND(100\*(SUM(F.ACTPROFIT)/SUM(F.ACTSALES)),2) AS PROFBYSALES,

ROUND(100\*(SUM(F.ACTMARKCOST)/SUM(F.ACTPROFIT)),2) AS COSTBYPROF

FROM STATES S, FACTCOFFEE F, AREACODE A

WHERE S.STATEID = A.STATEID AND F.AREAID = A.AREAID

GROUP BY S.STATENAME

ORDER BY COSTBYSALES DESC)

WHERE ROWNUM <=5)

SELECT \* FROM (

SELECT Q1.StateName, A.AreaID, ROUND(100\*(SUM(F.ActMarkCost))/SUM(F.ActSales), 2) AS COSTBYSALES,

ROUND(100\*(SUM(F.ActProfit)/SUM(F.ActSales)),2) AS PROFITBYSALES

FROM Q1, AreaCode A, FactCoffee F, States S

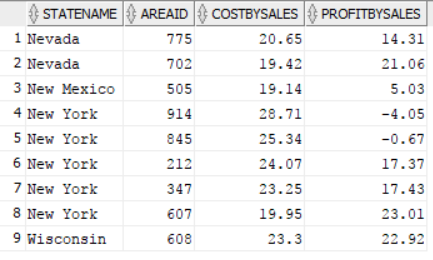
WHERE Q1.StateName = S.StateName AND S.StateID = A.StateID AND A.AreaID = F.AreaID

GROUP BY Q1.StateName, A.AreaID

ORDER BY Q1.STATENAME, COSTBYSALES DESC

)

WHERE PROFITBYSALES < 25;



**Insights:** It is clear that a few of these area codes spend too much on marketing expenses relative to others. New York (914) and New York (845) both spend such high amounts on their sales on marketing but earn negative profits. So, we do have such cases.

1. **STRATEGY:**
   1. **You are in a high-level strategy meeting to discuss how to improve performance. This may involve shutting down stores in losing area codes and/or expanding in very profitable/high growth area. Evaluate the data and recommend which stores to close and where?**

**Solution:**

SELECT X.AREAID, X.STATENAME, P\_2012, P\_2013, ROUND(100\*(P\_2013-P\_2012)/abs(P\_2012),2) as PerInc

FROM (

SELECT \* FROM (

Select A.AREAID, S.STATENAME, EXTRACT(YEAR FROM F.FACTDATE) AS YEARS, SUM(F.ACTPROFIT) AS TOTPROFIT

FROM AREACODE A, FACTCOFFEE F, STATES S

WHERE A.AREAID = F.AREAID AND S.STATEID = A.STATEID

GROUP BY A.AREAID, S.STATENAME, EXTRACT(YEAR FROM F.FACTDATE)

)

PIVOT (

SUM(TOTPROFIT)

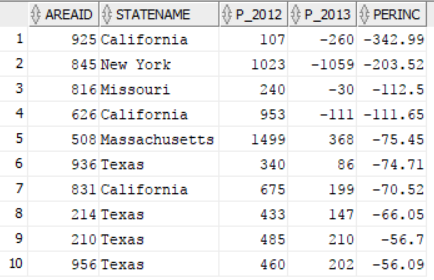
FOR Years IN (2012 as P\_2012,2013 as P\_2013)

)) X

WHERE P\_2012 <> 0

ORDER BY PERINC

FETCH FIRST 10 ROWS ONLY;



**Insights:** If we look at the relative decline of profits from 2012 to 2013, we see that California (925), New York (845) are not performing well and are showing declining trends of performance. We should probably shut down these.

* 1. **Where should the firm focus on expanding?**

**Solution:**

SELECT \* FROM (

Select A.AREAID, S.STATENAME, EXTRACT(YEAR FROM F.FACTDATE) AS YEARS, SUM(F.ACTPROFIT) AS TOTPROFIT

FROM AREACODE A, FACTCOFFEE F, STATES S

WHERE A.AREAID = F.AREAID AND S.STATEID = A.STATEID

GROUP BY A.AREAID, S.STATENAME, EXTRACT(YEAR FROM F.FACTDATE)

)

PIVOT (

SUM(TOTPROFIT)

FOR Years IN (2012 as P\_2012,2013 as P\_2013)

)

ORDER BY P\_2013 DESC

FETCH FIRST 10 ROWS ONLY;

SELECT X.AREAID, X.STATENAME, P\_2012, P\_2013, ROUND(100\*(P\_2013-P\_2012)/abs(P\_2012),2) as PerInc

FROM (

SELECT \* FROM (

Select A.AREAID, S.STATENAME, EXTRACT(YEAR FROM F.FACTDATE) AS YEARS, SUM(F.ACTPROFIT) AS TOTPROFIT

FROM AREACODE A, FACTCOFFEE F, STATES S

WHERE A.AREAID = F.AREAID AND S.STATEID = A.STATEID

GROUP BY A.AREAID, S.STATENAME, EXTRACT(YEAR FROM F.FACTDATE)

)

PIVOT (

SUM(TOTPROFIT)

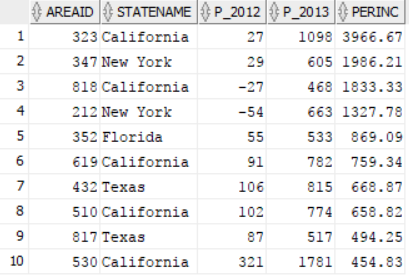
FOR Years IN (2012 as P\_2012,2013 as P\_2013)

)) X

WHERE P\_2012 <> 0

ORDER BY PERINC DESC

FETCH FIRST 10 ROWS ONLY;



**Insights:** Just doing the reverse of first part, we see that California (323), New York (347) are showing positive trends of profits indicating that they must be doing well hence we could expand these.

# PART B: Office Product

**QUESTION 1: Create the 5 tables given above. You should define primary keys, foreign keys, and other CHECK constraints. And, load the data from Excel spreadsheet.**

**Solution:**

CREATE TABLE MANAGERS (

RegID NUMBER,

Region VARCHAR2(10 BYTE),

RegManager VARCHAR2(10 BYTE),

PRIMARY KEY (RegID),

CONSTRAINT ch\_reg CHECK (Region IN ('East', 'South', 'Central', 'West')));

CREATE TABLE PRODUCTS (

ProdID NUMBER,

ProdName VARCHAR2(100 BYTE),

ProdCat VARCHAR2(30 BYTE),

ProdSubCat VARCHAR2(30 BYTE),

ProdCont VARCHAR2(30 BYTE),

ProdUnitPrice NUMBER(7,2),

ProdMargin NUMBER(5,3),

PRIMARY KEY (ProdID),

CONSTRAINT ch\_cat CHECK (ProdCat IN ('Technology', 'Furniture', 'Office Supplies')),

CONSTRAINT ch\_cont CHECK (ProdCont IN ('Jumbo Drum', 'Medium Box', 'Jumbo Box', 'Wrap Bag',

'Large Box', 'Small Box', 'Small Pack')));

CREATE TABLE ORDERS (

OrderID NUMBER,

Status VARCHAR2(10 BYTE),

PRIMARY KEY (OrderID));

CREATE TABLE CUSTOMERS (

CustID NUMBER,

CustName VARCHAR2(35 BYTE),

CustReg NUMBER(1,0),

CustState VARCHAR2(20 BYTE),

CustCity VARCHAR2(20 BYTE),

CustZip NUMBER(5,0),

CustSeg VARCHAR2(15 BYTE),

PRIMARY KEY (CustID),

FOREIGN KEY (CustReg) REFERENCES Managers(RegID) ON DELETE CASCADE,

CONSTRAINT ch\_seg CHECK (CustSeg IN ('Home Office', 'Corporate', 'Small Business', 'Consumer')));

CREATE TABLE ORDERDET (

OrderID NUMBER,

CustID NUMBER,

ProdID NUMBER,

OrdPriority VARCHAR2(15 BYTE),

OrdDiscount NUMBER(3,2),

OrdShipMode VARCHAR2(15 BYTE),

OrdDate DATE,

OrdShipDate DATE,

OrdShipCost NUMBER(5,2),

OrdQty NUMBER,

OrdSales NUMBER(8,2),

CONSTRAINT PK PRIMARY KEY (OrderID, CustID, ProdID),

CONSTRAINT FK1 FOREIGN KEY (OrderID) REFERENCES Orders(OrderID),

CONSTRAINT FK2 FOREIGN KEY (CustID) REFERENCES Customers(CustID),

CONSTRAINT FK3 FOREIGN KEY (ProdID) REFERENCES Products(ProdID),

CONSTRAINT ch\_priority CHECK (OrdPriority IN ('Low', 'Medium', 'High', 'Critical', 'Not Specified')),

CONSTRAINT ch\_mode CHECK (OrdShipMode IN ('Regular Air', 'Delivery Truck', 'Express Air')));

**QUESTION 2: ORDER Cancellations**

1. **What fraction of the orders was cancelled?**

**Solution:**

WITH Q1 AS (

SELECT COUNT(O.ORDERID) AS CANCELLED

FROM ORDERS O

WHERE STATUS LIKE 'Returned'),

Q2 AS (

SELECT COUNT(O.ORDERID) AS TOTAL

FROM ORDERS O)

SELECT Q1.CANCELLED, Q2.TOTAL, ROUND(Q1.CANCELLED/Q2.TOTAL,5) AS FRACTION

FROM Q1, Q2;



1. **What were the sales from cancelled orders?**

**Solution:**

SELECT SUM(D.OrdSales) AS CANCELLEDSALES

FROM Orders O, OrderDet D

WHERE O.OrderID = D.OrderID AND O.Status LIKE 'Returned';



1. **Who are the top five customers in terms of cancelled orders?**

**Solution:**

SELECT C.CUSTNAME, COUNT(O.ORDERID) AS CANCELLED

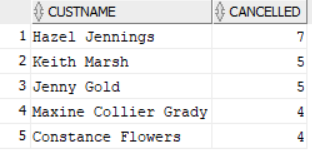
FROM CUSTOMERS C, ORDERS O, ORDERDET D

WHERE O.ORDERID=D.ORDERID AND D.CUSTID=C.CUSTID AND O.STATUS LIKE 'Returned'

GROUP BY C.CUSTNAME

ORDER BY COUNT(O.ORDERID) DESC

FETCH FIRST 5 ROWS ONLY;



**QUESTION 3: CUSTOMER related:**

1. **Who are the top 10 customers in terms of revenues generated?**

**Solution:**

SELECT C.CUSTNAME, ROUND(SUM(D.ORDSALES),2) AS TOTREVENUE

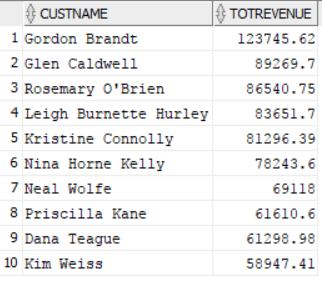
FROM CUSTOMERS C, ORDERDET D

WHERE C.CUSTID=D.CUSTID

GROUP BY C.CUSTNAME

ORDER BY ROUND(SUM(D.ORDSALES),2) DESC

FETCH FIRST 10 ROWS ONLY;



***Question 3 a –REDO—***

SELECT \* FROM(

SELECT C.CUSTNAME AS CUSTNAME, STATUS, SUM(ORDSALES)

FROM ORDERS O, ORDERDET D, CUSTOMERS C

WHERE O.ORDERID=D.ORDERID AND C.CUSTID=D.CUSTID

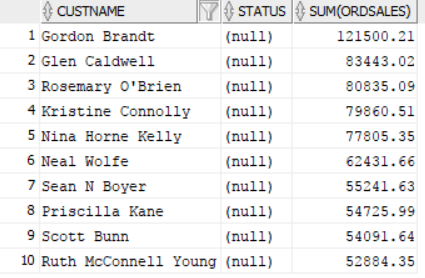
AND STATUS IS NULL

GROUP BY STATUS,C.CUSTNAME

ORDER BY SUM(ORDSALES) DESC)

WHERE ROWNUM<=10;

**Solution:**



1. **Are there customers who buy mostly some categories of products and there is a potential for them to buy other product categories?**

**Solution:**

**Code 1:**

WITH Q1 AS (

SELECT C.CustName, SUM(D.OrdSales) AS CustSales

FROM Customers C, OrderDet D

WHERE C.CustID = D.CustID

GROUP BY C.CustName),

Q2 AS (

SELECT C.CustName, P.ProdCat, SUM(D.OrdSales) AS CatSales

FROM Products P, OrderDet D, Customers C

WHERE C.CustID = D.CustID AND P.ProdID = D.ProdID

GROUP BY C.CustName, P.ProdCat)

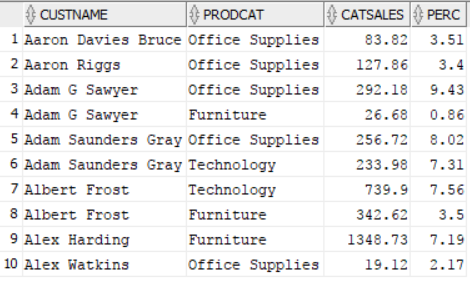
SELECT Q1.CustName, Q2.ProdCat, Q2.CatSales, ROUND(100\*(Q2.CatSales/Q1.CustSales),2) AS PERC

FROM Q1, Q2

WHERE Q1.CustName = Q2.CustName AND ROUND(100\*(Q2.CatSales/Q1.CustSales),2) <10

ORDER BY Q1.CustName

FETCH FIRST 10 ROWS ONLY;



**Insights:**

Here we look at the percentage of sales of each customer spent of certain categories of products.

we see that for, say Aaron Davies Bruce (row 1) he spends only 3.51% of his total sales on Office Supplies, that means he is mostly spending on some other category of product.

We could confirm by changing Perc<10 to >50

**Code 2:**

WITH Q1 AS (

SELECT C.CustName, SUM(D.OrdSales) AS CustSales

FROM Customers C, OrderDet D

WHERE C.CustID = D.CustID

GROUP BY C.CustName),

Q2 AS (

SELECT C.CustName, P.ProdCat, SUM(D.OrdSales) AS CatSales

FROM Products P, OrderDet D, Customers C

WHERE C.CustID = D.CustID AND P.ProdID = D.ProdID

GROUP BY C.CustName, P.ProdCat)

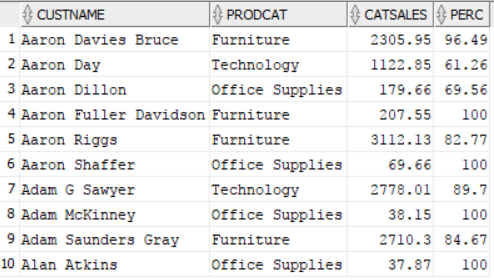
SELECT Q1.CustName, Q2.ProdCat, Q2.CatSales, ROUND(100\*(Q2.CatSales/Q1.CustSales),2) AS PERC

FROM Q1, Q2

WHERE Q1.CustName = Q2.CustName AND ROUND(100\*(Q2.CatSales/Q1.CustSales),2) >50

ORDER BY Q1.CustName

FETCH FIRST 10 ROWS ONLY;



**Insights:** Like we guessed, Aaron mostly buys (96%) on Furniture. So, there is scope for him to buy Technology too. Right now, he is spending 97% on Furniture and 3% on Office Supplies. He could buy a new category Technology too.

We see that many customers follow this trend.

**QUESTION 4: There are differences in the actual (theoretical) price ((unit price \* number of units\*(1-discount) + shipping cost) and the actual sales for all products. There are some discounts and shipping costs. Yet, there are discrepancies in the theoretical sales and actual sales.**

1. **How much more or less are the actual sales value compared to the theoretical sales value?**

**Solution:**

WITH Q1 AS(

SELECT D.OrderID, ((P.ProdUnitPrice\*D.OrdQty)\*(1-D.OrdDiscount)+OrdShipCost) AS Theoretical, D.OrdSales AS Actual

FROM OrderDet D, Products P

WHERE D.ProdID = P.ProdID)

SELECT SUM(Actual) - SUM(Theoretical) AS Difference

FROM Q1;



***Question 4 a--- REDO ----***

WITH Q1 AS (SELECT PRODUCTS.PRODNAME, SUM(OrderDet.OrdSales) ActualSales

FROM PRODUCTS

INNER JOIN ORDERDET

ON ORDERDET.PRODID = PRODUCTS.PRODID

GROUP BY PRODUCTS.PRODNAME),

Q2 AS (SELECT PRODUCTS.PRODNAME,

SUM(PRODUCTS.ProdUnitPrice \* ORDERDET.OrdQty \* (1-ORDERDET.OrdDiscount)+ ORDERDET.OrdShipCost) TheoreticalSales

FROM PRODUCTS

INNER JOIN ORDERDET

ON PRODUCTS.PRODID = ORDERDET.PRODID

GROUP BY PRODUCTS.PRODNAME)

SELECT Q1.PRODNAME,

Q1.ActualSales,

Q2.TheoreticalSales,

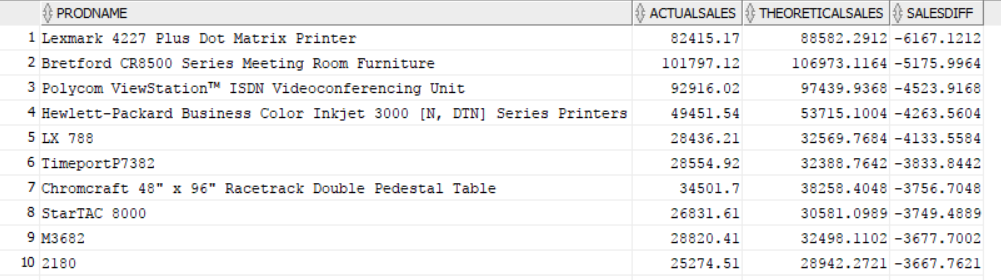
(Q1.ActualSales - Q2.TheoreticalSales) SalesDiff

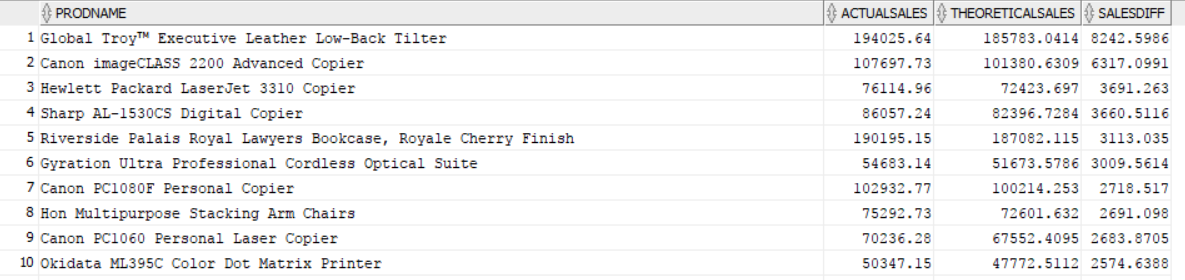
FROM Q1, Q2

WHERE Q1.PRODNAME = Q2.PRODNAME

ORDER BY SalesDiff;

**Solution:**





**Insights:** Lexmark, Bretford, Polycom etc. have theoretical sales higher than the actual sales so they are underpriced. And on the other hand, products such as Global Troy, Canon etc. have actual sales much higher than their theoretical sales and hence are mostly priced higher than they should be.

1. **Are certain managers generally pricing more or less than theoretical sales? Analyze the differences based on the regions/managers.**

**Solutions:**

WITH Q1 AS(

SELECT D.OrderID, M.Region, M.RegManager,

((P.ProdUnitPrice\*D.OrdQty)\*(1-D.OrdDiscount)+OrdShipCost) AS Theoretical,

D.OrdSales AS Actual

FROM OrderDet D, Products P, Managers M, Customers C

WHERE D.ProdID = P.ProdID AND M.RegID = C.CustReg AND C.CustID = D.CustID)

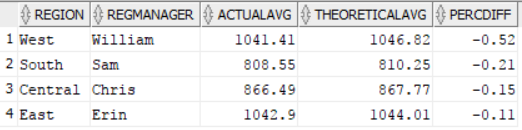
SELECT Region, RegManager, ROUND(AVG(Actual),2) AS ActualAvg, ROUND(AVG(Theoretical),2) AS TheoreticalAvg,

ROUND(100\*(AVG(Actual) - AVG(Theoretical))/AVG(Theoretical),2) AS PercDiff

FROM Q1

GROUP BY Region, RegManager

ORDER BY PercDiff;



**Insights:** Every region has only one manager so we could just do analysis based on either.

If we take the average of the total actual and average of the total theoretical sales for comparison, we see that all the four managers in the four regions are pricing lesser because as we see, the actual sales are lesser than the theoretically calculated sales.

**QUESTION 5: Product related questions:**

1. **Products have numbers within its name. Identify the product names with digits in their name. (hint: use REGEXP\_LIKE)**

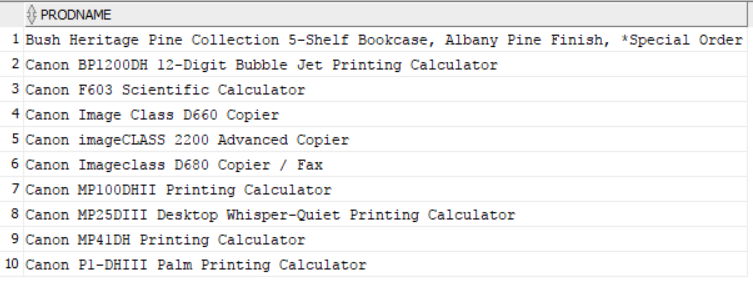
**Solution:**

SELECT P.ProdName

FROM PRODUCTS P

WHERE REGEXP\_LIKE (P.ProdName, '\d')

FETCH FIRST 10 ROWS ONLY;



1. **Which are the top 5 selling products during the year 2011?**

**Solution:**

SELECT Rank, ProdName, TotSales FROM (

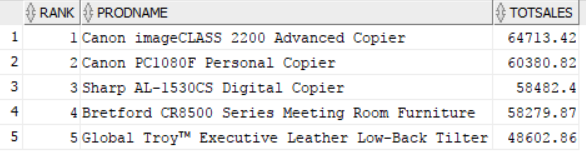
SELECT P.ProdName, SUM(D.OrdSales) AS TotSales, ROW\_NUMBER() OVER (ORDER BY SUM(D.OrdSales) DESC) AS Rank

FROM Products P, OrderDet D

WHERE P.ProdID = D.ProdID AND EXTRACT(YEAR FROM D.OrdDate) = 2011

GROUP BY P.ProdName)

WHERE Rank <=5;



**Insights:** The above five had the highest sales in 2011.

***Question 5 b –REDO--***

SELECT \* FROM(

SELECT P.PRODNAME, SUM(O.OrdQty) Quantity, D.STATUS

FROM ORDERS D, PRODUCTS P, ORDERDET O

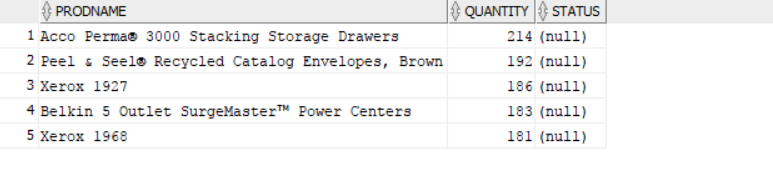
WHERE O.ORDERID=D.ORDERID AND P.PRODID = O.PRODID

AND extract(Year from O.ORDDATE) = 2011 AND D.STATUS IS NULL

GROUP BY D.STATUS,P.PRODNAME

ORDER BY SUM(O.OrdQty) DESC)

WHERE ROWNUM<=5;



**Insights:** These are the top 5 selling products in 2011

1. **Which are the top 10 products with greatest total profit margin? (i.e., sales\*margin).**

**Solution:**

WITH Q1 AS(

SELECT P.ProdName, SUM(D.OrdSales\*P.ProdMargin) AS TotMargin, ROW\_NUMBER() OVER (ORDER BY SUM(D.OrdSales\*P.ProdMargin) DESC) AS RANK

FROM Products P, OrderDet D

WHERE P.ProdID = D.ProdID

GROUP BY P.ProdName)

SELECT Rank, ProdName, TotMargin FROM Q1

WHERE Rank <= 10;



**Insights:**  Sales\*margin is the highest for the above mentioned ten products.

1. **Identify the worst five products in terms of sales?**

**Solution:**

WITH Q1 AS(

SELECT P.ProdName, SUM(D.OrdSales) AS TotSales, RANK() OVER (ORDER BY SUM(D.OrdSales)) AS Rank

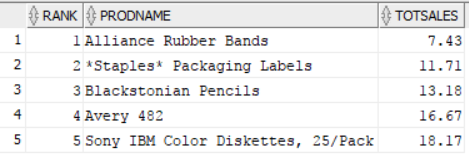
FROM Products P, OrderDet D

WHERE P.ProdID = D.ProdID

GROUP BY P.ProdName)

SELECT Rank, ProdName, TotSales FROM Q1

WHERE Rank <= 5;



**Insights:** These above five products have the worst total sales.

***Question 5d –REDO—***

SELECT \* FROM(

SELECT P.PRODNAME, SUM(O.Ordsales), D.STATUS

FROM ORDERS D, PRODUCTS P, ORDERDET O

WHERE O.ORDERID=D.ORDERID AND P.PRODID = O.PRODID

AND extract(Year from O.ORDDATE) = 2011 AND D.STATUS IS NULL

GROUP BY D.STATUS,P.PRODNAME

ORDER BY SUM(O.Ordsales))

WHERE ROWNUM<=5;

