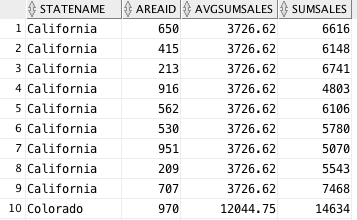
Assignment 2

# PART A – Coffee Sales

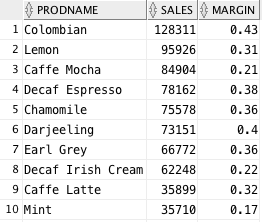
The following questions are based on the Coffee store sales data. Answer the following queries and you must use SQL to extract data and not eyeball some data to answer the questions.

1. Just for starters - SQL questions:
   1. In each state, find the area codes with sales more than 10% the average sales of all area codes within that state for the year 2013.
2. **Create** **view** q1sub **as**
3. **Select** statename, areacode.areaid, sum(Actsales) **as** sumsales
4. **from** states, areacode, factcoffee
5. **where** states.stateid=areacode.stateid and factcoffee.areaid=areacode.areaid and extract(year **from** factdate)=2013
6. **group** **by** statename, areacode.areaid;
8. **SELECT** X.statename, areaid, Avgsumsales, SUmsales
9. **FROM**
10. (**select** statename, Round(Avg(sumsales),2) **as** Avgsumsales
11. **from** q1sub
12. **group** **by** statename) X, q1sub
13. **WHERE** x.statename = q1sub.statename AND q1sub.sumsales > 1.1\*X.Avgsumsales
14. **Order** **by** statename;



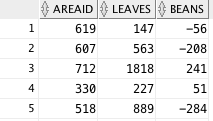
* 1. Find the products with profit margins as percentage of sales (profits/sales) of at least 15%. Display the results in descending order of total actual sales. Round the percentage to two digits using ROUND(….,2) function.

1. **SELECT** Prodname, Round(Prof/sales,2) **as** Margin **from**(
2. **Select** \* **from**(
3. **SELECT** prodname, sum(actprofit) Prof, sum(actsales) Sales
4. **FROM** prodcoffee, factcoffee
5. **where** prodcoffee.productid=factcoffee.productid
6. **group** **by** prodname))
7. **where** Prof/Sales>=0.15
8. **Order** **by** Sales **DESC**;

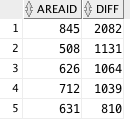


* 1. Find AreaIDs where the total profits from leaves in 2012 are two times greater than that from beans.

1. **SELECT** \* **from**(
2. **Select** \* **from**(
3. **SELECT** areaid, prodline **as** types, sum(actprofit) **as** TotProfits
4. **from** prodcoffee, factcoffee
5. **where** prodcoffee.productid=factcoffee.productid
6. and extract(year **from** factdate)=2012
7. **group** **by** prodline, areaid)
8. Pivot(sum(Totprofits) **for** types in ('Leaves' **as** Leaves, 'Beans' **as** Beans)))
9. **where** Leaves>=2\*Beans;

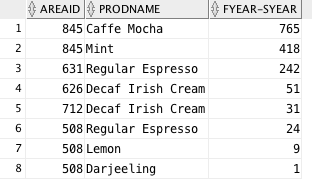


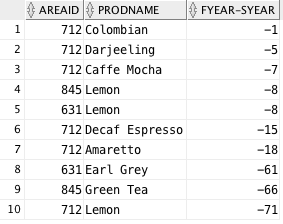
1. DECLINING PROFITS:
   1. Which are the top 5 area codes with declining profits and how much did the profits decline for these 5 area codes?
2. **Select** \* **from**(
3. **Select** areaid, Fyear-Syear **as** Diff **from**(
4. **SELECT** \* **from**(
5. **SELECT** areaid, extract(year **from** factdate) **as** Year, sum(actprofit) **as** Tot
6. **from** factcoffee
7. **group** **by** areaid, extract(year **from** factdate))
8. Pivot(sum(Tot) **for** year in (2012 **as** Fyear, 2013 **as** Syear)))
9. **where** Fyear-Syear>0
10. **order by diff DESC**)
11. **where** rownum<=5;



* 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.

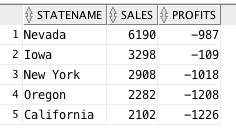
1. **Select** Areaid, Prodname, Fyear-Syear **from**(
2. **SELECT** \* **from**(
3. **SELECT** areaid,prodname, extract(year **from** factdate) **as** year, sum(actprofit) **as** tots
4. **From** prodcoffee, factcoffee
5. **where** prodcoffee.productid=factcoffee.productid
6. and areaid in (631,845,508,626,712)
7. **Group** **by** areaid, prodname, extract(year **from** factdate))
8. Pivot(sum(tots) **for** year in (2012 **as** Fyear, 2013 **as** Syear)))
9. **where** Fyear-Syear>0;  --second table Fyear-Syear<0





The top table shows the products with declining profits, and the bottom table shows those with increasing profits. Within the given area codes, we can see that cafe mocha, mint tea, regular espresso and decaf Irish cream substantially decreased in profits.

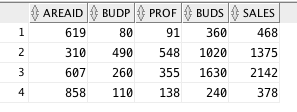
1. BUDGETED Numbers:
   1. All the budgeted numbers are expected targets for 2012 and 2013. Identify the top 5 states for the year 2012 that have substantially higher actual numbers relative to budgeted numbers for profits and sales.
2. **SELECT** \* **from**(
3. **Select** Statename, (Sact-Sbud) **as** Sales, (Pact-Pbud) **as** Profits **from**(
4. **Select** \* **from**(
5. **SELECT** statename, sum(budsales) **as** SBud, sum(budprofit) **as** PBud, sum(actsales) **as** SAct, sum(actprofit) **as** PAct
6. **from** states, areacode, factcoffee
7. **where** extract(year **from** factdate)=2012
8. and states.stateid=areacode.stateid and areacode.areaid=factcoffee.areaid
9. **Group** **by** statename)))
10. **order** **by** sales **DESC**
11. **fetch** **first** 5 **rows** **with** ties;



Interestingly, as a whole, no state surpassed its budgeted profits.

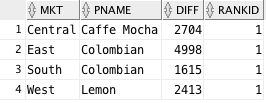
* 1. Identify area codes within these 5 states that beat budgeted sales and profits significantly (You need to define what significant means here).

1. **Select** \* **from**(
2. **Select** areaid, BudP,Prof, BudS,Sales **from**(
3. **SELECT** areacode.areaid, sum(budprofit) BudP, sum(budsales) BudS, sum(actprofit) Prof, sum(actsales) **as** Sales
4. **FROM** areacode, factcoffee, states
5. **where** areacode.areaid=factcoffee.areaid and states.stateid=areacode.stateid and extract(year **from** factdate)=2012
6. and statename in ('California','New York','Oregon','Iowa',’Nevada’)
7. **group** **by** areacode.areaid))
8. **where** Prof>1.1\*BudP and sales>1.1\*BudS;



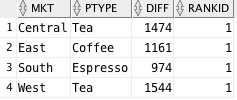
These area codes surpassed their budgeted profits and sales by 10%.

1. PRODUCT related:
   1. In each market, which products have the greatest increase in profits?
2. **Select** \* **from**(
3. **Select** Mkt, pname, Diff, row\_number() over (partition **by** Mkt **order** **by** Diff **DESC**) **as** rankid **from**(
4. **Select** Mkt, pname, Syear-Fyear **as** Diff **from**(
5. **SELECT** \* **FROM**(
6. **SELECT** statemkt mkt, prodname pname, sum(actprofit) SumProf, extract(year **from** factdate) year
7. **from** factcoffee, prodcoffee, states, areacode
8. **where** factcoffee.productid=prodcoffee.productid and areacode.areaid=factcoffee.areaid
9. and states.stateid=areacode.stateid
10. **group** **by** prodname, statemkt, extract(year **from** factdate))
11. pivot(sum(SumProf) **for** year in (2012 **as** Fyear, 2013 **as** Syear)))
12. **where** Syear-Fyear>0))
13. **where** rankid=1;



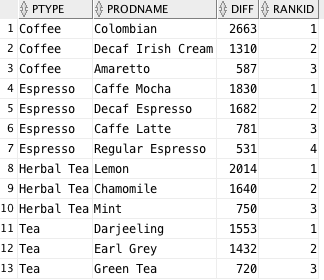
* 1. In each market, which **product types** have greatest increase in sales?

1. **Select** \* **from**(
2. **Select** ptype, prodname,Diff, row\_number() over (partition **by** ptype **order** **by** Diff **DESC**) **as** rankid **from**(
3. **Select** ptype, prodname, Syear-Fyear **as** Diff **from**(
4. **SELECT** \* **FROM**(
5. **SELECT** prodname, prodtype ptype, sum(actsales) SumSales, extract(year **from** factdate) year
6. **from** factcoffee, prodcoffee, areacode
7. **where** factcoffee.productid=prodcoffee.productid and areacode.areaid=factcoffee.areaid
8. **group** **by** prodtype, prodname, extract(year **from** factdate))
9. pivot(sum(SumSales) **for** year in (2012 **as** Fyear, 2013 **as** Syear)))
10. **where** Syear-Fyear>0))
11. **where** rankid=1;



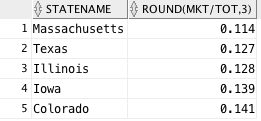
* 1. Do all products within the product types show similar behavior, or do some products within a product type have the greatest increase in sales?

1. **Select** Mkt, ptype, Diff, row\_number() over (partition **by** Mkt **order** **by** Diff **DESC**) **as** rankid **from**(
2. **Select** Mkt, ptype, Syear-Fyear **as** Diff **from**(
3. **SELECT** \* **FROM**(
4. **SELECT** statemkt mkt, prodtype ptype, sum(actsales) SumSales, extract(year **from** factdate) year
5. **from** factcoffee, prodcoffee, states, areacode
6. **where** factcoffee.productid=prodcoffee.productid and areacode.areaid=factcoffee.areaid
7. and states.stateid=areacode.stateid
8. **group** **by** prodtype, statemkt, extract(year **from** factdate))
9. pivot(sum(SumSales) **for** year in (2012 **as** Fyear, 2013 **as** Syear)))
10. **where** Syear-Fyear>0);



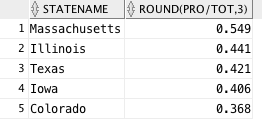
We can see from this chart that Colombian coffee, Caffe Mocha (espresso), Lemon Herbal Tea and Darjeeling Tea have a significant increase in sales compared to other products within their product types.

1. MARKETING EXPENSES (LOWEST):
   1. Which top 5 states have the lowest market expenses as a percentage of their sales?
2. **Select** \* **from**(
3. **SELECT** statename, Round(Mkt/TOT,3) **from**(
4. **SELECT** statename, sum(actsales) **as** Tot, sum(actmarkcost) **as** Mkt
5. **FROM** factcoffee, areacode, states
6. **where** factcoffee.areaid= areacode.areaid and areacode.stateid=states.stateid
7. **group** **by** statename)
8. **order** **by** Mkt/Tot)
9. **where** rownum<=5;



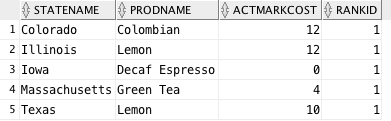
* 1. Do the above 5 states also have the highest profits as a percentage of sales?

1. **Select** \* **from**(
2. **SELECT** statename, Round(Pro/TOT,3) **from**(
3. **SELECT** statename, sum(actsales) **as** Tot, sum(actprofit) **as** Pro
4. **FROM** factcoffee, areacode, states
5. **where** factcoffee.areaid= areacode.areaid and areacode.stateid=states.stateid
6. **group** **by** statename)
7. **order** **by** Pro/Tot **DESC**)
8. **where** rownum<=5;



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

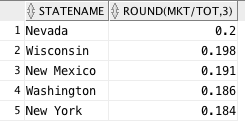
1. **Select** \* **from**(
2. **SELECT** statename,prodname, actmarkcost,
3. row\_number() over (partition **by** statename **order** **by** actmarkcost) **as** rankid
4. **from** states, areacode, factcoffee, prodcoffee
5. **where** states.stateid=areacode.stateid and factcoffee.areaid=areacode.areaid
6. and prodcoffee.productid=factcoffee.productid
7. and statename in ('Massachusetts','Texas','Illinois','Iowa','Colorado')
8. **group** **by** statename,prodname, actmarkcost)
9. **where** rankid=1;



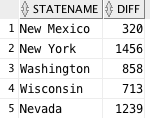
1. MARKETING EXPENSES (highest):
   1. Which 5 states have the highest marketing expenses as a percentage of sales?

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

1. **Select** \* **from**(
2. **SELECT** statename, Round(Mkt/TOT,3) **from**(
3. **SELECT** statename, sum(actsales) **as** Tot, sum(actmarkcost) **as** Mkt
4. **FROM** factcoffee, areacode, states
5. **where** factcoffee.areaid= areacode.areaid and areacode.stateid=states.stateid
6. **group** **by** statename)
7. **order** **by** Mkt/Tot **DESC**)
8. **where** rownum<=5;



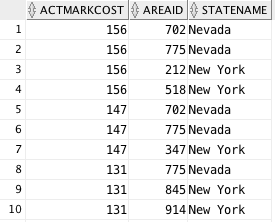
1. **SELECT** statename, Y2-Y1 **as** Diff **from**(
2. **SELECT** \* **from**(
3. **SELECT** statename, sum(actsales) sumsales, extract(year **from** factdate) year
4. **FROM** states, factcoffee, areacode
5. **where** states.stateid=areacode.stateid and factcoffee.areaid=areacode.areaid
6. and statename in ('Nevada','Wisconsin','New Mexico','Washington','New York')
7. **group** **by** statename, extract(year **from** factdate))
8. Pivot(sum(sumsales) **for** year in (2012 **as** Y1, 2013 **as** Y2)));



These marketing costs are justified because each of these states have grown significantly in sales from 2012 to 2013.

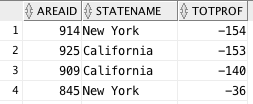
* 1. In each of these 5 states, do any area codes spend too much on marketing expenses relative to others?

1. **SELECT** actmarkcost, areacode.areaid, statename **from** factcoffee, areacode, states
2. **where** states.stateid=areacode.stateid and areacode.areaid=factcoffee.areaid
3. and statename in ('Nevada', 'Wisconsin', 'New Mexico', 'Washington', 'New York')
4. **group** **by** areacode.areaid, statename, actmarkcost
5. **order** **by** actmarkcost **DESC**;
6. --702 NV, 775 NV, 212 NY, 518 NY, 347 NY



These area codes in Nevada and New York spend the most on marketing.

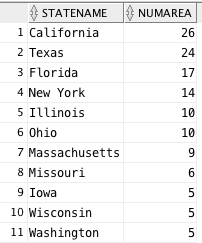
1. STRATEGY:
   1. You are in a high-level strategy meeting to discuss how to improve performance. This may involve shutting down stores in losing area codes and/or expanding in very profitable/high growth area. Evaluate the data and recommend which stores to close and where?
2. **SELECT** areacode.areaid, statename, sum(actprofit) **as** TotProf **from** factcoffee,states, areacode
3. **where** factcoffee.areaid=areacode.areaid and states.stateid=areacode.stateid
4. **group** **by** areacode.areaid,statename
5. **order** **by** TotProf;



Some area codes in larger states are not keeping up with other profit-bearing area codes within that state. It would be wise for the company to close these stores.

* 1. Where should the firm focus on expanding?

1. **select** statename, count(areacode.areaid) **as** NumArea **from** areacode, states
2. **where** areacode.stateid=states.stateid
3. **group** **by** statename
4. **order** **by** NumArea **DESC**;



The states listed in part F that are increasing their profits substantially (MA, WI, WA, etc.) should expand into more area codes.

# PART B: Office Product

The data files are available on Canvas. Here are the tables you need to create; Note: PK is primary key and FK is the foreign key.

**TASKS:**

Do the following and copy into Word document the DDL, DML, results, and any errors. Like in Part A, please copy and paste the first 10 rows if there are more than 10 rows in the answer.

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.

1. **CREATE** **Table** Managers (
2. RegID NUMBER **PRIMARY** **KEY**,
3. Region  **varchar**(10),
4. Regmanager  varchar2(10),
5. **CONSTRAINT** con\_region **CHECK** (Region IN ('East', 'South','Central','West')));

8. **CREATE** **Table** Products (
9. ProdID   number **PRIMARY** **KEY**,
10. Prodname varchar2(100),
11. Prodcat  varchar2(30) **CONSTRAINT** ch\_cat **CHECK** (prodcat IN ('Technology', 'Furniture', 'Office Supplies')),
12. Prodsubcat varchar2(30),
13. Prodcont varchar2(20) **CONSTRAINT** ch\_cont **CHECK** (prodcont IN ('Jumbo Drum', 'Medium Box','Jumbo Box','Wrap Bag','Large Box','Small Box','Small Pack')),
14. Produnitprice number(8,2),
15. prodmargin number(5,3));
17. **CREATE** **Table** Orders (
18. OrderID   number **PRIMARY** **KEY**,
19. Status varchar2(10));
21. **CREATE** **Table** Customers (
22. CustID  number **PRIMARY** **KEY**,
23. Custname    varchar2(35),
24. CustReg     number(1,0),
25. Custstate   varchar2(20),
26. Custcity    varchar2(20),
27. Custzip     number(5,0),
28. CustSeg     varchar2(15) **CONSTRAINT** ch\_seg **CHECK** (CustSeg in ('Home Office','Corporate','Small Business','Consumer')),
29. **CONSTRAINT** fk\_CustReg **FOREIGN** **KEY** (CustReg) **REFERENCES** Managers (RegID) **on** **DELETE** **CASCADE**);
31. **CREATE** **Table** OrderDet (
32. OrderID    number,
33. CustID     number,
34. ProdID     number,
35. Ordpriority varchar2(15) **CONSTRAINT** ch\_pri **CHECK** (Ordpriority in ('Low','Medium','High','Critical','Not Specified')),
36. Orddiscount number(3,2),
37. Ordshipmode varchar2(15) **CONSTRAINT** ch\_ship **CHECK** (Ordshipmode in ('Regular Air','Delivery Truck','Express Air')),
38. orddate     **DATE**,
39. Ordshipdate **Date**,
40. Ordshipcost number(5,2),
41. OrdQty      number,
42. OrdSales    number(8,2),
43. **CONSTRAINT** pk\_orderdet **PRIMARY** **KEY** (CustID, OrderID, ProdID),
44. **CONSTRAINT** fk\_cust **FOREIGN** **KEY** (CustID) **REFERENCES** Customers (CustID),
45. **CONSTRAINT** fk\_order **FOREIGN** **KEY** (OrderID) **REFERENCES** Orders (OrderID),
46. **CONSTRAINT** fk\_prod **FOREIGN** **KEY** (ProdID) **REFERENCES** Products (ProdID));

QUESTION 2: ORDER Cancellations

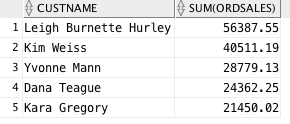
1. What fraction of the orders was cancelled?
2. **Select** Round(Count(Status)/sum(Total),4) **as** FracCancelled **from**(
3. **SELECT** Count(OrderID) **as** Total, status **from** orders
4. **group** **by** status)
5. **where** status='Returned';



1. What were the sales from cancelled orders?
2. **Select** status, sum(ordsales) **as** CancelledSales **from** orderdet, orders
3. **where** status='Returned' and orderdet.orderid=orders.orderid
4. **Group** **by** status;

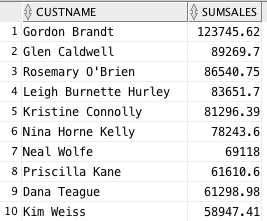


1. Who are the top five customers in terms of cancelled orders?
2. **SELECT** \* **from**(
3. **select** custname, sum(ordsales) **from** orderdet, orders, customers
4. **where** customers.custid=orderdet.custid and orderdet.orderid=orders.orderid
5. and status='Returned'
6. **group** **by** custname
7. **order** **by** sum(ordsales) **DESC**)
8. **where** rownum<=5;

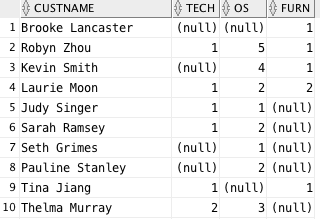


QUESTION 3: CUSTOMER related:

1. Who are the top 10 customers in terms of revenues generated?
2. **select** custname, sum(ordsales) **as** SumSales **from** orderdet, orders, customers
3. **where** customers.custid=orderdet.custid and orderdet.orderid=orders.orderid
4. **group** **by** custname
5. **order** **by** sum(ordsales) **DESC**;



1. Are there customers who buy mostly some categories of products and there is a potential for them to buy other product categories?
2. **select** \* **from**(
3. **select** custname, prodcat, count(prodcat) **as** Cat **from** products, customers, orderdet
4. **where** products.prodid=orderdet.prodid and customers.custid=orderdet.custid
5. **group** **by** custname, prodcat
6. **order** **by** custname)
7. pivot(sum(Cat) **for** prodcat in ('Technology' **as** tech, 'Office Supplies' **as** os, 'Furniture' **as** furn));



It looks like most customers have preference in categories, but there could be some overlap in office supplies and the combination of technology and furniture. When people buy a computer, they typically need a desk or chair to go with it, which encompasses furniture or office supplies.

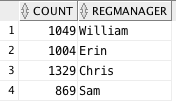
QUESTION 4: There are differences in the actual (theoretical) price ((unit price \* number of units\*(1-discount) + shipping cost) and the actual sales for all products. There are some discounts and shipping costs. Yet, there are discrepancies in the theoretical sales and actual sales.

1. How much more or less are the actual sales value compared to the theoretical sales value?
2. **SELECT** round(AVG(diff),4) **as** AverageDiff **FROM**(
3. **Select** prodid, Theo-Totsales **as** Diff **FROM**(
4. **SELECT** prodid, ((produnitprice\*ordqty)\*(1-orddiscount)+ordshipcost) **as** theo, Totsales **FROM**(
5. **SELECT** orderdet.prodid, orddiscount, sum(ordsales) TotSales, produnitprice, ordshipcost, ordqty
6. **FROM** products, orderdet
7. **where** products.prodid=orderdet.prodid
8. **group** **by** orderdet.prodid, orddiscount, produnitprice, ordqty, ordshipcost)
9. **group** **by** prodid, ((produnitprice\*ordqty)\*(1-orddiscount)+ ordshipcost), TotSales));



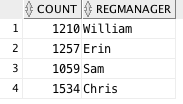
On average, actual sales were $11.59 less than the theoretical value.

1. Are certain managers generally pricing more or less than theoretical sales? Analyze the differences based on the regions/managers.
2. **SELECT** COUNT(REGMANAGER) **AS** COUNT, REGMANAGER **FROM**(
3. **SELECT** \* **From**(
4. **Select** regmanager, prodid, Theo-Totsales **as** Diff **FROM**(
5. **SELECT** regmanager, prodid, ((produnitprice\*ordqty)\*(1-orddiscount)+ordshipcost) **as** theo, Totsales **FROM**(
6. **SELECT** orderdet.prodid, regmanager,orddiscount, sum(ordsales) TotSales, produnitprice, ordshipcost, ordqty
7. **FROM** products, orderdet, customers, managers
8. **where** products.prodid=orderdet.prodid and managers.regid=customers.custreg and orderdet.custid=customers.custid
9. **group** **by** regmanager, orderdet.prodid, orddiscount, produnitprice, ordqty, ordshipcost)
10. **group** **by** regmanager, prodid, ((produnitprice\*ordqty)\*(1-orddiscount)+ ordshipcost), TotSales)
11. **order** **by** regmanager)
12. **where** Diff>0)
13. **GROUP** **BY** REGMANAGER;



This query illustrates the number of times each manager had overall sales for an item less than the theoretical sales.

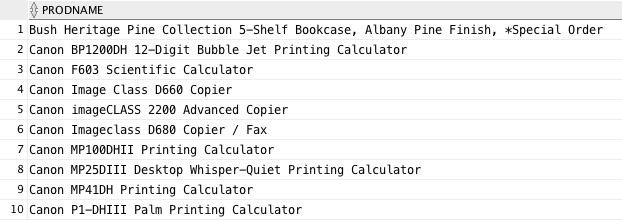
1. **SELECT** COUNT(REGMANAGER) **AS** COUNT, REGMANAGER **FROM**(
2. **SELECT** \* **From**(
3. **Select** regmanager, prodid, Theo-Totsales **as** Diff **FROM**(
4. **SELECT** regmanager, prodid, ((produnitprice\*ordqty)\*(1-orddiscount)+ordshipcost) **as** theo, Totsales **FROM**(
5. **SELECT** orderdet.prodid, regmanager,orddiscount, sum(ordsales) TotSales, produnitprice, ordshipcost, ordqty
6. **FROM** products, orderdet, customers, managers
7. **where** products.prodid=orderdet.prodid and managers.regid=customers.custreg and orderdet.custid=customers.custid
8. **group** **by** regmanager, orderdet.prodid, orddiscount, produnitprice, ordqty, ordshipcost)
9. **group** **by** regmanager, prodid, ((produnitprice\*ordqty)\*(1-orddiscount)+ ordshipcost), TotSales)
10. **order** **by** regmanager)
11. **where** Diff<0)
12. **GROUP** **BY** REGMANAGER;



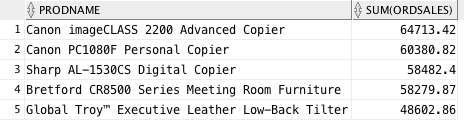
This query represents the number of times each manager had overall sales of a product that were greater than theoretical sales. Each manager seems to price items greater than their theoretical sales value.

QUESTION 5: these are product related questions:

1. Products have numbers within its name. Identify the product names with digits in their name. (hint: use REGEXP\_LIKE)
2. **Select** prodname **from** products
3. **where** REGEXP\_LIKE (prodname, '\d');



1. Which are the top 5 selling products during the year 2011?
2. **Select** \* **from**(
3. **select** prodname, sum(ordsales)
4. **from** orderdet, products
5. **where** orderdet.prodid=products.prodid and extract(year **from** orddate)=2011
6. **group** **by** prodname
7. **order** **by** sum(ordsales) **DESC**)
8. **where** rownum<=5;



1. Which are the top 10 products with greatest total profit margin? (i.e., sales\*margin).
2. **Select** \* **from**(
3. **Select** prodname, prodmargin\*Sales **from**(
4. **Select** prodname, prodmargin, sum(ordsales) **as** Sales
5. **from** products, orderdet
6. **where** products.prodid=orderdet.prodid
7. **group** **by** prodname, prodmargin)
8. **Order** **by** prodmargin\*Sales **DESC**)
9. **where** rownum<=10;



1. Identify the worst five products in terms of sales?
2. **select** products.prodid, prodname, sum(ordsales)
3. **from** orderdet, products
4. **where** orderdet.prodid=products.prodid
5. **group** **by** products.prodid, prodname
6. **order** **by** sum(ordsales)
7. **fetch** **first** 5 **rows** **only**;

