SQL Assignment 2 – Gihani Dissanayake

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.

A. 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.

```
with q1 as (
select s.statename, F.areaid, sum(actsales) as sumsales
from states S, areacode A, factcoffee F
where a.areaid =f.areaid
   and s.stateid = a.stateid
group by s.statename, f.areaid),

q2 as (
select statename, round(avg(sumsales),2) as avgsales
from q1
group by statename)

select q1.statename, q1.areaid, sumsales, avgsales
from q1, q2
where q1.statename = q2.statename
   and sumsales > 1.1*avgsales;
```

	AREAID	STATENAME	SUMSALES
1	775	Nevada	15890
2	702	Nevada	14809
3	435	Utah	13592
4	503	Oregon	8801
5	505	New Mexico	8106
6	603	New Hampshire	7596
7	563	Iowa	7557
8	970	Colorado	7423
9	773	Illinois	7107
10	515	Iowa	7062

2. ***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.***

```
WITH Q1 AS (
SELECT SUM(F.ACTPROFIT) AS TOTPROFITS, SUM(F.ACTSALES) AS SUMSALES, P.PRODNAME
FROM factcoffee F, prodcoffee P
WHERE P.PRODUCTID = F.PRODUCTID
GROUP BY P.PRODNAME),

Q2 as(
SELECT Round(Q1.TOTPROFITS/Q1.SUMSALES,2) AS ProfitMargin, Q1.PRODNAME
FROM Q1)

SELECT Q1.PRODNAME, Q2.PROFITMARGIN, Q1.SUMSALES, Q1.TOTPROFITS
FROM Q1, Q2
WHERE Q2.PROFITMARGIN >= .15
   AND Q1.PRODNAME = Q2.PRODNAME
ORDER BY SUMSALES DESC;
```

	PRODNAME	PROFITMARGIN	SUMSALES	TOTPROFITS
1	Colombian	0.43	128311	55804
2	Lemon	0.31	95926	29869
3	Caffe Mocha	0.21	84904	17678
4	Decaf Espresso	0.38	78162	29502
5	Chamomile	0.36	75578	27231
6	Darjeeling	0.4	73151	29053
7	Earl Grey	0.36	66772	24164
8	Decaf Irish Cream	0.22	62248	13989
9	Caffe Latte	0.32	35899	11375
10	Mint	0.17	35710	6154
11	Amaretto	0.19	26269	4890
12	Regular Espresso	0.42	24031	10065

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

```
With Leaves as (
SELECT F.areaid, P.prodline, SUM(F.actprofit) AS totprofits
FROM factcoffee F, prodcoffee P
WHERE F.productid = P.productid
   AND extract(year from F.factdate) = 2012
   AND P.prodline = 'Leaves'
GROUP BY F.areaid, P.prodline
ORDER BY F.AREAID ASC),

Beans as (
SELECT F.areaid, P.prodline, SUM(F.actprofit) AS totprofits
FROM factcoffee F, prodcoffee P
WHERE F.productid = P.productid
```

```
   AND extract(year from F.factdate) = 2012
   AND P.prodline = 'Beans'
GROUP BY F.areaid, P.prodline
ORDER BY F.AREAID ASC)

Select  B.areaid, round((L.totprofits-B.totprofits)/abs(B.totprofits),2) as ProfitRatio, L.totprofits as LeavesP, B.totprofits
as BeansP
FROM Leaves L, Beans B
Where B.areaid = L.areaid
   AND L.totprofits/B.totprofits > 1.2
Order by Profitratio DESC;
```

	AREAID	PROFITRATIO	LEAVESP	BEANSP
1	563	12.2	1808	137
2	641	11.89	1676	130
3	319	8.5	1577	166
4	515	7.03	1341	167
5	712	6.54	1818	241
6	775	4.27	2041	387
7	330	3.45	227	51
8	702	3.35	1557	358
9	516	2.69	866	235
10	315	2.56	267	75

B. 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?***

```
--just because the absolute profit decline is the greatest does not mean the percent decline is also highest
With q1 as (
Select A.areaid, statename, Extract(Year from factdate) as Years, sum(actprofit) as TotProfits
FROM areacode A, factcoffee F, states S
WHERE A.areaid = F.areaid
   and S.stateid = A.STATEID
GROUP BY A.AREAID, statename, Extract(Year from factdate)),

Q2 AS (
SELECT AREAID, STATENAME, NVL(FYEAR,0) AS FYEAR, NVL(SYEAR,0) AS SYEAR
FROM Q1
PIVOT (
 Sum(TotProfits)
 FOR Years IN (2012 as FYear,2013 as SYear)))

SELECT Areaid, statename, Fyear, SYear, round(100*(Syear-Fyear)/abs(Fyear),2) as PerInc, (Syear-Fyear) as TotInc
from Q2
WHERE  Fyear <> 0
 and syear <> 0
ORDER BY Totinc ASC
```

FETCH FIRST 5 ROWS ONLY;

	AREAID	STATENAME	FYEAR	SYEAR	PERINC	TOTINC
1	845	New York	1023	-1059	-203.52	-2082
2	508	Massachusetts	1499	368	-75.45	-1131
3	626	California	953	-111	-111.65	-1064
4	712	Iowa	2059	1020	-50.46	-1039
5	631	New York	1559	749	-51.96	-810

2. ***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.***

The decline in profit is not consistent between products in a single area code. I am assuming it is only worth considering products that were sold in 2012 and 2013. This is because a decline or increase in 100% should not influence the result unless the product was sold both years.

```
With q1 as (
Select A.areaid, statename, Extract(Year from factdate) as Years, sum(actprofit) as TotProfits
FROM areacode A, factcoffee F, states S
WHERE A.areaid = F.areaid
   and S.stateid = A.STATEID
GROUP BY A.AREAID, statename, Extract(Year from factdate)),

Q2 AS (
SELECT AREAID, STATENAME, NVL(FYEAR,0) AS FYEAR, NVL(SYEAR,0) AS SYEAR
FROM Q1
PIVOT (
  Sum(TotProfits)
  FOR Years IN (2012 as FYear,2013 as SYear))),

Q3 as (
SELECT Areaid, statename, Fyear, SYear, round(100*(Syear-Fyear)/abs(Fyear),2) as PerInc, (Syear-Fyear) as TotInc
from Q2
WHERE  Fyear <> 0
ORDER BY Totinc ASC
FETCH FIRST 5 ROWS ONLY),

q4 as (
Select A.areaid, statename, productid, Extract(Year from factdate) as Years, sum(actprofit) as TotProfits
FROM areacode A, factcoffee F, states S
WHERE A.areaid = F.areaid
   and S.stateid = A.STATEID
GROUP BY A.AREAID, statename, F.productid, Extract(Year from factdate)),

Q5 AS (
SELECT AREAID, STATENAME, productid, NVL(FYEAR,0) AS FYEAR, NVL(SYEAR,0) AS SYEAR
FROM Q4
PIVOT (
  Sum(TotProfits)
```

```
  FOR Years IN (2012 as FYear,2013 as SYear))),

Q6 as (
SELECT Areaid, statename, productid, Fyear, SYear, round(100*(Syear-Fyear)/abs(Fyear),2) as ProductPerInc,
(Syear-Fyear) as ProductTotInc
from Q5
WHERE  Fyear <> 0
   and  Syear <> 0
ORDER BY ProductTotInc ASC)

select q3.Areaid, q3.statename, productid, ProductPerInc, ProductTotInc
from q3, q6
where q3.areaid = q6.Areaid
   and q3.statename = q6.statename
ORDER BY ProductTotInc ASC;
```

As shown in the table, there are only 8 instances of products with declining profits in the area codes found in part 1. By far the worst decline is product 5 in NYC area code 845, maybe that product should not be sold there. The same can be said about product 10 in area code 845. Product 7 in area code 631 has a high absolute decrease but the percent change isn't as high as others. As such, maybe it is worth keeping the product around but keep a watch on it in that area code.

	AREAID	STATENAME	PRODUCTID	PRODUCTPERINC	PRODUCTTOTINC
1	845	New York	5	-378.71	-765
2	845	New York	10	-369.91	-418
3	631	New York	7	-49.09	-242
4	626	California	3	-37.23	-51
5	712	Iowa	3	-65.96	-31
6	508	Massachusetts	7	-7.97	-24
7	508	Massachusetts	9	-25	-9
8	508	Massachusetts	11	-1.54	-1
9	712	Iowa	2	11.11	1
10	712	Iowa	11	1.15	5

C. 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.***

```
SELECT S.statename, sum(F.actprofit - F.budprofit) as ProfitSurplus, sum(F.actsales - F.budsales) as SalesSurplus
FROM factcoffee F, states S, areacode A
WHERE S.stateid = A.stateid
   AND A.areaid = F.areaid
   AND extract(year from F.factdate) = 2012
GROUP BY S.statename
order by salessurplus desc;
```

Nevada was by far the best at surpassing sales expectations – nearly double Iowa's in second place. All states except Texas and Illinois were positive, which means that those two states did not sell as much as they expected.

	STATENAME	PROFITSURPLUS	SALESSURPLUS
1	Nevada	-987	6190
2	Iowa	-109	3298
3	New York	-1018	2908
4	Oregon	-1208	2282
5	California	-1226	2102

By sorting by profitsurplus, we can see that none of the states surpassed their budgeted profits, but Iowa was by far the most accurate in that regard. At the bottom of the list, Utah's profit estimate was nearly $2,000 greater than the reality. Sales surplus (previous pic) is a better indicator because profits surplus were closer to the rank bel

	STATENAME	PROFITSURPLUS	SALESSURPLUS
1	Iowa	-109	3298
2	Massachusetts	-588	1157
3	Louisiana	-858	746
4	Connecticut	-867	1491
5	Florida	-871	2041

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

I'm taking the top 5 found by sales surplus because it seems to be a more accurate figure than profit surplus, where everyone was overestimating.

```
with q1 as (
SELECT S.statename, sum(F.actprofit - F.budprofit) as StateProfitSurplus, sum(F.actsales - F.budsales) as
StateSalesSurplus
FROM factcoffee F, states S, areacode A
WHERE S.stateid = A.stateid
   AND A.areaid = F.areaid
   AND extract(year from F.factdate) = 2012
GROUP BY S.statename
order by Statesalessurplus desc
FETCH FIRST 5 ROWS ONLY),

q2 as (
SELECT S.statename, A.areaid, sum(F.actprofit - F.budprofit) as ProfitSurplus, sum(F.actsales - F.budsales) as
SalesSurplus
FROM factcoffee F, states S, areacode A
WHERE S.stateid = A.stateid
   AND A.areaid = F.areaid
   AND extract(year from F.factdate) = 2012
GROUP BY S.statename, A.areaid
order by salessurplus desc)
```

```
select q1.statename, q2.areaid, q2.ProfitSurplus, q2.SalesSurplus
from q1, q2
where q1.statename = q2.statename
order by q2.SalesSurplus desc;
```

The most significant absolute sales surpluses came from Nevada. Area codes 702 and 775, are the most significant, as they are several orders of magnitude larger than the mean sales surplus. To increase the significant values for a greater number or area codes, double the average sales surplus (671), so any value (the top 7) greater than that number is undoubtedly significant.

	STATENAME	AREAID	PROFITSURPLUS	SALESSURPLUS
1	Nevada	702	-425	3540
2	Nevada	775	-562	2650
3	Iowa	563	5	869
4	Oregon	503	-391	840
5	Oregon	971	-543	775
6	Iowa	641	6	685
7	Iowa	712	-41	671
8	Oregon	541	-274	667
9	California	562	70	641
10	Iowa	319	-47	562

By changing the order by to q2.profitsurplus, we get the following. Even though no state had a net positive profit surplus, when broken down into area code, there were 15 with a positive absolute profit surplus. The most significant are the top 3, as they are several orders of magnitude larger than the mean profit surplus. Since the average product surplus is -91, I would consider any positive profit surplus as significant since that is a rather rare outcome.

	STATENAME	AREAID	PROFITSURPLUS	SALESSURPLUS
1	New York	607	95	512
2	California	562	70	641
3	California	310	58	355
4	California	213	31	298
5	California	858	28	138
6	California	760	26	275
7	New York	516	21	266
8	California	831	15	206
9	California	530	11	202
10	California	619	11	108

D. PRODUCT related:

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

```
with q1 as (
select S.statemkt, F.productid, Extract(Year from factdate) as Years, sum(F.actprofit) as TotProdProfits
from factcoffee F, states S, areacode A
where A.areaid = F.areaid
   and A.stateid = S.stateid
group by S.statemkt, F.productid, Extract(Year from factdate)),

q2 as (
select statemkt, productid, nvl(FYear,0) AS FYear, nvl(SYear,0) AS SYear
FROM Q1
PIVOT (
  Sum(TotProdProfits)
  FOR Years IN (2012 as FYear,2013 as SYear))),

q3 as (
select statemkt, productid, Fyear, SYear, round(100*(Syear-Fyear)/abs(Fyear),2) as PerInc, (Syear-Fyear) as TotInc,
RANK() OVER (PARTITION BY statemkt
   ORDER BY (Syear-Fyear) DESC) as prodrank
from Q2
WHERE  Fyear <> 0
 and syear <> 0
ORDER BY Totinc ASC)

select statemkt, productid, PerInc, TotInc, prodrank
from q3
where prodrank = 1
order by totinc desc;
```

	STATEMKT	PRODUCTID	PERINC	TOTINC	PRODRANK
1	East	2	44.91	4998	1
2	Central	5	45.3	2704	1
3	West	9	45.07	2413	1
4	South	2	45.16	1615	1

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

```
with q1 as (
select S.statemkt, F.productid, Extract(Year from factdate) as Years, sum(F.actsales) as TotProdsales
from factcoffee F, states S, areacode A
where A.areaid = F.areaid
   and A.stateid = S.stateid
group by S.statemkt, F.productid, Extract(Year from factdate)),

q2 as (
select statemkt, productid, nvl(FYear,0) AS FYear, nvl(SYear,0) AS SYear
FROM Q1
PIVOT (
  Sum(TotProdsales)
```

```
  FOR Years IN (2012 as FYear,2013 as SYear))),

q3 as (
select statemkt, productid, Fyear, SYear, round(100*(Syear-Fyear)/abs(Fyear),2) as PerInc, (Syear-Fyear) as TotInc,
RANK() OVER (PARTITION BY statemkt
   ORDER BY (Syear-Fyear) DESC) as prodrank
from Q2
WHERE  Fyear <> 0
 and syear <> 0
ORDER BY Totinc ASC)

select statemkt, productid, PerInc, TotInc, prodrank
from q3
where prodrank = 1
order by totinc desc;
```

	STATEMKT	PRODUCTID	PERINC	TOTINC	PRODRANK
1	East	2	4.2	975	1
2	Central	8	4.48	801	1
3	West	9	4.22	667	1
4	South	2	4.29	455	1

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

```
with q1 as (
select P.prodtype, P.prodname, Extract(Year from factdate) as Years, sum(F.actsales) as TotProdsales
from factcoffee F, prodcoffee P
where P.productid = F.productid
group by P.prodtype, P.prodname, Extract(Year from factdate)),

q2 as (
select prodtype, prodname, nvl(FYear,0) AS FYear, nvl(SYear,0) AS SYear
FROM Q1
PIVOT (
  Sum(TotProdsales)
  FOR Years IN (2012 as FYear,2013 as SYear)))

select prodtype, prodname, Fyear, SYear, round(100*(Syear-Fyear)/abs(Fyear),2) as PerInc, (Syear-Fyear) as TotInc
from Q2
WHERE  Fyear <> 0
    and syear <> 0
ORDER BY prodtype;
```

Interestingly, there isn't much of a difference in the percent increase of sales of each product, even within each type, as they are all around 4-5%.
Corresponding to that, there are differences between the total increases in sales between products of the same type. Colombian coffee, the item with the greatest sales in 2012 also has the greatest sales in 2012 and the greatest total increase overall. This shows that there are definite differences between the total volume of each product in a category, but all products are increasing in popularly and sales in 2013 over 2012.

	PRODTYPE	PRODNAME	FYEAR	SYEAR	PERINC	TOTINC
1	Coffee	Amaretto	12841	13428	4.57	587
2	Coffee	Colombian	62824	65487	4.24	2663
3	Coffee	Decaf Irish Cream	30469	31779	4.3	1310
4	Espresso	Caffe Latte	17559	18340	4.45	781
5	Espresso	Caffe Mocha	41537	43367	4.41	1830
6	Espresso	Decaf Espresso	38240	39922	4.4	1682
7	Espresso	Regular Espresso	11750	12281	4.52	531
8	Herbal Tea	Chamomile	36969	38609	4.44	1640
9	Herbal Tea	Lemon	46956	48970	4.29	2014
10	Herbal Tea	Mint	17480	18230	4.29	750
11	Tea	Darjeeling	35799	37352	4.34	1553
12	Tea	Earl Grey	32670	34102	4.38	1432
13	Tea	Green Tea	16065	16785	4.48	720

E. MARKETING EXPENSES (LOWEST):

1. Which top 5 states have the lowest market expenses as a percentage of their sales?

```
select S.statename, round(sum(F.actmarkcost)/sum(F.actsales)*100,2) as PercMarketExp
from states S, factcoffee F, areacode A
where a.stateid = s.stateid
   and a.areaid = f.areaid
group by S.statename
order by PercMarketExp asc
fetch first 5 rows only;
```

	STATENAME	PERCMARKETEXP
1	Massachusetts	11.44
2	Texas	12.69
3	Illinois	12.82
4	Iowa	13.87
5	Colorado	14.13

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

```
select S.statename, round(sum(F.actprofit)/sum(F.actsales)*100,2) as PercProfits
from states S, factcoffee F, areacode A
where a.stateid = s.stateid
   and a.areaid = f.areaid
group by S.statename
order by PercProfits desc
```

fetch first 5 rows only;

The top 5 are the same, but the order is slightly different.

	STATENAME	PERCPROFITS
1	Massachusetts	54.87
2	Illinois	44.1
3	Texas	42.14
4	Iowa	40.57
5	Colorado	36.83

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

```
with q1 as (
select S.statename, round(sum(F.actmarkcost)/sum(F.actsales)*100,2) as PercMarketExp
from states S, factcoffee F, areacode A
where a.stateid = s.stateid
   and a.areaid = f.areaid
group by S.statename
order by PercMarketExp asc
fetch first 5 rows only),

q2 as (
select S.statename, F.productid, round(sum(F.actmarkcost)/sum(F.actsales)*100,2) as PercMarketExpProduct
from states S, factcoffee F, areacode A
where a.stateid = s.stateid
   and a.areaid = f.areaid
group by S.statename, F.productid)

select Q1.statename, Q2.productid, Q2.PercMarketExpProduct
from q1, q2
where q1.statename = q2.statename
order by Q2.PercMarketExpProduct asc;
```

Product 6 in Iowa and Product 2 in Massachusetts have by far the lowest percent of marketing expenses relative to sales.

	STATENAME	PRODUCTID	PERCMARKETEXPPRODUCT
1	Iowa	6	0.76
2	Massachusetts	2	3.79
3	Iowa	5	9.88
4	Massachusetts	13	9.94
5	Massachusetts	11	10.38

When I inverted the order, I found that product 5 in Massachusetts and product 9 in Colorado has percent of marketing expenses relative to sales that were significantly higher than the other products in those states at over 30%. Maybe there should be less marketing spent on those items in those areas.

	STATENAME	PRODUCTID	PERCMARKETEXPPRODUCT
1	Massachusetts	5	38.45
2	Colorado	9	36.73
3	Iowa	12	16.61
4	Colorado	11	16.26
5	Massachusetts	9	16.26
6	Colorado	13	13.84

F. MARKETING EXPENSES (highest):

1. Which 5 states have the highest marketing expenses as a percentage of sales?

```
select S.statename, round(sum(F.actmarkcost)/sum(F.actsales)*100,2) as PercMarketExp
from states S, factcoffee F, areacode A
where a.stateid = s.stateid
   and a.areaid = f.areaid
group by S.statename
order by PercMarketExp desc
fetch first 5 rows only;
```

	STATENAME	PERCMARKETEXP
1	Nevada	20.04
2	Wisconsin	19.81
3	New Mexico	19.14
4	Washington	18.61
5	New York	18.41

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

```
select S.statename, round(sum(F.actmarkcost)/sum(F.actsales)*100,2) as PercMarketExp,
round(sum(F.actprofit)/sum(F.actsales)*100,2) as PercProfits, round(sum(F.actmarkcost)/sum(F.actprofit)*100,2) as
PercMarketExpofProfit
from states S, factcoffee F, areacode A
where a.stateid = s.stateid
   and a.areaid = f.areaid
group by S.statename
order by PercMarketExp desc
fetch first 5 rows only;
```

High marketing expenses is most significant as a percentage of sales, and as a percentage of profit, because this accounts for high marketing expenses that in turn causes a huge jump in sales. We can clearly conclude from the below table that the profit from the sales is less than the marketing expenses to get to that sale in Nevada and New

Mexico. As such, high marketing expenses are not justified in these states. We can show a correlation between market expenses as a percent of sales and market expenses as a percent of profit

	STATENAME	PERCMARKETEXP	PERCPROFITS	PERCMARKETEXPOFPROFIT
1	Nevada	20.04	17.65	113.56
2	Wisconsin	19.81	26.31	75.29
3	New Mexico	19.14	5.03	380.73
4	Washington	18.61	29.3	63.52
5	New York	18.41	28.36	64.91

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

```
with q1 as (
select S.statename, round(sum(F.actmarkcost)/sum(F.actsales)*100,2) as PercMarketExp1,
round(sum(F.actprofit)/sum(F.actsales)*100,2) as PercProfits1, round(sum(F.actmarkcost)/sum(F.actprofit)*100,2) as
PercMarketExpofProfit1
from states S, factcoffee F, areacode A
where a.stateid = s.stateid
   and a.areaid = f.areaid
group by S.statename
order by PercMarketExp1 desc
fetch first 5 rows only),

q2 as (
select S.statename, a.areaid, round(sum(F.actmarkcost)/sum(F.actsales)*100,2) as PercMarketExp,
round(sum(F.actprofit)/sum(F.actsales)*100,2) as PercProfits, round(sum(F.actmarkcost)/sum(F.actprofit)*100,2) as
PercMarketExpofProfit
from states S, factcoffee F, areacode A
where a.stateid = s.stateid
   and a.areaid = f.areaid
group by S.statename, a.areaid
order by PercMarketExp desc)

select q1.statename, q2.areaid, PercMarketExp, PercProfits, PercMarketExpofProfit
from q1, q2
where q1.statename = q2.statename
order by PercMarketExp desc;
```

The average of market expenses as a percent of sales is about 19% and the average of profit as a percent of sales is about 26% for these 5 states. Within that set, area codes 914 and 845 are spending too much on marketing expenses. They have the highest market expenses as a percent of sales, but the most significant attribute is that even though their marketing expenses are high, their relative profitability is actually negative Though this shows correlation and not causation, these values are significant enough to reduce the marketing expenses in these area codes.

If even greater cost reductions were desired, I'd recommend reducing the marketing expenses of the top 6 rows. The market expenses as a percent of sales and as a percent of profit are high. The former value is greater than the mean of the sample set, which is already a skewed selection, and the absolute value of the latter is greater than 100%,

meaning that the marketing expenses are greater than the profits of selling in those area codes. Though this again shows correlation not causation, it could be the basis of a cost cutting strategy.

	STATENAME	AREAID	PERCMARKETEXP	PERCPROFITS	PERCMARKETEXPOFPROFIT
1	New York	914	28.71	-4.05	-709.74
2	New York	845	25.34	-0.67	-3763.89
3	New York	212	24.07	17.37	138.59
4	Wisconsin	608	23.3	22.92	101.7
5	New York	347	23.25	17.43	133.44
6	Nevada	775	20.65	14.31	144.34
7	Wisconsin	262	20	26.49	75.49
8	New York	585	19.95	27.17	73.44
9	New York	607	19.95	23.01	86.7
10	Wisconsin	715	19.91	26.57	74.94

G. 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?***

```
with q1 as (
select S.statename, A.areaid, Extract(Year from factdate) as Years, sum(F.actsales) as TotProdsales
from factcoffee F, states S, areacode A
where A.areaid = F.areaid
   and A.stateid = S.stateid
group by S.statename, A.areaid, Extract(Year from factdate)),

q2 as (
select statename, areaid, nvl(FYear,0) AS FYear, nvl(SYear,0) AS SYear
FROM Q1
PIVOT (
  Sum(TotProdsales)
  FOR Years IN (2012 as FYear,2013 as SYear))),

q3 as (
select statename, areaid, Fyear, SYear, round(100*(Syear-Fyear)/abs(Fyear),2) as PerIncSales, (Syear-Fyear) as
TotIncSales
from Q2
WHERE  Fyear <> 0
 and syear <> 0)

select statename, areaid, PerIncSales, TotIncSales
from q3
order by totincSales asc;
```

Area codes with the least total increase in sales (most negative)

	STATENAME	AREAID	PERINCSALES	TOTINCSALES
1	Iowa	712	-62.87	-3666
2	California	707	-52.47	-2656
3	California	626	-82.99	-2420
4	Illinois	217	-69.81	-2349
5	Massachusetts	508	-70.63	-2333
6	New York	845	-57.31	-2148
7	New York	646	-54.14	-1891
8	New York	631	-48.33	-1870
9	Illinois	630	-24.63	-1583
10	California	916	-48.68	-1545

The tables above show the area codes and the corresponding states that have the lowest increase in sales from 2012 to 2013. In an ideal world, more than two years of data would be considered when evaluating whether to expand stores and/or close down stores, but that data is not available. The average total increase in sales is $106, and the average percent increase in sales is 35%. As such, the top 6 in this table have total declines in sales that are very substantial, exponentially more than the mean and have high percent declines as well. I would recommend closing those stores.

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

Area codes with the highest total increases in sales

	STATENAME	AREAID	PERINCSALES	TOTINCSALES
1	Illinois	773	134.86	4081
2	California	530	233.93	3116
3	Massachusetts	774	220.16	2697
4	Iowa	515	48.64	2311
5	New York	914	287.45	2245
6	California	323	724.16	2158
7	New York	518	91.35	2143
8	Colorado	720	42.1	2016
9	New York	516	73.54	1946
10	Connecticut	203	60.47	1842

Area codes with the highest percent increase in sales.

	STATENAME	AREAID	PERINCSALES	TOTINCSALES
1	California	323	724.16	2158
2	Florida	352	705.66	1122
3	California	510	600	1404
4	Texas	432	387.11	1231
5	Texas	806	353.16	671
6	New York	914	287.45	2245
7	Texas	817	235.22	748
8	California	530	233.93	3116
9	Massachusetts	774	220.16	2697
10	California	619	217.31	1017

I would recommend expanding stores in area codes 773, 530, 774, 914, and 323 which are area codes with percent increase of sales over 100% and very high total increases in sales. If you only have to pick one, I would recommend either the 323 area code which ash the highest percent increase in sales or area code 773, where the total increase of sales was exceptionally higher than the rest. The former could be best because a boom in demand or a drop in competition could push the store past capacity. The latter could be best because high growth continues to grow and would thus justify the high expense of opening a new store. Despite these conclusions, it's quite hard to make this decision with only two years of data.

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.

TABLE: MANAGERS (REGID is the PK)

COLUMN_NAME	DATA_TYPE
REGID	NUMBER
REGION	VARCHAR2(10 BYTE)
REGMANAGER	VARCHAR2(10 BYTE)

CONSTRAINT:
REGION can be only 'East', 'South', 'Central', 'West'.

TABLE: PRODUCTS (ProdID is the PK)

COLUMN_NAME	DATA_TYPE
PRODID	NUMBER
PRODNAME	VARCHAR2(100 B...
PRODCAT	VARCHAR2(30 BYTE)
PRODSUBCAT	VARCHAR2(30 BYTE)
PRODCONT	VARCHAR2(20 BYTE)
PRODUNITPRICE	NUMBER(7,2)
PRODMARGIN	NUMBER(5,3)

CONSTRAINTS:
PRODCAT can only be 'Technology' 'Furniture' or 'Office Supplies'
PRODCONT take on only 'Jumbo Drum', 'Medium Box', 'Jumbo Box', 'Wrap Bag', 'Large Box', 'Small Box', 'Small Pack'

TABLE: ORDERS (OrderID is the PK)

COLUMN_NAME	DATA_TYPE
ORDERID	NUMBER
STATUS	VARCHAR2(10 BYTE)

TABLE: CUSTOMERS (CustID is the PK; CustReg is the FK on delete cascade)

COLUMN_NAME	DATA_TYPE
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)

CONSTRAINT:

CUSTSEG can be only Home Office 'Corporate', 'Small Business', 'Consumer'.

TABLE: ORDERDET (OrderID (FK), CustID (FK), ProdID (FK) are together a PK; All FK are on delete restrict)

COLUMN_NAME	DATA_TYPE
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(7,2)

CONSTRAINTS
ORDPRIORITY can be 'Low', 'Medium', 'High', 'Critical', 'Not Specified'

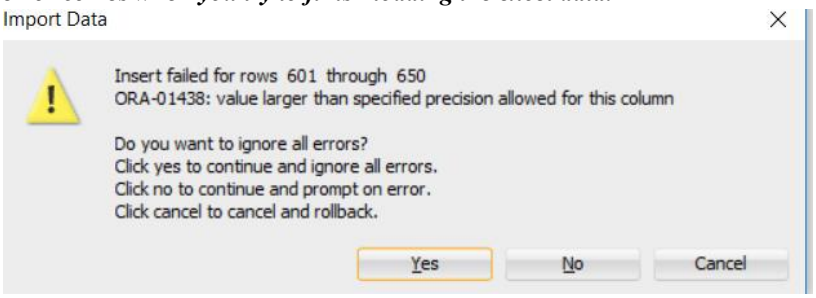
ORDSHIPMODE can be 'Regular Air', 'Delivery Truck', 'Express Air'

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.

The original Orderdet table didn't work because ordsales(7,2) was not enough so changed it to ordsales(8,2). The error comes when you try to finish loading the excel data.




```
CREATE TABLE Managers (
   regid number primary key,
   region varchar2(10 byte) CONSTRAINT ch_region CHECK (region IN ('East', 'West', 'Central', 'South')),
   regmanager varchar2(10 byte));

create table products (
   prodid number primary key,
   prodname varchar2(100 byte),
   prodcat varchar2(30 byte) constraint ch_prodcat check (prodcat in ('Technology', 'Furniture', 'Office Supplies')),
   prodsubcat varchar2(30 byte),
   prodcont varchar2(20 byte) constraint ch_prodcont check (prodcont in ('Jumbo Drum', 'Medium Box', 'Jumbo Box',
'Wrap Bag', 'Large Box', 'Small Box', 'Small Pack')),
   produnitprice number(7,2),
   prodmargin number(5,3));

create table orders(
   orderid number primary key,
   status varchar2(10 byte));

create table customers(
   custid number primary key,
   custname varchar2(35 byte),
   custreg number(1,0),
   custstate varchar2(20 byte),
   custcity varchar2(20 byte),
   custzip number(5,0),
   custseg varchar2(15 byte) constraint ch_custseg check (custseg in ('Home Office', 'Corporate', 'Small Business',
'Consumer')),
   CONSTRAINT fk_custreg FOREIGN KEY (custreg) REFERENCES managers (regid) on DELETE CASCADE);

create table orderdet(
   orderid number,
   custid number,
   prodid number,
   ordpriority varchar2(15 byte) constraint ch_ordpriority check (ordpriority in ('Low', 'Medium', 'High', 'Critical', 'Not
Specified')),
   orddiscount number(3,2),
   ordshipmode varchar2(15 byte) constraint ch_ordshipmode check (ordshipmode in ('Regular Air', 'Delivery Truck',
'Express Air')),
   orddate date,
   ordshipdate date,
```

```
    ordshipcost number(5,2),
    ordqty number,
    ordsales number(8,2),
    CONSTRAINT pk_orderdet PRIMARY KEY (orderid, custid, prodid),
    CONSTRAINT fk_orderid FOREIGN KEY (orderid) REFERENCES orders (orderid),
    CONSTRAINT fk_custid FOREIGN KEY (custid) REFERENCES customers (custid),
    CONSTRAINT fk_prodid FOREIGN KEY (prodid) REFERENCES products (prodid));
```

QUESTION 2: ORDER Cancellations

a) What fraction of the orders was cancelled?

```
--Assumption is that a cancelled order = returned order
with q1 as(
select *
from orders

pivot (
count(orderid)
for status in ('Returned' as NumReturned, null as NumNull)))
select round(NumReturned/(NumReturned+NumNull),4) as PercentReturned
from q1;
```

	PERCENTRETURNED
1	0.0093

By doing 'select NumReturned, NumNull' instead, I got the count of returned vs not

	NUMRETURNED	NUMNULL
1	60	6368

b) What were the sales from cancelled orders?

```
select * from(
select sum(D.ordsales) as TotReturnedSales
from orders O, orderdet D
where O.orderid = D.orderid
   and O.status = 'Returned');
```

	TOTRETURNEDSALES
1	308455.12

c) Who are the top five customers in terms of cancelled orders?

```
select * from(
select D.custid, sum(D.ordsales) as TotReturnedSales, count(O.status) as NumReturned
from orders O, orderdet D
where O.orderid = D.orderid
   and O.status = 'Returned'
group by D.custid
```

order by TotReturnedSales DESC
FETCH FIRST 5 ROWS ONLY);

Ordered by the returned sales volume, could have alternatively sorted by count of returned orders but since the range was only 0-7, I felt that total returned sales volume was more representative of the broad range.

	CUSTID	TOTRETURNEDSALES	NUMRETURNED
1	2107	56387.55	4
2	349	40511.19	2
3	2670	28779.13	2
4	2867	24362.25	1
5	2403	21450.02	2

QUESTION 3: CUSTOMER related:

a) Who are the top 10 customers in terms of revenues generated?

*select * from (*
select C.custname, sum(D.ordsales) as CustomerRevenue
from Orderdet D, customers C
where D.custid = C.custid
group by custname
order by sum(D.ordsales) DESC
FETCH FIRST 10 ROWS ONLY);

	CUSTNAME	CUSTOMERREVENUE
1	Gordon Brandt	123745.62
2	Glen Caldwell	89269.7
3	Rosemary O'Brien	86540.75
4	Leigh Burnette Hurley	83651.7
5	Kristine Connolly	81296.39
6	Nina Horne Kelly	78243.6
7	Neal Wolfe	69118
8	Priscilla Kane	61610.6
9	Dana Teague	61298.98
10	Kim Weiss	58947.41

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

with q1 as (
select D.custid, sum(D.ordqty) as TotQuantityS
from orderdet D, products P
where D.prodid = P.prodid
and P.prodcat = 'Office Supplies'
group by D.custid
order by custid),

```
q2 as (select D.custid, sum(D.ordqty) as TotQuantityT
from orderdet D, products P
where D.prodid = P.prodid
   and P.prodcat = 'Technology'
group by D.custid
order by custid),

q3 as (
select D.custid, sum(D.ordqty) as TotQuantityF
from orderdet D, products P
where D.prodid = P.prodid
   and P.prodcat = 'Furniture'
group by D.custid
order by custid)

select q1.custid, TotQuantityS, TotQuantityT, TotQuantityF, sum(TotQuantityS+TotQuantityT+TotQuantityF) as Allsum
from q1, q2, q3
where q1.custid = q2.custid
   and q2.custid = q3.custid
   and (abs(TotQuantityS-TotQuantityT) >= 100 or abs(TotQuantityS-TotQuantityF) >= 100 or abs(TotQuantityF-
TotQuantityT) >= 100)
group by q1.custid, TotQuantityS, TotQuantityT, TotQuantityF
order by allsum desc;
```

	CUSTID	TOTQUANTITYS	TOTQUANTITYT	TOTQUANTITYF	ALLSUM
1	2107	657	182	153	992
2	1999	374	412	162	948
3	1723	466	261	174	901
4	3079	601	235	58	894
5	699	500	164	229	893
6	1402	322	348	209	879
7	640	496	189	179	864
8	1193	601	125	129	855
9	2491	424	130	233	787
10	1106	505	80	196	781
11	3075	370	207	171	748
12	94	199	404	137	740
13	1413	481	249	8	738
14	2756	234	400	92	726
15	2882	410	204	95	709
16	272	394	197	109	700
17	3011	485	157	56	698

Customers who have bought from all three categories in the past are more likely to buy from all three in the future. Additionally, customers who have bought more total items overall are more likely to increase the number of goods that they buy. Finally, the highlighted customers are the ones who heavily buy two of the three product categories here but likely opt for a different supplier for one of the categories (either furniture or technology, as office supplies seems to be by far the most popular thing to buy here). Since it's more convenient to have a single supplier and because they already buy heavily from this company, these highlighted customers are the most likely to substantially

increase their purchases in their low category, assuming they can be persuaded to change from their status quo. As such, these customers would be great targets for a marketing or advertising campaign.

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.

a) How much more or less are the actual sales value compared to the theoretical sales value?

```
select P.prodname, sum(D.ordsales) as ActSumSales, sum(P.Produnitprice*D.ordqty*(1-D.orddiscount) +
D.ordshipcost) as ExpSumSales, (sum(D.ordsales)-(sum(P.Produnitprice*D.ordqty*(1-D.orddiscount) +
D.ordshipcost))) as DollarDiff
from orderdet D, products P
where P.prodid = D.prodid
group by P.prodname
order by DollarDiff DESC;
```

This shows the items that bring in more profit than expected.

	PRODNAME	ACTSUMSALES	EXPSUMSALES	DOLLARDIFF
1	Global Troy™ Executive Leather Low-Back Tilter	194025.64	185783.0414	8242.5986
2	Canon imageCLASS 2200 Advanced Copier	107697.73	101380.6309	6317.0991
3	Hewlett Packard LaserJet 3310 Copier	76114.96	72423.697	3691.263
4	Sharp AL-1530CS Digital Copier	86057.24	82396.7284	3660.5116
5	Riverside Palais Royal Lawyers Bookcase, Royale Cherry Finish	190195.15	187082.115	3113.035
6	Gyration Ultra Professional Cordless Optical Suite	54683.14	51673.5786	3009.5614
7	Canon PC1080F Personal Copier	102932.77	100214.253	2718.517
8	Hon Multipurpose Stacking Arm Chairs	75292.73	72601.632	2691.098
9	Canon PC1060 Personal Laser Copier	70236.28	67552.4095	2683.8705
10	Okidata ML395C Color Dot Matrix Printer	50347.15	47772.5112	2574.6388

By reversing DESC to ASC you can view the items that are the most underpriced.

	PRODNAME	ACTSUMSALES	EXPSUMSALES	DOLLARDIFF
1	Lexmark 4227 Plus Dot Matrix Printer	82415.17	88582.2912	-6167.1212
2	Bretford CR8500 Series Meeting Room Furniture	101797.12	106973.1164	-5175.9964
3	Polycom ViewStation™ ISDN Videoconferencing Unit	92916.02	97439.9368	-4523.9168
4	Hewlett-Packard Business Color Inkjet 3000 [N, DTN] Series Printers	49451.54	53715.1004	-4263.5604
5	LX 788	28436.21	32569.7684	-4133.5584
6	TimeportP7382	28554.92	32388.7642	-3833.8442
7	Chromcraft 48" x 96" Racetrack Double Pedestal Table	34501.7	38258.4048	-3756.7048
8	StarTAC 8000	26831.61	30581.0989	-3749.4889
9	M3682	28820.41	32498.1102	-3677.7002
10	2180	25274.51	28942.2721	-3667.7621

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

```
select M.regmanager, (sum(D.ordsales)-(sum(P.Produnitprice*D.ordqty*(1-D.orddiscount) + D.ordshipcost))) as
DollarDiff
from orderdet D, products P, managers M, customers C
where P.prodid = D.prodid
   and M.regid = C.custreg
   and C.custid = D.custid
group by M.regmanager
order by DollarDiff ASC;
```

Seems that the net impact of William on price is much more than the net impact of Erin. All of the managers overall bring in less than theoretical sales, resulting in lower profits, but William does so far more than the others. Another way of interpreting this result is that the amounts here show lost revenue.

	REGMANAGER	DOLLARDIFF
1	William	-12282.4158
2	Chris	-3687.4457
3	Sam	-3300.2021
4	Erin	-2521.193

QUESTION 5: these are product related questions:

***a) Products have numbers within its name. Identify the product names with digits in their name. (hint: use* REGEXP_LIKE)**

```
SELECT P.prodname
FROM products P
WHERE REGEXP_LIKE(P.prodname, '[0-9]')
```

	PRODNAME
1	Accessory36
2	Accessory37
3	Accessory39
4	Accessory4
5	Accessory41
6	Accessory6
7	Accessory8
8	Accessory9
9	Acco 3-Hole Punch
10	Acco 6 Outlet Guardian Premium Surge Suppressor

b) Which are the top 5 selling products during the year 2011?

```
select P.prodname, (D.ordqty) as NumberUnits, (D.ordsales) as SumSales
from orderdet D, products P
where D.prodid=P.prodid
   and extract(year from D.orddate) = 2011
order by numberunits desc
FETCH FIRST 5 ROWS ONLY;
```

The top 5 products by volume were not the most profitable products.

	PRODNAME	NUMBERUNITS	SUMSALES
1	Acco Perma® 3000 Stacking Storage Drawers	160	3607.56
2	Belkin 5 Outlet SurgeMaster™ Power Centers	146	8262.68
3	Xerox 1968	139	883.63
4	Xerox 1927	129	568.31
5	Newell 315	112	706.11

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

```
select P.prodname, sum(D.ordsales*P.prodmargin) as TotProfitMargin, sum(D.ordsales) as SumSales,
round(avg(P.prodmargin),2) as PercMargin
```

from products P, orderdet D
where P.prodid = D.prodid
group by P.prodname
*order by sum(D.ordsales*P.prodmargin) DESC*
FETCH FIRST 10 ROWS ONLY;

	PRODNAME	TOTPROFITMARGIN	SUMSALES	PERCMARGIN
1	Riverside Palais Royal Lawyers Bookcase, Royale Cherry Finish	117920.993	190195.15	0.62
2	Global Troy™ Executive Leather Low-Back Tilter	116415.384	194025.64	0.6
3	Bretford CR8500 Series Meeting Room Furniture	71134.7789	101797.12	0.69
4	Chromcraft Bull-Nose Wood 48" x 96" Rectangular Conference Tables	63456.0774	92208.46	0.7
5	Bretford CR4500 Series Slim Rectangular Table	54207.7652	84024.22	0.63
6	Canon PC1080F Personal Copier	51466.385	102932.77	0.5
7	Hon 2090 "Pillow Soft" Series Mid Back Swivel/Tilt Chairs	48947.0592	62752.64	0.78
8	BoxOffice By Design Rectangular and Half-Moon Meeting Room Tables	48811.1858	63825.89	0.76
9	Lexmark 4227 Plus Dot Matrix Printer	45328.3435	82415.17	0.55
10	Hon Multipurpose Stacking Arm Chairs	44422.7107	75292.73	0.59

d) Identify the worst five products in terms of sales?

select P.prodname, sum(D.ordqty) as NumberUnits, sum(D.ordsales) as SumSales
from orderdet D, products P
where D.prodid=P.prodid
group by P.prodname
order by sumsales asc
FETCH FIRST 5 ROWS ONLY;

	PRODNAME	NUMBERUNITS	SUMSALES
1	Alliance Rubber Bands	3	7.43
2	*Staples* Packaging Labels	4	11.71
3	Blackstonian Pencils	5	13.18
4	Avery 482	6	16.67
5	Sony IBM Color Diskettes, 25/Pack	15	18.17