

Assignment 2-SQL

Davis Townsend



September 16, 2016

Intro to Data management (MIS 381N)

Professor Konana

Assignment 2

**Due**: September 16, 2016 (Submit through Canvas by 11:59pm)

**Type**: Individual (OK to ask others for simple clarification question)

**Grade**: 8%

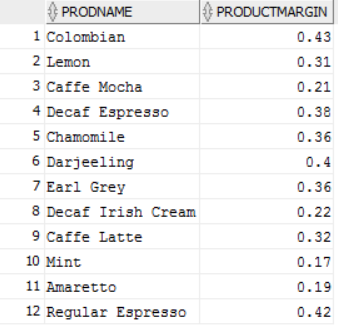
**WARNING**: There are two parts to this assignment – Part A and Part B. The assignment is not something that be done in one sitting or the day before the deadline. You need to really step back and ask what problem you are trying to address and how to break the bigger problem into smaller questions to solve using SQL. Given the nature of questions, there will be subjectivity and variance in the answers.

Please submit SQL statement (well formatted) with results in a Word document. Whenever there are more than 10 records in the result, please copy and paste the first 10 records.

# PART A – Coffee Sales

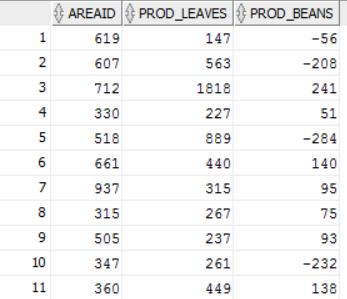
The following questions are based on the Coffee store sales data. Answer the following queries and you must use SQL to extract data and not eyeball some data to answer the questions.

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. **SELECT**
3. STATES.STATENAME
4. ,AREACODE.AREAID
5. ,sum(FACTCOFFEE.ACTSALES) "Area Code Sales"
6. ,round(AvgSales.SalesPerCode,2) "State Average"
7. ,round(100\*sum(FACTCOFFEE.ACTSALES)/AvgSales.SalesPerCode,2) "%OverStateAvg"
9. **FROM**
10. FACTCOFFEE
11. **inner** join AREACODE **on** FACTCOFFEE.AREAID = AREACODE.AREAID
12. **inner** join STATES **on** AREACODE.STATEID = STATES.STATEID
13. **inner** join
14. (
15. **SELECT**
16. STATES.STATENAME
17. ,sum(FACTCOFFEE.ACTSALES)/count(**distinct** FACTCOFFEE.AREAID) SalesPerCode
18. **FROM**
19. FACTCOFFEE
20. **inner** join AREACODE **on** FACTCOFFEE.AREAID = AREACODE.AREAID
21. **inner** join STATES **on** AREACODE.STATEID = STATES.STATEID
22. **WHERE**
23. EXTRACT(YEAR **from** FACTCOFFEE.FACTDATE) = 2013
24. **GROUP** **BY**
25. STATES.STATENAME
26. ) AvgSales **on** AvgSales.STATENAME = STATES.STATENAME
28. **WHERE**
29. EXTRACT(YEAR **from** FACTCOFFEE.FACTDATE) = 2013
31. **GROUP** **BY**
32. STATES.STATENAME
33. ,AREACODE.AREAID
34. ,AvgSales.SalesPerCode
36. **HAVING**
37. sum(FACTCOFFEE.ACTSALES)/AvgSales.SalesPerCode >= 1.1
39. **Order** **by**
40. sum(FACTCOFFEE.ACTSALES)/AvgSales.SalesPerCode **desc**
41. ;
    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.
42. **SELECT** ProdCoffee.ProdName, ROUND(SUM(ActProfit)/SUM(ActSales),2) ProductMargin
43. **FROM** FactCoffee
44. **INNER** JOIN ProdCoffee
45. **ON** ProdCoffee.ProductID = FactCoffee.ProductID
46. **GROUP** **BY** ProdName
47. **HAVING** sum(actProfit)/sum(actsales) >=0.15
48. **ORDER** **BY** sum(actSales) **DESC**;d\_Leaves >= Prod\_Beans \* 2.0;

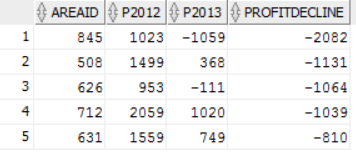


* 1. *Find AreaIDs where the* total profits from leaves in 2012 are two times greater than that from beans.

1. **SELECT** \* **FROM**
2. (**SELECT** FactCoffee.AreaID, ProdLine, sum(actprofit) **as** TotProfit
3. **FROM** FactCoffee
4. **INNER** JOIN ProdCoffee
5. **ON** FactCoffee.ProductID = ProdCoffee.ProductID
6. **WHERE** extract(year **from** factdate)=2012
7. **GROUP** **BY** FactCoffee.AREAID, ProdCoffee.ProdLine)
8. PIVOT
9. (
10. SUM(TotProfit)
11. **FOR** Prodline in ('Leaves' **as** Prod\_Leaves, 'Beans' **as** Prod\_Beans)
12. )
13. **WHERE** Prod\_Leaves >= Prod\_Beans \* 2.0;

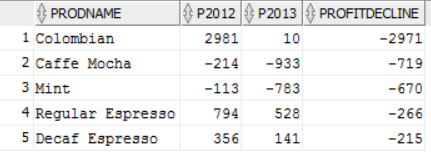
(25 total area codes)

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** X12.areaid,
4. x12.P2012,
5. X13.P2013,
6. (X13.P2013 - X12.P2012) ProfitDecline
7. **FROM**
8. (**SELECT** Areaid,
9. SUM(actprofit) P2012
10. **FROM** factcoffee
11. **WHERE** extract(YEAR **FROM** factdate) = 2012
12. **GROUP** **BY** areaid
13. ) X12,
14. (**SELECT** Areaid,
15. SUM(actprofit) P2013
16. **FROM** factcoffee
17. **WHERE** extract(YEAR **FROM** factdate) = 2013
18. **GROUP** **BY** areaid
19. ) X13
20. **WHERE** x12.areaid = x13.areaid
21. **ORDER** **BY** ProfitDecline)
22. **WHERE** ROWNUM <=5;



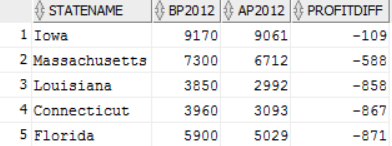
* 1. Among the five profit-declining area codes, are the profits consistently declining for all products? If not, identify the products for which they had significantly higher profit decline.

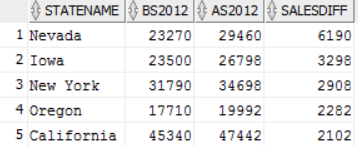
1. **SELECT** \* **FROM**
2. (
3. **SELECT**
4. x12.ProdName,
5. x12.P2012,
6. x13.P2013,
7. (x13.p2013 - X12.p2012) ProfitDecline
8. **FROM**
9. (**SELECT** ProdName,
10. SUM(actprofit) P2012
11. **FROM** factcoffee
12. **INNER** JOIN ProdCoffee
13. **ON** Factcoffee.ProductID = ProdCoffee.ProductID
14. **WHERE** extract(YEAR **FROM** factdate) = 2012
15. AND Areaid IN (845, 508, 626, 712, 631)
16. **GROUP** **BY** ProdName
17. ) X12,
18. (**SELECT** ProdName,
19. SUM(actprofit) P2013
20. **FROM** factcoffee
21. **INNER** JOIN ProdCoffee
22. **ON** Factcoffee.ProductID = ProdCoffee.ProductID
23. **WHERE** extract(YEAR **FROM** factdate) = 2013
24. AND Areaid IN (845, 508, 626, 712, 631)
25. **GROUP** **BY** ProdName
26. ) X13
27. **WHERE** x12.ProdName = x13.ProdName
28. **ORDER** **BY** ProfitDecline)
29. **WHERE** ROWNUM <=5;



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** X12.Statename,
4. x12.BP2012,
5. X13.AP2012,
6. (X13.AP2012 - X12.BP2012) ProfitDiff
7. **FROM**
8. (**SELECT** Statename,
9. SUM(Budprofit) BP2012
10. **FROM** factcoffee
11. **INNER** JOIN Areacode
12. **ON** Factcoffee.AreaID = Areacode.AreaID
13. **INNER** JOIN States
14. **ON** Areacode.stateid = States.StateID
15. **WHERE** extract(YEAR **FROM** factdate) = 2012
16. **GROUP** **BY** Statename
17. ) X12,
18. (**SELECT** Statename,
19. SUM(Actprofit) AP2012
20. **FROM** factcoffee
21. **INNER** JOIN Areacode
22. **ON** Factcoffee.AreaID = Areacode.AreaID
23. **INNER** JOIN States
24. **ON** Areacode.stateid = States.StateID
25. **WHERE** extract(YEAR **FROM** factdate) = 2012
26. **GROUP** **BY** Statename
27. ) X13
28. **WHERE** x12.Statename = x13.Statename
29. **ORDER** **BY** ProfitDiff **DESC**)
30. **WHERE** ROWNUM <=5;

33. --Now do the same for Sales
34. **SELECT** \* **FROM** (
35. **SELECT** X12.Statename,
36. x12.BS2012,
37. X13.AS2012,
38. (X13.AS2012 - X12.BS2012) SalesDiff
39. **FROM**
40. (**SELECT** Statename,
41. SUM(BudSales) BS2012
42. **FROM** factcoffee
43. **INNER** JOIN Areacode
44. **ON** Factcoffee.AreaID = Areacode.AreaID
45. **INNER** JOIN States
46. **ON** Areacode.stateid = States.StateID
47. **WHERE** extract(YEAR **FROM** factdate) = 2012
48. **GROUP** **BY** Statename
49. ) X12,
50. (**SELECT** Statename,
51. SUM(Actsales) AS2012
52. **FROM** factcoffee
53. **INNER** JOIN Areacode
54. **ON** Factcoffee.AreaID = Areacode.AreaID
55. **INNER** JOIN States
56. **ON** Areacode.stateid = States.StateID
57. **WHERE** extract(YEAR **FROM** factdate) = 2012
58. **GROUP** **BY** Statename
59. ) X13
60. **WHERE** x12.Statename = x13.Statename
61. **ORDER** **BY** SalesDiff **DESC**)
62. **WHERE** ROWNUM <=5;





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

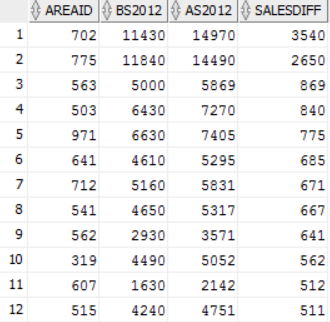
First We’ll look at Profit

1. **SELECT** \* **FROM** (
2. **SELECT** X12.AreaID,
3. x12.BP2012,
4. X13.AP2012,
5. (X13.AP2012 - X12.BP2012) ProfitDiff
6. **FROM**
7. (**SELECT** FactCoffee.AreaID,
8. SUM(Budprofit) BP2012
9. **FROM** factcoffee
10. **INNER** JOIN Areacode
11. **ON** Factcoffee.AreaID = Areacode.AreaID
12. **INNER** JOIN States
13. **ON** Areacode.stateid = States.StateID
14. **WHERE** extract(YEAR **FROM** factdate) = 2012
15. AND StateName IN ('Iowa', 'Massachusetts', 'Louisiana', 'Connecticut', 'Florida')
16. **GROUP** **BY** FactCoffee.AreaID
17. ) X12,
18. (**SELECT** FactCoffee.AreaID,
19. SUM(Actprofit) AP2012
20. **FROM** factcoffee
21. **INNER** JOIN Areacode
22. **ON** Factcoffee.AreaID = Areacode.AreaID
23. **INNER** JOIN States
24. **ON** Areacode.stateid = States.StateID
25. **WHERE** extract(YEAR **FROM** factdate) = 2012
26. AND StateName IN ('Iowa', 'Massachusetts', 'Louisiana', 'Connecticut', 'Florida')
27. **GROUP** **BY** FactCoffee.AreaID
28. ) X13
29. **WHERE** x12.AreaID = x13.AreaID
30. **ORDER** **BY** ProfitDiff **DESC**)
31. **WHERE** ROWNUM <=5;

*These are the only Area Codes within these 5 states that beat budgeted profit*

And we’ll do the same for Sales now:

1. **SELECT** \* **FROM** (
2. **SELECT** X12.AreaID,
3. x12.BS2012,
4. X13.AS2012,
5. (X13.AS2012 - X12.BS2012) SalesDiff
6. **FROM**
7. (**SELECT** FactCoffee.AreaID,
8. SUM(Budsales) BS2012
9. **FROM** factcoffee
10. **INNER** JOIN Areacode
11. **ON** Factcoffee.AreaID = Areacode.AreaID
12. **INNER** JOIN States
13. **ON** Areacode.stateid = States.StateID
14. **WHERE** extract(YEAR **FROM** factdate) = 2012
15. AND StateName IN ('Iowa', 'Nevada', 'New York', 'California', 'Oregon')
16. **GROUP** **BY** FactCoffee.AreaID
17. ) X12,
18. (**SELECT** FactCoffee.AreaID,
19. SUM(Actsales) AS2012
20. **FROM** factcoffee
21. **INNER** JOIN Areacode
22. **ON** Factcoffee.AreaID = Areacode.AreaID
23. **INNER** JOIN States
24. **ON** Areacode.stateid = States.StateID
25. **WHERE** extract(YEAR **FROM** factdate) = 2012
26. AND StateName IN ('Iowa', 'Nevada', 'New York', 'California', 'Oregon')
27. **GROUP** **BY** FactCoffee.AreaID
28. ) X13
29. **WHERE** x12.AreaID = x13.AreaID
30. **ORDER** **BY** SalesDiff **DESC**)
31. **WHERE** ROWNUM <=15;

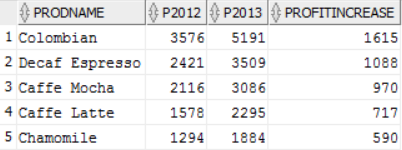
*These are the area codes for these 5 states that beat budgeted sales by over 500*

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

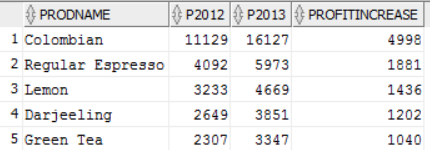
Note: I will only show the code once but all the outputs since the only thing that changes is the market in the where clause 4 different times

1. **SELECT** \* **FROM** (
2. **SELECT** X12.ProdName,
3. x12.P2012,
4. X13.P2013,
5. (X13.P2013 - X12.P2012) ProfitIncrease
6. **FROM**
7. (**SELECT** ProdCoffee.ProdName,
8. SUM(actprofit) P2012
9. **FROM** factcoffee
10. **INNER** JOIN ProdCoffee
11. **ON** factcoffee.ProductID = ProdCoffee.PRODUCTID
12. **INNER** JOIN AreaCode
13. **ON** AreaCode.AreaID = FactCoffee.AreaID
14. **INNER** JOIN States
15. **ON** States.StateID = AreaCode.StateID
16. **WHERE** extract(YEAR **FROM** factdate) = 2012
17. AND States.StateMkt = 'South'
18. **GROUP** **BY** ProdName
19. ) X12,
20. (**SELECT** ProdCoffee.ProdName,
21. SUM(actprofit) P2013
22. **FROM** factcoffee
23. **INNER** JOIN ProdCoffee
24. **ON** factcoffee.ProductID = ProdCoffee.PRODUCTID
25. **INNER** JOIN AreaCode
26. **ON** AreaCode.AreaID = FactCoffee.AreaID
27. **INNER** JOIN States
28. **ON** States.StateID = AreaCode.StateID
29. **WHERE** extract(YEAR **FROM** factdate) = 2013
30. AND States.StateMkt = 'South'
31. **GROUP** **BY** ProdName
32. ) X13
33. **WHERE** x12.ProdName = x13.ProdName
34. **ORDER** **BY** ProfitIncrease **DESC**)
35. **WHERE** ROWNUM <=5;

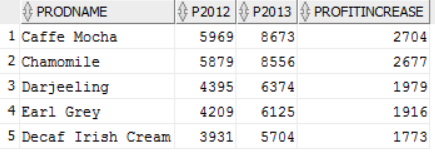
Southern Market: 5 products with largest increase in profits from 2013 to 2013



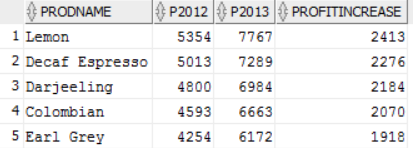
Eastern Market: 5 products with largest increase in profits from 2013 to 2013



Central Market: 5 products with largest increase in profits from 2013 to 2013



Western Market: 5 products with largest increase in profits from 2013 to 2013

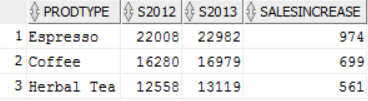


* 1. In each market, which **product types** have greatest increase in sales?

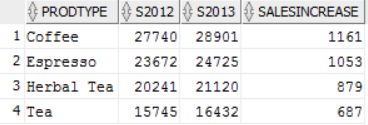
Just as in D1 I posted the code once and the outputs for each market I will do the same here as it is the same idea

1. **SELECT** \* **FROM** (
2. **SELECT** X12.ProdType,
3. x12.S2012,
4. X13.S2013,
5. (X13.S2013 - X12.S2012) SalesIncrease
6. **FROM**
7. (**SELECT** ProdCoffee.ProdType,
8. SUM(actsales) S2012
9. **FROM** factcoffee
10. **INNER** JOIN ProdCoffee
11. **ON** factcoffee.ProductID = ProdCoffee.PRODUCTID
12. **INNER** JOIN AreaCode
13. **ON** AreaCode.AreaID = FactCoffee.AreaID
14. **INNER** JOIN States
15. **ON** States.StateID = AreaCode.StateID
16. **WHERE** extract(YEAR **FROM** factdate) = 2012
17. AND States.StateMkt = 'South'
18. **GROUP** **BY** ProdType
19. ) X12,
20. (**SELECT** ProdCoffee.ProdType,
21. SUM(actsales) S2013
22. **FROM** factcoffee
23. **INNER** JOIN ProdCoffee
24. **ON** factcoffee.ProductID = ProdCoffee.PRODUCTID
25. **INNER** JOIN AreaCode
26. **ON** AreaCode.AreaID = FactCoffee.AreaID
27. **INNER** JOIN States
28. **ON** States.StateID = AreaCode.StateID
29. **WHERE** extract(YEAR **FROM** factdate) = 2013
30. AND States.StateMkt = 'South'
31. **GROUP** **BY** ProdType
32. ) X13
33. **WHERE** x12.ProdType = x13.ProdType
34. **ORDER** **BY** SalesIncrease **DESC**)
35. **WHERE** ROWNUM <=5;

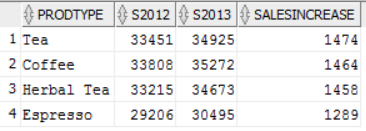
Southern Market: *Espresso has highest sales increase*



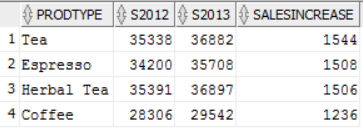
Eastern Market: *Coffee has highest sales increase*



Central Market: *Tea has highest sales increase*



Western Market: *Tea has highest sales increase*

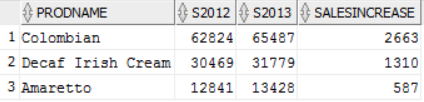


* 1. Have all products within the product types show similar behavior, or some products within a product type have greatest increase in sales?

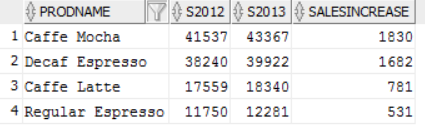
I will again use the same intuition as the other parts and post 1 code but all outputs of the different Product Types since only the where clauses changes

1. **SELECT** \* **FROM** (
2. **SELECT** X12.ProdName,
3. x12.S2012,
4. X13.S2013,
5. (X13.S2013 - X12.S2012) SalesIncrease
6. **FROM**
7. (**SELECT** ProdCoffee.ProdName,
8. SUM(actsales) S2012
9. **FROM** factcoffee
10. **INNER** JOIN ProdCoffee
11. **ON** factcoffee.ProductID = ProdCoffee.PRODUCTID
12. **WHERE** extract(YEAR **FROM** factdate) = 2012
13. AND ProdCoffee.ProdType = 'Coffee'
14. **GROUP** **BY** ProdName
15. ) X12,
16. (**SELECT** ProdCoffee.ProdName,
17. SUM(actsales) S2013
18. **FROM** factcoffee
19. **INNER** JOIN ProdCoffee
20. **ON** factcoffee.ProductID = ProdCoffee.PRODUCTID
21. **WHERE** extract(YEAR **FROM** factdate) = 2013
22. AND ProdCoffee.ProdType = 'Coffee'
23. **GROUP** **BY** ProdName
24. ) X13
25. **WHERE** x12.ProdName = x13.ProdName
26. **ORDER** **BY** SalesIncrease **DESC**)
27. **WHERE** ROWNUM <=5;

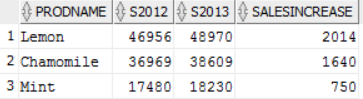
Change in Sales for Coffee:



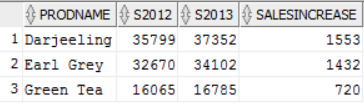
Change in Sales for Espresso:



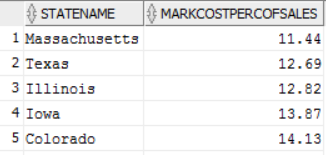
Change in Sales for Herbal Tea:



Change in Sales for Tea:

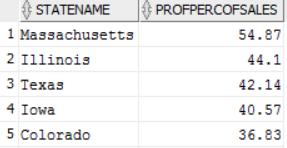


1. MARKETING EXPENSES (LOWEST):
   1. Which top 5 states have the lowest market expenses as a percentage of their sales?
2. **SELECT** States.StateName,ROUND((SUM(ActMarkCost)/SUM(ActSales))\*100,2) MarkCostPercofSales
3. **FROM** FactCoffee
4. **INNER** JOIN Areacode
5. **ON** FactCoffee.AreaID = Areacode.AreaID
6. **INNER** JOIN States
7. **ON** States.StateID = Areacode.StateID
8. **GROUP** **BY** StateName
9. **ORDER** **BY** MarkCostPercofSales;



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

1. **SELECT** States.StateName,ROUND((SUM(ActProfit)/SUM(ActSales))\*100,2) ProfPercofSales
2. **FROM** FactCoffee
3. **INNER** JOIN Areacode
4. **ON** FactCoffee.AreaID = Areacode.AreaID
5. **INNER** JOIN States
6. **ON** States.StateID = Areacode.StateID
7. **GROUP** **BY** StateName
8. **ORDER** **BY** ProfPercofSales **DESC**;



*Turns out that they do!!!*

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

First we’ll look at the overall marketing expenses of the products

1. **SELECT** ProdCoffee.ProdName, SUM(ActMarkCost) MarkExp
2. **FROM** FactCoffee
3. **INNER** JOIN ProdCoffee
4. **ON** FactCoffee.ProductID = ProdCoffee.ProductID
5. **GROUP** **BY** ProdName
6. **ORDER** **BY** MarkExp;

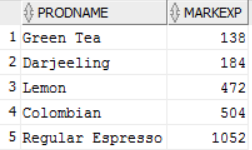


*It’s lowest for regular espresso*

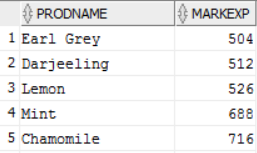
Now we’ll look at the products by cheapest marketing costs for each of the 5 states we found in the previous problem (Note: the code will be listed once since only the state name will change every time):

1. **SELECT** ProdCoffee.ProdName, SUM(ActMarkCost) MarkExp
2. **FROM** FactCoffee
3. **INNER** JOIN ProdCoffee
4. **ON** FactCoffee.ProductID = ProdCoffee.ProductID
5. **INNER** JOIN Areacode
6. **ON** FactCoffee.AreaID = Areacode.AreaID
7. **INNER** JOIN States
8. **ON** States.StateID = Areacode.StateID
9. **WHERE** States.StateName = 'Massachusetts'
10. **GROUP** **BY** ProdName
11. **ORDER** **BY** MarkExp;

Massachusetts:



Illinois:



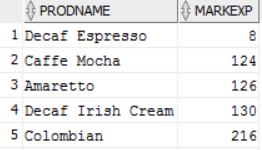
Texas:



Iowa:



Colorado:



*Decaf Espresso seems to have very little marketing expense in the Midwest*

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

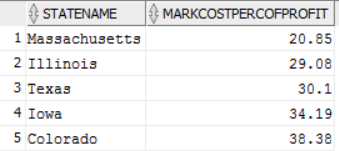
Are these marketing expenses justified? (Note: you need to think how you will justify high marketing expenses)?

1. **SELECT** States.StateName,ROUND((SUM(ActMarkCost)/SUM(ActSales))\*100,2) MarkCostPercofSales
2. **FROM** FactCoffee
3. **INNER** JOIN Areacode
4. **ON** FactCoffee.AreaID = Areacode.AreaID
5. **INNER** JOIN States
6. **ON** States.StateID = Areacode.StateID
7. **GROUP** **BY** StateName
8. **ORDER** **BY** MarkCostPercofSales **DESC**;



These high marketing costs would be justified if the extra expense on marketing led to higher profits in these states, so we should see if in these states there is a small percentage of marketing expenses as a percentage of profits. If a state listed here appears in the top 5 of lowest marketing expense as a percentage of profits then I consider the marketing expenses justified.

1. **SELECT** States.StateName,ROUND((SUM(ActMarkCost)/SUM(ActProfit))\*100,2) MarkCostPercofProfit
2. **FROM** FactCoffee
3. **INNER** JOIN Areacode
4. **ON** FactCoffee.AreaID = Areacode.AreaID
5. **INNER** JOIN States
6. **ON** States.StateID = Areacode.StateID
7. **GROUP** **BY** StateName
8. **ORDER** **BY** MarkCostPercofProfit;

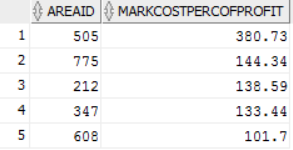


Seen here, I would not justify any of those states’ marketing expenses, however if we had more info we could look at market share and competition in the markets. Maybe it’s just that those markets are more competitive, but we still want to sell there despite a lower margin.

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

Yes, these area codes are the ones that have higher marketing costs than they have profits

1. **SELECT** AreaCode.AreaID,ROUND((SUM(ActMarkCost)/SUM(ActProfit))\*100,2) MarkCostPercofProfit
2. **FROM** FactCoffee
3. **INNER** JOIN Areacode
4. **ON** FactCoffee.AreaID = Areacode.AreaID
5. **INNER** JOIN States
6. **ON** States.StateID = Areacode.StateID
7. **WHERE** States.StateName In ('Nevada', 'Wisconsin', 'New Mexico', 'Washington', 'New York')
8. **GROUP** **BY** AreaCode.AreaID
9. **HAVING** ROUND((SUM(ActMarkCost)/SUM(ActProfit))\*100,2) > 100
10. **ORDER** **BY** MarkCostPercofProfit **DESC**;



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. --Find States with highest percent expenses with respect to profits
3. --(i.e. they spend the least to make the most money)
4. **SELECT** States.StateName,ROUND((SUM(ActExpenses)/SUM(ActProfit))\*100,2) ExpensePercofProfit
5. **FROM** FactCoffee
6. **INNER** JOIN Areacode
7. **ON** FactCoffee.AreaID = Areacode.AreaID
8. **INNER** JOIN States
9. **ON** States.StateID = Areacode.StateID
10. **GROUP** **BY** States.StateName
11. **ORDER** **BY** ExpensePercofProfit **DESC**;
12. --Now within this state, use the same idea but for the area codes within this state
13. **SELECT** AreaCode.AreaID,ROUND((SUM(ActExpenses)/SUM(ActProfit))\*100,2) ExpensePercofProfit
14. **FROM** FactCoffee
15. **INNER** JOIN Areacode
16. **ON** FactCoffee.AreaID = Areacode.AreaID
17. **INNER** JOIN States
18. **ON** States.StateID = Areacode.StateID
19. **WHERE** States.StateName = 'New Mexico'
20. **GROUP** **BY** AreaCode.AreaID
21. **ORDER** **BY** ExpensePercofProfit **DESC**;

We find that the state that spends the most relative to how much they make is New Mexico and the area code within this state where we should look to close is the following:

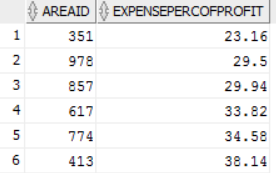


* 1. Where should the firm focus on expanding?

1. **SELECT** States.StateName,ROUND((SUM(ActExpenses)/SUM(ActProfit))\*100,2) ExpensePercofProfit
2. **FROM** FactCoffee
3. **INNER** JOIN Areacode
4. **ON** FactCoffee.AreaID = Areacode.AreaID
5. **INNER** JOIN States
6. **ON** States.StateID = Areacode.StateID
7. **GROUP** **BY** States.StateName
8. **ORDER** **BY** ExpensePercofProfit;
9. --Now within this state, use the same idea but for the area codes within this state
10. **SELECT** AreaCode.AreaID,ROUND((SUM(ActExpenses)/SUM(ActProfit))\*100,2) ExpensePercofProfit
11. **FROM** FactCoffee
12. **INNER** JOIN Areacode
13. **ON** FactCoffee.AreaID = Areacode.AreaID
14. **INNER** JOIN States
15. **ON** States.StateID = Areacode.StateID
16. **WHERE** States.StateName = 'Massachusetts'
17. **GROUP** **BY** AreaCode.AreaID
18. **ORDER** **BY** ExpensePercofProfit;

I do the exact same thing, except find the state with the lowest expense as a percent of profit instead of the highest

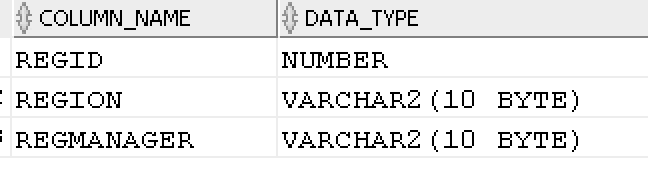
I Find the state of Massachusetts has the lowest expenses as a percent of profit, and following the same idea, here are the area codes within Massachusetts where we should look at expanding(their percentages are less than 50%):



# 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.

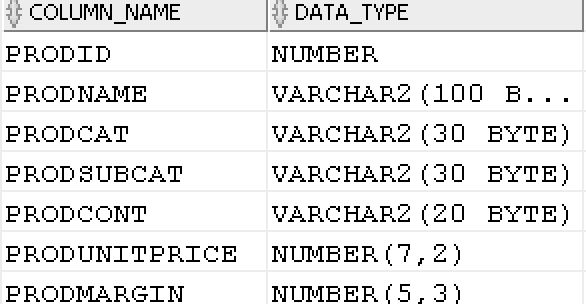
**TABLE: MANAGERS (REGID is the PK)**



CONSTRAINT:

REGION can be only ‘East’, ‘South’, ‘Central’, ‘West’.

**TABLE: PRODUCTS (ProdID is the PK)**

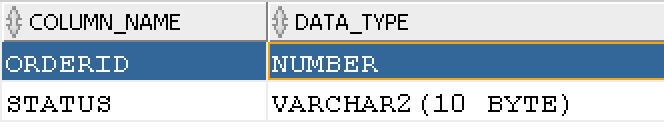


CONSTRAINTS:

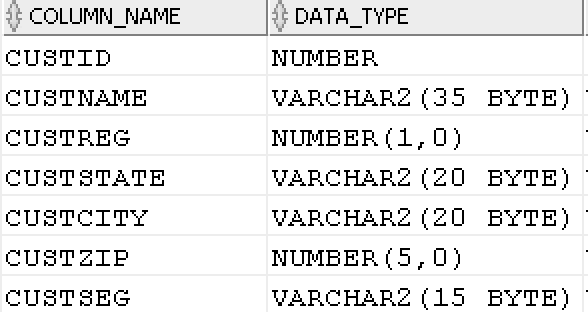
PRODCAT can only be ‘Technology’ ‘Furniture’ or ‘Office Supplies’

PRODCONT take on only ‘Jumbo Drum’, ‘Medium Box’, ‘Jumbo Box’, ‘Wrap Bag’, ‘Large Box’, ‘Small Box’, ‘Small Pack’

**TABLE: ORDERS (OrderID is the PK)**



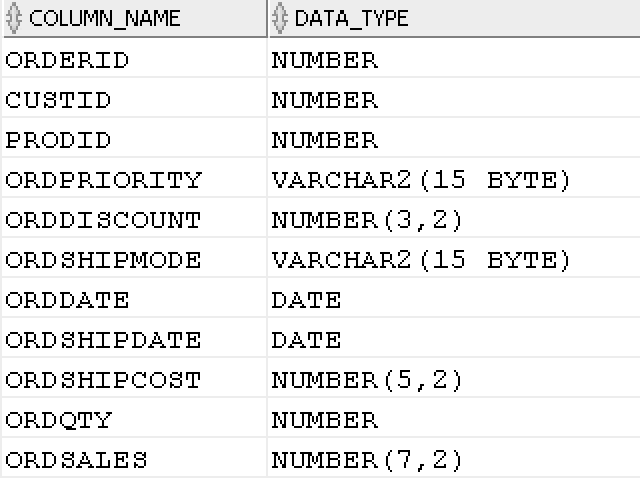
**TABLE: CUSTOMERS (CustID is the PK; CustReg is the FK on delete cascade)**



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)**



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.

1. **DROP** **TABLE** Customers;
2. **DROP** **TABLE** Managers;
3. **DROP** **TABLE** Products;
4. **DROP** **TABLE** Orders;
5. **DROP** **TABLE** OrderDet;

8. **CREATE** **Table** Managers
9. (
10. RegID NUMBER,
11. Region  varchar2(10BYTE),
12. RegManager  varchar2(10BYTE),
13. **CONSTRAINT** pk\_RegID **Primary** **Key**(RegID),
14. **CONSTRAINT** ch\_Region **CHECK** (Region IN ('South', 'Central', 'West', 'East'))
15. );
17. **CREATE** **TABLE** Products
18. (
19. ProdID NUMBER,
20. ProdName VARCHAR2(100BYTE),
21. ProdCat VARCHAR2(30BYTE),
22. ProdSubCat VARCHAR2(30BYTE),
23. ProdCont VARCHAR2(20BYTE),
24. ProdUnitPrice NUMBER (7,2),
25. ProdMargin NUMBER (5,3),
26. **CONSTRAINT** pk\_ProdID **Primary** **Key**(ProdID),
27. **CONSTRAINT** ch\_ProdCat **CHECK** (ProdCat IN ('Technology', 'Furniture', 'Office Supplies')),
28. **CONSTRAINT** ch\_ProdCont **CHECK** (ProdCont IN ('Jumbo Drum', 'Medium Box', 'Jumbo Box', 'Wrap Bag', 'Large Box', 'Small Box', 'Small Pack'))
29. );
31. **CREATE** **TABLE** Orders
32. (
33. OrderID Number,
34. Status VARCHAR2(10 BYTE),
35. **CONSTRAINT** pk\_OrderID **Primary** **Key**(OrderID)
36. );
38. **CREATE** **TABLE** Customers
39. (
40. CustID Number,
41. CustName VARCHAR2(35BYTE),
42. CustReg Number(1,0),
43. CustState VARCHAR2(20BYTE),
44. CustCity VARCHAR2(20BYTE),
45. CustZIP NUMBER (5,0),
46. CustSeg VARCHAR2(15BYTE),
47. **CONSTRAINT** pk\_CustID **Primary** **Key**(CustID),
48. **CONSTRAINT** fk\_CustReg **Foreign** **Key** (CustReg) **REFERENCES** Managers (RegID) **on** **DELETE** **CASCADE**,
49. **CONSTRAINT** ch\_CustSeg  **CHECK** (CustSeg IN ('Home Office', 'Corporate', 'Small Business', 'Consumer'))
50. );
52. **CREATE** **TABLE** ORDERDET
53. (
54. OrderID Number,
55. CustID Number,
56. ProdID Number,
57. OrdPriority VARCHAR2(15 BYTE),
58. OrdDiscount NUMBER(3,2),
59. OrdShipMode VARCHAR2(15 BYTE),
60. OrdDate **Date**,
61. OrdShipDate **Date**,
62. OrdShipCost Number(5,2),
63. OrdQty Number,
64. OrdSales Number(7,2),
65. **CONSTRAINT** ch\_OrdPriority  **CHECK** (OrdPriority IN ('Low', 'Medium', 'High', 'Critical', 'Not Specified')),
66. **CONSTRAINT** ch\_OrdShipMode  **CHECK** (OrdShipMode IN ('Regular Air', 'Delivery Truck', 'Express Air')),
67. **CONSTRAINT** fk\_dOrderID **Foreign** **Key** (OrderID) **REFERENCES** Orders (OrderID),
68. **CONSTRAINT** fk\_dCustID **Foreign** **Key** (CustID) **REFERENCES** Customers (CustID),
69. **CONSTRAINT** fk\_dProdID **Foreign** **Key** (ProdID) **REFERENCES** Products (ProdID)
70. );

I inserted using the import function, there were a few rows dropped during the data import to the orderdet table probably due to the check constraints

QUESTION 2: ORDER Cancellations

1. What fraction of the orders was cancelled?

Here I am assuming cancelled means the product has returned status in the Orders table. Almost 1% if the orders were cancelled(returned)

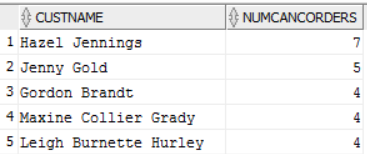
1. **SELECT** CanO, FillO, (CanO/FillO)\*100 FractionCancelled
2. **FROM**
3. (
4. (**SELECT** SUM(test1) CanO
5. **FROM**
6. (**SELECT** OrderID, COUNT(OrderID) **as** test1
7. **FROM** Orders
8. **WHERE** Status = 'Returned'
9. **GROUP** **BY** OrderID))
10. ),
11. (
12. (**SELECT** SUM(test2) FillO
13. **FROM**
14. (**SELECT** OrderID, COUNT(OrderID) **as** test2
15. **FROM** Orders
16. **GROUP** **BY** OrderID)))
17. ;



1. What were the sales from cancelled orders?
2. **SELECT** SUM(OrdSales) TotCancOrdSales
3. **FROM** OrderDet
4. **INNER** JOIN Orders
5. **ON** Orders.OrderID = OrderDet.OrderID
6. **WHERE** Orders.Status = 'Returned';

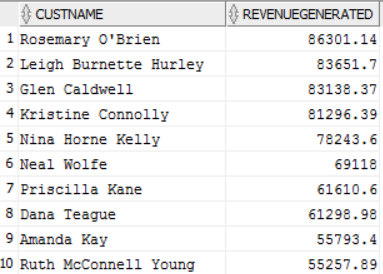


1. Who are the top five customers in terms of cancelled orders?
2. **SELECT** Customers.CustName, COUNT(Orders.OrderID) NumCancOrders
3. **FROM** Customers
4. **INNER** JOIN ORderDet
5. **ON** OrderDet.CustID = Customers.CustID
6. **INNER** JOIN Orders
7. **ON** Orders.OrderID = OrderDet.OrderID
8. **WHERE** Orders.Status = 'Returned'
9. **GROUP** **BY** CustName
10. **ORDER** **BY** NumCancOrders **DESC**
11. ;



QUESTION 3: CUSTOMER related:

1. Who are the top 10 customers in terms of revenues generated?
2. **SELECT** Customers.CustName, SUM(OrderDet.OrdSales) RevenueGenerated
3. **FROM** Customers
4. **INNER** JOIN OrderDet
5. **ON** OrderDet.CustID = Customers.CustID
6. **GROUP** **BY** CustName
7. **ORDER** **BY** RevenueGenerated **DESC**
8. ;



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** Customers.CustID, Customers.CustName, Products.ProdCat, RANK()
3. OVER (PARTITION **BY** Customers.CustID **ORDER** **BY** Customers.CustName) **AS** Prodcategory
4. **FROM** Customers
5. **INNER** JOIN OrderDet
6. **ON** OrderDet.CustID = Customers.CustID
7. **INNER** JOIN Products
8. **ON** OrderDet.ProdID = Products.ProdID
9. ;

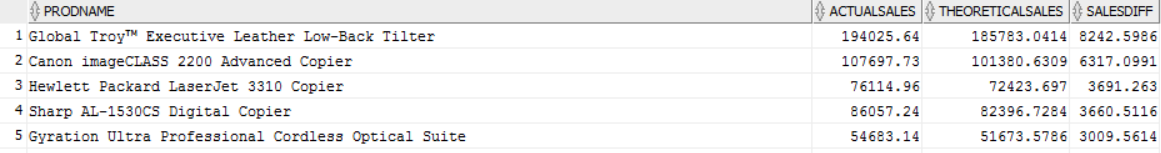
From this output we can find the customers who buy a lot of a certain type of good and combine it with our selections above to discover which customers would be in the market for other types of product categories they aren’t currently buying

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?

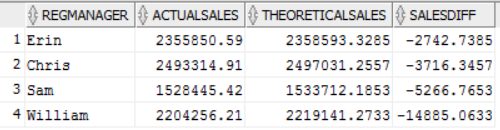
I took the Difference of theoretical sales from actual sales for each product and show the top 5 and then reran it but only selected the total summed sales difference

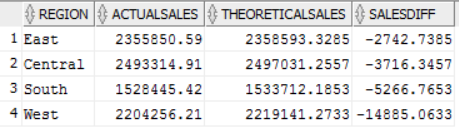
1. **SELECT** \* **FROM** (
2. **SELECT** A.ProdName,
3. A.ActualSales,
4. T.TheoreticalSales,
5. (A.ActualSales - T.TheoreticalSales) SalesDiff
6. **FROM**
7. (**SELECT** Products.ProdName, SUM(OrderDet.OrdSales) ActualSales
8. **FROM** Products
9. **INNER** JOIN OrderDet
10. **ON** OrderDet.ProdID = Products.ProdID
11. **GROUP** **BY** ProdName) A,
12. (**SELECT** Products.ProdName, SUM(Products.ProdUnitPrice \* OrderDet.OrdQty \* (1-OrderDet.OrdDiscount)+ OrderDet.OrdShipCost) TheoreticalSales
13. **FROM** Products
14. **INNER** JOIN OrderDet
15. **ON** Products.ProdID = OrderDet.ProdID
16. **GROUP** **BY** ProdName) T
17. **WHERE** A.ProdName = T.ProdName
18. **ORDER** **BY** SalesDiff **DESC**)
19. ;





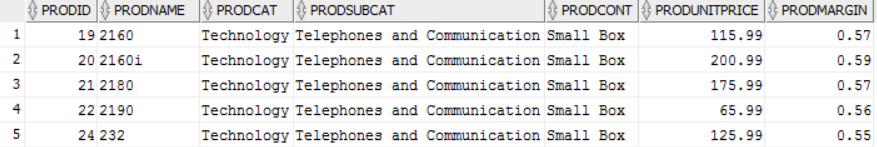
1. Are certain managers generally pricing more or less than theoretical sales? Analyze the differences based on the regions/managers.
2. --By Manager
3. **SELECT** \* **FROM** (
4. **SELECT** A.RegManager,
5. A.ActualSales,
6. T.TheoreticalSales,
7. (A.ActualSales - T.TheoreticalSales) SalesDiff
8. **FROM**
9. (**SELECT** Managers.RegManager, SUM(OrderDet.OrdSales) ActualSales
10. **FROM** Managers
11. **INNER** JOIN Customers
12. **ON** Customers.CustReg = Managers.RegID
13. **INNER** JOIN OrderDet
14. **ON** Customers.CustID = OrderDet.CustID
15. **GROUP** **BY** RegManager) A,
16. (**SELECT** Managers.RegManager, SUM(Products.ProdUnitPrice \* OrderDet.OrdQty \* (1-OrderDet.OrdDiscount)+ OrderDet.OrdShipCost) TheoreticalSales
17. **FROM** Managers
18. **INNER** JOIN Customers
19. **ON** Customers.CustReg = Managers.RegID
20. **INNER** JOIN OrderDet
21. **ON** Customers.CustID = OrderDet.CustID
22. **INNER** JOIN Products
23. **ON** OrderDet.ProdID = Products.ProdID
24. **GROUP** **BY** RegManager) T
25. **WHERE** A.RegManager = T.RegManager
26. **ORDER** **BY** SalesDiff **DESC**)
27. ;
29. --By Region
30. **SELECT** \* **FROM** (
31. **SELECT** A.Region,
32. A.ActualSales,
33. T.TheoreticalSales,
34. (A.ActualSales - T.TheoreticalSales) SalesDiff
35. **FROM**
36. (**SELECT** Managers.Region, SUM(OrderDet.OrdSales) ActualSales
37. **FROM** Managers
38. **INNER** JOIN Customers
39. **ON** Customers.CustReg = Managers.RegID
40. **INNER** JOIN OrderDet
41. **ON** Customers.CustID = OrderDet.CustID
42. **GROUP** **BY** Region) A,
43. (**SELECT** Managers.Region, SUM(Products.ProdUnitPrice \* OrderDet.OrdQty \* (1-OrderDet.OrdDiscount)+ OrderDet.OrdShipCost) TheoreticalSales
44. **FROM** Managers
45. **INNER** JOIN Customers
46. **ON** Customers.CustReg = Managers.RegID
47. **INNER** JOIN OrderDet
48. **ON** Customers.CustID = OrderDet.CustID
49. **INNER** JOIN Products
50. **ON** OrderDet.ProdID = Products.ProdID
51. **GROUP** **BY** Region) T
52. **WHERE** A.Region = T.Region
53. **ORDER** **BY** SalesDiff **DESC**)
54. ;



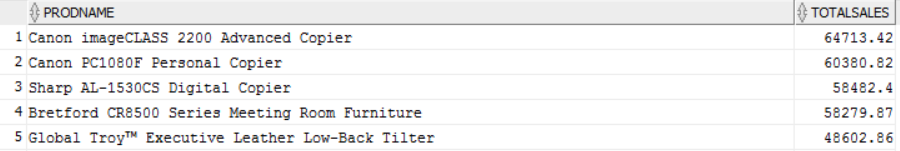


QUESTION 5: these are product related questions:

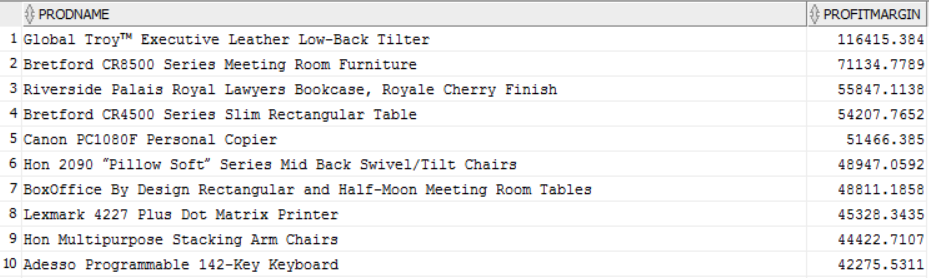
1. Products have numbers within its name. Identify the product names with digits in their name. (hint: use REGEXP\_LIKE)
2. **SELECT** \*
3. **FROM** Products
4. **WHERE** REGEXP\_LIKE(ProdName, '^[A-Za-z0-9]+$');



1. Which are the top 5 selling products during the year 2011?
2. **SELECT** Products.ProdName, SUM(OrderDet.OrdSales) TotalSales
3. **FROM** Products
4. **INNER** JOIN OrderDet
5. **ON** Products.ProdID = OrderDet.ProdID
6. **WHERE** Extract(Year **From** OrdDate) = 2011
7. **GROUP** **BY** ProdName
8. **ORDER** **BY** TotalSales **DESC**
9. ;
11. --c)    Which are the top 10 products with greatest total profit margin? (i.e., sales\*margin).
12. **SELECT** Products.ProdName, SUM(OrderDet.OrdSales\*Products.ProdMargin) ProfitMargin
13. **FROM** Products
14. **INNER** JOIN OrderDet
15. **ON** Products.ProdID = OrderDet.ProdID
16. **GROUP** **BY** ProdName
17. **ORDER** **BY** ProfitMargin **DESC**
18. ;



1. Which are the top 10 products with greatest total profit margin? (i.e., sales\*margin).
2. **SELECT** Products.ProdName, SUM(OrderDet.OrdSales\*Products.ProdMargin) ProfitMargin
3. **FROM** Products
4. **INNER** JOIN OrderDet
5. **ON** Products.ProdID = OrderDet.ProdID
6. **GROUP** **BY** ProdName
7. **ORDER** **BY** ProfitMargin **DESC**
8. ;



1. Identify the worst five products in terms of sales?
2. **SELECT** Products.ProdName, SUM(OrderDet.OrdSales) TotalSales
3. **FROM** Products
4. **INNER** JOIN OrderDet
5. **ON** Products.ProdID = OrderDet.ProdID
6. **GROUP** **BY** ProdName
7. **ORDER** **BY** TotalSales
8. ;

