Assignment 3

**Due**: September 2,4 2016 (midnight)

**Type**: Individual (PLEASE DO NOT COLLABORATE except for clarification question)

**Grade**: 7%

This assignment will use the same data set as Assignment 2.

**WARNING**: The assignment is not something that be done in one sitting and 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

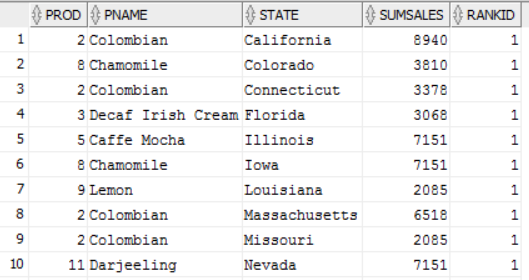
The following questions are based on the Coffee store sales data. Answer the following queries.

* 1. Extract the total sales for each product for each month. List all months (like January, February, etc) in the columns.

FINISH THIS Q

* 1. In each state, identify the product with greatest sales for the year 2012.

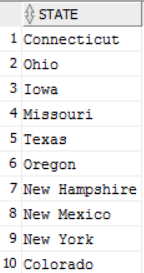
1. **SELECT** \* **FROM** (
2. **SELECT** factcoffee.productid prod,  prodname pname,  statename state,
3. SUM(actsales) Sumsales,
4. ROW\_NUMBER() OVER (PARTITION **BY** statename **ORDER** **BY** SUM(ACTSALES) **DESC**) **AS** RANKID
5. **FROM** factcoffee,   prodcoffee,  states,  areacode
6. **WHERE** factcoffee.productid     = prodcoffee.productid
7. AND areacode.AREAID            = factcoffee.AREAID
8. AND states.STATEID             = areacode.stateid
9. AND extract(YEAR **FROM** factdate)=2012
10. **GROUP** **BY** factcoffee.productid,  prodname, statename)
11. **WHERE** rankid = 1;



* + 1. Identify the states where the best selling product remained the same in 2013 (compared to best selling product in 2012)

1. **SELECT** **DISTINCT** State **FROM**
2. (
3. **SELECT** \* **FROM** (
4. **SELECT** factcoffee.productid prod,  prodname pname,  statename state,
5. SUM(actsales) Sumsales,
6. ROW\_NUMBER() OVER (PARTITION **BY** statename **ORDER** **BY** SUM(ACTSALES) **DESC**) **AS** RANKID
7. **FROM** factcoffee,   prodcoffee,  states,  areacode
8. **WHERE** factcoffee.productid     = prodcoffee.productid
9. AND areacode.AREAID            = factcoffee.AREAID
10. AND states.STATEID             = areacode.stateid
11. AND extract(YEAR **FROM** factdate)=2012
12. **GROUP** **BY** factcoffee.productid,  prodname, statename)
13. **WHERE** rankid = 1
14. **UNION**
15. **SELECT** \* **FROM** (
16. **SELECT** factcoffee.productid prod,  prodname pname,  statename state,
17. SUM(actsales) Sumsales,
18. ROW\_NUMBER() OVER (PARTITION **BY** statename **ORDER** **BY** SUM(ACTSALES) **DESC**) **AS** RANKID
19. **FROM** factcoffee,   prodcoffee,  states,  areacode
20. **WHERE** factcoffee.productid     = prodcoffee.productid
21. AND areacode.AREAID            = factcoffee.AREAID
22. AND states.STATEID             = areacode.stateid
23. AND extract(YEAR **FROM** factdate)=2013
24. **GROUP** **BY** factcoffee.productid,  prodname, statename)
25. **WHERE** rankid = 1
26. );

Turned out to be all of them



* + 1. Identify the states where the best selling product has changed.

Based on last question, the best selling product did not change in any state.

1. **SELECT** **DISTINCT** State **FROM**
2. (
3. **SELECT** \* **FROM** (
4. **SELECT** factcoffee.productid prod,  prodname pname,  statename state,
5. SUM(actsales) Sumsales,
6. ROW\_NUMBER() OVER (PARTITION **BY** statename **ORDER** **BY** SUM(ACTSALES) **DESC**) **AS** RANKID
7. **FROM** factcoffee,   prodcoffee,  states,  areacode
8. **WHERE** factcoffee.productid     = prodcoffee.productid
9. AND areacode.AREAID            = factcoffee.AREAID
10. AND states.STATEID             = areacode.stateid
11. AND extract(YEAR **FROM** factdate)=2012
12. **GROUP** **BY** factcoffee.productid,  prodname, statename)
13. **WHERE** rankid = 1
14. **INTERSECT**
15. **SELECT** \* **FROM** (
16. **SELECT** factcoffee.productid prod,  prodname pname,  statename state,
17. SUM(actsales) Sumsales,
18. ROW\_NUMBER() OVER (PARTITION **BY** statename **ORDER** **BY** SUM(ACTSALES) **DESC**) **AS** RANKID
19. **FROM** factcoffee,   prodcoffee,  states,  areacode
20. **WHERE** factcoffee.productid     = prodcoffee.productid
21. AND areacode.AREAID            = factcoffee.AREAID
22. AND states.STATEID             = areacode.stateid
23. AND extract(YEAR **FROM** factdate)=2013
24. **GROUP** **BY** factcoffee.productid,  prodname, statename)
25. **WHERE** rankid = 1
26. );

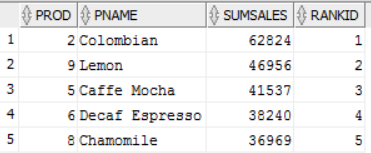


* + 1. Identify the products that were best in 2012 but not in 2013.

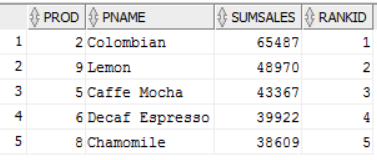
1. **SELECT** \* **FROM** (
2. **SELECT** factcoffee.productid prod,  prodname pname,
3. SUM(actsales) Sumsales,
4. ROW\_NUMBER() OVER (**ORDER** **BY** SUM(actsales) **DESC**) **AS** RANKID
5. **FROM** factcoffee,   prodcoffee
6. **WHERE** factcoffee.productid     = prodcoffee.productid
7. AND extract(YEAR **FROM** factdate)=2012
8. **GROUP** **BY** factcoffee.productid,  prodname)
9. **WHERE** RANKID IN (1, 2, 3, 4, 5)
10. **ORDER** **BY** Sumsales **DESC**
11. ;
13. **SELECT** \* **FROM** (
14. **SELECT** factcoffee.productid prod,  prodname pname,
15. SUM(actsales) Sumsales,
16. ROW\_NUMBER() OVER (**ORDER** **BY** SUM(actsales) **DESC**) **AS** RANKID
17. **FROM** factcoffee,   prodcoffee
18. **WHERE** factcoffee.productid     = prodcoffee.productid
19. AND extract(YEAR **FROM** factdate)=2013
20. **GROUP** **BY** factcoffee.productid,  prodname)
21. **WHERE** RANKID IN (1, 2, 3, 4, 5)
22. **ORDER** **BY** Sumsales **DESC**
23. ;

Again we see they are all the same for 2012 and 2013

*2012*



*2013*



* + 1. Identify the top two best selling products that are common to both 2012 and 2013.

As seen from the previous problem it is Colombian and Lemon

* 1. What fraction of the top selling states contributes to at least 50% of the total sales? Do they also contribute to 50% of the profit share as well? (Please note that you won’t likely get exact 50% when you do your analysis)

1. --Sales Run
2. **With** Cumsale **as** (**SELECT** Statename, SUM(actsales) Sumsales, Row\_number() OVER (**ORDER** **BY** sum(actsales) **DESC**) Rowsales
3. **FROM** factcoffee
4. **INNER** JOIN Areacode **ON** Factcoffee.areaid = areacode.areaid
5. **INNER** JOIN States **On** areacode.stateid = states.stateid
6. **GROUP** **BY** Statename),
8. totalcount **AS** (
9. **SELECT** count(\*) **as** totcount
10. **FROM** Cumsale),
12. totsales **as** (**SELECT** sum(sumsales) totsumsales **FROM** cumsale),
14. Cumtotsales **as** (**SELECT** rowsales, sum(sumsales) over (**order** **by** rowsales) Csales
15. **FROM** cumsale)
17. **SELECT** totcount, totsumsales, rowsales, csales, (rowsales/totcount) **as** statepercentageSales
18. **FROM** totalcount, totsales, cumtotsales
19. **WHERE** csales >= 0.5\*totsumsales AND rownum =1;
21. --Profits Run
22. **With** Cumprofit **as** (**SELECT** Statename, SUM(actprofit) Sumprofit, Row\_number() OVER (**ORDER** **BY** sum(actprofit) **DESC**) Rowprofit
23. **FROM** factcoffee
24. **INNER** JOIN Areacode **ON** Factcoffee.areaid = areacode.areaid
25. **INNER** JOIN States **On** areacode.stateid = states.stateid
26. **GROUP** **BY** Statename),
28. totalcount **AS** (
29. **SELECT** count(\*) **as** totcount
30. **FROM** Cumprofit),
32. totprofit **as** (**SELECT** sum(sumprofit) totsumprofit **FROM** cumprofit),
34. Cumtotprofit **as** (**SELECT** rowprofit, sum(sumprofit) over (**order** **by** rowprofit) Cprofit
35. **FROM** cumprofit)
37. **SELECT** totcount, totsumprofit, rowprofit, cprofit, (rowprofit/totcount) **as** statepercentageProfit
38. **FROM** totalcount, totprofit, cumtotprofit
39. **WHERE** cprofit >= 0.5\*totsumprofit AND rownum =1;





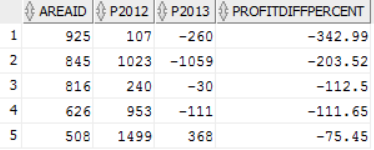
For sales: 35%, there are 7 contributing states

For Profits: 30%, there are 6 contributing states

So the states are different

* 1. Identify the area codes with greatest decline in profits from the year 2012 to 2013. List the profits for 2012 and 2013 in the columns and display the percentage decline.

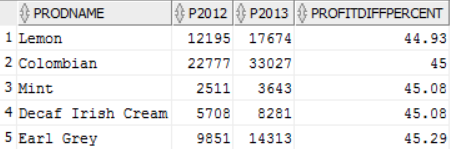
1. **SELECT** \* **FROM** (
2. **SELECT** X12.areaid,
3. x12.P2012,
4. X13.P2013,
5. ROUND(((X13.p2013 - X12.p2012)/x12.p2012)\*100,2) ProfitDiffPercent
7. **FROM**
8. (**SELECT** Areaid,
9. SUM(actprofit) P2012
10. **FROM** factcoffee
11. **WHERE** extract(YEAR **FROM** factdate) = 2012
12. **GROUP** **BY** areaid
13. ) X12,
15. (**SELECT** Areaid,
16. SUM(actprofit) P2013
17. **FROM** factcoffee
18. **WHERE** extract(YEAR **FROM** factdate) = 2013
19. **GROUP** **BY** areaid
20. ) X13
21. **WHERE** x12.areaid = x13.areaid
22. --have to remove these area codes due to how the answers come out given the
23. --multiplication of a negative number
24. AND x12.areaid NOT IN (818, 212, 914, 909)
25. **ORDER** **BY** ProfitDiffPercent)
26. **WHERE** ROWNUM <=5;



* 1. If you have to discontinue some product, which one would you suggest and why? Formulate your questions.

I will answer this question by discontinuing the product which had the largest percentage decrease in profits from 2012 to 2013

1. **SELECT** \* **FROM** (
2. **SELECT** X12.prodname,
3. x12.P2012,
4. X13.P2013,
5. ROUND(((X13.p2013 - X12.p2012)/x12.p2012)\*100,2) ProfitDiffPercent
7. **FROM**
8. (**SELECT** Prodname,
9. SUM(actprofit) P2012
10. **FROM** factcoffee
11. **INNER** JOIN ProdCoffee
12. **ON** ProdCoffee.ProductID = FactCoffee.ProductID
13. **WHERE** extract(YEAR **FROM** factdate) = 2012
14. **GROUP** **BY** ProdName
15. ) X12,
17. (**SELECT** Prodname,
18. SUM(actprofit) P2013
19. **FROM** factcoffee
20. **INNER** JOIN ProdCoffee
21. **ON** ProdCoffee.ProductID = FactCoffee.ProductID
22. **WHERE** extract(YEAR **FROM** factdate) = 2013
23. **GROUP** **BY** ProdName
24. ) X13
25. **WHERE** x12.prodname = x13.prodname
26. **ORDER** **BY** ProfitDiffPercent)
27. **WHERE** ROWNUM <=5;



*Turns out to be Lemon*

* 1. Where should the marketing expenses be increased and reduced? You need to structure the queries appropriately.

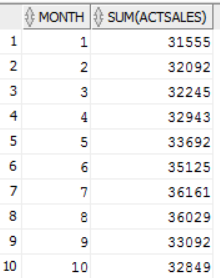
*Marketing expenses should be increased for Green Tea and reduced for Caffe Mocha*

1. **SELECT** ProdCoffee.ProdName,ROUND((SUM(ActMarkCost)/SUM(ActProfit))\*100,2) MarkCostPercofProfit
2. **FROM** FactCoffee
3. **INNER** JOIN ProdCoffee
4. **ON** FactCoffee.ProductID = ProdCoffee.ProductID
5. **GROUP** **BY** ProdName
6. **ORDER** **BY** MarkCostPercofProfit;





* 1. The overall sales per month could be seasonal. That is, you will find greater sales in some months than others and this may be consistent with 2012 and 2013. Identify if there are seasonal trends. Plot month vs. sales for each year.

1. **SELECT** extract(Month **from** FactDate) Month, SUM(actsales)
2. **FROM** FactCoffee
3. **WHERE** extract(Year **from** FactDate) = 2012
4. **GROUP** **BY** extract(Month **from** FactDate)
5. **ORDER** **BY** Month
6. ;
8. **SELECT** extract(Month **from** FactDate) Month, SUM(actsales)
9. **FROM** FactCoffee
10. **WHERE** extract(Year **from** FactDate) = 2013
11. **GROUP** **BY** extract(Month **from** FactDate)
12. **ORDER** **BY** Month
13. ;

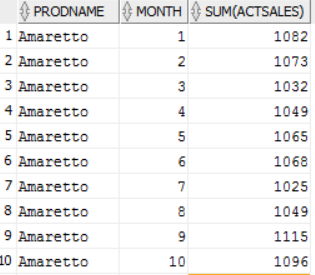
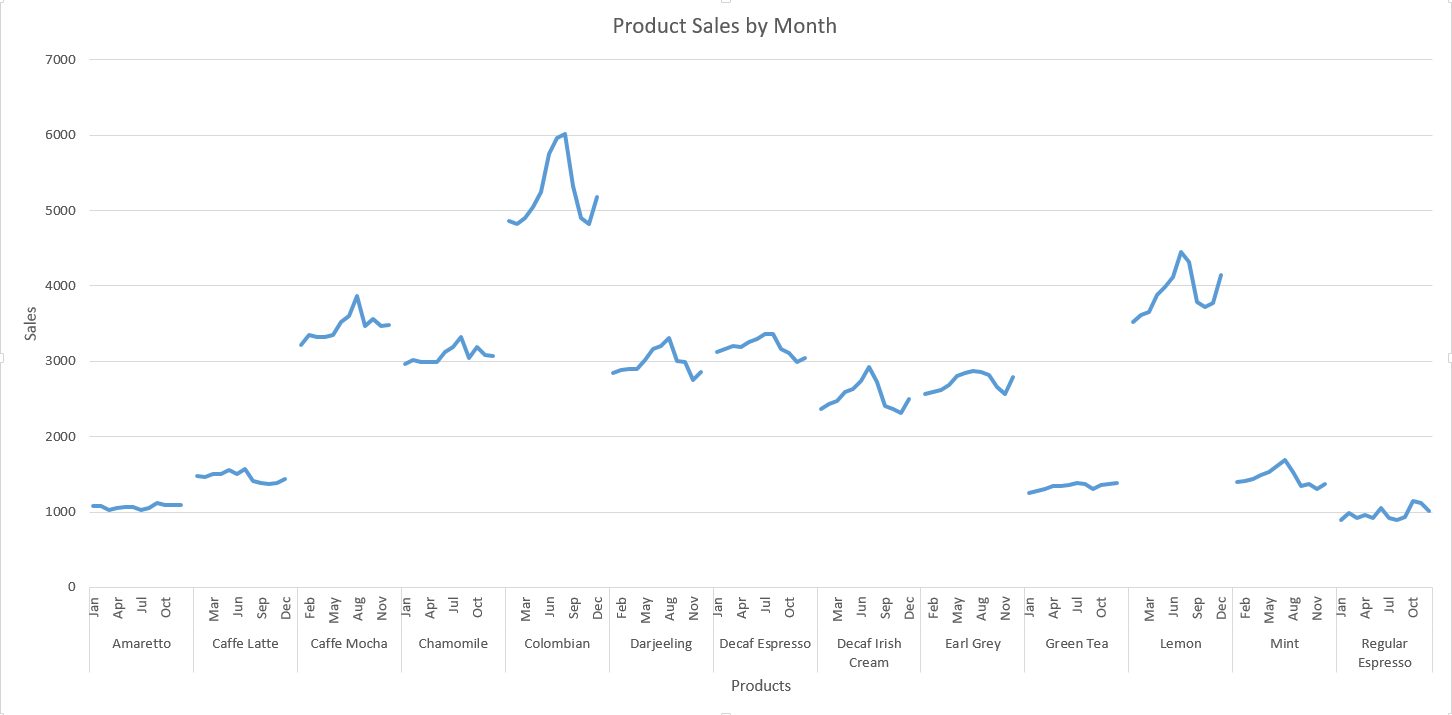
I used the output data from these queries to make graphs

**I found that sales tend to increase in the summer months and around the holidays as seen below:**

* + 1. Are there trends for any particular product?

Same idea as before

1. **SELECT** ProdName, extract(Month **from** FactDate) Month, SUM(actsales)
2. **FROM** FactCoffee
3. **INNER** JOIN ProdCoffee
4. **ON** Prodcoffee.ProductID = FActCoffee.productID
5. **WHERE** extract(Year **from** FactDate) = 2012
6. **GROUP** **BY** ProdName, extract(Month **from** FactDate)
7. **ORDER** **BY** ProdName, Month
8. ;

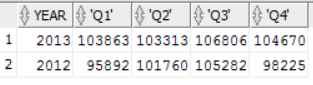


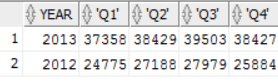
From this graph we see there are some products which aren’t affected much by seasonality while others like Colombian and Lemon seems to be affected by seasonality quite a bit

* + 1. Are there trends in any particular state for any product?

CHECK THIS

1. **SELECT** StateName, ProdName, extract(Month **from** FactDate) Month, SUM(actsales)
2. **FROM** FactCoffee
3. **INNER** JOIN ProdCoffee
4. **ON** Prodcoffee.ProductID = FActCoffee.productID
5. **INNER** JOIN Areacode
6. **ON** Areacode.AreaID = FactCoffee.AreaID
7. **INNER** JOIN States
8. **ON** States.stateid = areacode.stateid
9. **WHERE** extract(Year **from** FactDate) = 2012
10. **GROUP** **BY** StateName, ProdName, extract(Month **from** FactDate)
11. **ORDER** **BY** StateName, ProdName, Month
12. ;
    1. Insert a new column into Factcoffee table called Quarter. Now depending on the month, update the quarter number as Q1, Q2, Q3, or Q4 for each row.
       1. Now find the total sales for years 2012 and 2013 for each quarter. Display quarter in columns.

Sales: 

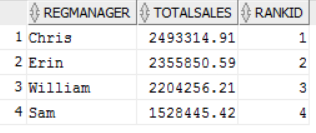
Profits: 

* + 1. Which quarter has the greatest sales and profits? **Q3 for both**

1. --8.    Insert a new column into Factcoffee table called Quarter.
2. **ALTER** **TABLE** Factcoffee
3. **ADD**
4. (Quarter VARCHAR2(30));
5. --ALTER TABLE Factcoffee
6. --drop column Quarter;
8. -- Now depending on the month, update the quarter number as Q1, Q2, Q3, or Q4 for each row.
9. **UPDATE** Factcoffee
10. **SET** Quarter = 'Q1'
11. **WHERE** extract(Month **from** FactDate) IN (1, 2, 3);
12. **UPDATE** Factcoffee
13. **SET** Quarter = 'Q2'
14. **WHERE** extract(Month **from** FactDate) IN (4, 5, 6);
15. **UPDATE** Factcoffee
16. **SET** Quarter = 'Q3'
17. **WHERE** extract(Month **from** FactDate) IN (7, 8, 9);
18. **UPDATE** Factcoffee
19. **SET** Quarter = 'Q4'
20. **WHERE** extract(Month **from** FactDate) IN (10, 11, 12);
22. **select** \* **FROM** Factcoffee;
23. --i.    Now find the total sales for years 2012 and 2013 for each quarter. Display quarter in columns.
24. **SELECT** \* **FROM**
25. (
26. (**SELECT** extract(YEar **FROM** FactDate) Year, Quarter, SUM(Actsales) TotalSales
27. **FROM** FactCoffee
28. **GROUP** **BY** extract(YEar **FROM** FactDate), Quarter)
29. PIVOT
30. (SUM(TotalSales) **FOR** (Quarter) In ('Q1', 'Q2', 'Q3', 'Q4'))
31. )
32. ;
33. --ii.   Which quarter has the greatest sales and profits?
34. --From last problem we see that Q3 has greatest sales
35. **SELECT** \* **FROM**
36. (
37. (**SELECT** extract(YEar **FROM** FactDate) Year, Quarter, SUM(Actprofit) TotalProfit
38. **FROM** FactCoffee
39. **GROUP** **BY** extract(YEar **FROM** FactDate), Quarter)
40. PIVOT
41. (SUM(TotalProfit) **FOR** (Quarter) In ('Q1', 'Q2', 'Q3', 'Q4'))
42. )
43. ;
44. --Now for profits: We see that again Q3 has greatest profits
    1. CREATE a TABLE that captures for each state, product, and quarter combination, the following measures - the total sales, total profits, percentage margin, total marketing expenses, and rank order of sales for each quarter. You may use many different queries to INSERT or UPDATE using a single query or union of many different queries.

PART B: Office Product

1. Rank managers based on the sales generated.
2. **SELECT** \* **FROM** (
3. **SELECT** Managers.RegManager, SUM(OrdSales) TotalSales,
4. ROW\_NUMBER() OVER (**Order** **BY** SUM(Ordsales) **DESC**) **AS** RANKID
5. **FROM** Managers
6. **INNER** JOIN Customers
7. **ON** Managers.RegID = Customers.CustReg
8. **INNER** JOIN OrderDet
9. **ON** Customers.CustID = OrderDet.CustID
10. **GROUP** **BY** RegManager)



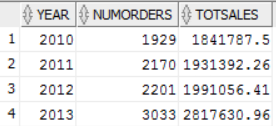
1. Find the products that had the worst average shipping times.
2. **SELECT** ProdName, AVG(OrdShipDate - OrdDate) AVGSHipTIme,
3. ROW\_NUMBER() OVER(**ORDER** **BY** (AVG(OrdShipDate - OrdDate)) **DESC**) **AS** RANKID
4. **FROM** Products
5. **INNER** JOIN OrderDet
6. **ON** Products.ProdID = OrderDet.ProdID
7. **GROUP** **BY** ProdName
8. ;



1. What fraction of the revenues is generated from the top 10% of the customers?

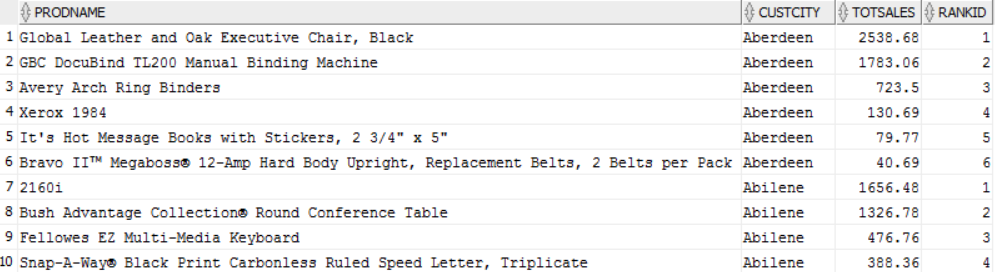
Are these 10% of the customers also the leaders in the number of orders placed?

1. List the number of orders, number of returns, total sales and any other metric for each year. List the years or measures in the columns.
2. **SELECT** extract(Year **FROM** ORdDate) Year, COUNT(Orderdet.OrderID) NumOrders, SUM(Orderdet.OrdSales) TotSales
3. **FROM** ORderDet
4. **INNER** JOIN Orders
5. **ON** Orderdet.OrderID = Orders.OrderID
6. **GROUP** **BY** extract(Year **FROM** ORdDate)
7. **ORDER** **BY** Year;
8. ------------------------------------------------------------------------------------------------------------------------------
9. --THis finds returned orders by year
10. **SELECT** Year, SUM(Ord)**as** TotORd
11. **FROM**
12. (
13. **SELECT** extract(Year **FROM** OrderDet.OrdDate) Year, Orders.OrderID, COUNT(Orders.OrderID) Ord
14. **FROM** Orders
15. **INNER** JOIN ORderDet
16. **ON** OrderDet.ORderID = Orders.ORderID
17. **WHERE** Status = 'Returned'
18. **GROUP** **BY** extract(Year **FROM** OrderDet.OrdDate), Orders.OrderID
19. )
20. **GROUP** **BY** Year;

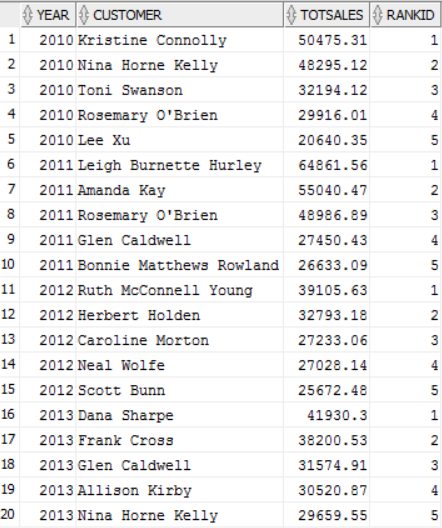




1. For each city and product combination, list the total sales and rank order in each city by total sales.
2. **SELECT** ProdName, CustCity, SUM(OrdSales) TotSales, Row\_number() OVER (PARTITION **BY** CustCIty **ORDER** **BY** sum(OrdSales) **DESC**) **AS** RANKID
3. **FROM** Products
4. **INNER** JOIN ORderDet
5. **ON** OrderDet.PRodID = Products.ProdID
6. **INNER** JOIN Customers
7. **ON** OrderDet.CustID = Customers.CustID
8. **GROUP** **BY** ProdNAme, CustCity;

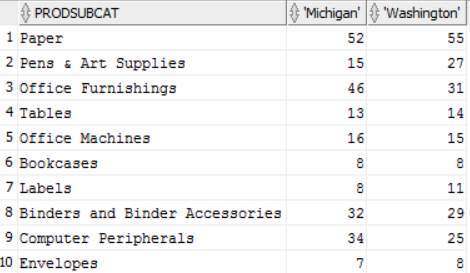


1. Which are the top 5 customers for each of the years?
2. **SELECT** \* **FROM**
3. (
4. **SELECT** extract(Year **from** OrdDate) Year, CustName Customer, SUM(OrdSales) TotSales,
5. ROW\_number() OVER(PARTITION **BY** extract(Year **from** OrdDate) **Order** **By** sum(OrdSales) **DESC**) **AS** RANKID
6. **FROM** OrderDet
7. **INNER** JOIN Customers
8. **ON** Orderdet.custid = customers.custid
9. **GROUP** **BY** extract(Year **from** OrdDate), CustName
10. )
11. **WHERE** RANKID IN (1, 2, 3, 4, 5);

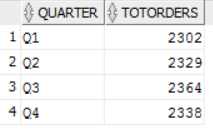


* 1. Who are common customers across all years?
  2. Are there some customers in any year that are distinct?

1. Find the number of orders in each subcategory in states Michigan and Washington. List Washington and Michigan in different columns.
2. **SELECT** \* **FROM**
3. (
4. **SELECT** ProdSubCat, Customers.CustState, COUNT(OrderDet.OrderID) NumOrders
5. **FROM** Products
6. **INNER** JOIN OrderDet
7. **ON** Products.ProdID = OrderDet.ProdID
8. **INNER** JOIN Customers
9. **ON** Customers.CustID = OrderDet.CustID
10. **WHERE** Customers.CustState In ('Michigan', 'Washington')
11. **GROUP** **BY** ProdSubCat, CustState
12. )
13. PIVOT
14. (SUM(NumOrders) **FOR** CustState IN ('Michigan', 'Washington'))
15. ;



1. Find total orders in each quarter.
2. --8.    Find total orders in each quarter.
4. --first I'll add a quarter column to OrderDet table
5. **ALTER** **TABLE** OrderDet
6. **ADD**
7. (Quarter VARCHAR2(30));
8. --ALTER TABLE Factcoffee
9. --drop column Quarter;
11. -- Now depending on the month, update the quarter number as Q1, Q2, Q3, or Q4 for each row.
12. **UPDATE** OrderDet
13. **SET** Quarter = 'Q1'
14. **WHERE** extract(Month **from** OrdDate) IN (1, 2, 3);
15. **UPDATE** OrderDet
16. **SET** Quarter = 'Q2'
17. **WHERE** extract(Month **from** OrdDate) IN (4, 5, 6);
18. **UPDATE** OrderDet
19. **SET** Quarter = 'Q3'
20. **WHERE** extract(Month **from** OrdDate) IN (7, 8, 9);
21. **UPDATE** OrderDet
22. **SET** Quarter = 'Q4'
23. **WHERE** extract(Month **from** OrdDate) IN (10, 11, 12);
25. **SELECT** \* **FROM** OrderDet;
26. --Now find total orders per quarter
27. **SELECT** Quarter, COUNT(OrderID) totOrders
28. **FROM** ORderDET
29. **GROUP** **BY** Quarter
30. **ORDER** **BY** QUARTER
31. ;



1. For each quarter and customer segment, find the total sales. Display data for quarters in column.
2. **SELECT** \* **FROM**
3. (**SELECT** Custseg, Quarter, SUM(OrdSales) TotSales
4. **FROM** Customers
5. **INNER** JOIN ORderDet
6. **On** ORderdet.CustID = Customers.CustID
7. **GROUP** **BY** CUstSeg, Quarter
8. **ORDER** **BY** TotSales **DESC**)
9. PIVOT
10. (SUM(TotSales) **FOR** Quarter In ('Q1', 'Q2', 'Q3', 'Q4'))
11. ;

