Assignment 3

**Due**: September 24, 2016 (midnight)

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

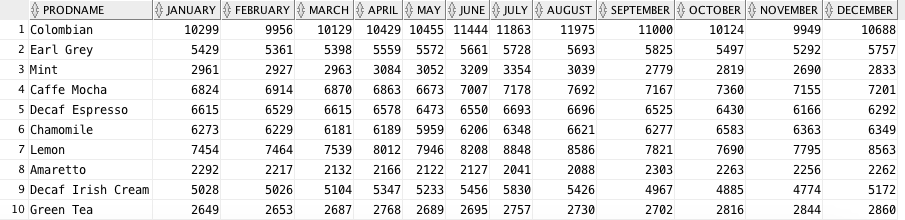
**Grade**: 7%

# PART A – Coffee Sales

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

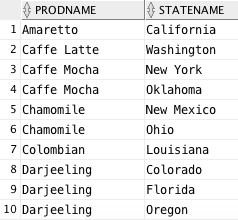
* 1. Extract the total sales for each product for each month. List all months (like January, February, etc) in the columns.

1. SELECT \* FROM (
2. SELECT prodname, extract(month from factdate) as monthname, sum(actsales) Sumsales
3. from prodcoffee, factcoffee WHERE factcoffee.productid=prodcoffee.productid
4. GROUP BY prodname, extract(month from factdate))
5. PIVOT (
6. SUM(Sumsales)
7. FOR monthname IN
8. (1 as January, 2 as February, 3 as March, 4 as April, 5 as May, 6 as June, 7 as July, 8 as August, 9 as September, 10 as October, 11 as November, 12 as December));



* 1. In each state, identify the product with greatest sales for the year 2012.
     1. Identify the states where the best selling product remained the same in 2013 (compared to best selling product in 2012)

1. **SELECT** prodname, statename **from**(
2. **SELECT** \* **FROM**(
3. **SELECT** prodname, statename, sum(actsales) **as** totsales,--, extract(year from factdate),
4. row\_number() over (partition **by** statename **order** **by** sum(actsales)) **as** rankid
5. **FROM** prodcoffee, factcoffee, areacode, states
6. **WHERE** prodcoffee.productid=factcoffee.productid and factcoffee.areaid=areacode.areaid and states.stateid=areacode.stateid
7. and extract(year **from** factdate)=2012
8. **GROUP** **BY** statename, prodname, extract(year **from** factdate)
9. **ORDER** **BY** totsales **DESC**)
10. **WHERE** rankid=1)
11. **INTERSECT**
12. **Select** prodname, statename **from**(
13. **SELECT** \* **FROM**(
14. **SELECT** prodname, statename, sum(actsales) **as** totsales,
15. row\_number() over (partition **by** statename **order** **by** sum(actsales)) **as** rankid
16. **FROM** prodcoffee, factcoffee, areacode, states
17. **WHERE** prodcoffee.productid=factcoffee.productid and factcoffee.areaid=areacode.areaid and states.stateid=areacode.stateid
18. and extract(year **from** factdate)=2013
19. **GROUP** **BY** statename, prodname, extract(year **from** factdate)
20. **ORDER** **BY** totsales **DESC**)
21. **WHERE** rankid=1);



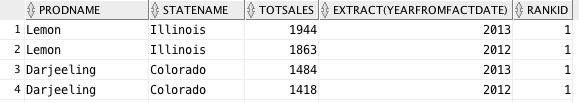
* + 1. Identify the states where the best selling product has changed.

None of the best selling products changed.

1. **SELECT** prodname, statename **from**(
2. **SELECT** \* **FROM**(
3. **SELECT** prodname, statename, sum(actsales) **as** totsales,--, extract(year from factdate),
4. row\_number() over (partition **by** statename **order** **by** sum(actsales)) **as** rankid
5. **FROM** prodcoffee, factcoffee, areacode, states
6. **WHERE** prodcoffee.productid=factcoffee.productid and factcoffee.areaid=areacode.areaid and states.stateid=areacode.stateid
7. and extract(year **from** factdate)=2013
8. **GROUP** **BY** statename, prodname, extract(year **from** factdate)
9. **ORDER** **BY** totsales **DESC**)
10. **WHERE** rankid=1)
11. MINUS
12. **Select** prodname, statename **from**(
13. **SELECT** \* **FROM**(
14. **SELECT** prodname, statename, sum(actsales) **as** totsales,--, extract(year from factdate),
15. row\_number() over (partition **by** statename **order** **by** sum(actsales)) **as** rankid
16. **FROM** prodcoffee, factcoffee, areacode, states
17. **WHERE** prodcoffee.productid=factcoffee.productid and factcoffee.areaid=areacode.areaid and states.stateid=areacode.stateid
18. and extract(year **from** factdate)=2012
19. **GROUP** **BY** statename, prodname, extract(year **from** factdate)
20. **ORDER** **BY** totsales **DESC**)
21. **WHERE** rankid=1);
    * 1. Identify the products that were best in 2012 but not in 2013.

None

1. **SELECT** prodname, statename **from**(
2. **SELECT** \* **FROM**(
3. **SELECT** prodname, statename, sum(actsales) **as** totsales,--, extract(year from factdate),
4. row\_number() over (partition **by** statename **order** **by** sum(actsales)) **as** rankid
5. **FROM** prodcoffee, factcoffee, areacode, states
6. **WHERE** prodcoffee.productid=factcoffee.productid and factcoffee.areaid=areacode.areaid and states.stateid=areacode.stateid
7. and extract(year **from** factdate)=2012
8. **GROUP** **BY** statename, prodname, extract(year **from** factdate)
9. **ORDER** **BY** totsales **DESC**)
10. **WHERE** rankid=1)
11. MINUS
12. **Select** prodname, statename **from**(
13. **SELECT** \* **FROM**(
14. **SELECT** prodname, statename, sum(actsales) **as** totsales,--, extract(year from factdate),
15. row\_number() over (partition **by** statename **order** **by** sum(actsales)) **as** rankid
16. **FROM** prodcoffee, factcoffee, areacode, states
17. **WHERE** prodcoffee.productid=factcoffee.productid and factcoffee.areaid=areacode.areaid and states.stateid=areacode.stateid
18. and extract(year **from** factdate)=2013
19. **GROUP** **BY** statename, prodname, extract(year **from** factdate)
20. **ORDER** **BY** totsales **DESC**)
21. **WHERE** rankid=1);
    * 1. Identify the top two best selling products that are common to both 2012 and 2013.
22. **SELECT** \* **FROM**(
23. **SELECT** prodname, statename, sum(actsales) **as** totsales,extract(year **from** factdate),--, extract(year from factdate),
24. row\_number() over (partition **by** statename, extract(year **from** factdate) **order** **by** sum(actsales)) **as** rankid
25. **FROM** prodcoffee, factcoffee, areacode, states
26. **WHERE** prodcoffee.productid=factcoffee.productid and factcoffee.areaid=areacode.areaid and states.stateid=areacode.stateid
27. --and extract(year from factdate)=2013
28. **GROUP** **BY** statename, prodname, extract(year **from** factdate)
29. **ORDER** **BY** totsales **DESC**)
30. **WHERE** rankid=1
31. **Order** **by** totsales **DESC**;



Lemon tea and Darjeeling are the top products in Illinois and Colorado, and lead in overall sales.

* 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. **With** Cumsale **as** (**SELECT** statename, SUM(actsales) Sumsales,
2. Row\_number() OVER (**ORDER** **BY** sum(actsales) **DESC**) Rowsales
3. **FROM** factcoffee, areacode, states
4. **WHere** factcoffee.areaid=areacode.areaid and states.stateid=areacode.stateid
5. **GROUP** **BY** statename),
6. totalcount **AS** ( **SELECT** count(\*) **as** totcount **FROM** Cumsale),
7. totsales **as** (**SELECT** sum(sumsales) totsumsales **FROM** cumsale),
8. Cumtotsales **as** (**SELECT** rowsales, sum(sumsales) over (**order** **by** rowsales) Csales **FROM** cumsale)
9. **SELECT** totcount, totsumsales, rowsales, csales, rowsales/totcount
10. **FROM** totalcount, totsales, cumtotsales
11. **WHERE** csales >= 0.5\*totsumsales AND rownum =1;

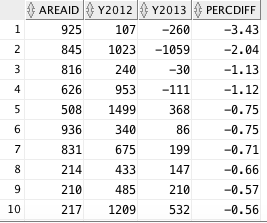


1. **With** CumProfit **as** (**SELECT** statename, SUM(actprofit) Sumprofit,
2. Row\_number() OVER (**ORDER** **BY** sum(actprofit) **DESC**) Rowprofit
3. **FROM** factcoffee, areacode, states
4. **WHere** factcoffee.areaid=areacode.areaid and states.stateid=areacode.stateid
5. **GROUP** **BY** statename),
6. totalcount **AS** ( **SELECT** count(\*) **as** totcount **FROM** CumProfit),
7. totprof **as** (**SELECT** sum(sumprofit) totsumprof **FROM** cumprofit),
8. Cumtotprof **as** (**SELECT** rowprofit, sum(sumprofit) over (**order** **by** rowprofit) Cprof **FROM** cumProfit)
9. **SELECT** totcount, totsumprof, rowprofit, cprof, rowprofit/totcount
10. **FROM** totalcount, totprof, cumtotprof
11. **WHERE** cprof >= 0.5\*totsumprof AND rownum =1;



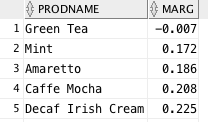
* 1. Identify the area codes with greatest decline in profits from the year 2012 to 2013. List the profits for 2012 and 2013 in the columns and display the percentage decline.

1. **SELECT** areaid, Y2012, Y2013, Round((Y2013-Y2012)/abs(Y2012),2) **as** PercDiff **from**(
2. **Select** \* **from**(
3. **SELECT** areaid, extract(year **from** factdate) year, sum(actprofit) **as** Profit
4. **FROM** factcoffee
5. **group** **by** areaid, extract(year **from** factdate))
6. Pivot (sum(Profit) **for** year in (2012 **as** Y2012, 2013 **as** Y2013)))
7. **Order** **by** PercDiff;



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

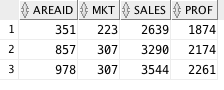
1. **SELECT** prodname, Round(Prof/Sales,3) Marg **from**(
2. **SELECT** prodname, sum(actsales) Sales, sum(actprofit) Prof
3. **FROM** factcoffee, prodcoffee
4. **WHERE** prodcoffee.productid=factcoffee.productid
5. **GROUP** **BY** prodname)
6. **order** **by** Marg;



I would discontinue Green Tea, since it has a negative profit margin.

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

1. **SELECT** \* **FROM**(
2. **SELECT** areaid, sum(actmarkcost) MKT, sum(actsales) Sales, sum(actprofit) **as** Prof
3. **from** factcoffee
4. **group** **by** areaid
5. **order** **by** MKT)
6. **Where** MKT>0.25\*Sales;
7. --areaids 914 and 845, also making negative profits
8. --marketing expenses should be reduced here
9. **SELECT** \* **FROM**(
10. **SELECT** areaid, sum(actmarkcost) MKT, sum(actsales) Sales, sum(actprofit) **as** Prof
11. **from** factcoffee
12. **group** **by** areaid
13. **order** **by** MKT)
14. **Where** MKT<0.1\*Sales;
15. --areaids 351,857 and 978 should increase marketing expenses

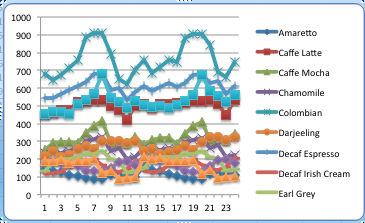


From these two tables, we see that areaids 914 and 845 are spending more than 25% of their sales revenue on marketing, and are having negative profits, so they should think about decreasing marketing costs. Areaids 351, 857, and 978, on the other hand, are spending less than 10% of their sales revenue on marketing, and potentially increase their profits if they spent more on marketing.

* 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.

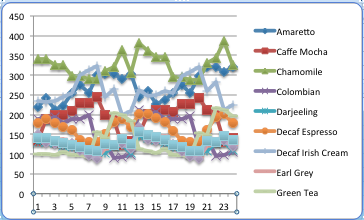
* + 1. Are there trends for any particular product?
    2. Are there trends in any particular state for any product?

California



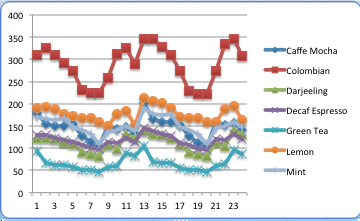
Colombian coffee is the most cyclical in terms of sales.

Colorado

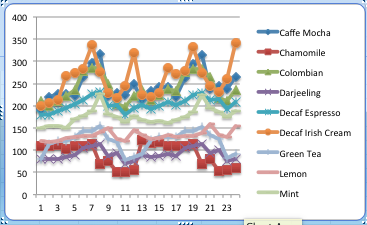


Amarettos are increasing in popularity, and decaf irish cream and chamomile are the most cyclical.

Connecticut

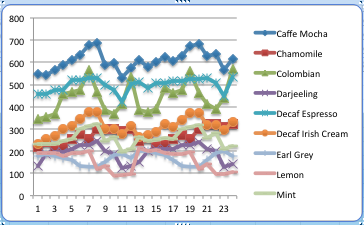
  
Colombian coffee is very cyclical.

Florida



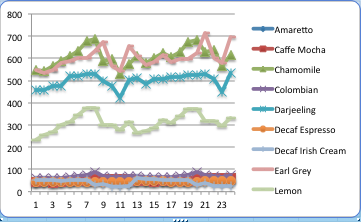
Decaf Irish cream, cafe mocha, and Colombian coffee follow similar trends.

Illinois



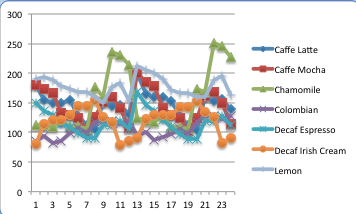
Colombian coffee once again has the most severe seasonality.

Iowa



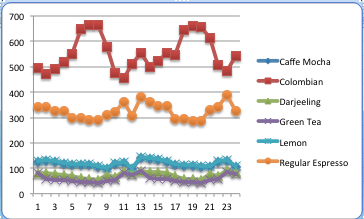
The top 3 products in Iowa are Darjeeling, chamomile, and earl grey teas.

Louisiana



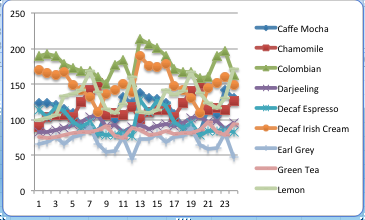
Chamomile has had an upward trend.

Massachusetts



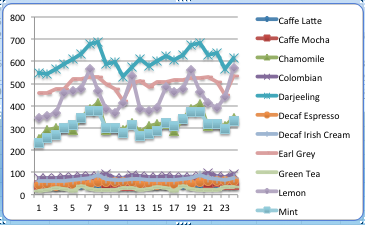
Regular espresso and Colombian coffee are much more popular, and Colombian coffee is once again very seasonal.

Missouri



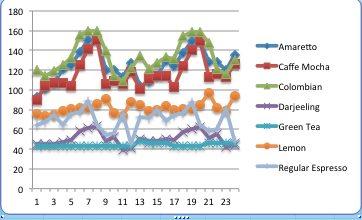
Chamomile, Colombian coffee, lemon tea and decaf Irish cream are very seasonal.

Nevada



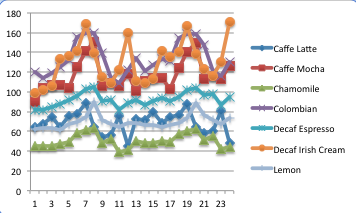
Lemon tea has a harsher seasonal trend, while mint tea, chamomile, and Darjeeling teas have a less severe cyclical trend.

New Hampshire



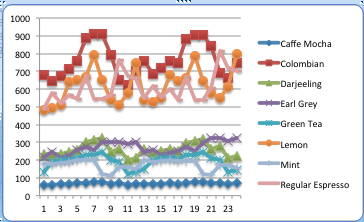
Colombian coffee, cafe mochas, regular espresso, and amarettos follow similar trends.

New Mexico



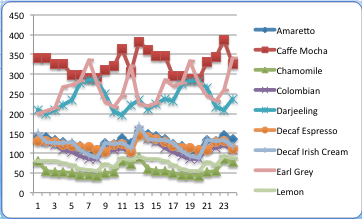
Decaf Irish cream has spikes every few months, while Colombian has a more continuous cycle. Cafe Lattes have had a somewhat steady downward trend.

New York



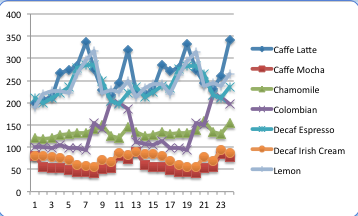
Colombian coffee, regular espresso, and lemon tea have cyclical trends almost inverses of each other.

Ohio



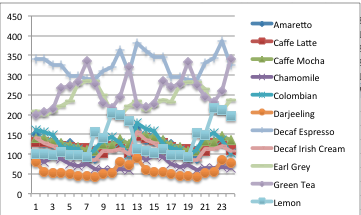
Earl grey tea, Colombian, and Darjeeling have the most cyclical patterns.

Oklahoma



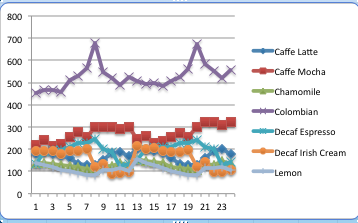
Cafe lattes, lemon tea, and decaf espresso are the most popular and very cyclical; Colombian coffee has a very dramatic cyclical trend.

Oregon



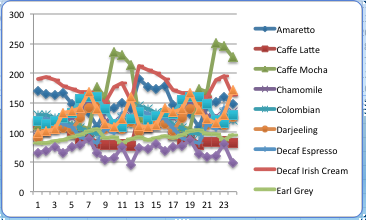
Green tea, earl grey, lemon tea and decaf espresso are very cyclical.

Texas



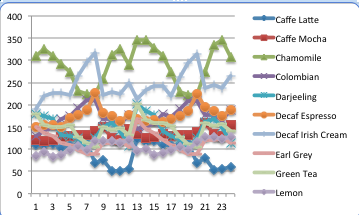
Colombian coffee peaks in august in both years.

Utah



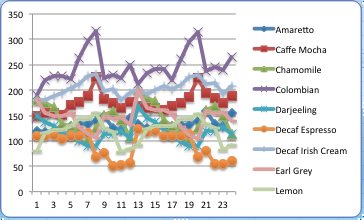
Cafe mochas are increasing in popularity; decaf Irish cream, amarettos, and Darjeeling teas have a cyclical pattern.

Washington



Chamomile has a steep drop in August, and cafe lattes are decreasing in popularity. Decaf Irish cream and chamomile seem inversely related.

Wisconsin



Decaf espressos have a pretty consistent cyclical, and seems to be inversely related to Colombian coffee and cafe mochas.

* 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. **ALTER** **table** factcoffee
2. **Add** Quarter **char**(2);
4. **UPDATE** factcoffee
5. **SET** Quarter='Q4'
6. **where** to\_char(factdate, 'Q')=4;

(completed for each quarter)

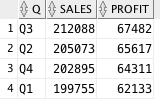
* + 1. Now find the total sales for years 2012 and 2013 for each quarter. Display quarter in columns.

1. **Select** \* **from**(
2. **SELECT** Quarter **as** Q, sum(actsales) **as** sales, extract(year **from** factdate) year
3. **From** factcoffee
4. **group** **by** extract(year **from** factdate), Quarter)
5. Pivot (sum(sales) **for** Q in ('Q1' **as** Q1, 'Q2' **as** Q2, 'Q3' **as** Q3, 'Q4' **as** Q4));



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

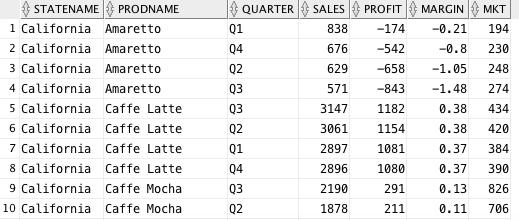
1. **SELECT** Quarter **as** Q, sum(actsales) **as** Sales, sum(actprofit) as Profit
2. **from** factcoffee
3. **group** **by** quarter;



Quarter 3 has the highest sales and profits combined for both years.

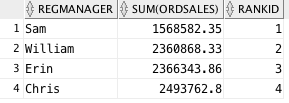
* 1. CREATE a TABLE that captures for each state, product, and quarter combination, the following measures - the total sales, total profits, percentage margin, total marketing expenses, and rank order of sales for each quarter. You may use many different queries to INSERT or UPDATE using a single query or union of many different queries.

1. **CREATE** **table** Q9 **as** (
2. **SELECT** statename, prodname, quarter, SALES, PROFIT, Round(Profit/Sales,2) Margin, MKT **FROM**(
3. **SELECT** statename, prodname, sum(actsales) SALES, sum(actprofit) PROFIT, sum(actmarkcost) MKT, quarter,
4. row\_number() over (partition **by** statename, prodname **order** **by** sum(actsales) **DESC**) **as** rankid
5. **FROM** states, areacode, factcoffee, prodcoffee
6. **WHERE** states.stateid=areacode.stateid and areacode.areaid=factcoffee.areaid and prodcoffee.productid=factcoffee.productid
7. **GROUP** **BY** statename, prodname, quarter));

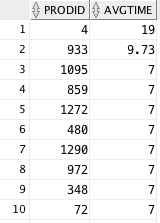


PART B: Office Product

1. Rank managers based on the sales generated.
2. **SELECT** Regmanager, sum(ordsales),
3. row\_number() over (**order** **by** sum(ordsales)) **as** rankid
4. **FROM** managers, orderdet, customers
5. **where** managers.regid=customers.custreg and customers.custid=orderdet.custid
6. **group** **by** regmanager;



1. Find the products that had the worst average shipping times.
2. **SELECT** prodid, Round(**time**/Num,2) **as** avgtime **FROM**(
3. **SELECT** prodid, sum(NumOrder) Num, (sum(ordshipdate-orddate)) **as** **time**
4. **FROM**(
5. **SELECT** prodid, count(custid) count, orddate, ordshipdate, count(orderid) NumOrder
6. **FROM** orderdet
7. **GROUP** **BY** prodid, orddate, ordshipdate
8. **order** **by** count **DESC**)
9. **group** **by** prodid)
10. **Order** **by** avgtime **DESC**;



1. What fraction of the revenues is generated from the top 10% of the customers?
2. **Create** **view** sales **as**
3. **SELECT** sum(sales) totsales **FROM**(
4. **SELECT** \* **From**(
5. **Select** custname, sum(ordsales) sales
6. **from** customers, orderdet
7. **where** customers.custid=orderdet.custid
8. **group** **by** custname
9. **order** **by** sum(ordsales) **DESC**)
10. **where** rownum<=270);
12. **With** totalsales **as** (**Select** totsales **from** sales),
13. total **as** (**Select** sum(ordsales) **as** tot **from** orderdet)
14. **Select** Round(totsales/tot,4) **as** Perc
15. **from** totalsales, total;

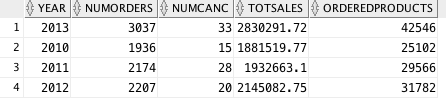


1. Are these 10% of the customers also the leaders in the number of orders placed?
2. **Select** \* **from**(
3. **SELECT** custname, sum(ordsales) sales
4. **from** customers, orderdet
5. **where** customers.custid=orderdet.custid
6. **group** **by** custname
7. **order** **by** sales **DESC**)
8. **where** rownum<=270
9. **INTERSECT**
10. **SELECT** custname **from**(
11. **select** \* **from**(
12. **SELECT** custname, count(orderid) orders
13. **from** customers, orderdet
14. **where** customers.custid=orderdet.custid
15. **group** **by** custname
16. **order** **by** orders **DESC**)
17. **where** rownum<=270);

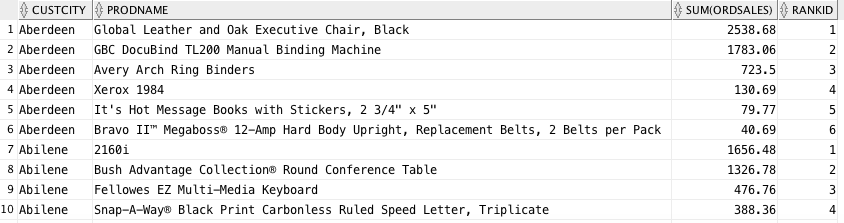


Yes, quite a few customers are in the list of top 10% revenue-bringing customers and top number of orders placed.

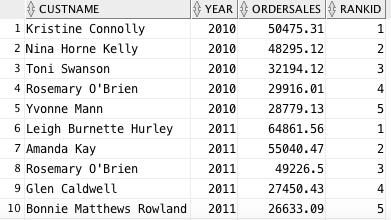
1. List the number of orders, number of returns, total sales and any other metric for each year. List the years or measures in the columns.
2. **SELECT** extract(year **from** orddate) **as** Year, count(orders.orderid) **as** NumOrders, count(status) **as** NumCanc, sum(ordsales) **as** totsales, sum(ordqty) as OrderedProducts
3. **FROM** orderdet, orders
4. **WHERE** orderdet.orderid=orders.orderid
5. **GROUP** **BY** extract(year **from** orddate);



1. For each city and product combination, list the total sales and rank order in each city by total sales.
2. **SELECT** custcity, prodname, sum(ordsales),
3. row\_number() over (partition **by** custcity **order** **by** sum(ordsales) **DESC**) **as** rankid
4. **FROM** customers, products, orderdet
5. **WHERE** customers.custid=orderdet.custid and orderdet.prodid=products.prodid
6. **GROUP** **BY** custcity, prodname;



1. Which are the top 5 customers for each of the years?
2. **Select** \* **from**(
3. **SELECT** custname, extract(year **from** orddate) year, sum(ordsales) **as** OrderSales,
4. row\_number() over (partition **by** extract(year **from** orddate) **order** **by** sum(ordsales) **DESC**) **as** rankid
5. **FROM** customers, orderdet
6. **where** customers.custid=orderdet.custid
7. **GROUP** **by** custname, extract(year **from** orddate))
8. **where** rankid<=5;



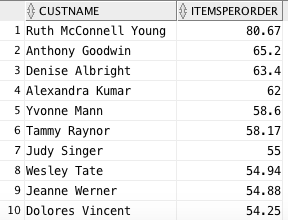
* 1. Who are common customers across all years?

1. **SELECT** custname
2. **FROM** customers, orderdet
3. **where** customers.custid=orderdet.custid and extract(year **from** orddate)=2010
4. **INTERSECT**
5. **SELECT** custname
6. **FROM** customers, orderdet
7. **where** customers.custid=orderdet.custid and extract(year **from** orddate)=2011
8. **INTERSECT**
9. **SELECT** custname
10. **FROM** customers, orderdet
11. **where** customers.custid=orderdet.custid and extract(year **from** orddate)=2012
12. **INTERSECT**
13. **SELECT** custname
14. **FROM** customers, orderdet
15. **where** customers.custid=orderdet.custid and extract(year **from** orddate)=2013;



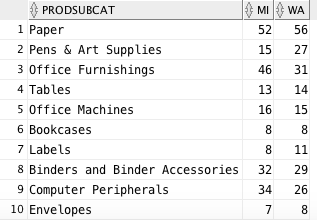
* 1. Are there some customers in any year that are distinct?

1. **SELECT** custname, Round(Qty/NumOrder,2) ItemsPerOrder **from**(
2. **SELECT** custname, count(orderid) **as** NumOrder, sum(ordqty) Qty
3. **from** orderdet, customers
4. **where** customers.custid=orderdet.custid
5. **group** **by** custname
6. **order** **by** Qty **DESC**)
7. **Order** **by** itemsperorder **DESC**;



These customers ordered on average more than 50 items per order.

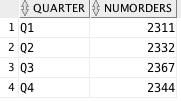
1. Find the number of orders in each subcategory in states Michigan and Washington. List Washington and Michigan in different columns.
2. **SELECT** \* **FROM** (
3. **SELECT** custstate State, prodsubcat, count(orderid) **as** NumOrders
4. **FROM** products, orderdet, customers
5. **WHERE** products.prodid=orderdet.prodid and customers.custid=orderdet.custid
6. and custstate in ('Michigan','Washington')
7. **GROUP** **BY** custstate, prodsubcat)
8. PIVOT (sum(NumOrders) **for** state in ('Michigan' **as** MI, 'Washington' **as** WA));



1. Find total orders in each quarter.

First, I altered the table for each quarter.

1. **ALTER** **table** orderdet
2. **Add** Quarter **char**(2);
4. **UPDATE** orderdet
5. **SET** Quarter='Q4'
6. **where** to\_char(orddate, 'Q')=4;



1. For each quarter and customer segment, find the total sales. Display data for quarters in column.
2. **SELECT** \* **FROM**(
3. **SELECT** Quarter Q, Custseg, sum(ordsales) TotSales
4. **FROM** customers, orderdet
5. **where** customers.custid=orderdet.custid
6. **GROUP** **BY** Quarter, Custseg)
7. PIVOT (sum(TotSales) **for** Q in ('Q1' **as** Q1, 'Q2' **as** Q2, 'Q3' **as** Q3, 'Q4' **as** Q4));

