SQL Assignment 3 – Gihani Dissanayake

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.***

with q1 as(
SELECT to_char(factdate, 'mm') as Month, prodname, sum(actsales) as SumSales
FROM factcoffee F, prodcoffee P
where f.productid = P.productid
group by to_char(factdate, 'mm'), extract(month from factdate), prodname
order by extract(month from factdate) asc, sumsales desc)

select prodname, nvl(Jan,0) as Jan,
nvl(Feb,0) as Feb, nvl(Mar,0) as Mar, nvl(Apr,0) as Apr,
nvl(May,0) as May, nvl(Jun,0) as Jun, nvl(Jul,0) as Jul,
nvl(Aug,0) as aug, nvl(Sep,0) as sep, nvl(Oct,0) as Oct,
nvl(Nov,0) as nov, nvl(Dec,0) as Dec
from q1
PIVOT (
  sum(sumsales)
  for month in ('01' as Jan,
  '02' as Feb, '03' as Mar,
  '04' as Apr, '05' as May,
  '06' as Jun, '07' as Jul,
  '08' as Aug, '09' as Sep,
  '10' as Oct, '11' as Nov,
  '12' as Dec));

PRODNAME	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1 Colombian	10299	9956	10129	10429	10455	11444	11863	11975	11000	10124	9949	10688
2 Earl Grey	5429	5361	5398	5559	5572	5661	5728	5693	5825	5497	5292	5757
3 Mint	2961	2927	2963	3084	3052	3209	3354	3039	2779	2819	2690	2833
4 Caffe Mocha	6824	6914	6870	6863	6673	7007	7178	7692	7167	7360	7155	7201
5 Lemon	7454	7464	7539	8012	7946	8208	8848	8586	7821	7690	7795	8563
6 Decaf Espresso	6615	6529	6615	6578	6473	6550	6693	6696	6525	6430	6166	6292
7 Chamomile	6273	6229	6181	6189	5959	6206	6348	6621	6277	6583	6363	6349
8 Amaretto	2292	2217	2132	2166	2122	2127	2041	2088	2303	2263	2256	2262
9 Darjeeling	6017	5946	5980	5973	5996	6289	6382	6592	6218	6177	5671	5910
10 Decaf Irish Cream	5028	5026	5104	5347	5233	5456	5830	5426	4967	4885	4774	5172
11 Green Tea	2649	2653	2687	2768	2689	2695	2757	2730	2702	2816	2844	2860
12 Caffe Latte	3131	3022	3096	3117	3091	2988	3126	2809	2854	2831	2850	2984
13 Regular Espresso	1899	2040	1906	1970	1825	2092	1843	1789	1923	2361	2301	2082

Assume that the desired result is the sum of sales for each month between multiple years for each product.

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

with q1 as (
select S.statename, F.productid, sum(F.actsales) as SumSales,
RANK() OVER (PARTITION BY S.statename
   ORDER BY (sum(F.actsales)) DESC) as prodrank
from states S, factcoffee F, areacode A
where S.stateid = A.stateid
  and A.areaid = F.areaid
  and extract(year from F.factdate) = 2012
group by S.statename, F.productid)

select statename, productid, sumsales, prodrank
from q1
where q1.prodrank = 1
order by statename;

	STATENAME	PRODUCTID	SUMSALES	PRODRANK
1	California	2	8940	1
2	Colorado	8	3810	1
3	Connecticut	2	3378	1
4	Florida	3	3068	1
5	Illinois	5	7151	1
6	Iowa	8	7151	1
7	Louisiana	9	2085	1
8	Massachusetts	2	6518	1
9	Missouri	2	2085	1
10	Nevada	11	7151	1

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

with q1 as (
select S.statename, F.productid, sum(F.actsales) as SumSales12,
RANK() OVER (PARTITION BY S.statename
   ORDER BY (sum(F.actsales)) DESC) as prodrank2012
from states S, factcoffee F, areacode A
where S.stateid = A.stateid
  and A.areaid = F.areaid
  and extract(year from F.factdate) = 2012
group by S.statename, F.productid),

q2 as (
select S.statename, F.productid, sum(F.actsales) as SumSales13,
RANK() OVER (PARTITION BY S.statename
   ORDER BY (sum(F.actsales)) DESC) as prodrank2013
from states S, factcoffee F, areacode A

where S.stateid = A.stateid
   and A.areaid = F.areaid
   and extract(year from F.factdate) = 2013
group by S.statename, F.productid)

select q1.statename, q1.productid, sumsales12, sumsales13, prodrank2012, prodrank2013
from q1, q2
where q1.statename = q2.statename
   and q1.productid = q2.productid
   and prodrank2012 = 1
   and prodrank2013 = 1
order by statename;

	STATENAME	PRODUCTID	SUMSALES12	SUMSALES13	PRODRANK2012	PRODRANK2013
1	California	2	8940	9305	1	1
2	Colorado	8	3810	3988	1	1
3	Connecticut	2	3378	3545	1	1
4	Florida	3	3068	3194	1	1
5	Illinois	5	7151	7456	1	1
6	Iowa	8	7151	7456	1	1
7	Louisiana	9	2085	2182	1	1
8	Massachusetts	2	6518	6783	1	1
9	Missouri	2	2085	2182	1	1
10	Nevada	11	7151	7456	1	1
11	New Hampshire	2	1570	1634	1	1
12	New Mexico	2	1570	1634	1	1
13	New York	2	8940	9305	1	1
14	Ohio	5	3810	3988	1	1
15	Oklahoma	4	3068	3194	1	1
16	Oregon	6	3810	3988	1	1
17	Texas	2	6210	6471	1	1
18	Utah	3	2085	2182	1	1
19	Washington	8	3378	3545	1	1
20	Wisconsin	2	2890	3008	1	1

ii. **Identify the states where the best selling product has changed.**

with q1 as (
select S.statename, F.productid, sum(F.actsales) as SumSales12,
RANK() OVER (PARTITION BY S.statename
   ORDER BY (sum(F.actsales)) DESC) as prodrank2012
from states S, factcoffee F, areacode A
where S.stateid = A.stateid
   and A.areaid = F.areaid
   and extract(year from F.factdate) = 2012
group by S.statename, F.productid),

q2 as (
select S.statename, F.productid, sum(F.actsales) as SumSales13,
RANK() OVER (PARTITION BY S.statename
   ORDER BY (sum(F.actsales)) DESC) as prodrank2013
from states S, factcoffee F, areacode A
where S.stateid = A.stateid

and A.areaid = F.areaid
and extract(year from F.factdate) = 2013
group by S.statename, F.productid)

select q1.statename, q1.productid, sumsales12, sumsales13, prodrank2012, prodrank2013
from q1, q2
where q1.statename = q2.statename
and q1.productid = q2.productid
and prodrank2012 <> prodrank2013
order by statename;

There were none that met this criteria. In fact, the product ranking within a state did not change at all from 2012 to 2013.

STATENAME	PRODUCTID	SUMSALES12	SUMSALES13	PRODRANK2012	PRODRANK2013

.

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

with q1 as (
select S.statename, F.productid, sum(F.actsales) as SumSales12,
RANK() OVER (PARTITION BY S.statename
ORDER BY (sum(F.actsales)) DESC) as prodrank2012
from states S, factcoffee F, areacode A
where S.stateid = A.stateid
and A.areaid = F.areaid
and extract(year from F.factdate) = 2012
group by S.statename, F.productid),

q2 as (
select S.statename, F.productid, sum(F.actsales) as SumSales13,
RANK() OVER (PARTITION BY S.statename
ORDER BY (sum(F.actsales)) DESC) as prodrank2013
from states S, factcoffee F, areacode A
where S.stateid = A.stateid
and A.areaid = F.areaid
and extract(year from F.factdate) = 2013
group by S.statename, F.productid)

select q1.statename, q1.productid, sumsales12, sumsales13, prodrank2012, prodrank2013
from q1, q2
where q1.statename = q2.statename
and q1.productid = q2.productid
and prodrank2012 = 1
and prodrank2013 <> 1
order by statename;

There were none that met this criteria.

STATENAME	PRODUCTID	SUMSALES12	SUMSALES13	PRODRANK2012	PRODRANK2013

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

with q1 as (
select P.prodname, sum(F.actsales) as SumSales12,
ROW_NUMBER() OVER (ORDER BY sum(F.actsales) DESC) as prodrank2012
from factcoffee F, prodcoffee P
where extract(year from F.factdate) = 2012
*   and p.productid = f.productid*
group by P.prodname
order by sumsales12 desc),

q2 as (
select P.prodname, sum(F.actsales) as SumSales13,
ROW_NUMBER() OVER (ORDER BY sum(F.actsales) DESC) as prodrank2013
from factcoffee F, prodcoffee P
where extract(year from F.factdate) = 2013
*   and p.productid = f.productid*
group by P.prodname
order by sumsales13 desc)

select q1.prodname, SumSales12, SumSales13, prodrank2012, prodrank2013
from q1, q2
where q1.prodname = q2.prodname
fetch first 2 rows only;

	PRODNAME	SUMSALES12	SUMSALES13	PRODRANK2012	PRODRANK2013
1	Colombian	62824	65487	1	1
2	Lemon	46956	48970	2	2

All the products had the same sum of sales rank in 2012 and 2013. As such, the top 2 sellers are the top two common to both years.

4. **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)**

with q1 as (
select S.statename, sum(actsales) as SumSales, sum(actprofit) as sumProfit,
ROW_NUMBER() OVER (ORDER BY sum(F.actsales) DESC) as prodranksales,
ROW_NUMBER() OVER (ORDER BY sum(F.actprofit) DESC) as prodrankprofit
from states S, factcoffee F, areacode A
where A.areaid = F.areaid
*   and S.stateid = A.stateid*
group by S.statename),

q2 as (
*select statename, sumsales, round(100*ratio_to_report(sum(sumsales)) over (), 2) as PercTotalSales*
from q1
group by statename, sumsales),

q3 as (
*select statename, sumprofit, round(100*ratio_to_report(sum(sumprofit)) over (), 2) as PercTotalProfit*
from q1

group by statename, sumprofit),

q4 as (select q2.statename, sumsales, sumprofit, PercTotalSales, PercTotalProfit
from q3, q2
where q2.statename = q3.statename
order by perctotalsales desc),

q5 as (select statename, sumsales, sumprofit, PercTotalSales, PercTotalProfit,
sum(PercTotalSales) over (order by rownum rows unbounded preceding) as cumsalesP,
sum(PercTotalProfit) over (order by rownum rows unbounded preceding) as cumprofitP
from q4)

select *
from q5
where cumsalesP <=55;

	STATENAME	SUMSALES	SUMPROFIT	PERCTOTALSALES	PERCTOTALPROFIT	CUMSALESP	CUMPROFITP
1	California	96892	31785	11.82	12.25	11.82	12.25
2	New York	70852	20096	8.64	7.74	20.46	19.99
3	Illinois	69883	30821	8.52	11.88	28.98	31.87
4	Nevada	60159	10616	7.34	4.09	36.32	35.96
5	Iowa	54750	22212	6.68	8.56	43	44.52
6	Colorado	48179	17743	5.88	6.84	48.88	51.36
7	Oregon	40899	12439	4.99	4.79	53.87	56.15

The above table shows the percent of total sales that each state contributes to the whole. The top 6 make up about 48% of sales, but the 7th state, Oregon, pushes it past the 50th percentile to nearly 54%. These top 7 states make up about 56% of the company's profits, which means the actual sales is spread out of a bit more between states than profits. Though there is clear relationship between sales and profit, the states with the most profits are not quite the states the most sales.

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

with q1 as (
select P.prodname, sum(actsales) as SumSales, sum(actprofit) as sumProfit,
ROW_NUMBER() OVER (ORDER BY sum(F.actsales) DESC) as prodranksales,
ROW_NUMBER() OVER (ORDER BY sum(F.actprofit) DESC) as prodrankprofit
from prodcoffee P, factcoffee F
where P.productid = F.productid
group by P.prodname),

q2 as (
select prodname, sumsales, round(100*ratio_to_report(sum(sumsales)) over (), 2) as PercTotalSales
from q1
group by prodname, sumsales),

q3 as (
select prodname, sumprofit, round(100*ratio_to_report(sum(sumprofit)) over (), 2) as PercTotalProfit

from q1
group by prodname, sumprofit
)

select q2.prodname, sumsales, sumprofit, PercTotalSales, PercTotalProfit
from q3, q2
where q2.prodname = q3.prodname
order by perctotalprofit asc;

PRODNAME	SUMSALES	SUMPROFIT	PERCTOTALSALES	PERCTOTALPROFIT
1 Green Tea	32850	-231	4.01	-0.09
2 Amaretto	26269	4890	3.2	1.88
3 Mint	35710	6154	4.36	2.37
4 Regular Espresso	24031	10065	2.93	3.88
5 Caffe Latte	35899	11375	4.38	4.38

Above, we looked to the products that are the least profitable overall. Green tea and Amaretto are the worse in this regard and are top contenders for being discontinued. However, we should also look to see how these products are doing over time…

with q1 as (
select P.prodname, Extract(Year from factdate) as Years, sum(F.actprofit) as TotProdprofit
from factcoffee F, prodcoffee P
where p.productid = f.productid
group by P.prodname, Extract(Year from factdate)),

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

select prodname, Fyear, SYear, round(100(Syear-Fyear)/abs(Fyear),2) as PerInc, (Syear-Fyear) as TotInc*
from q2
order by totinc asc;

PRODNAME	FYEAR	SYEAR	PERINC	TOTINC
1 Green Tea	-90	-141	-56.67	-51
2 Amaretto	1983	2907	46.6	924
3 Mint	2511	3643	45.08	1132
4 Regular Espresso	4092	5973	45.97	1881
5 Caffe Latte	4636	6739	45.36	2103

The above table shows that not only is green tea not profitable overall, losses actually increase from 2012 to 2013. Assuming that there can only be a single product that is discontinued country-wide, I recommend discontinuing the

sale of green tea throughout the company. This could be the first step of a death spiral in the company, so they need to be quite careful of that.

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

```
with q1 as (
select P.prodname, to_char(factdate, 'mm') as Month, sum(F.actsales) as TotProdsales12
from factcoffee F, prodcoffee P
where p.productid = f.productid
   and extract(Year from factdate) = 2012
group by P.prodname, to_char(factdate, 'mm'))

select prodname, nvl(Jan12,0) as Jan12, nvl(Feb12, 0) as Feb12,
nvl(Mar12,0) as Mar12, nvl(Apr12, 0) as Apr12,
nvl(May12,0) as May12, nvl(Jun12, 0) as Jun12,
nvl(Jul12,0) as Jul12, nvl(Aug12, 0) as Aug12,
nvl(Sep12,0) as Sep12, nvl(Oct12, 0) as Oct12,
nvl(Nov12,0) as Nov12, nvl(Dec12, 0) as Dec12

FROM Q1
PIVOT (
   Sum(TotProdsales12)
   FOR Month IN ('01' as Jan12, '02' as Feb12, '03' as Mar12,
   '04' as Apr12, '05' as May12, '06' as Jun12, '07' as Jul12,
   '08' as Aug12, '09' as Sep12, '10' as Oct12, '11' as Nov12, '12' as Dec12));
```

	PRODNAME	JAN12	FEB12	MAR12	APR12	MAY12	JUN12	JUL12	AUG12	SEP12	OCT12	NOV12	DEC12
1	Colombian	4860	4821	4904	5048	5252	5748	5959	6014	5325	4902	4817	5174
2	Earl Grey	2562	2596	2613	2691	2798	2843	2877	2860	2820	2662	2562	2786
3	Mint	1397	1417	1434	1493	1533	1612	1685	1526	1345	1365	1302	1371
4	Caffe Mocha	3220	3348	3326	3322	3351	3519	3605	3864	3470	3564	3463	3485
5	Decaf Espresso	3122	3161	3203	3184	3251	3290	3362	3364	3159	3114	2985	3045
6	Chamomile	2960	3016	2993	2996	2992	3117	3188	3326	3039	3188	3081	3073
7	Lemon	3517	3613	3650	3878	3992	4123	4445	4311	3785	3723	3774	4145
8	Amaretto	1082	1073	1032	1049	1065	1068	1025	1049	1115	1096	1092	1095
9	Decaf Irish Cream	2372	2433	2471	2588	2628	2741	2929	2724	2404	2365	2311	2503
10	Darjeeling	2839	2879	2896	2891	3011	3159	3206	3311	3010	2992	2745	2860
11	Green Tea	1250	1284	1301	1340	1350	1353	1384	1370	1308	1364	1377	1384
12	Caffe Latte	1478	1463	1499	1509	1552	1501	1570	1411	1381	1371	1380	1444
13	Regular Espresso	896	988	923	954	917	1051	926	899	931	1143	1114	1008

2012 Sales Volume
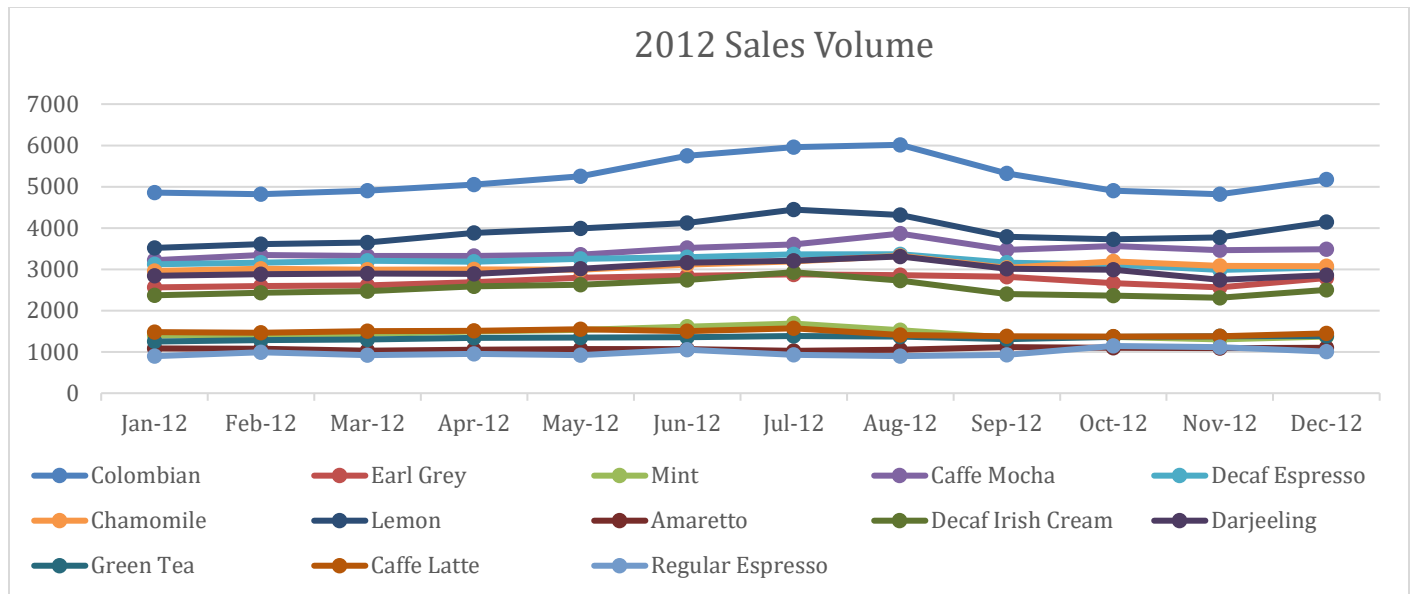
By changing the values to 2013, instead of 2012, we get the following:

	PRODNAME	JAN13	FEB13	MAR13	APR13	MAY13	JUN13	JUL13	AUG13	SEP13	OCT13	NOV13	DEC13
1	Colombian	5439	5135	5225	5381	5203	5696	5904	5961	5675	5222	5132	5514
2	Earl Grey	2867	2765	2785	2868	2774	2818	2851	2833	3005	2835	2730	2971
3	Mint	1564	1510	1529	1591	1519	1597	1669	1513	1434	1454	1388	1462
4	Caffe Mocha	3604	3566	3544	3541	3322	3488	3573	3828	3697	3796	3692	3716
5	Decaf Espresso	3493	3368	3412	3394	3222	3260	3331	3332	3366	3316	3181	3247
6	Chamomile	3313	3213	3188	3193	2967	3089	3160	3295	3238	3395	3282	3276
7	Lemon	3937	3851	3889	4134	3954	4085	4403	4275	4036	3967	4021	4418
8	Amaretto	1210	1144	1100	1117	1057	1059	1016	1039	1188	1167	1164	1167
9	Decaf Irish Cream	2656	2593	2633	2759	2605	2715	2901	2702	2563	2520	2463	2669
10	Darjeeling	3178	3067	3084	3082	2985	3130	3176	3281	3208	3185	2926	3050
11	Green Tea	1399	1369	1386	1428	1339	1342	1373	1360	1394	1452	1467	1476
12	Caffe Latte	1653	1559	1597	1608	1539	1487	1556	1398	1473	1460	1470	1540
13	Regular Espresso	1003	1052	983	1016	908	1041	917	890	992	1218	1187	1074

2013 Sales Volume
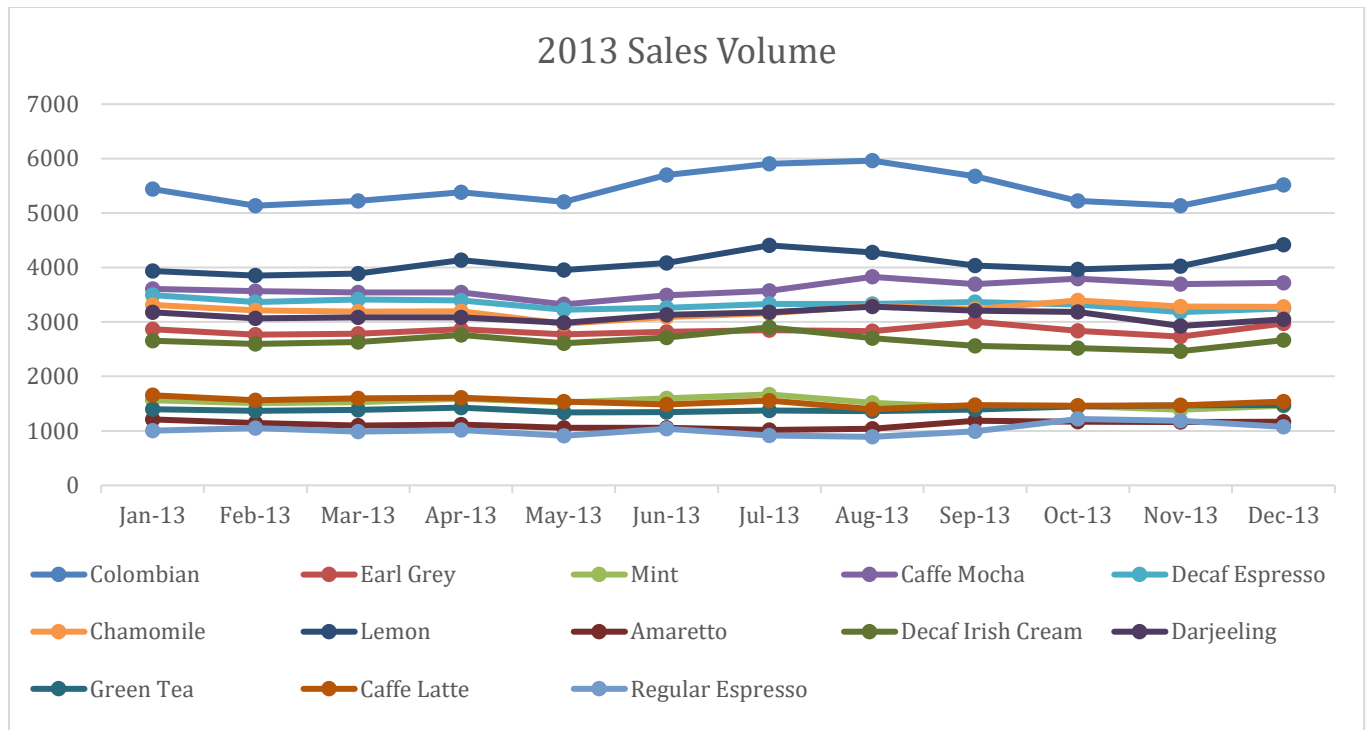
Overall there are two key trends among all the products. First, the winter months (January, December) correspond to higher sales than other times of the year. This could be related to the weather, as more people drink heated beverages to stay warm. Additionally, the summer months of July and August also see an increase in sales. This could be a result of classes starting again for students, or maybe it corresponds with growth/production seasons. Regardless of the potential cause, sales are typically higher for a given product in the winter and summer. The trends seem very similar for each year, which we take into account below.
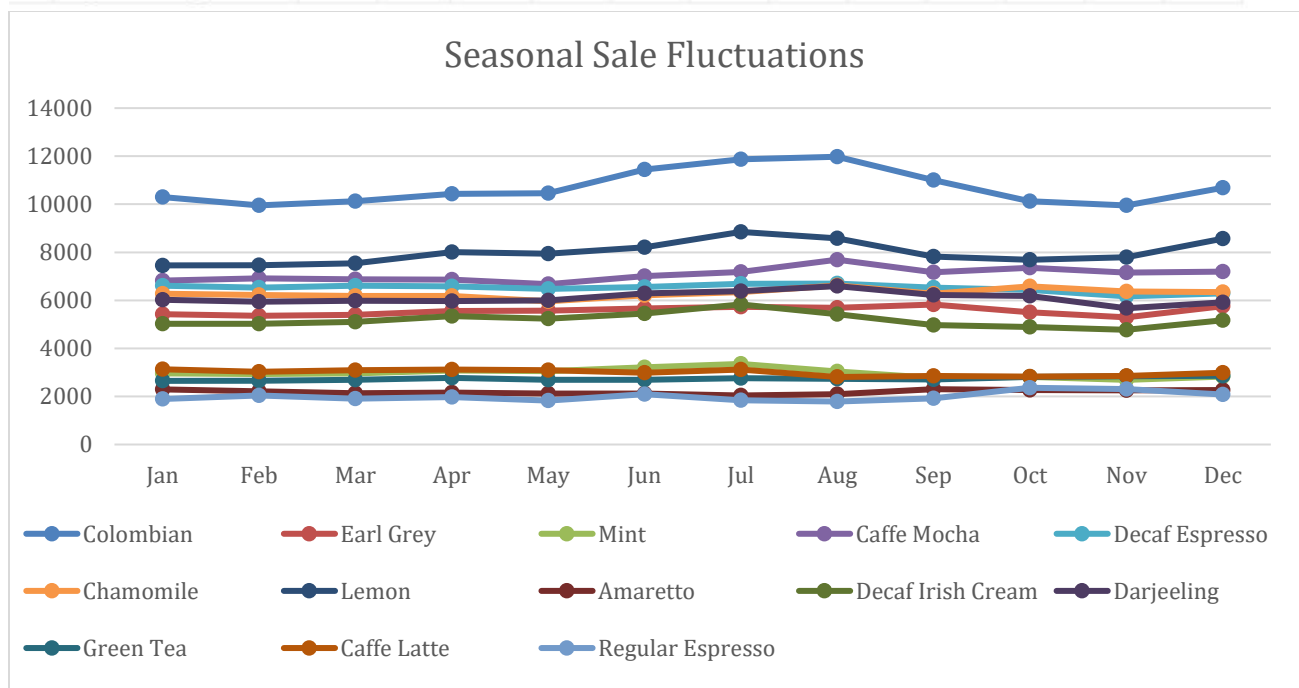
i. **Are there trends for any particular product?**

```
with q1 as (
select P.prodname, to_char(factdate, 'mm') as Month, sum(F.actsales) as TotProdsales
from factcoffee F, prodcoffee P
where p.productid = f.productid
group by P.prodname, to_char(factdate, 'mm'))

select prodname, nvl(Jan,0) as Jan, nvl(Feb, 0) as Feb,
nvl(Mar,0) as Mar, nvl(Apr, 0) as Apr,
nvl(May,0) as May, nvl(Jun, 0) as Jun,
nvl(Jul,0) as Jul, nvl(Aug, 0) as Aug,
nvl(Sep,0) as Sep, nvl(Oct, 0) as Oct,
nvl(Nov,0) as Nov, nvl(Dec, 0) as Dec

FROM Q1
PIVOT (
 Sum(TotProdsales)
 FOR Month IN ('01' as Jan, '02' as Feb, '03' as Mar, '04' as Apr, '05' as May, '06' as Jun, '07' as Jul, '08' as Aug, '09'
as Sep, '10' as Oct, '11' as Nov, '12' as Dec));
```

PRODNAME	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1 Colombian	10299	9956	10129	10429	10455	11444	11863	11975	11000	10124	9949	10688
2 Earl Grey	5429	5361	5398	5559	5572	5661	5728	5693	5825	5497	5292	5757
3 Mint	2961	2927	2963	3084	3052	3209	3354	3039	2779	2819	2690	2833
4 Caffe Mocha	6824	6914	6870	6863	6673	7007	7178	7692	7167	7360	7155	7201
5 Decaf Espresso	6615	6529	6615	6578	6473	6550	6693	6696	6525	6430	6166	6292
6 Chamomile	6273	6229	6181	6189	5959	6206	6348	6621	6277	6583	6363	6349
7 Lemon	7454	7464	7539	8012	7946	8208	8848	8586	7821	7690	7795	8563
8 Amaretto	2292	2217	2132	2166	2122	2127	2041	2088	2303	2263	2256	2262
9 Decaf Irish Cream	5028	5026	5104	5347	5233	5456	5830	5426	4967	4885	4774	5172
10 Darjeeling	6017	5946	5980	5973	5996	6289	6382	6592	6218	6177	5671	5910
11 Green Tea	2649	2653	2687	2768	2689	2695	2757	2730	2702	2816	2844	2860
12 Caffe Latte	3131	3022	3096	3117	3091	2988	3126	2809	2854	2831	2850	2984
13 Regular Espresso	1899	2040	1906	1970	1825	2092	1843	1789	1923	2361	2301	2082

Seasonal Sale Fluctuations

To consider this, I took the sum of each month's sales in 2012 and 2013 to make the magnitude of seasonal change is more apparent. Caffe Mocha's peak month is August. Colombian fluctuated greatly, even though it is the product with the most overall sales. It's peak months were August and July, while February and November had the lowest monthly sales. Mint and Decaf Irish Cream

ii. *Are there trends in any particular state for any product?*
   New York:
   with q1 as (
   select P.prodname, to_char(factdate, 'mm') as Month, sum(F.actsales) as TotProdsales
   from factcoffee F, prodcoffee P, areacode A
   where p.productid = f.productid
   *   and A.areaid = F.areaid*
   *   and A.stateid = 1013*

11

group by P.prodname, to_char(factdate, 'mm'))

select prodname, nvl(Jan,0) as Jan, nvl(Feb, 0) as Feb,
nvl(Mar,0) as Mar, nvl(Apr, 0) as Apr,
nvl(May,0) as May, nvl(Jun, 0) as Jun,
nvl(Jul,0) as Jul, nvl(Aug, 0) as Aug,
nvl(Sep,0) as Sep, nvl(Oct, 0) as Oct,
*nvl(Nov,0) as Nov, nvl(Dec, 0) as Dec*

FROM Q1
PIVOT (
 Sum(TotProdsales)
 FOR Month IN ('01' as Jan, '02' as Feb, '03' as Mar, '04' as Apr, '05' as May, '06' as Jun, '07' as Jul, '08' as Aug, '09'
*as Sep, '10' as Oct, '11' as Nov, '12' as Dec));*

	PRODNAME	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1	Colombian	1437	1332	1394	1471	1505	1772	1816	1812	1632	1343	1287	1444
2	Earl Grey	464	502	440	461	510	550	510	599	624	624	599	624
3	Mint	381	372	376	390	394	394	394	247	231	339	331	306
4	Caffe Mocha	129	126	130	136	137	143	153	155	140	143	126	136
5	Lemon	1023	1023	1060	1318	1296	1332	1579	1300	1118	1060	1190	1543
6	Darjeeling	496	479	483	506	510	599	621	643	506	547	409	434
7	Green Tea	284	390	409	409	418	458	458	488	411	386	254	275
8	Regular Espresso	1038	1198	1080	1165	1081	1344	1087	1081	1171	1580	1394	1357

Monthly Sales Fluctuations in New York
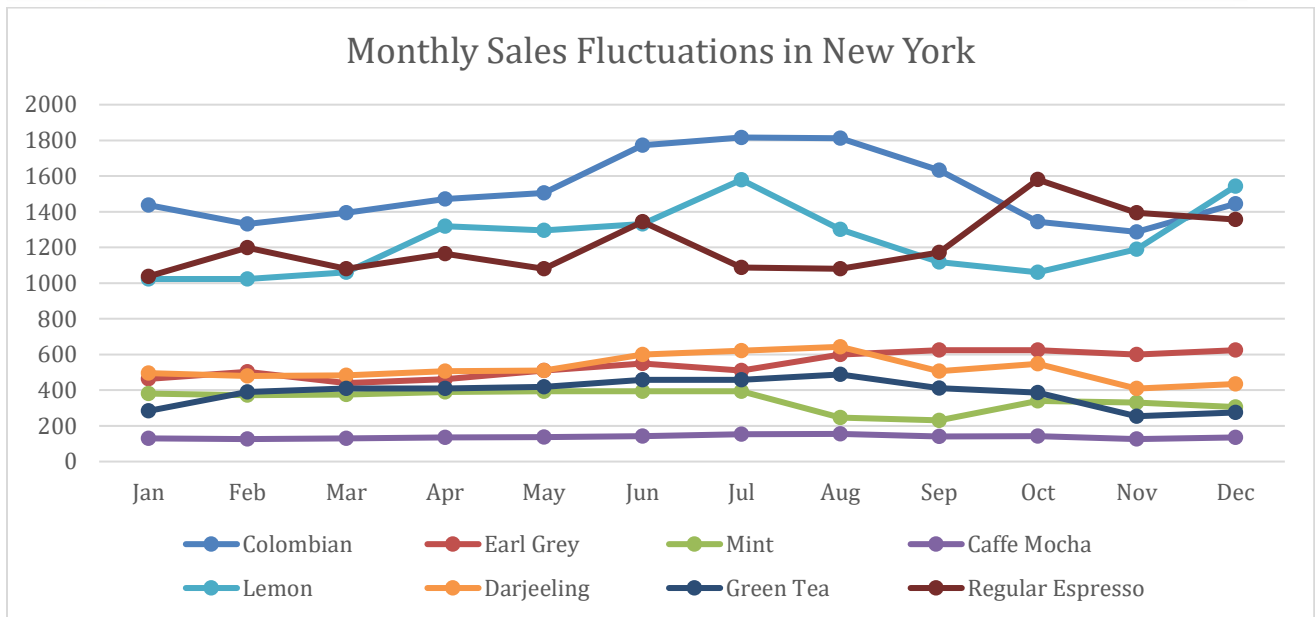
Oklahoma:
By changing the stateID from 1013 to 1015, we get the data for a very different state.

	PRODNAME	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1	Colombian	212	207	202	219	197	197	187	309	289	423	413	382
2	Caffe Mocha	170	116	109	109	98	88	88	82	103	109	165	151
3	Decaf Espresso	445	413	434	459	468	554	569	569	514	423	407	457
4	Chamomile	254	244	248	262	259	265	265	279	308	260	250	300
5	Lemon	403	454	469	469	440	526	591	631	461	475	463	516
6	Decaf Irish Cream	170	165	163	155	141	119	113	110	149	136	182	169
7	Caffe Latte	424	426	442	552	544	561	669	552	475	450	506	661

Monthly Sales Fluctuations in Oklahoma
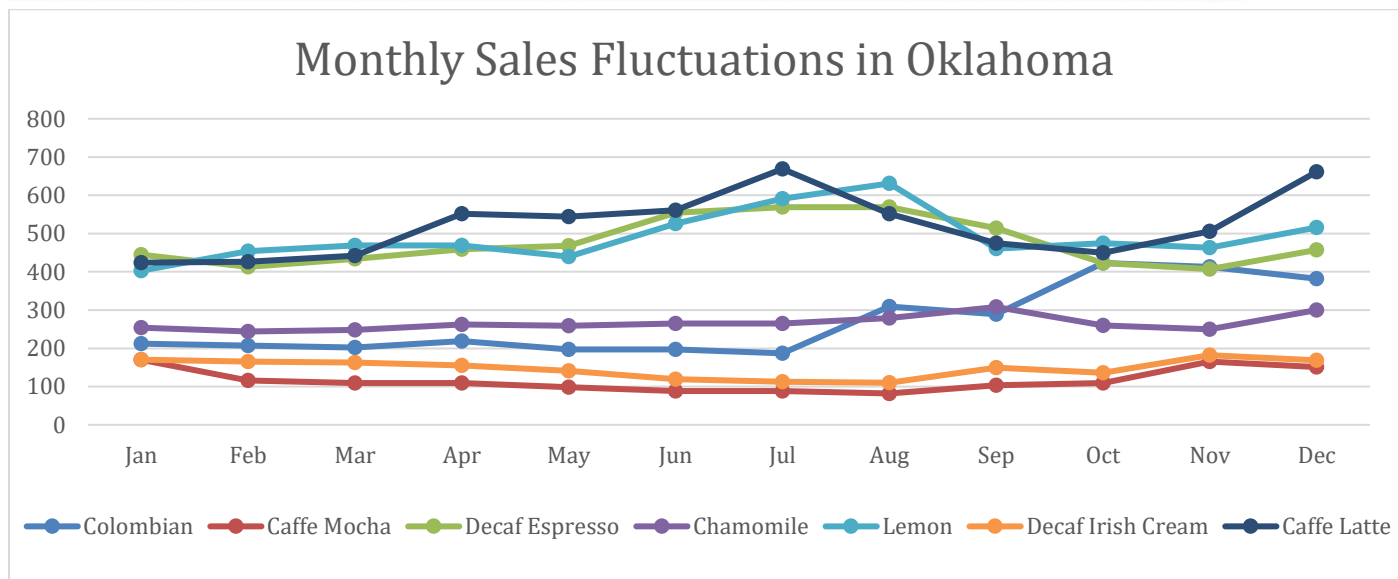
Neither state sold every product. It would be interesting to see why that is the case. Maybe the stores don't have the capacity to stock other items, maybe they were not worth stocking, or maybe it was some other reason altogether. The total volume of sales was much higher in New York than in Oklahoma, not surprising considering the population size difference.

Café Latte in July was the most popular product for any month in Oklahoma, followed by the same product in Dec. This is which is interesting because it isn't a stand out product in the countrywide plots, and it is not even sold in New York. It highlights that regional preferences can vary greatly.

Espresso, Lemon Tea, and Colombian Coffee were notably more popular in New York than the other products that were sold there. Through the trends of these three varied, Colombian was consistently a favorite, though regular espresso was substantially more popular in October.

New York and Oklahoma were arbitrarily chosen, more accuracy analysis could be done by comparing two states that sell the same set of products and view fluctuations as percentages to better understand the proportional impact of seasonal change.

7. *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.*

ALTER TABLE Factcoffee
ADD Quarter Char(2);

UPDATE Factcoffee
  SET Quarter = 'Q3'
  WHERE (to_char(factdate, 'Mon') = 'Jul' or to_char(factdate, 'Mon') = 'Aug' or to_char(factdate, 'Mon') = 'Sep');

And then I repeated the above for Q1, Q2, and Q4 with the appropriate months.

i. **Now find the total sales for years 2012 and 2013 for each quarter. Display quarter in columns.**
with q1 as(
SELECT extract(year from factdate) as Years, F.quarter, SUM(actsales) as sumsales
FROM factcoffee F
GROUP by GROUPING SETS (
   (extract(year from factdate), F.quarter),
   extract(year from factdate)))

select Years, nvl(Q1,0) as Q1, nvl(Q2,0) as Q2, nvl(Q3,0) as Q3, nvl(Q4,0) as Q4
from q1
pivot(
   sum(sumsales)
   for quarter in ('Q1' as Q1, 'Q2' as Q2, 'Q3' as Q3, 'Q4' as Q4))
order by years asc;

	YEARS	Q1	Q2	Q3	Q4
1	2012	95892	101760	105282	98225
2	2013	103863	103313	106806	104670

2013 was notably more successful in sales volume than 2012 in Q1 and Q4, but Q2 and Q3 were fairly close to each other.

ii. **Which quarter has the greatest sales and profits?**

SELECT F.quarter as Q2012, SUM(actsales) as sumsales, sum(actprofit) as sumprofit,
ROW_NUMBER() OVER (ORDER BY sum(F.actsales) DESC) as SalesRank,
ROW_NUMBER() OVER (ORDER BY sum(F.actprofit) DESC) as ProfitRank
FROM factcoffee F
where extract(year from factdate) = 2012
group by F.quarter;

	Q2012	SUMSALES	SUMPROFIT	SALESRANK	PROFITRANK
1	Q3	105282	27979	1	1
2	Q2	101760	27188	2	2
3	Q4	98225	25884	3	3
4	Q1	95892	24775	4	4

SELECT F.quarter as Q2013, SUM(actsales) as sumsales, sum(actprofit) as sumprofit,
ROW_NUMBER() OVER (ORDER BY sum(F.actsales) DESC) as SalesRank,
ROW_NUMBER() OVER (ORDER BY sum(F.actprofit) DESC) as ProfitRank
FROM factcoffee F
where extract(year from factdate) = 2013
group by F.quarter;

	Q2013	SUMSALES	SUMPROFIT	SALESRANK	PROFITRANK
1	Q3	106806	39503	1	1
2	Q2	103313	38429	4	2
3	Q4	104670	38427	2	3
4	Q1	103863	37358	3	4

In both years, Q3 provided both the greatest sales and the greatest profit. If you compare Q3 in 2012 and 2013, then you can see that sales were slightly higher in 2013 over 2013, and profit was substantially higher, increasing by over $10,000.

8. ***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.***

create table Question8 as (
select S.statename, F.productid, F.quarter, sum(f.actsales) as sumsales,
sum(f.actprofit) as sumprofit, round((sum(actprofit)/sum(actsales)),2) as Margin, sum(actmarkcost) as summark,
RANK() OVER (PARTITION BY F.quarter
   ORDER BY (sum(f.actsales)) DESC) as prodrank
from factcoffee F, areacode A, states S
where F.areaid = A.areaid
   and A.stateid = S.stateid
group by statename, productid, quarter);

	STATENAME	PRODUCTID	QUARTER	SUMSALES	SUMPROFIT	MARGIN	SUMMARK	PRODRANK
1	California	2	Q1	4163	1922	0.46	556	1
2	New York	2	Q1	4163	1922	0.46	556	1
3	Illinois	5	Q1	3454	1542	0.45	468	3
4	Nevada	11	Q1	3454	1548	0.45	468	3
5	Iowa	8	Q1	3454	1543	0.45	468	3
6	California	6	Q1	3454	1545	0.45	468	3
7	Iowa	12	Q1	3388	1303	0.38	554	7
8	New York	7	Q1	3316	1586	0.48	384	8
9	New York	9	Q1	3106	959	0.31	554	9
10	Massachusetts	2	Q1	3034	2693	0.89	140	10

I ranked the products by sale volume for each quarter. I could have done it further by considering each product for each quarter or each state for each quarter, but I felt this was a better way of comparing on a national level. There is significance in knowing that in Q1, product 2 in both new York and California were exceptionally successful. Additionally, rank() instead of dense_rank() is useful for understanding how many places are superior to that row.

PART B: Office Product

1. *Rank managers based on the sales generated.*

   select M.regmanager, sum(D.ordsales) as Sumsales,
   row_number() over (order by sum(D.ordsales) desc) as salesrank
   from managers M, customers C, orderdet D
   where C.custreg = M.regid
   and D.custid = C.custid
   group by M.regmanager;

	REGMANAGER	SUMSALES	SALESRANK
1	Chris	2493762.8	1
2	Erin	2366343.86	2
3	William	2360868.33	3
4	Sam	1568582.35	4

2. *Find the products that had the worst average shipping times.*

   For this problem, I am considering shipping time to be the difference between when the order was made and when the order was shipped.

   select d.prodid, round(avg(D.ORDSHIPDATE - D.ORDDATE),2) as shiptime, count(d.prodid) numshipments
   from orderdet D
   group by d.prodid
   order by shiptime desc;

	PRODID	SHIPTIME	NUMSHIPMENTS
1	4	19	5
2	933	9.91	11
3	28	9.57	7
4	62	9.22	9
5	1095	7	3
6	859	7	1
7	1272	7	2
8	181	7	4
9	348	7	2
10	1290	7	1

   It is uncommon for an item with a high ship time to be ordered more than a dozen times. The first to do this is the 17th item, with 13 shipments.

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

with q1 as (
select C.custname, sum(ordqty) as Ordqty, sum(ordsales) as sumsales
from customers C, orderdet D
where C.custid = D.custid
group by C.custname
order by sumsales desc),

q2 as (
select custname, ordqty, sumsales, round(100*ratio_to_report(sum(sumsales)) over (), 2) as PercTotalSales
from q1
group by custname, ordqty, sumsales
order by sumsales desc)

select custname, ordqty, sumsales, PercTotalSales, sum(PercTotalSales) over (order by ROWNUM rows unbounded preceding) as cumsales
from q2
fetch FIRST 10 PERCENT ROWS ONLY;

	CUSTNAME	ORDQTY	SUMSALES	PERCTOTALSALES	CUMSALES
1	Gordon Brandt	748	123745.62	1.41	1.41
2	Glen Caldwell	986	89269.7	1.02	2.43
3	Rosemary O'Brien	693	86540.75	0.98	3.41
4	Leigh Burnette Hurley	992	83651.7	0.95	4.36
5	Kristine Connolly	616	81296.39	0.92	5.28
269	Ricky Dougherty	74	7181.43	0.08	55.62
270	Lloyd Fuller	76	7179.92	0.08	55.7

The first 10% of customers, which I am assuming is based on the sum of their sales, make up over 55% of total sales! I was initially sorting by order quantity and taking the top 10% of items ordered, and then I realized that was the below question. However, if you do sort by order quantity, then 44% of all total sales come from them. This is still quite high, it is just lower relative to the previous bit.

268	April Hu	80	2973.76	0.03	44.24
269	Carlos Hess	80	4924.94	0.06	44.3
270	Constance Robertson	80	1052.78	0.01	44.31

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

with q1 as (
select C.custname, sum(ordqty) as Ordqty, sum(ordsales) as sumsales,
row_number() over (order by sum(ordqty) desc) as qtyrank,
row_number() over (order by sum(ordsales) desc) as salesrank
from customers C, orderdet D
where C.custid = D.custid

group by C.custname
order by sumsales desc),

q2 as (
select custname, ordqty, sumsales, qtyrank, salesrank, round(100*ratio_to_report(sum(sumsales)) over (), 2) as
PercTotalSales, round(100*ratio_to_report(sum(ordqty)) over (), 2) as PercTotalQty
from q1
group by custname, ordqty, sumsales, qtyrank, salesrank
order by sumsales desc),

q3 as (select custname, ordqty, sumsales, PercTotalQty, PercTotalSales, qtyrank, salesrank
from q2
fetch first 10 percent rows only)

select round(avg(qtyrank),2), avg(salesrank)
from q3;

	ROUND(AVG(QTYRANK),2)	AVG(SALESRANK)
1	436.98	135.5

	CUSTNAME	ORDQTY	SUMSALES	PERCTOTALQTY	PERCTOTALSALES	QTYRANK	SALESRANK
1	Gordon Brandt	748	123745.62	0.58	1.41	12	1
2	Glen Caldwell	986	89269.7	0.76	1.02	2	2
3	Rosemary O'Brien	693	86540.75	0.54	0.98	19	3
4	Leigh Burnette Hurley	992	83651.7	0.77	0.95	1	4
5	Kristine Connolly	616	81296.39	0.48	0.92	21	5
6	Nina Horne Kelly	223	78243.6	0.17	0.89	77	6
7	Neal Wolfe	864	69118	0.67	0.79	8	7
8	Priscilla Kane	948	61610.6	0.73	0.7	3	8
9	Dana Teague	311	61298.98	0.24	0.7	62	9
10	Kim Weiss	341	58947.41	0.26	0.67	58	10

There is a relationship between quantity of items purchased and average sales rank, but it's not as strong as one might assume. Overall, one cannot assume quantity purchased based on their sales volume or rank. This comes from people ordering a lower volume of higher priced goods, and others ordering a high volume of low priced goods. We can see the top 10 ranked people for sales volume above, but their quantity purchased rank varies quite extensively.

However, if they are on either end of a spectrum (e.g. extremely low or high sales volume) then it is extremely unlikely that they would be on the complete opposite end of the other spectrum (high or low quantity ordered). For example, the person who orders the most items is not likely to have one of the lowest sales volume.

5. *For each city and product combination, list the total sales and rank order in each city by total sales.*
   select C.custstate, C.custcity, D.prodid, sum(ordsales) as sumsales,
   RANK() OVER (PARTITION BY D.prodid
   ORDER BY (sum(ordsales)) DESC) as prodrank
   from customers C, orderdet D
   where C.custid = D.custid
   group by C.custstate, C.custcity, D.prodid;

	CUSTSTATE	CUSTCITY	PRODID	SUMSALES	PRODRANK
1	California	Los Angeles	1	95.75	1
2	Nebraska	Grand Island	1	52.93	2
3	Texas	Pharr	1	39.64	3
4	Virginia	Tysons Corner	1	30.01	4
5	California	Saratoga	1	22.1	5
6	Colorado	Northglenn	1	18.88	6
7	Florida	Coral Springs	2	219.73	1
8	California	Los Angeles	2	179.33	2
9	Texas	Victoria	2	149.42	3
10	Massachusetts	Billerica	2	138.16	4

We need to consider each city, state, and product combination, because the same city name exists in multiple states. The ranks shown here are the ranks of each product in the various cities by sales. Since there are so many products, this is an excellent way of determining where products are most successful.

6. *Which are the top 5 customers for each of the years?*
   select extract(year from D.orddate) as years, C.custname, sum(ordsales) as sumsales,
   rank() over (partition by extract(year from D.orddate)
       order by sum(ordsales)desc) as custrank
   from customers C, orderdet D
   where C.custid = D.custid
   group by extract(Year from D.orddate), C.custname
   order by custrank
   fetch first 20 rows only;

	YEARS	CUSTNAME	SUMSALES	CUSTRANK
1	2010	Kristine Connolly	50475.31	1
2	2013	Dana Sharpe	41930.3	1
3	2012	Gordon Brandt	114184.81	1
4	2011	Leigh Burnette Hurley	64861.56	1
5	2010	Nina Horne Kelly	48295.12	2
6	2013	Frank Cross	38200.53	2
7	2012	Kim Weiss	53178.92	2
8	2011	Amanda Kay	55040.47	2
9	2010	Toni Swanson	32194.12	3
10	2013	Glen Caldwell	37425.62	3
11	2012	Ruth McConnell Young	39105.63	3
12	2011	Rosemary O'Brien	49226.5	3
13	2010	Rosemary O'Brien	29916.01	4
14	2013	Allison Kirby	30520.87	4
15	2012	Herbert Holden	32793.18	4
16	2011	Glen Caldwell	27450.43	4
17	2010	Yvonne Mann	28779.13	5
18	2013	Nina Horne Kelly	29659.55	5
19	2012	Caroline Morton	27233.06	5
20	2011	Bonnie Matthews Rowland	26633.09	5

This view prioritizes the customer rank, so it is clear to view who was the most important customer in each year. Another view could have been a pivot table, with rank on the side, years as the columns, and customer names as the data. However, then the sales volume wouldn't have a place, and I thought that was crucial information.

a. **Who are common customers across all years?**

with q1 as (
select extract(year from D.orddate) as years, C.custname, sum(ordsales) as sumsales,
rank() over (partition by extract(year from D.orddate)
*   order by sum(ordsales)desc) as prodrank*
from customers C, orderdet D
where C.custid = D.custid
group by extract(Year from D.orddate), C.custname
order by prodrank
fetch first 20 rows only),

q2 as (
select custname, nvl(Y2010,0) as Y2010, nvl(Y2011,0) as Y2011, nvl(Y2012,0) as Y2012, nvl(Y2013,0) as Y2013
from q1
pivot(
*   sum(sumsales)*
*   for years in ('2010' as Y2010, '2011' as Y2011, '2012' as Y2012, '2013' as Y2013))*
order by custname asc)

select custname, sum(Y2010) as Y2010, sum(Y2011) as Y2011, sum(Y2012) as Y2012, sum(Y2013) as Y2013
from q2
group by custname
order by custname;

	CUSTNAME	Y2010	Y2011	Y2012	Y2013
1	Allison Kirby	0	0	0	30520.87
2	Amanda Kay	0	55040.47	0	0
3	Bonnie Matthews Rowland	0	26633.09	0	0
4	Caroline Morton	0	0	27233.06	0
5	Dana Sharpe	0	0	0	41930.3
6	Frank Cross	0	0	0	38200.53
7	Glen Caldwell	0	27450.43	0	37425.62
8	Gordon Brandt	0	0	114184.81	0
9	Herbert Holden	0	0	32793.18	0
10	Kim Weiss	0	0	53178.92	0
11	Kristine Connolly	50475.31	0	0	0
12	Leigh Burnette Hurley	0	64861.56	0	0
13	Nina Horne Kelly	48295.12	0	0	29659.55
14	Rosemary O'Brien	29916.01	49226.5	0	0
15	Ruth McConnell Young	0	0	39105.63	0
16	Toni Swanson	32194.12	0	0	0
17	Yvonne Mann	28779.13	0	0	0

As shown in the table above, there are no customers in the top 5 for all years. I'm assuming the question is limiting the scope to the top 5 for each year.

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

   Yes, it is interesting that #8, Gordon Brandt bought a great deal in Y2012, substantially more than any single person in any given year, but he only purchased substantially that one time. Maybe he was testing out suppliers and decided not to go with this company anymore.

7. *Find the number of orders in each subcategory in states Michigan and Washington. List Washington and Michigan in different columns.*

   with q1 as(
   select P.prodsubcat, sum(d.ordqty) as Ordqty_M
   from orderdet D, products P, customers C
   where c.custid = d.custid
   *   and d.prodid = p.prodid*
   *   and C.CUSTSTATE = 'Michigan'*
   group by prodsubcat),

   q2 as (
   select P.prodsubcat, sum(d.ordqty) as Ordqty_W
   from orderdet D, products P, customers C
   where c.custid = d.custid
   *   and d.prodid = p.prodid*
   *   and C.CUSTSTATE = 'Washington'*
   group by prodsubcat)

   select q1.prodsubcat, Ordqty_M, Ordqty_W
   from q1
   full join q2 on q1.prodsubcat = q2.prodsubcat
   order by ordqty_M desc;

	PRODSUBCAT	ORDQTY_M	ORDQTY_W
1	Paper	807	834
2	Office Furnishings	767	403
3	Computer Peripherals	549	407
4	Telephones and Communication	523	1000
5	Binders and Binder Accessories	466	372
6	Storage & Organization	381	219
7	Appliances	264	214
8	Tables	259	263
9	Office Machines	259	182
10	Pens & Art Supplies	203	601
11	Chairs & Chairmats	174	394
12	Envelopes	135	268
13	Bookcases	98	113
14	Labels	74	296
15	Scissors, Rulers and Trimmers	63	41
16	Copiers and Fax	45	131
17	Rubber Bands	37	104

8. ***Find total orders in each quarter.***
   --to create a quarter column
   ALTER TABLE OrderDet
   ADD Quarter Char(2);

   UPDATE OrderDet
   * SET Quarter = 'Q2'*
   * WHERE (to_char(orddate, 'Mon') = 'Apr' or to_char(orddate, 'Mon') = 'May' or to_char(orddate, 'Mon') = 'Jun');*

   --query
   with q1 as (
   select quarter, sum(ordqty) as Y2010
   from orderdet D
   where extract(year from orddate) = 2010
   group by quarter),

   q2 as (
   select quarter, sum(ordqty) as Y2011
   from orderdet D
   where extract(year from orddate) = 2011
   group by quarter),

   q2 as (
   select quarter, sum(ordqty) as Y2011
   from orderdet D
   where extract(year from orddate) = 2011
   group by quarter),

q3 as (
select quarter, sum(ordqty) as Y2012
from orderdet D
where extract(year from orddate) = 2012
group by quarter),

q4 as (
select quarter, sum(ordqty) as Y2013
from orderdet D
where extract(year from orddate) = 2013
group by quarter)

select q1.quarter, Y2010, Y2011, Y2012, Y2013
from q1, q2, q3, q4
where q1.quarter=q2.quarter
  and q2.quarter = q3.quarter
  and q2.quarter = q4.quarter
order by quarter asc;

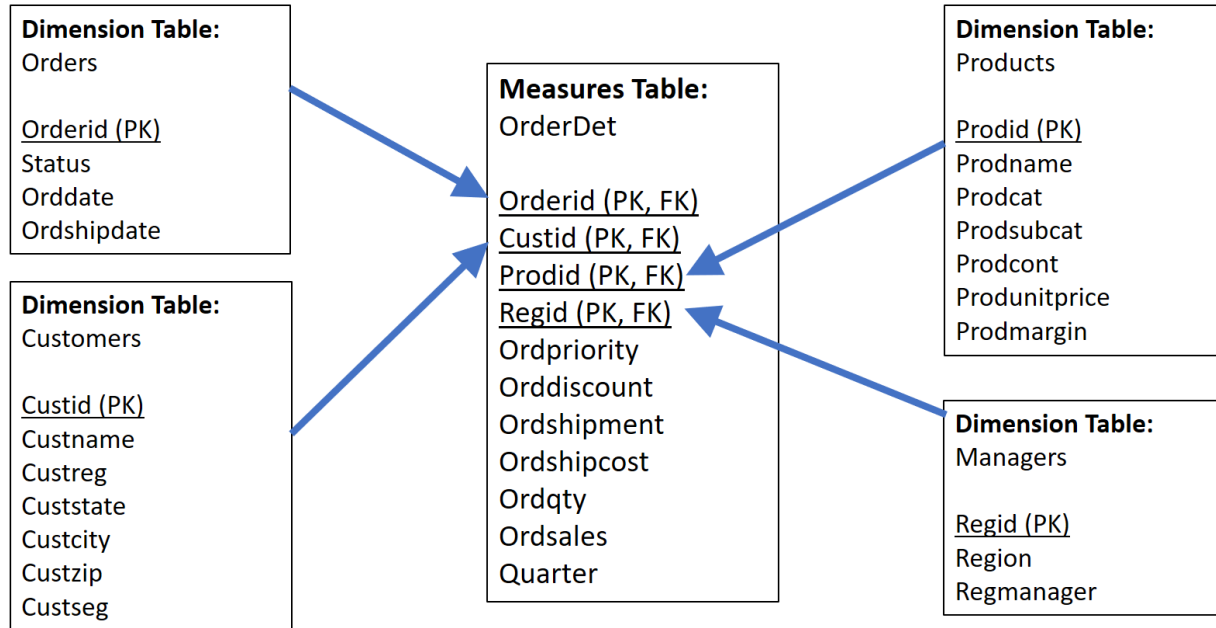|	QUARTER	Y2010	Y2011	Y2012	Y2013
1	Q1	4481	4500	4890	6488
2	Q2	5550	5926	7551	8324
3	Q3	6624	8254	7055	12708
4	Q4	8447	10886	12286	15026

9. **For each quarter and customer segment, find the total sales. Display data for quarters in column.**

with q1 as(
select D.quarter, C.custseg, sum(D.ordsales) as sumsales
from orderdet D, customers C
where C.custid = d.custid
group by D.quarter, C.custseg)

select custseg, nvl(Q1,0) as q1, nvl(Q2,0) as q2, nvl(q3,0) as q3, nvl(q4,0) as q4
from q1
PIVOT (
  sum(sumsales)
  for quarter in ('Q1' as q1, 'Q2' as q2, 'Q3' as q3, 'Q4' as q4))
order by q4 desc;

	CUSTSEG	Q1	Q2	Q3	Q4
1	Corporate	579294.28	512958.62	784153.53	1400335.59
2	Small Business	286493.19	323872.68	315099.62	718491.07
3	Home Office	330047.96	417618.47	632944.31	711499.27
4	Consumer	296515.61	343673.26	449666.56	686893.32

10. *If you are creating a data warehouse for Office Product, what would be your dimensions and measures? Develop star schema for the data warehouse.*

Dimension Table:
Orders

Orderid (PK)
Status
Orddate
Ordshipdate

Dimension Table:
Customers

Custid (PK)
Custname
Custreg
Custstate
Custcity
Custzip
Custseg

Measures Table:
OrderDet

Orderid (PK, FK)
Custid (PK, FK)
Prodid (PK, FK)
Regid (PK, FK)
Ordpriority
Orddiscount
Ordshipment
Ordshipcost
Ordqty
Ordsales
Quarter

Dimension Table:
Products

Prodid (PK)
Prodname
Prodcat
Prodsubcat
Prodcont
Produnitprice
Prodmargin

Dimension Table:
Managers

Regid (PK)
Region
Regmanager

There are Four dimensions and thus dimension tables, and they each are tied to the measures table via Foreign Key. Instead of adding the managers table to the customers table, which would have reduced dimensionality, or having the table in 2NF, which is slower and typically more common in snowflake schemas than star schemas, I felt that adding RegID as a primary key and foreign key to the Order Details measures table was less expensive and more accurate.

The measures left in the measures table are all features that can be aggregated for comparison and analysis, such as order discount, the shipping cost, the quarter the transaction occurred in, etc. The dimension tables can change without the measures table changing. E.g. a customer can be added to the customers dimension table without them being added to the measure table, until they make a purchase, which reduces the impact of single changes on the data warehouse, but maintains accuracy and simplicity.