DSO 552 Final Project Queries and Output

1. How big is the customer base of Parch and Posey (i.e. how many customers/accounts does the company have?) (1 point)
   select count(distinct(id)) from accounts;

	no_of_customers bigint
1	351

2. How many areas do they sell at? (1 point)
   select count(distinct(id)) from region;

	no_of_regions bigint
1	7

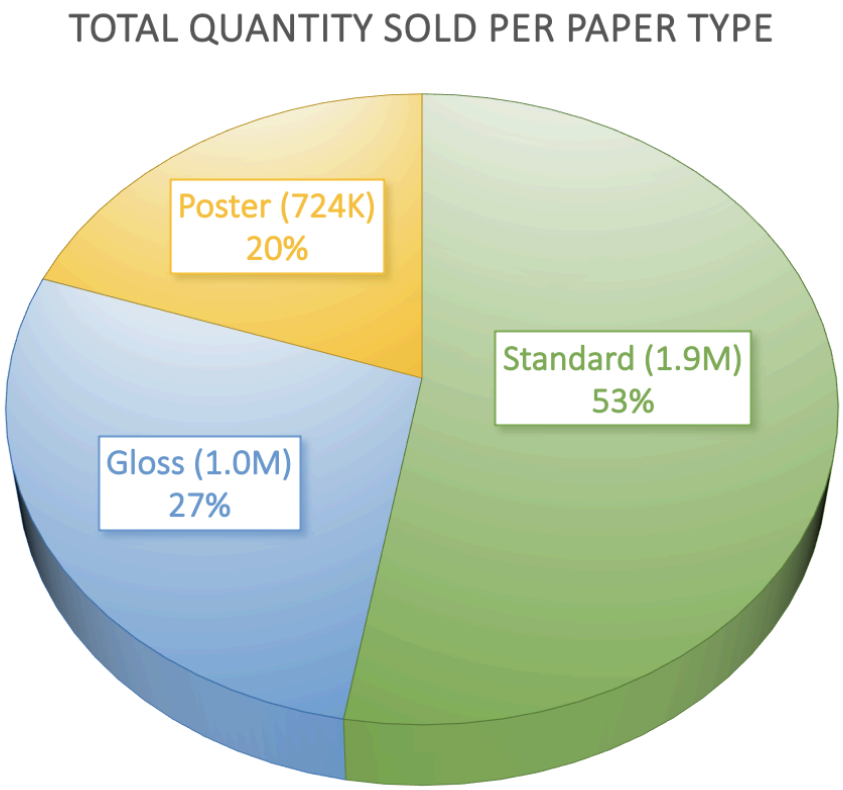
   select name from region order by name;

	name character
1	International
2	Midwest
3	North
4	Northeast
5	South
6	Southeast
7	West

3. Look into the revenue streams:
   a. How many types of paper do they sell and what percentage each one of them makes out of the total quantity sold? Provide a visualization that illustrates the results (e.g. pie chart, bar plot, or any chart of your choice)(1.5 point)

      select round(100*sum(standard_qty)/sum(total),2) as standard_qty_pct,
      round(100*sum(gloss_qty)/sum(total),2) as gloss_qty_pct,
      round(100*sum(poster_qty)/sum(total),2) as poster_qty_pct
      from orders;

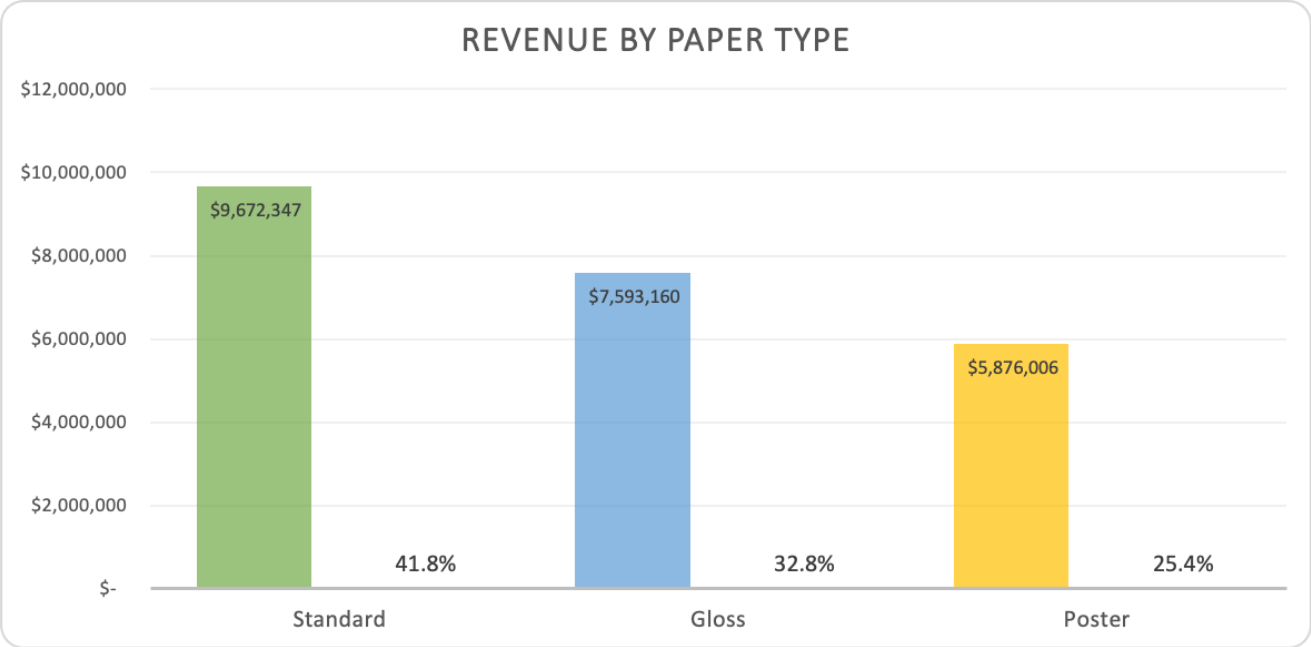
	standard_qty_pct numeric	gloss_qty_pct numeric	poster_qty_pct numeric
1	52.00	27.00	19.00



b. What percentage of revenues comes from which type of paper? Provide a visualization that illustrates the results (e.g. pie chart, bar plot, or any chart of your choice)

```
select round(100*sum(standard_amt_usd)/sum(total_amt_usd),2) as standard_rev_pct,
round(100*sum(gloss_amt_usd)/sum(total_amt_usd),2) as gloss_rev_pct,
round(100*sum(poster_amt_usd)/sum(total_amt_usd),2) as poster_rev_pct
from orders;
```

	standard_rev_pct numeric	gloss_rev_pct numeric	poster_rev_pct numeric
1	41.80	32.81	25.39



4. Is the business growing?
    a. How have revenues been year over year? For this, only take into account years with full data (2017 just started, so we don't know how yearly revenues will be and 2013 seems to have data only from December). Provide a visualization that illustrates the results (e.g. line chart, bar plot, or any chart of your choice). (1.5 point)

    ```
    select extract(year from occurred_at) as year, sum(total_amt_usd) total_revenue
    from orders
    where extract(year from occurred_at) in (select extract(year from occurred_at) as
    year from orders
            group by year
            having count(distinct(extract(month from occurred_at)))=12)
    group by year
    order by year;
    ```

	year numeric	total_revenue numeric
1	2014	4069106.54
2	2015	5752004.94
3	2016	12864917.92

    b. How have units sold evolved year over year? Here too, only take into account the past years' data. Provide a visualization that illustrates the results (e.g. line chart, bar plot, or any chart of your choice) (1.5 point)

```
select extract(year from occurred_at) as year, sum(standard_qty)
standard_qty_total, sum(gloss_qty) gloss_qty_total, sum(poster_qty)
poster_qty_total,sum(total) total_qty_sold
from orders
where extract(year from occurred_at) in (select extract(year from occurred_at) as
year from orders
        group by year
         having count(distinct(extract(month from occurred_at)))=12)
group by year
order by year;
```

	year numeric	standard_qty_total bigint	gloss_qty_total bigint	poster_qty_total bigint	total_qty_sold bigint
1	2014	353967	171829	125100	650896
2	2015	477789	263251	171932	912972
3	2016	1074751	553815	413034	2041600

5. How many sales reps do they have in each region?
   Sort the result by alphabetical order and include the regions that do not have any sales reps (1.5 point)

```
select r.name, count(distinct(sr.id)) from region r
left join sales_reps sr on r.id=sr.region_id
group by r.name
order by r.name;
```

	name character	count bigint
1	International	1
2	Midwest	9
3	North	0
4	Northeast	21
5	South	0
6	Southeast	10
7	West	10

6.

a. From Parch and Posey's leadership team you know that North, South and International are 3 newly added regions. If Dunder Mifflin decided to buy Parch and Posey, they would need to jump start sales in those areas. How would you suggest reallocating sales reps from the old to the new regions to cover the needs of the latter, i.e. which old regions would you recommend to pull sales reps from?

```
select r.name, count(distinct o.id) as Total_Orders,
count(distinct sr.id) as Total_Reps, count(distinct a.id) as Total_Accounts,
sum(o.total_amt_usd) as Total_Rev, avg(o.total_amt_usd) as Average_Rev
from region r
right join sales_reps sr on r.id=sr.region_id
right join accounts a on a.sales_rep_id=sr.id
full join orders o on o.account_id=a.id
where extract (year from o.occurred_at) = '2016'
group by r.name;
```

	name character	total_orders bigint	total_reps bigint	total_accounts bigint	total_rev numeric	average_rev numeric
1	Midwest	483	9	41	1711747.25	3543.9901656314699793
2	Northeast	1196	21	97	3999036.82	3343.6762709030100334
3	Southeast	1110	10	86	3545487.49	3194.1328738738738739
4	West	968	10	93	3608646.36	3727.9404545454545455

b. Based on the previous result, compute also by region: number of orders per representative in that region, number of accounts handled per representative, revenues per representative

```
select r.name, count(distinct o.id)/count(distinct sales_rep_id) as
orders_per_representative,
count (distinct a.id)/count(distinct sales_rep_id) as
accounts_per_representative,
```

```
sum(o.total_amt_usd)/count(distinct sales_rep_id) as
revenues_per_representative

from region r

right join sales_reps sr on r.id=sr.region_id

right join accounts a on a.sales_rep_id=sr.id

full join orders o on o.account_id=a.id

where extract(year from o.occurred_at) = '2016'

group by r.name

order by revenues_per_representative ASC;
```

	name character	orders_per_representative bigint	accounts_per_representative bigint	revenues_per_representative numeric
1	Midwest	53	4	190194.138888888889
2	Northeast	56	4	190430.324761904762
3	Southeast	111	8	354548.749000000000
4	West	96	9	360864.636000000000

c. Based on the calculation above, it is evident that the Midwest region is currently underperforming compared to other regions. To optimize overall sales efficiency and distribution, I recommend a strategic reallocation of sales representatives from the Midwest to other regions.

	name character	total_orders bigint	total_reps bigint	total_accounts bigint	total_rev numeric	average_rev numeric
1	Midwest	483	9	41	1711747.25	3543.9901656314699793
2	Northeast	1196	21	97	3999036.82	3343.6762709030100334
3	Southeast	1110	10	86	3545487.49	3194.1328738738738739
4	West	968	10	93	3608646.36	3727.9404545454545455

7. To answer if this is true, create a new column in your output that is:
‘group’ if the name of the account ends with the word ‘group’-‘not group’ otherwise
Then, based on the above result, compute the average (per account) revenues that came respectively from ‘group’ and from ‘not group’ accounts. (Hint: Here we would need 2 numbers, the average revenues for ‘group’ accounts and the average revenues for ‘not group’ accounts).

Finally, comment on the result and on whether your assumption was correct. (2 points)

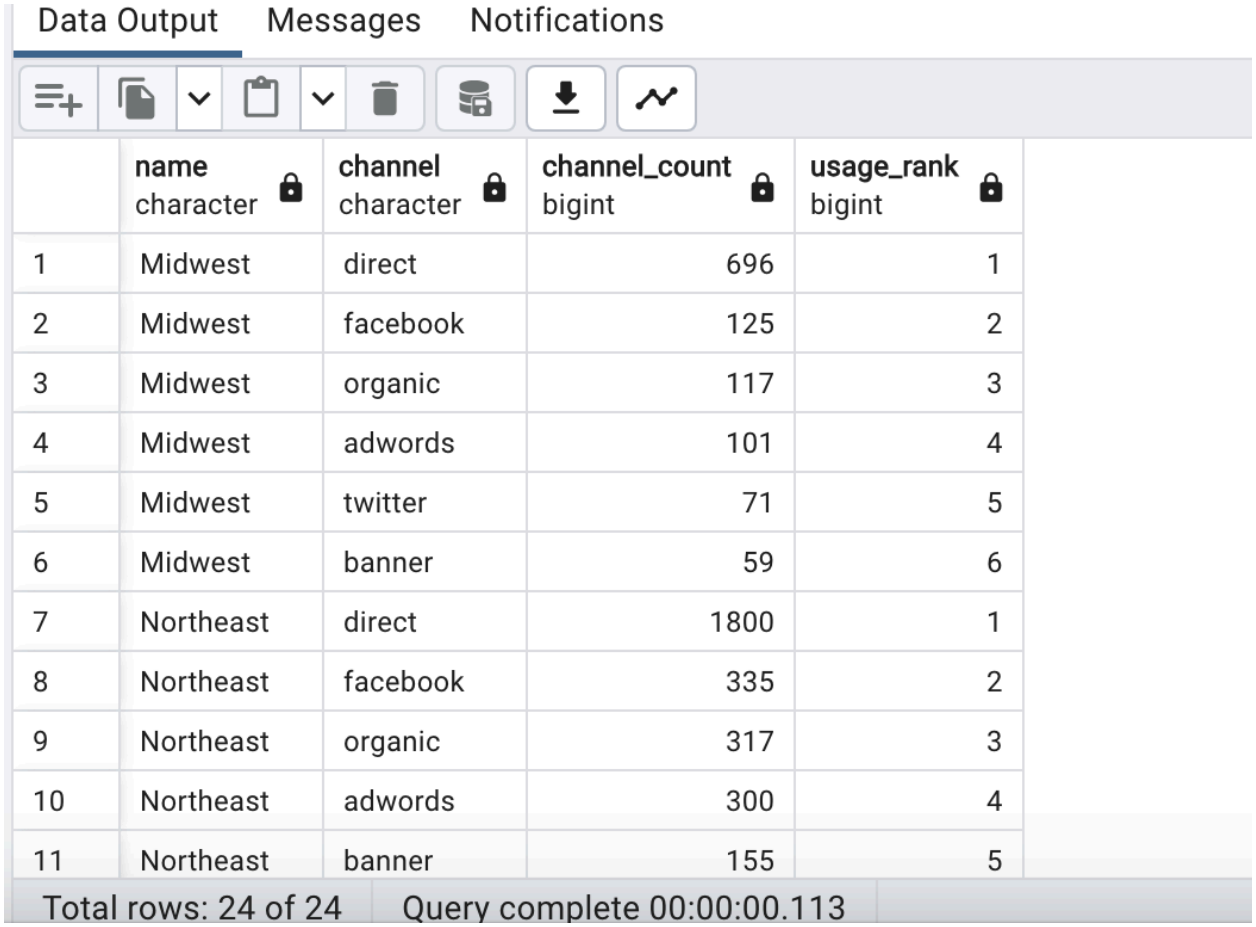
```
select gt.group_type, avg(total_revenue)
from (select o.account_id, sum(o.total_amt_usd) avg_total_revenue
        from orders o
        group by o.account_id) tr
left join (select a.id,
                case
                    when lower(right(a.name,5)) = 'group' then 'group'
                    else 'not_group'
                end group_type
                from accounts a) as gt on tr.account_id=gt.id
group by gt.group_type;
```

	group_type text	avg_total_revenue numeric
1	group	61831.742777777778
2	not_group	66351.025481927711

8. The Marketing team needs to focus on channels for the newly added sales regions, and because of its limited resources, it will have to deprioritize/deactivate temporarily some channels in the old areas. Specifically it decided to deactivate, for every old region, the channel that is used the least for web events in that region. Which channels should they deactivate in each region? Use a window function to give the answer here.

```
select gt.group_type, avg(total_revenue)
from (select o.account_id, sum(o.total_amt_usd) avg_total_revenue
        from orders o
        group by o.account_id) tr
left join (select a.id,
                case
                    when lower(right(a.name,5)) = 'group' then 'group'
                    else 'not_group'
                end group_type
                from accounts a) as gt on tr.account_id=gt.id
group by gt.group_type;
```

–They should deactivate banner in Midwest, Twitter in Northeast, Twitter in Southeast, and Banner in West.



Data Output | Messages | Notifications

	name character	channel character	channel_count bigint	usage_rank bigint
1	Midwest	direct	696	1
2	Midwest	facebook	125	2
3	Midwest	organic	117	3
4	Midwest	adwords	101	4
5	Midwest	twitter	71	5
6	Midwest	banner	59	6
7	Northeast	direct	1800	1
8	Northeast	facebook	335	2
9	Northeast	organic	317	3
10	Northeast	adwords	300	4
11	Northeast	banner	155	5

Total rows: 24 of 24 Query complete 00:00:00.113