

Week 5: Subqueries, Views and Data Cleaning in SQL

Parch and Posey Database

Subqueries and Views

1. What is the lifetime average amount spent in USD for the top 10 total spending accounts?

```
select a.name, sum(total_amt_usd) total_spent
from accounts a
join orders o
  on o.account_id = a.id
group by a.name
order by total_spent desc
limit 10
```

name	total_spent
EOG Resources	382873.3
Mosaic	345618.6
IBM	326819.5
General Dynamics	300694.8
Republic Services	293861.1
Leucadia National	291047.2
Arrow Electronics	281018.4
Sysco	278575.6
Supervalu	275288.3
Archer Daniels Midland	272672.8

```
select round(avg(total_spent),0)
from
  (select a.name, sum(total_amt_usd) total_spent
   from accounts a
   join orders o
     on o.account_id = a.id
   group by a.name
   order by total_spent desc
   limit 10
  ) tbl1
```

round

304847

2. For the customer/account that spent the most (in total over their lifetime as a customer) `total_amt_usd`, how many `web_events` did they have for each channel?

```
select name
from (
  select
```

```

    a.name,
    sum(total_amt_usd) total_spent
from accounts a
  join orders o
    on a.id = o.account_id
group by a.name
order by total_spent desc
limit 1
) tbl1

```

name
EOG Resources

```

select a.name, w.channel, count(*)
from accounts a
  join web_events w
    on a.id = w.account_id
where a.name = (
    select name
    from(
        select name ,sum(o.total_amt_usd) total_spent
        from orders o
        join accounts a
          on a.id = o.account_id
        join web_events w
          on w.account_id = a.id
        group by a.name
        order by total_spent desc
        limit 1
    ) tbl1
    )
group by a.name, w.channel
order by 3 desc

```

name	channel	count
EOG Resources	direct	44
EOG Resources	organic	13
EOG Resources	adwords	12
EOG Resources	facebook	11
EOG Resources	twitter	5
EOG Resources	banner	4

3. Which channel was the most frequently used by different accounts?

```

select tbl1.name, tbl1.channel, tbl2.max
from(
    select a.name as name, w.channel as channel, count(*) total
    from accounts a
    join web_events w
      on w.account_id = a.id
    group by a.name, w.channel
    order by a.name

```

```

    ) tbl1
join(
    select name, max(total)
    from tbl1
    group by name
    ) tbl2
on tbl1.name = tbl2.name and tbl1.total = tbl2.max
limit 5;

```

name	channel	max
3M	direct	19
ADP	direct	48
AECOM	direct	8
AES	direct	2
AIG	facebook	1

Or, we can create two views and join them as follows:

```

create view tbl1 as
    select a.name as name, w.channel as channel, count(*) total
    from accounts a
    join web_events w
    on w.account_id = a.id
    group by a.name, w.channel
    order by a.name;

create view tbl2 as
    select name, max(total)
    from tbl1
    group by name;

select tbl1.name, tbl1.channel, tbl2.max
from tbl1
join tbl2
on tbl1.name = tbl2.name and tbl1.total = tbl2.max
order by tbl1.name

DROP VIEW IF EXISTS tbl1, tbl2;

```

4. **** YOUR TURN**** Provide the name of the sales_rep in each region with the largest amount of total_amt_usd sales.

```

SELECT s.name rep_name, r.name region_name, SUM(o.total_amt_usd) total_amt
FROM sales_reps s
JOIN accounts a
ON a.sales_rep_id = s.id
JOIN orders o
ON o.account_id = a.id
JOIN region r
ON r.id = s.region_id
GROUP BY 1,2
ORDER BY 3 DESC
limit 5;

```

rep_name	region_name	total_amt
Earlie Schleusner	Southeast	1098137.7
Tia Amato	Northeast	1010690.6
Vernita Plump	Southeast	934212.9
Georgianna Chisholm	West	886244.1
Arica Stoltzfus	West	810353.3

```

SELECT region_name, MAX(total_amt) max_total_amt
FROM (SELECT s.name rep_name, r.name region_name, SUM(o.total_amt_usd) total_amt
      FROM sales_reps s
      JOIN accounts a
      ON a.sales_rep_id = s.id
      JOIN orders o
      ON o.account_id = a.id
      JOIN region r
      ON r.id = s.region_id
      GROUP BY 1, 2) tbl1
GROUP BY 1;

```

region_name	max_total_amt
Midwest	675637.2
Southeast	1098137.7
West	886244.1
Northeast	1010690.6

```

SELECT tbl3.region_name, tbl3.rep_name, tbl3.total_amt
FROM (SELECT region_name, MAX(total_amt) max_total_amt
      FROM (SELECT s.name rep_name, r.name region_name, SUM(o.total_amt_usd) total_amt
            FROM sales_reps s
            JOIN accounts a
            ON a.sales_rep_id = s.id
            JOIN orders o
            ON o.account_id = a.id
            JOIN region r
            ON r.id = s.region_id
            GROUP BY 1, 2) tbl1
      GROUP BY 1) tbl2
JOIN (SELECT s.name rep_name, r.name region_name, SUM(o.total_amt_usd) total_amt
      FROM sales_reps s
      JOIN accounts a
      ON a.sales_rep_id = s.id
      JOIN orders o
      ON o.account_id = a.id
      JOIN region r
      ON r.id = s.region_id
      GROUP BY 1, 2
      ORDER BY 3 DESC) tbl3
ON tbl3.region_name = tbl2.region_name AND tbl3.total_amt = tbl2.max_total_amt
limit 5;

```

region_name	rep_name	total_amt
Southeast	Earlie Schleusner	1098137.7
Northeast	Tia Amato	1010690.6
West	Georgianna Chisholm	886244.1
Midwest	Charles Bidwell	675637.2

Subqueries using ‘WITH’

5. What is the lifetime average amount spent in terms of total_amt_usd for the top 10 total spending accounts?

```
with tbl1 as(
  select a.name, sum(total_amt_usd) total_spent
  from accounts a
  join orders o
    on a.id = o.account_id
  group by a.name
  order by total_spent desc
  limit 10
)
select avg(total_spent)
from tbl1
```

avg
304847

6. For the customer/account that spent the most (in total over their lifetime as a customer) total_amt_usd, how many web_events did they have for each channel?

```
with tbl1 as (
  select a.name, sum(total_amt_usd) total_spent
  from orders o
  join accounts a
    on o.account_id = a.id
  group by a.name
  order by total_spent desc
  limit 1)
select a.name, w.channel, count(*) num_events
from web_events w
join accounts a
on a.id = w.account_id and a.name = (select name from tbl1)
group by a.name, w.channel
order by num_events desc
```

name	channel	num_events
EOG Resources	direct	44
EOG Resources	organic	13
EOG Resources	adwords	12
EOG Resources	facebook	11
EOG Resources	twitter	5
EOG Resources	banner	4

7. Provide the name of the sales_rep in each region with the largest amount of total_amt_usd sales.

```

with tbl1 as(
    select s.name sales_person, r.name region_name, sum(total_amt_usd) total_spent
    from region r
    join sales_reps s
        on s.region_id = r.id
    join accounts a
        on a.sales_rep_id = s.id
    join orders o
        on o.account_id = a.id
    group by s.name, r.name
    order by r.name, total_spent desc
),
tbl2 as(
    select region_name, max(total_spent) max_total_spent
    from tbl1
    group by region_name
)
select tbl1.region_name, max_total_spent
from tbl1
join tbl2
    on tbl1.region_name = tbl2.region_name and tbl1.total_spent = tbl2.max_total_spent
order by max_total_spent desc

```

region_name	max_total_spent
Southeast	1098137.7
Northeast	1010690.6
West	886244.1
Midwest	675637.2