SQL Data Analyst Interview

Q&A

You have the following table t1:

```
-- one row per ad per day

(date, ad_id, account_id, spend)
2021-05-01, 123, 001101, 100.00
2021-05-02, 123, 001101, 128.00
```

NOTE: We can create this table as follows:

```
-- create table

create table t1 (
 date date,
 ad_id int,
 account_id int,
 spend float);
```

```
 insert into t1
  (date, ad_id, account_id, spend)
 values
 ('2021-05-01', 123, 001101, 100.00),
 ('2021-05-02', 123, 001101, 128.00);
```

Q1. Your task is to analyze how the ad business is performing over time.

How many advertisers and how much total spend do we have this year?

We will use count() and distinct to answer first question:

```sql
select count(distinct account_id) as "number of advertisers"
from t1;
```

We use sum() to answer part 2 of the question:

```sql
select sum(spend) as "total spend"
from t1;
```

Note: If the table contains many years, you can use a where and date_part() to filter only records from this year:

```sql
select sum(spend) from t1
where date_part('year', date::timestamp) ='2021';
```

Q2. To check for anomalous behavior e.g., big drops/gains, you want to compare week-over-week changes. Write a query to show, on each day, total spend and the week-over-week change, since the beginning of the year.

We use again date_part() , group by and sum():

```sql
-- week-over-week

select date_part('week',date::timestamp) as week, sum(spend) as "WOW"
from t1
where date_part('year', date::timestamp) ='2021'
group by 1;
```

Q3) To understand the distribution of spend per advertiser (account), write a query to populate the histogram of last 7 days of spend using $1000 as bin size.

Distribution table:

```
select account_id, sum(spend) as "total spend"
from t1
group by 1
order by 2;
```

This will give us a noisy hard-to-interpret.

The question is asking to group the accounts into buckets of total spend. We can a subquery like this:

```
-- hist with bins

select ((total_spend/1000)*1000, -2)as bucket,
count(account_id) as accounts
    from(
    select account_id, sum(spend) as total_spend from t1
    where date > now()- interval '7 day'
    group by 1
    order by 2) a
group by 1;
```

The subquery creates the distribution table for the last 7 days of spend:

```
select account_id, sum(spend) as total_spend from t1
where date > now()- interval '7 day'
group by 1
order by 2;
```