Mindy Ng

Q1. SQL Technique Question
Problem: We have 2 tables available below -

Table A: This is a transaction table containing details of transactions from 2018 to 2021.

transaction_id	seller_id	buyer_id	item_id	quantity	item_price	payment_success_yn	Transaction_date (in dd/mm/yyyy) format

Note: payment_success_yn is a Yes / No flag to tell whether this transaction is successful with payment
cleared.

Table B: This is a lookup table telling which country the seller is from.

seller_id	country_name

Note: The list of sellers in table A is a subset of Table B.

1. Using the below two tables, write two queries to answer the following-

--assuming DB: PostgreSQL v15 for queries

1.1. What is the weekly revenue for 2020?

ALTER TABLE table_A
   ALTER COLUMN "Transaction_date" TYPE DATE USING TO_DATE("Transaction_date", 'DD/MM/YYYY'); --Transaction_date column is string type. TO_CHAR function will be used later to convert date to ISO, but only takes non-string. So TO_DATE used here to produce non-string data type.

ALTER TABLE table_A
   ALTER COLUMN "Transaction_date" TYPE DATE USING TO_DATE((TO_CHAR("Transaction_date", 'YYYY/MM/DD')), 'YYYY/MM/DD'); --TO_DATE converts string into a date format. Before doing that, needed to convert our data into ISO format.

SELECT DATE_PART('week', "Transaction_date") AS week --With data in ISO and date format, we extract week value.
, SUM(CASE WHEN payment_process_yn = 'y' THEN quantity * item_price ELSE 0 END) AS revenue

```
FROM table_A
WHERE "Transaction_date" BETWEEN '2020/01/01' AND '2020/12/31'
GROUP BY 1
ORDER BY 1;
```

1.2. What is the highest actual revenue for each country? (The final result of your query has to show a record for each country and the record must have all details (i.e. all fields in Table A below) of the transaction)

```
WITH base AS (

SELECT table_B.country_name
, table_A.transaction_id
, table_A.seller_id
, table_A.buyer_id
, table_A.item_id
, table_A.quantity
, table_A.item_price
, table_A.payment_process_yn
, table_A."Transaction_date"
, SUM(CASE WHEN table_A.payment_process_yn = 'y' THEN table_A.quantity *
table_A.item_price ELSE 0 END) AS revenue
FROM table_A
RIGHT JOIN table_B
ON table_A.seller_id = table_B.seller_id
GROUP BY 1
, 2
, 3
, 4
, 5
, 6
, 7
, 8
, 9

)

SELECT country_name
, transaction_id
, seller_id
, buyer_id
, item_id
, quantity
, item_price
```

```
, payment_process_yn
, "Transaction_date"
, revenue AS highest_revenue
FROM

(

SELECT country_name
, transaction_id
, seller_id
, buyer_id
, item_id
, quantity
, item_price
, payment_process_yn
, "Transaction_date"
, revenue
, DENSE_RANK() OVER (PARTITION BY country_name ORDER BY revenue DESC,
"Transaction_date" DESC) AS revenue_rank
FROM base

) AS subq

WHERE revenue_rank = 1
ORDER BY revenue DESC
```

** Recommendation to DE team: Follow column naming principles:
1. Leave data description out of column names and put in documentation (payment_process_yn -> payment_process).
2. Use all lower-case for column names (Transaction_date -> transaction_date).
3. Have date formatted as ISO ('DD/MM/YYYY' -> 'YYYY/MM/DD').
Making changes upstream by following data engineering best practices would help analysts provide answers to business stakeholders faster.

Q2. Diagnose a national education system

Description of situation

Country X is a country located in Africa with a population of 20 million. The government of X wants to achieve major improvements in both the quantity and quality of education for its children. The country's free-market economy is still developing, having emerged from many decades under communism. Recently, the government of X put in place a new economic plan, with aspirations to transform its economy and "turbocharge" its development so that it is well

positioned to compete with its African neighbors. The government of X realizes that the education of its children is a critical factor in meeting its economic development goals. It intends to transform its school system over the next 10 years so that it is able to support its economic aspirations.

Schooling in country X is completely public, and is provided by a network of government run schools which admit children from ages 5 through 18.

The first stage of this effort is to diagnose the current state of education in schools to determine how best to meet the government's future aspirations.

QUESTION 1

What are the issues you would want to investigate in diagnosing the condition of the current school system in X?

The government wants to improve the quantity and quality of the education system. And by improving their education system, X's economy will improve/"turbocharge".

To diagnose the school system, I would look at metrics of success for schools broken down by quantity and quality of education as well as how those measurements will meet economic development goals:

Quantity of education:
1. How many public schools in X
2. Available schools by school age, neighborhood/income level
3. Number of available teachers at different geographic levels, e.g. national, state, city
4. Funds at national, state, city levels

Quality of education:
1. Type of curriculum and learning goals by grade
2. Highest educational background of teachers

Economic development:
1. Determine sort of industry does X want to be known for, e.g. Textiles/technology/diverse amount of industries
2. Skills needed in order to produce marketable skills for that industry
3. Current availability of teachers and classes that would foster those skills
4. Would different sorts of systems be better to develop these skills in the targeted time frame of 10 years: home-schooling, private schools?

QUESTION 2

The table below shows some important education-related measures for country X, and also for

some comparison countries. Three sets of comparison countries have been used. In the first set are some of X's neighboring countries in Africa. In the second set are some of the most developed economies in Africa. Finally, in the third set are some countries that have similar sized economies to country X on a per person basis (similar GDP per capita).

What can you observe from this chart?

	Students per School	Students per Teacher	Government Spending on Education per Student ($)	Average International-Assessment Score
Country X	500	18	5,000	41
Close Neighbors				
Neighbor A	400	16	3,000	38
Neighbor B	700	20	4,000	43
Neighbor C	800	20	5,500	48
Developed African Countries				
Developed A	500	21	8,000	54
Developed B	350	23	8,500	55
Developed C	400	20	9,000	53

Similar African Countries				
Similar A	600	22	3,500	45
Similar B	500	20	4,000	44
Similar C	300	19	4,500	46

I observe from this chart:
1. Neighboring cities have relatively more crowded schools and have less financial aid from government despite similar test performance scores
2. X is more similar to countries with same economy than developed countries
3. There is no correlation between student enrollment, student/teacher ratio and student test scores
4. The major differences between X and developed countries were government budget and student assessment scores. So there had to be something in quality metric that differed for the gap in the last two metrics despite similarities in enrolled students and student/teacher ratio
5. Therefore, if looking at these metrics alone, looks like number students enrolled and student-to-teacher ratio are not enough to factor into improved student test performance. There could be stronger signal with where else the government funds could be used for, e.g. hiring teachers with high degrees/exceptional backgrounds/high teacher assessment scores, raising the curriculum criteria bar to pass a certain grade

Q3. What key metrics would you propose to monitor how well B-Stock is doing? What, according to you, are the north star metrics for our company?

USER ADOPTION
- TAM - total addressable market
- SAM - serviceable addressable market
- SOM - share of market

MARKET
- installs/TAM
- MAU (monthly active users)/total users

GROWTH
- MAU, WAU (weekly active users), DAU (daily active users)
- D/D, W/W, M/M, Y/Y

- Quick Ratio (ratio of sum of new and resurrected to churned)

RETENTION
- Dn/Mn/Wn where n is day(s) since first time on website

STICKINESS
- Open Rate
- Lness (number of days visited in a given time frame, e.g. L5/7 = five days out of 7 days)
- Sticky DoD, MoM, YoY retention

ENGAGEMENT
- Time spent in product/DAU
- # sessions
- Time spent/session
- Top searched items
- Most bought items broken down by impression indices (top10/top20)
- Sell Through Rate
- Number of impressions, bids, add-to-cart, purchases
- Available Inventory

SHIPMENT
- Total time from purchase to received pallets

PAYMENT
- Time from choosing payment method to making purchase
- Fraudulent accounts

CUSTOMER SATISFACTION
- NPS score
- Number of customer service tickets per week

Metric that Aligns with Mission:
- How Much Money That Could have been lost by partner with non-sold inventory + revenue from palettes sold

Northstar metrics for B-stock:

- Total subscriptions
- GMV - gross merchandise value
- ARR/MRR - Average Recurring Revenue/Monthly Recurring Revenue
- GMV/X-amount of search impressions
- AOV - Average Order Value