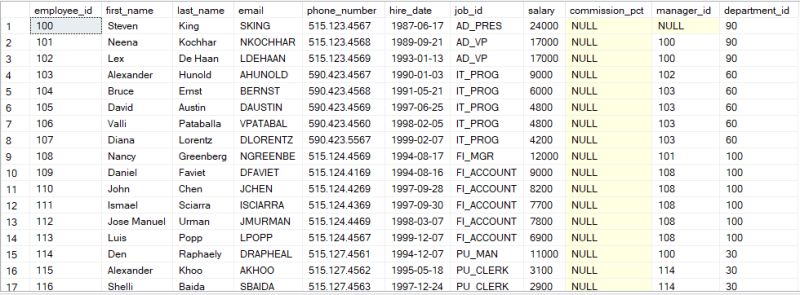
1. Given the table below, write a query to return employees that have spent at least 25 years in the company.

Link to data [here](https://lnkd.in/daTBPMrw)

Link to query data [here](https://lnkd.in/d-BUCy5W)



1. Using the employee’s table as provided above, write a query to extract information of employees where job\_id like CLERK and salary above or equal average employee salary in their department.
2. So an E-commerce company has some customers they intend to track their performance for the month. Here is the business question:

We want to track the days at which customers buy more compared to the preceding days in the month

Table schema looks like this;

transaction\_table(order\_id, product\_id, quantity, unit\_price, order\_date, customer\_id)

NB: order\_date is a datetime field

1. A bank has a table called transaction\_table (transaction\_id, customer\_id, account\_id, amount, transaction\_type, narration, resp\_code, transaction\_date) that stores information about customer's financial transactions.

The bank wants to know the monthly volume and value of successful credit and debit transactions of each customer.

The final table should look like this;

CUSTOMER\_ID, TRANX\_YEAR, TRANX\_MONTH, CREDIT\_VOLUME, CREDIT\_VALUE, DEBIT\_VOLUME, DEBIT\_VALUE

NB:

transaction\_date is a timestamp

transaction\_type is: C for Credit & D for Debit

resp\_code is the response code for transactions. '00' for successful transactions

1. There are two tables:

scores (student\_id, course\_id, full\_name, mark)

grades (grade, lower\_limit, upper\_limit)

Write a query to categorize the marks in the scores table into grades given that the marks fall between the lower and upper limit inclusive.

1. Given an employee attendance table, kindly provide a query that shows if an employee was absent 3 days consecutively in a week.

The table structure is given below:

employee\_attendance (ID VARCHAR(10), EMPLOYEE\_ID INT, TIME\_IN DATETIME, TIME\_OUT DATETIME, FIRST\_NAME VARCHAR(100), LAST\_NAME VARCHAR(100))

1. The sales of an E-commerce store are recorded in the table called sales\_fact (order\_id varchar(200), order\_date datetime, amount float, customer\_id varchar(100), order\_status varchar(200)). The business wants to find the 7 days moving average of delivered orders recorded from January 2022 till date.

NB:

order\_status has values ('Delivered', 'Cancelled', 'In progress')

Write an SQL query that solves the business problem.

1. The business wants to know how long it took a customer to perform his/her first transaction from the time the customer was registered on the platform.

Customers (id INT, join\_date datetime, Full\_name Varchar(200));

Sales (order\_id INT, customer\_id INT, order\_date datetime, Amount INT, order\_status VARCHAR(50));

NB: Join date is the date the customer registered on the platform

1. The business requires the highest performing employee per department each month. The tables required to solve this problem is as highlighted below:

department (id, name, location);

employee (id, first\_name, last\_name, manager\_id, dept\_id);

sales\_table (order\_id, customer\_id, employee\_id, amount, order\_date, order\_status);

NB: order\_status: {'delivered', 'cancelled'}

1. Given an employee table, write a query to generate the last name of employees that start with R and end with e.
2. Given a football stat table, where the time is stored as varchar, write a query to add a new column to store the result as integer.

The sample rows of the time column is given below:

'30', '44', '90+3', '89', '105+1', '120+2', '67'

The result column should be:

30, 44, 93, 89, 106, 122, 67

1. Bank XYZ wants to campaign its customers on the available transaction channels, so it needed to know what channel(s) its customers transact on. The given table to get this metric is provided below:

TRANX\_TABLE (tranx\_id, customer\_id, account\_id, amount, tranx\_date, channel, resp\_code, resp\_mssg)

The result table should consist of just two columns, the customer\_id, followed by the channel(s) used by the customer a comma-separated list in ascending order of channel(s).

NB:

channels: {'POS', 'ATM', 'MOBILE', 'WEB', 'INTERNET BANKING', 'OTC', 'USSD'}

Successful transactions are with resp\_mssg = 'successful'

1. Company ABC posted a Hackathon on Zindi with data professionals all over the world participating. The Hackathon was planned to run for 30 days. After the end of the 30 days, company ABC is interested in extracting the first time a hacker had its maximum score for all submissions. The table to answer this metric is as provided below:

submissions (submission\_id varchar(10), hacker\_id varchar(10), submission\_date datetime, score int);

NB:

A Hacker can make multiple submissions in a day

Sample input

1, 1, 2022-07-01 12:30:00, 40

2, 1, 2022-07-01 16:29:31, 41

3, 2, 2022-07-03 13:01:56, 35

4, 1, 2022-07-02 15:45:00, 44

5, 1, 2022-07-02 14:29:06, 44

6, 3, 2022-07-03 10:56:40, 39

7, 2, 2022-07-04 12:43:55, 42

8, 2, 2022-07-10 09:36:11, 46

9, 4, 2022-07-30 23:30:00, 41

10, 3, 2022-07-10 12:36:43, 40

Sample Output

5, 1, 2022-07-02 14:29:06, 44

8, 2, 2022-07-10 09:36:11, 46

9, 4, 2022-07-30 23:30:00, 41

10, 3, 2022-07-10 12:36:43, 40

1. An E-commerce company Radical is planning to measure the effectiveness of its campaigns on customer transactions. It is called a campaign evaluation.

The given table is given below to provide an SQL query that answers the metric:

customers (id varchar(40), first\_name varchar(100), last\_name varchar(100), age int, dob date, gender varchar(1), mobile varchar(20));

campaign (campaign\_id varchar(40), customer\_id varchar(40), campaign\_manager varchar(200), campaign\_name varchar(200), start\_date date, end\_date date);

transaction (id varchar(40), customer\_id varchar(40), amount int, channel varchar(100), tran\_date date);

Metric(s) to track

The revenue generated pre-campaign (14 days before the campaign start date), during the campaign (between start\_date and end\_date), and post-campaign (14 days after the campaign end date).

1. An E-commerce company wants to know how much revenue they make for every order each month. Here are the tables that are needed for this analysis:

orders (order\_id (PK), order\_date, customer\_id, product\_id, location, description)

products (id (PK), name, description, supplier\_id, available\_quantity, unit\_price)

order\_details (order\_id, product\_id, quantity)

1. A retail company is interested in categorizing its customers into four groups based on monthly customer transaction revenue. The four groups are as highlighted below:

1. Bronze: (<=50,000)

2. Silver: (>50,000 & <=250,000)

3. Gold: (>250,000 & <=1,000,000)

4. Platinum: (>1,000,000)

The transaction table required for this metric is as highlighted below:

transaction (id, customer\_id, tranx\_date, amount, channel, status)

NB:

status = 'successful' indicates a successful transaction

1. A company wants to build an SQL-based loyalty model which tracks the activities of customers using two months intervals (current month and previous month). The company wants to know customers that fall in these categories:

1. Customers that subscribed to a product in both months

2. Customers that subscribed in the previous month but not in the current month

3. Customers that subscribed in the current month but not in the previous month

4. Customers that didn't subscribe in both months

Assumptions

Current month - August 2022

Previous month - July 2022

The table(s) to answer these metrics is as highlighted below:

customers (id, firstName, lastName, phone, age, address, gender)

subscriptions (id, customer\_id, sub\_date, amount, plan, description)

1. Given a sales table as described below, write a query that fetches the daily cumulative sales for the sales team.

sales (id, sales\_date, amount, order\_id, customer\_id, description);

1. Consider a Lending company interested in tracking some business metrics as highlighted below

# Number of loan applications by customers

# Number of active and cleared loan applications per customer

# Number of default loan applications per customer (Only cleared loan applications)

# Number of late payments per application and customer

Using the tales highlighted below:

loan\_application(id, customer\_id, first\_name, last\_name, amount, duration, start\_date, end\_date, pay\_day, status);

loan\_repayment(id, application\_id, customer\_id, amount, expected\_amount, payment\_date, expected\_payment\_date, pay\_num);

Write a single query that answers the above metrics.

NB:

status: {active, cleared}.

pay\_day: The day in the month expected to pay.

duration: integer value specifying the duration of the loan in months.

pay\_num: Integer value corresponding to the nth payment of the applicant.

default loan means that the applicant cleared the loan after the expected end date.

late payment means paying after the expected payment date.

1. A transportation company is interested in answering some business metrics in its data.

Highlighted below are some of the metrics that are needed to be captured.

1a. Waiting time (in minutes) for each customer ride order.

1b. Speed in m/s of the driver between arrival time and request time.

The table to capture this metric is as highlighted below:

rides (id VARCHAR(20) PK, trip\_date DATETIME, passenger\_id VARCHAR(20), request\_time DATETIME, driver\_arrived DATETIME, cancelled\_at DATETIME, driver\_id VARCHAR(20), started\_at DATETIME, drivers\_distance\_at\_acceptance\_KM FLOAT, dropoff\_at DATETIME, region VARCHAR(200));

NB:

trip\_date: Time at which the trip was initiated

request\_time: Time at which passenger successfully requested a ride

driver\_arrived: Time the driver got to the pickup location

cancelled\_at: Time at which the passenger cancelled the ride started\_at: Time at which the driver started the trip

dropoff\_at: Time the trip was completed

1. Use table from 20 to answer this metric: Number of successful/cancelled trips for each month as a pivot
2. Use table from 20 to answer this metric: Number of successful trips and average travel time per driver
3. Use table from 20 to answer this metric: Top 10 drivers (percentage of successful trips compared to total ride requests).

Only consider total trips per driver that exceeds the average number of trips for all drivers.

1. Use table from 20 to answer this metric: Waiting time between a passenger canceled order and the next order request.
2. A transportation business just lauched its mobile app for customers to order rides to diffferent destinations. This application is been monitored for quality assurance purpose and logs of the ride orders are logged into ride\_logs table. Highlighted below is the structure of the table:

ride\_logs (log\_id, trip\_id, trip\_date, passenger\_id, passenger\_cancelled, pickup, destination);

The business needs to be aware based on historical log data available, is the distribution of monthly cancelled rides a decreasing sequence? Write a query to answer the business question.

NB: passenger\_cancelled is the time at which a passenger calcelled the order else NULL

1. A financial institution is planning a reward model for its customers that do not spend up to 50% of their inflow in the previous month. Write an SQL query to identify such customers from the transaction table.

Constraints: The customer credit transaction for the month must be at least 50,000.

NB: A customer can have more than 1 account

1. As an analyst for a sport company, you are requested to present the football data available in the database as: Form of last five matches of each football team in the football table as a pivot table.

Form implies either W, D or L. Where W: win, D: Draw, L: Lose

1. A transportation business is interested in understanding its customers so as to provide a more personalized experience. The business needs to identify customers that ordered a return ride from their pick up location.
2. You have been provided with employees table, and you are charged to create an Org chart with it. How do you write a query to identify the hierarchy of employees from the employees table?

The sample of employees data is attached to this post.

1. Write a query to return employees with lowest salary in each department

**Bonus Questions**

1. Is there anything wrong with the SQL statement below?

SELECT customer\_id, sum(amount) revenue

FROM customer\_tranx

GROUP BY customer\_id

WHERE sum(amount) > 20000;

1. Given that an employee table has details of about 50 employees with an ID ranging from 1 - 50, what is the appropriate operator in the query below to select details of all employees whose salary is above all employees with an ID above 40?

SELECT \*

FROM EMPLOYEES E1

WHERE SALARY \_\_\_\_\_ (SELECT SALARY FROM EMPLOYEES E2 WHERE ID > 40);

1. Is this query valid? If yes, what is the behaviour?

SELECT first\_name, last\_name, SALARY/(SELECT SUM(SALARY) FROM EMPLOYEES)

FROM EMPLOYEES

WHERE SALARY/(SELECT SUM(SALARY) FROM EMPLOYEES) > 0.005;

1. Given a table cust\_tranx with 1,000,000 records, what is the result/behaviour of the query below and why?

SELECT SUM(1)

FROM cust\_tranx;

1. What is the behavior of the UPDATE statement below?

UPDATE EMPLOYEES

SET SALARY = (CASE WHEN ID > 3 THEN SALARY\*2 END);

1. What is the behavior of this query?

SELECT MAX(CASE WHEN SALARY > 300000 THEN 1 ELSE 0 END)

FROM EMPLOYEES;

1. What is the result of the query below?

SELECT NAME, SALARY

FROM EMPLOYEES

HAVING SALARY > 10000;

1. Given table TRANX having 50000 records, what is the result of the query below?

SELECT 2\*COUNT(2), 2\*SUM(2) FROM TRANX;

1. A phone call is considered an international call when the person calling is in a different country than the person receiving the call.

What percentage of phone calls are international? Round the result to 1 decimal.

Assumption:

The caller\_id in phone\_info table refers to both the caller and receiver.

phone\_calls(caller\_id, receiver\_id, call\_time)

phone\_info(caller\_id, country\_id, network, phone\_no)

1. Your team at JPMorgan Chase is soon launching a new credit card. You are asked to estimate how many cards you'll issue in the first month.

Before you can answer this question, you want to first get some perspective on how well new credit card launches typically do in their first month.

Write a query that outputs the name of the credit card, and how many cards were issued in its launch month. The launch month is the earliest record in the monthly\_cards\_issued table for a given card. Order the results starting from the biggest issued amount.

monthly\_cards\_issued(issue\_year, issue\_month, card\_name, issued\_amount)