Q1. **Spotify Penetration Analysis**

Market penetration is an important metric for understanding Spotify's performance and growth potential in different regions. You are part of the analytics team at Spotify and are tasked with calculating the active user penetration rate in specific markets.

For this task, 'active\_users' are defined based on the following criterias:

• last\_active\_date: The user must have interacted with Spotify within the last 30 days.

• monthly\_active\_sessions: The user must have engaged with Spotify for at least 5 sessions in the past month.

• listening\_hours: The user must have spent at least 10 hours listening on Spotify in the past month.

Based on the condition above, calculate the active 'user\_penetration\_rate' by using the following formula.

• Active User Penetration Rate = (Number of Active Spotify Users in the Market / Total Population of the Market)

Total Population of the market is based on both active and passive users. ​ The output should contain 'country' and 'active\_user\_penetration\_rate'. Make sure that all countries that appear in the dataset are also present in the output of your solution. Ensure there are 10 decimal places in your solution.

Let us assume the current\_day is 2024-01-31.

Q2. **Population Density**

You are working on a data analysis project at Deloitte where you need to analyze a dataset containing information about various cities. Your task is to calculate the population density of these cities, rounded to the nearest integer, and identify the cities with the minimum and maximum densities. The population density should be calculated as (Population / Area).

The output should contain 'city', 'country', 'density'.

Q3. **Aggregate Listening Data**

You are tasked with analyzing a Spotify-like dataset that captures user listening habits. For each user, calculate the total listening time and the count of unique songs they have listened to. In the database duration values are displayed in seconds. Round the total listening duration to the nearest whole minute.

The output should contain three columns: 'user\_id', 'total\_listen\_duration', and 'unique\_song\_count'.

Q4. **Customer Feedback Analysis**

Capital One's marketing team is working on a project to analyze customer feedback from their feedback surveys.

The team sorted the words from the feedback into three different categories;

• short\_comments • mid\_length\_comments • long\_comments

The team wants to find comments that are not short and come from social media. The output should include 'feedback\_id,' 'feedback\_text,' 'source\_channel,' and a calculated category.

Q5. **Common Friends Script**

You are analyzing a social network dataset at Google. Your task is to find mutual friends between two users, Karl and Hans. There is only one user named Karl and one named Hans in the dataset.

The output should contain 'user\_id' and 'user\_name' columns.

Q6. **Friday's Likes Count**

You have access to Facebook's database which includes several tables relevant to user interactions. For this task, you are particularly interested in tables that store data about user posts, friendships, and likes. Calculate the total number of likes made on friend posts on Friday.

The output should contain two different columns 'likes' and 'date'.