SQL Case Study Activity: Act Like a Data Analyst

"Think Like a Data Analyst: Real-World SQL Case Study Challenge"

**Activity Description**:

Welcome to the SQL Case Study Challenge!

This activity is designed to simulate real-world scenarios where a data analyst must explore data, identify key patterns, and generate actionable insights. You will be given **datasets**, **problem statements**, and **data descriptions**. Your task is to **select any ONE case study**, perform SQL analysis, and present your findings.

Use your analytical thinking, SQL knowledge, and real-world logic to solve the problem. You are **free to use** any tools or platforms such as **Google**, **ChatGPT**, or **other learning resources** during this activity — just like in a real data analyst role.

**Instructions:**

* Choose **one problem statement** from the list below.
* Explore the provided dataset.
* Write SQL queries to solve the questions or extract insights.
* Document your approach and conclusions in a structured format.
* You may visualize insights (optional) using Excel/Power BI/Tableau.

**Rules:**

* You must **submit your SQL queries + final insights**.
* You are **encouraged to use external help** (search engines, AI tools), but ensure you understand and can explain your solution.
* Deadline: [Insert Date]
* Submission Format: A zipped folder containing:
  + SQL script or screenshots of queries
  + A PDF or Word report with conclusions
  + (Optional) Excel/Power BI visuals

**Available Case Studies (Choose One)**

Case Study 1: Financial Consumer Complaints – Bank of America

**Dataset**: [Dataset link](https://drive.google.com/drive/folders/1m8fCnhLiks6XnAD1KHpAs41uhtjNvnxu?usp=sharing)   
**Description**:  
This dataset contains consumer complaints related to financial products and services offered by **Bank of America** between **2017 and 2023**. Each complaint includes:

* Date submitted to the Consumer Financial Protection Bureau (CFPB)
* Product type (e.g., credit card, mortgage, checking account)
* Reported issue
* Response timeline and type of resolution by the company

This dataset provides valuable insights into customer dissatisfaction and how the organization handles complaints.

### ****Problem Statement****:

You have been hired as a **Data Analyst** for a consumer protection organization. Your task is to explore complaints lodged against Bank of America and identify areas that need attention from both the business and policy-making perspectives.

**Your Analysis Should Include:**

1. **Seasonal Trends**:
   * Are there any months or seasons when consumer complaints spike?
   * Can you identify patterns across years?
2. **Product-Level Insights**:
   * Which financial products receive the **most complaints**?
   * What are the **top 3 issues** associated with the top product?
3. **Resolution and Timeliness**:
   * How are complaints typically resolved (e.g., closed with explanation, relief provided)?
   * What percentage of complaints receive **untimely responses**?
4. **Deeper Insights** *(Optional)*:
   * Are untimely responses more frequent for specific product types or issues?
   * How does resolution type vary across complaint categories?

**Case Study 2: Personal Music Trends – Spotify Streaming History**

**Dataset**: [Dataset link](https://drive.google.com/drive/folders/1sg41KxSm-J8k2kmSxC42AECpRIJ9HXol?usp=sharing)  
**Description**:  
This dataset captures a user's **complete music streaming history** on Spotify. It includes:

* Timestamps of when each track was played
* Track name, artist, and album
* Reasons for playing and stopping the track
* Duration played and whether the track was skipped

This data reflects real listening behavior and patterns over time, ideal for personal or user-level music analytics.

### ****Problem Statement****:

You are acting as a **Data Analyst** for a music behavior research project. Your task is to analyze a user's Spotify streaming history and generate insights into their listening habits, preferences, and behavioral patterns.

**Your Analysis Should Include:**

1. **Top Artists & Songs**:
   * Which artists were most listened to **this year**?
   * How does that compare to **last year**?
   * What are the most-played songs overall?
2. **Skipping Behavior**:
   * Which songs are **most frequently skipped**?
   * Are favorite songs also being skipped sometimes?
3. **Listening Time Analysis**:
   * What are the most common **times of day** or **days of the week** for music streaming?
   * Identify any clear trends in listening behavior (e.g., late-night listening, weekend bingeing).
4. **Discovery vs. Loyalty**:
   * How often does the user listen to **new artists** vs. repeating familiar ones?
   * Can you identify patterns of **exploration** vs. **routine**?

### ****Case Study 3: Himalayan Expeditions – A Century of Mountaineering****

**Dataset**: [Dataset link](https://drive.google.com/drive/folders/1v3g5b81-vyYSXYtpLnwK_nNhd9xV1wXc?usp=sharing)  
**Description**:  
This dataset is derived from the **archives of Elizabeth Hawley**, documenting **Himalayan expeditions in Nepal from 1905 to 2024**. It features over **11,000 expeditions**, **89,000+ climbers**, and records of **480 peaks**, including the dates, outcomes, and notable occurrences (e.g., fatalities, summit success, injuries).

### ****Problem Statement****:

You are working as a **Data Analyst for a Himalayan Mountaineering Federation**. Your task is to analyze historical expedition data to generate actionable insights about mountaineering trends, expedition risks, and popular peaks.

**Recommended Analysis Questions:**

1. **Summit Success Trends**:
   * What is the **average summit success rate** across all peaks?
   * Which **peaks have the highest** and **lowest** summit success rates?
   * How has the **success rate changed over decades**?
2. **Expedition Volume & Timing**:
   * What are the **most popular years** and **seasons** for expeditions?
   * How many expeditions happened **before vs. after 2000**?
3. **Peak-Level Insights**:
   * Which **peaks attract the most expeditions**?
   * Are certain peaks more **dangerous** (higher death or injury rates)?
4. **Member Demographics & Outcomes**:
   * What are the most **common nationalities** of climbers?
   * What percentage of climbers **reached the summit**, were **injured**, or **died**?
5. **Climber Repetition Analysis** *(Bonus)*:
   * Are there climbers who have **attempted multiple expeditions**?
   * How many people have **summited more than one peak**?