Assignment 2: SQL and APIs

Available Date: February 5th
Deadline: February 12th, 11:59 PM PST

Assignment Overview

This assignment is about using **SQL and APIs**. You will first extract the **S&P 500 ESG Risk Ratings dataset** and then answer SQL questions based on it.

Part 1: Extract Data and Create Tables (60%)

Instructions

1. **Download the S&P 500 ESG Risk Ratings Dataset**
   Use the following Python code to download the dataset from Kaggle (**Refer to Lecture 7.ipynb**):

   ```
   import kagglehub

   # Download latest version
   path = kagglehub.dataset_download("pritish509/s-and-p-500-esg-risk-ratings")

   print("Path to dataset files:", path)
   ```

2. **Preprocess the Data**

   - Remove the column named "Description" before proceeding to create tables.

3. **Create a Database and Tables**

   - Use the **same method from Lecture 9 Part 2** to connect with your localhost.
   - Create a database named `fre521da2`.
   - Create a table called `rating`.
   - Ensure the database is structured properly before inserting data.

4. **Verify Data Insertion**

    - Use the last cell in your **Jupyter Notebook** to check if the data has been successfully inserted.

Part 2: SQL Queries (40%)

Instructions

- Answer the following **five SQL questions** using the extracted dataset.
- Execute the queries and verify the outputs in your Data Studio.

SQL Questions

1. Retrieve the top 10 companies with the highest total ESG risk score.
2. Find the company with the lowest environmental risk score.
3. List all companies in the **"Technology"** sector, sorted by their governance risk score in descending order.
4. Find the average social risk score for companies in the **"Healthcare"** industry.
5. Identify companies that have a controversy score greater than 3 and categorize them as **"High-Risk"**.

Final Submission Requirements

- **Jupyter Notebook** (`.ipynb`):
  - Demonstrates the extraction, preprocessing, and table creation process.
- **SQL File** (`.sql`):
  - Contains all SQL queries used in **Part 2**.

Marks Distribution

Task	Marks
Part 1: Extract Data and Create Tables	60%
Download dataset and remove "Description" column	15%
Create database and table	15%
Insert and verify data in the table	15%
Submission of `.ipynb` and `.sql` files	15%
Part 2: SQL Queries	40%
SQL Query 1	8%
SQL Query 2	8%
SQL Query 3	8%
SQL Query 4	8%
SQL Query 5	8%