World Mining Commodities Project Overview

This project is a comprehensive SQL-based analysis of the mining industry, ideal for beginners and intermediate learners.
It provides hands-on experience with real-world data and covers skills like data cleaning, querying, and visualization.

About the Project

This project is centered around analyzing the mining industry using real-world datasets from the
World Mining Commodities dataset repository. It involves exploring the mining companies, their operational data,
and country-level mining statistics to generate insights and develop SQL proficiency.

Objective:
To analyze global mining trends, understand commodity production levels, and evaluate

company performance using SQL.
The project also includes crafting interview-level SQL queries (basic, intermediate, and advanced) for practice.

Datasets:
Available on GitHub: https://github.com/AnchalDayal/World-Mining-Commodities---SQL

Table 1: 116_world_mining_companies_clean.csv
Contains data on 116 mining companies, their operations, commodities, and project stages.

Table 2: world_mining_commodities_clean.csv
Covers country-level production data for 65 commodities across 169 countries from 2018 to 2022.

Project Deliverables

1. Database:
   - Two SQL tables loaded into a database.
   - Cleaned and structured data ready for analysis.

2. SQL Queries:
   - Basic, intermediate, and advanced queries answering 30 interview-style questions.

Basic SQL Questions

1. List all unique commodities mined by companies in Table 1.
2. Retrieve the names and websites of all companies operating in Canada.
3. List the countries from Table 2 where 'Gold' was mined in 2022.
4. Find the total production of 'Copper' for the year 2020 across all countries.
5. Display the name and project stage of companies involved in Diamond mining.
6. Identify the unit of measurement used for 'Nickel' in Table 2.
7. List all companies in Table 1 that are in the 'Production' stage.
8. Find the top 3 commodities mined in Zimbabwe based on total production in 2021.
9. Retrieve records where the production of any commodity exceeded 1 million tons in 2022.
10. List the countries where 'Talc' production decreased between 2018 and 2022.

Intermediate SQL Questions

1. Calculate the total revenue generated by all companies mining 'Gold' from 2018 to 2022.

   Clarification: "In this context, the market capitalization value in Table 1 (column: 'Ticker / Market Capitalization') will be used as a proxy for the company's potential revenue. Please assume that market capitalization is directly proportional to the revenue generated from mining 'Gold'."

2. Find the top 5 countries with the highest total production of all commodities combined for 2022.
3. Retrieve the names of companies operating in both 'Canada' and 'Peru'.
4. Identify the commodity with the highest average production per country between 2018 and 2022.
5. Write a query to list the top 3 project stages (by count of companies) from Table 1.
6. Find the total production of 'Copper' in countries where the leading company is based in Africa.
7. Retrieve all commodities mined by companies with a market capitalization exceeding $100M.
8. Calculate the year-over-year growth rate for 'Gold' production in Australia.
9. Find all companies involved in the mining of at least 3 different commodities.
10. List the countries that consistently mined 'Platinum' every year from 2018 to 2022.

Advanced SQL Questions

1. Identify the top 3 companies contributing to the highest revenue for 'Diamonds' in 2021.
2. Write a query to rank all countries based on their average production of 'Silver' between 2018 and 2022.
3. Find all companies involved in mining commodities that were never mined in their operating locations.
4. Calculate the percentage contribution of each company to the global production of 'Nickel' in 2020.

5. List the commodities mined by companies operating in more than 3 countries and producing over 100,000 tons annually.

   Clarification: "To address this, you will likely need to JOIN the two tables (companies and country-level production data), aggregate by country and company, and filter results based on the number of countries and production thresholds."

6. Retrieve the countries where the total production of 'Platinum' exceeded the combined production of 'Gold' and 'Silver' for any year.

7. Write a query to generate a summary of mining activity by continent using the country and location information.

8. Find the correlation between production trends of 'Copper' and 'Iron' from 2018 to 2022.
   (*This will likely involve calculating correlation manually or exporting data for visualization.*)

   Clarification: "To calculate correlation, you can either calculate the correlation coefficient in SQL using appropriate aggregation techniques or export the data for visualization and analysis using an external tool such as Python or Excel."

9. List the top 5 countries by growth rate in 'Rare Earths' production between 2018 and 2022, along with their respective rates.

10. Identify the company with the broadest operational scope (most locations) and the most diversified commodity portfolio.