1. Data Loading

The dataset was loaded into Excel using **Power Pivot → Get Data → From Text/CSV**. The raw data contained **304,632 rows and 22 columns**.

---

2. Data Cleaning and Preparation

2.1 Handling Missing Data

Rows with incomplete critical information were removed to ensure data quality:

- **11 rows** with empty site names

- **81 rows** with no commodity information

- **10 sites** with missing latitude or longitude

- **19 rows** with no country or region data

2.2 Removing Irrelevant Columns

Columns that were not relevant to the objectives of the analysis were removed. These included:

- state

- county

- com_type

- dep_type

- prod_size

- ore

- gangue

- work_type

- names

- ore_ctrl

- hrock_type

- arock_type

This helped reduce noise and improve performance during analysis.

2.3 Commodity Normalization

Some mining pits produced multiple commodities, which were originally stored across different columns. To standardize the data:

- All commodity columns were **combined into a single column**

- The combined values were **split into rows** so that each row represented one commodity per mining unit

- Extra spaces were trimmed

- The original separate commodity columns were then removed

This ensured consistency and made commodity-based analysis accurate.

2.4 Resolving Duplicate Site Names

Some sites had duplicate or inconsistent naming. Since Excel's default **Remove Duplicates** feature could not be used directly (due to differences across related columns), a new approach was applied:

- A **new unique identifier column** was created using a combination of **Site Name, Latitude, and Longitude**

- Any records sharing the same values for all three fields were treated as the **same mining unit**

- If any of the three values differed, the records were treated as **separate mining units**

This ensured accurate identification of unique mining units.

2.5 Classifying Operation Type

Some site names represented **processing facilities** rather than actual mining pits. To address this:

- A new column called **Operation Type** was created

- Mining pits and processing facilities were clearly categorized

This distinction was important for meaningful operational analysis.

2.6 Removing Location Coordinates

The **latitude and longitude** columns were removed because the analysis focused on **country-level insights**, not exact pit locations.

2.7 Final Dataset Structure

After completing the cleaning process, the remaining columns were:

- MiningUnit_ID

- Operation Category

- Continent

- Country

- Commodity

- Operation Type

- Development Status

For clarity and readability, the following columns were renamed:

- site_name → **MiningUnit_ID**

- region → **Continent**

- country → **Country**

- dev_stat → **Development Status**

---

3. Data Analysis

The following analyses were performed using **distinct counts** to avoid over-counting mining units that appear multiple times due to different commodities:

1. **Distinct count of mining units by country**

   o Ensured that a mining unit producing multiple commodities in the same country was counted once

   o The **top five countries** were identified

2. **Distinct count of mining units by continent**

   o   Further split into **mining pits** and **processing facilities**

3. **Unique commodities produced per country**

   o   Each commodity was counted once per country, regardless of how many pits produced it

   o   Identified the **top five countries** by commodity diversity

4. **Distinct count of countries producing each commodity**

   o   Each country was counted once per commodity

   o   Highlighted the **top five most globally produced commodities**

5. **Distinct count of mining units per commodity**

   o   Measured how intensively each commodity is produced worldwide

6. **Distinct count of mining units by development status**

   o   Covered categories such as producers, occurrences, past producers, prospects, and plants

7. **Distinct count of mining units by operation type**

   o   Included surface mining, underground mining, combined methods, and others

8. **Distinct count of commodities**

   o   Determined the total number of commodities produced worldwide

9. **Distinct count of mining units**

   o   Calculated the total number of unique mining units globally, including both mining pits and processing facilities

---

4. Dashboard Design

An interactive dashboard was created to visualize insights clearly and intuitively.

4.1 Charts Used

- **Column charts** to display:

   o   Top commodities by number of mining units involved

   o   Top commodities by global country operation

       o   Top countries by number of unique mining units

       o   Top operation types

- **Pie chart** showing the distribution of development status categories

- **Bar chart and live map** highlighting the top countries by total number of unique mining units

---

4.2 Interactivity

- **Slicers** were added to enable dynamic filtering by:

       o   Continent

       o   Country

       o   Commodity

The dashboard updates automatically based on selected filters.

4.3 Key Performance Indicators (KPIs)

The dashboard includes the following KPIs:

- **Total number of mining units**

- **Total number of commodities**

- **Top commodity produced worldwide**

---

5. Conclusion

This structured data cleaning, transformation, and analysis process ensured accurate insights into global mining operations. The final dashboard provides a clear, interactive overview of mining activity by commodity, geography, development status, and operation type, making it suitable for academic reporting and professional portfolio presentation.