Occurrence of Crystals in Switzerland

CIP02 – Data Collection, Integration & Preprocessing

Lucerne University of Applied Sciences & Arts

MSc in Applied Information and Data Science

Spring Semester 2025

Authors:

Andreas Goerre

Barbara Maier

Sheena Walker

Lecturer Info

Andreas Melillo

Ramón Christen

TABLE OF CONTENTS

1 Introduction, Scope & Purpose

Switzerland's diverse geology and alpine terrain host a rich variety of crystals, from quartz to rare minerals. These formations are influenced by temperature, pressure and mineral composition, making altitude a key factor in their distribution. This project investigates the correlation between crystal occurrence and elevation in Switzerland. We aim to determine whether certain crystals are more abundant at certain altitudes and how geological formations influence their diversity.

Using public datasets (Mindat.org, Swiss Geological Survey, scientific publications), we will apply Python-based data analysis to identify patterns.Our approach includes data collection, cleaning, (statistical) analysis, and visualization to uncover relationships between elevation and crystal formation.

The results will be valuable for mineralogists, geologists, and collectors, with potential applications in other mountainous regions.

2 Research Questions

To structure our analysis, we propose the following key research questions:

1) **What is the spatial distribution of crystal occurrences across different regions in Switzerland?**
   To achieve this we will:
   - Collect geolocation data on known crystal sites from mineralogical databases and geological surveys.
   - Use GIS tools (e.g., GeoPandas, Folium) to map occurrences across different regions.
   - Categorize sites by geographic features (e.g., Alpine vs. Jura regions).
   - Identify spatial clustering patterns and regional mineral diversity.
2) **What are the most common crystal types found in Switzerland?**
   For this we will:
   - Extract and classify crystal occurrences by **mineral type**.
   - Analyze frequency distributions of different crystal types across Switzerland.
   - Compare findings with **geological literature** to validate classification.
   - Identify **potential geological factors** influencing crystal variety.
3) **Is there a statistically significant correlation between the occurrences of crystals and the elevation of the place found?**
   We will do this:
   - Categorize crystal occurrences by mineral type.
   - Perform statistical correlation analysis (e.g., Pearson or Spearman correlation) between mineral type and elevation.
   - Use data visualization techniques (e.g., histograms, box plots, scatter plots) to identify trends.
   - Compare results with geological literature to interpret possible scientific reasons for observed patterns.

3 DATA SOURCES

To support our analysis, we will obtain data among others from the following main source:

- Mindat.org - A well-established online mineral database with location-specific information on crystal occurrences.
- There are also other websites like Swiss Geological Survey or bfs.admin.ch that provide further data insights.

Data is collected through web scraping, API requests, and direct downloads from official geological sources.

4 POTENTIAL RISKS AND MITIGATION STRATEGIES

Data Completeness and Availability

- Risk: Some crystal locations may be missing or inaccurately recorded in available data sets.
- Mitigation: We will use reliable data sources and cross-check records to improve completeness. Where gaps exist, we will acknowledge them in our analysis.

Data Accuracy and Consistency

- Risk: Different sources may provide conflicting elevation data or use different coordinate formats.
- Mitigation: We will standardize datasets using Python (Pandas, NumPy) and GIS tools to ensure consistency in format and units.

Legal and Ethical Considerations

- Risk: Web scraping from some websites may be restricted by terms of service.
- Mitigation: We will review robots.txt files and ensure compliance with ethical data collection practices. Where APIs are available, we will use them instead of direct scraping.

Computational

- Risk: Large datasets with lots of crystal occurrences and elevation points may require significant computing power.
- Mitigation: We will use efficient data processing techniques, optimize Python scripts, and use cloud computing resources if needed.

Statistical and Scientific Validity

- Risk: Correlations found may be coincidental rather than scientifically meaningful.
- Mitigation: We will check the results with scientific sources to ensure plausible outcomes.

5 CONCLUSION

This study explores the relationship between crystal deposits in Switzerland and altitude using data science techniques. Our findings can provide insights for geologists and mineralogists while honing our Python skills in a research-oriented environment. Through careful data selection, statistical analysis, and visualization, we aim to contribute to both scientific research and data-driven geoscience.