****

***Progressive Education Society’s***

**Modern College of Arts, Science and Commerce, (Autonomous)**

**Shivajinagar, Pune – 411005.**

**Department of Statistics**

Academic Year 2025-2026

Project Report on

**A Study of Rare Earth Element Abundance in Various Locations Worldwide; Statistical Approach**

*Submitted by*

**Kiran Vishnu Sanap (2534304)**

*Under the guidance of*

**Dr. Mohan Manohar Kale**

Head of department,

Department of Statistics,

Savitribai Phule Pune University.

* **Acknowledgement**

We express our sincere gratitude to all those who supported and guided us throughout the completion of this project. We deeply thankful to **Dr. Mohan Manohar Kale**, our project guide and mentor, for his valuable guidance, encouragement, and continuous support throughout the duration of this project. His expertise, insightful suggestions, and patient supervision played a crucial role in shaping the direction and outcome of this research on Rare Earth Elements (REE).

We sincerely thankful to our institutions internal mentor **Miss Vaishnavi Mane**, for creating a supportive academic environment and offering access to software, research databases, and computational tools essential for carrying out this project.

Finally, we acknowledge the constant support of our friends and peers, whose discussions, feedback, and motivation helped us refine our ideas and complete this work successfully. We deeply appreciate the collective effort that contributed to the successful completion of this group project.

* **Aim/Objectives**

The aim of this project is to study the abundance, distribution, and geochemical behavior of Rare Earth Elements (REE) across different rock types from various geological sites around the world, using statistical methods to analyze REE patterns and assess their potential significance.

* **Abstract**

Rare Earth Elements (REEs) are essential for clean energy, electronics, and advanced technologies. Although Asia currently dominates global REE production, regions such as Greenland, Africa, and Europe hold significant untapped potential. This project presents a comprehensive geochemical and statistical analysis of REE concentrations in geological samples collected from multiple sites worldwide.Multivariate techniques including **clustering, cluster profiling, Elastic Net regression, and ANOVA** are applied to evaluate the variation, distribution, and relationships among Light and Heavy Rare Earth Elements (LREEs and HREEs). These methods help identify natural geochemical groupings, quantify influential elements, and assess patterns of enrichment and depletion across different rock types.

* **Introduction**

Rare Earth Elements (REEs) are critically important to the advancement of modern technologies, particularly those aligned with sustainable and environmentally friendly practices. Their unique electronic, magnetic, and catalytic properties make them indispensable in a wide range of high-tech and green energy applications including electric vehicles, wind turbines, solar panels, energy-efficient lighting, and electronic devices. As such, REEs are increasingly regarded as strategic materials in the transition towards low-carbon energy systems.

**Rare Earth Elements:** Despite their name, rare earth metals are not rare at all. Their unique magnetic, luminescent, and electrochemical properties make them essential. Rare Earth Elements (REEs) have become indispensable in a wide array of modern technological applications due to their unique magnetic, optical, and catalytic properties. They are extensively used in the electronics industry for the manufacture of smartphones, laptops, LED screens, and high-performance speakers, where their ability to enhance sound and image quality is critical. In the field of clean energy, REEs such as Neodymium and Dysprosium are key components in the permanent magnets used in wind turbines and electric vehicle (EV) motors, contributing significantly to the development of sustainable energy solutions. In the defence sector, REEs are employed in advanced military technologies including radar systems, missile guidance systems, and night-vision equipment, making them strategically important materials. Furthermore, REEs play a vital role in the healthcare sector, being utilized in magnetic resonance imaging (MRI) contrast agents and X-ray imaging equipment. They are also crucial in industrial applications such as petroleum refining and automotive catalytic converters, where they act as efficient catalysts. Given their widespread usage across critical industries, REEs are considered essential for economic development, technological advancement, and national security.

Rare Earth Elements (REEs) refer to a group of 17 chemically similar elements that include the 15 lanthanides from lanthanum (La) to lutetium (Lu) along with yttrium (Y) and scandium (Sc). Although the term "rare" suggests scarcity, most REEs are relatively abundant in the Earth’s crust; however, they rarely occur in concentrated and economically viable forms, which justifies the label. REEs are commonly classified into two categories Light Rare Earth Elements (LREEs) and Heavy Rare Earth Elements (HREEs) based on their atomic number and geochemical behaviour. Light Rare Earth Elements (LREEs) include lanthanum (La), cerium (Ce), praseodymium (Pr), neodymium (Nd), promethium (Pm), samarium (Sm), europium (Eu). LREEs generally occur in higher abundance than HREEs and tend to concentrate in early-crystallizing minerals during magmatic differentiation. Heavy Rare Earth Elements (HREEs) include gadolinium (Gd), terbium (Tb), dysprosium (Dy), holmium (Ho), erbium (Er), thulium (Tm), ytterbium (Yb), lutetium (Lu), yttrium (Y), which behaves similarly due to its comparable ionic radius. HREEs are less abundant, more valuable, and often more challenging to extract due to their association with resistant minerals. In the periodic table, the 15 lanthanides are placed in the f-block (atomic numbers 57–71), often shown separately below the main table. Scandium (Sc, atomic number 21) and Yttrium (Y, atomic number 39), though d-block elements, are included in the REE group due to their occurrence with lanthanides in nature and similar trivalent (+3) oxidation state.

REEs possess a range of magnetic, optical, and catalytic properties that make them strategic and technologically critical. Their applications span across:

* Clean energy technologies: Neodymium (Nd), praseodymium (Pr), dysprosium (Dy), and terbium (Tb) are key components in high-strength permanent magnets used in wind turbines and electric vehicle motors.
* Electronics and communication: Europium (Eu) and yttrium (Y) are used in phosphors for LED and LCD screens, while lanthanum (La) is used in camera and telescope lenses.
* Defence systems: REEs are essential for precision-guided munitions, radar systems, and night vision devices.
* Catalysis and metallurgy: Cerium (Ce) is widely used in catalytic converters and glass polishing compounds, and scandium (Sc) improves the strength and corrosion resistance of aluminium alloys.

The study of Rare Earth Elements (REEs) is of growing importance due to their critical role in modern technology, economic development, and strategic resource planning. As global demand for REE-based products continues to rise particularly in sectors such as renewable energy, electronics, and defence there is an urgent need to explore and evaluate new sources of these elements. Understanding the geochemical behaviour, distribution, and concentration of REEs in specific geological settings can provide valuable insights into their potential for sustainable extraction and utilization. Ultimately, the objective of this study is to generate reliable geochemical and statistical data that can guide further exploration, promote resource security, and contribute to environmentally responsible mining practices.

The global distribution of REE deposit types is summarized in Fig-1 which also provides an overview of REE production sites, advanced REE projects and important REE occurrences.



**Figure 1.** *Overview of REE mining operations, advanced projects and known deposits and occurrences according to deposit type (British Geological Survey 2011).*

To analyse Rare Earth Elements (REEs) in geological samples, geologists employ a range of advanced geochemical and mineralogical techniques to determine their concentration, distribution, and host phases. One of the most widely used methods is Inductively Coupled Plasma Mass Spectrometry (ICP-MS), which offers high sensitivity and precision for detecting trace concentrations of REEs in rock, soil, and mineral samples (Long et al., 1990). Another commonly applied technique is Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES), particularly effective for analysing higher concentrations in bulk samples. X-ray Fluorescence (XRF) is often used for rapid, non-destructive analysis of major and trace elements, although its sensitivity for REEs is limited compared to ICP-MS. Additionally, Electron Probe Microanalysis (EPMA) and Scanning Electron Microscopy with Energy Dispersive Spectroscopy (SEM-EDS) are utilized for studying the mineralogical associations and micro-scale distribution of REEs within host minerals (Williams-Jones et al., 2012). For isotopic studies and detailed phase identification, Laser Ablation ICP-MS (LA-ICP-MS) is a preferred tool, enabling in-situ analysis with spatial resolution. These analytical techniques, when integrated with petrographic observations and statistical modelling, provide a comprehensive understanding of REE geochemistry and help identify economically viable deposits (Bau & Dulski, 1996). Together, these methods form the backbone of REE exploration and characterization in modern geoscientific research.

This research focuses on the geochemical assessment of Rare Earth Elements (REEs) in geological samples collected from multiple sites across the world. REEs are critical for clean energy, advanced technologies, and strategic mineral supply chains. Although Asia remains the leading producer, regions such as Greenland, Africa, and Europe show significant potential, provided environmental and processing challenges are addressed.

* **Data Description**

1. **Bayan Obo**



The Bayan Obo mine located in the Inner Mongolia region of China is the world’s biggest rare earth element (REE) mine both by recoverable reserves and production. It accounts for more than 40% of the total known REE reserves in the world and nearly half of the global rare earth production.

**Source of Data****:** “A Geochemical Study of an REE-rich Carbonatite Dyke at Bayan Obo, Inner Mongolia, Northern China. (YANG Xueming1, ZHENG Yongfei1, YANG Xiaoyong1, ZHANG Peishan2 and M.J. LE BAS3)”

**Sample and Dataset Overview:** The dataset includes four rock types from the Bayan Obo Carbonatite Complex, with their respective rock samples:

Cbt (carbonatite): (90/39, 90/43, 90/44, 90/48, 93/149, 93/151)

Ft (fenite): (90/49, 90/51, CD13‑3, CD13‑4, CD13‑5)

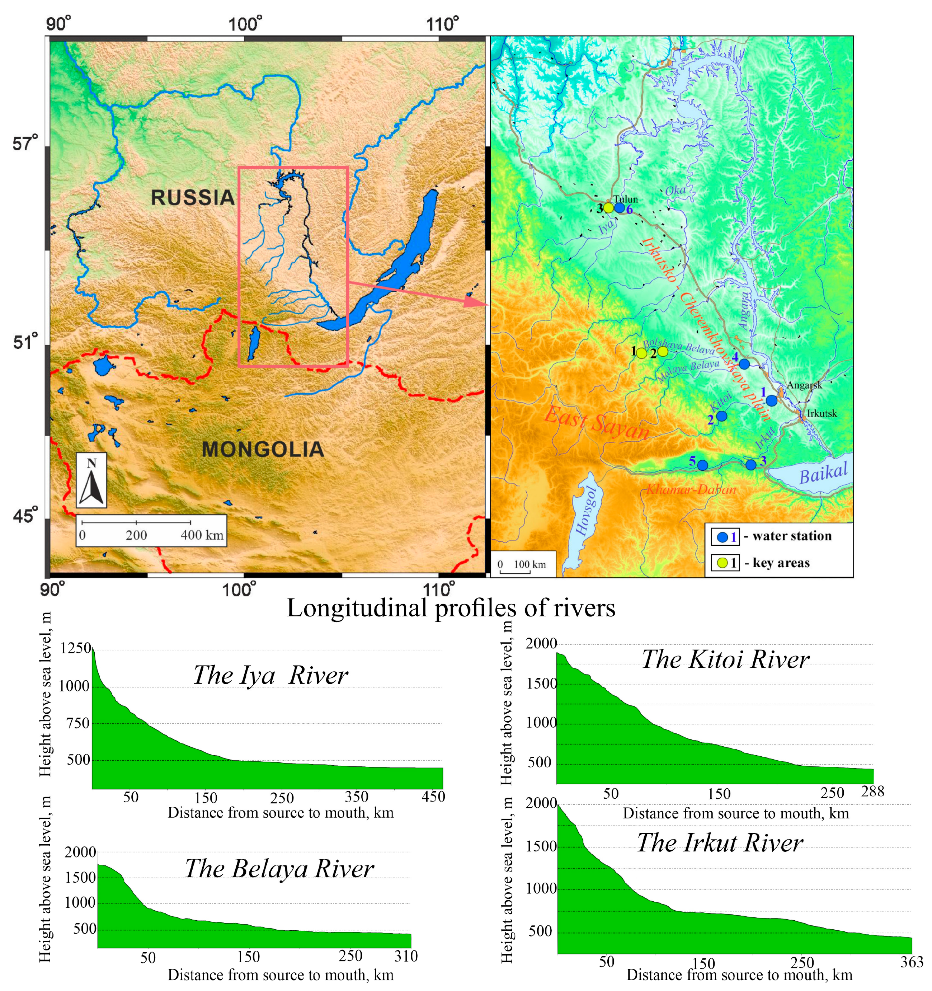
**Elements Measured:** The dataset includes 14 elements: La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu

**Observations:**

Given observations = 149, Missing observations = 19, Total observations = 168

1. **Belaya zima**

The Belaya Zima (Beloziminsky) site is a carbonatite complex it is located in the northern foothills of the Eastern Sayan Mountains, approximately 140 km south of the town of Tulun. in the Eastern Sayan Mountains, Irkutsk region, in southern Siberia, Russia. The intrusion is situated on the left bank of the Zima River, in the river's left tributary valleys.



The Belaya Zima (Beloziminsky) site is a carbonatite complex it is located in the northern foothills of the Eastern Sayan Mountains, approximately 140 km south of the town of Tulun. in the Eastern Sayan Mountains, Irkutsk region, in southern Siberia, Russia. The intrusion is situated on the left bank of the Zima River, in the river's left tributary valleys.

**Source of Data:** **“**Trace-element composition of minerals and rocks in the Belaya Zima carbonatite complex (Russia): Implications for the mechanisms of magma evolution and carbonatite formation**.”** Anna G. Doroshkevich, Ilya V. Veksler, Reiner Klemd , Elena A. Khromova, Ivan A. Izbrodin

**Sample and Dataset Overview:** The dataset includes four rock types from the Belaya zima Carbonatite Complex, with their respective rock samples:

M (Melteigite) : 2309/109-114, 01/151, 2092-761-763

I (Ijolite): 2092-761-763, 2095-226-231, 520-95-99

A (Alkaline syenite): C1-135-144, C1-131-135, 2095-325

CD (Calcite–dolomite carbonatite): c411/25-34, c-257, 2096-257, BZOl, 2050-119-121, 2099- 75-80, 2350, 2098-679-665

AC (Ankerite carbonatite): 2099-297-300, 2098-679, 2098

**Element in data set: Include Oxide, Trace Element, and REE:** SiO₂, TiO₂, Al₂O₃, Fe₂O₃, FeO, MnO, MgO, CaO, Na₂O, K₂O, P₂O₅, LOI, Total, CO₂, S, F, Li, Be, Sc, V, Cr, Co, Ni, Cu, Zn, Ga, Ge, Rb, Sr, Y, Zr, Nb, Mo, Ag, Cd, Sn, Sb, Cs, Ba, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Hf, Ta, W, Tl, Pb, Bi, Th, U

**Observations:**

Total Observation: 1,220, Given Observation: 1,104 , Missing Value: 116

1. **Fen**



The Fen Carbonatite Complex, located in Telemark, Norway, is a unique geological site and the type locality for carbonatite rocks. Formed around 580 million years ago, it hosts a variety of rare igneous rocks and is enriched in elements such as niobium, thorium, and rare-earth elements, making it important for both scientific research and mineral exploration.

**Source of Data:** “The hydrothermal alteration of carbonatite in the Fen Complex, Norway: mineralogy, geochemistry, and implications for rare-earth element resource formation” by C. Marien, A. H. Dijkstra, and C. Wilkins, School of Geography, Earth and Environmental Sciences, Plymouth University, UK.

**Sample and Dataset Overview:** The dataset includes four rock types from the Fen Carbonatite Complex, with their respective rock samples:

Carbonatite – 15-82, 15-83, 15-84, 16-21, 15-85, 16-18

Transitional Carbonatite W – 16-20, 16-20/2, 16-19, 16-17

Rødbergite – 15-86, 16-104, 15-88, 15-88/2, 16-23, 16-22, 15-89, 15-90, 16-112, 15-94

Transitional Rødbergite – 15-91, 15-92, 16-110, 15-93, 16-111

**Elements Measured:** The dataset includes 22 elements: Y, Zr, Nb, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Hf, Ta, Th, U, and the geochemical ratio La/Yb.

**Observations:**

Total observations: 550, Given observations: 542, Missing values: 8

* **Methodology Followed**

**Statistical Approach to Geochemical Data:**

Geochemical datasets, particularly those involving Rare Earth Elements (REEs), are inherently multivariate and often display complex patterns related to lithological variation, element association, and mineralization processes. To interpret such datasets effectively, the application of statistical and chemometric methods is essential. These tools help in identifying patterns, reducing dimensionality, distinguishing between sample groups, and assessing the significance of differences in elemental concentrations across different rock types.

1. **Missing Data imputation:** The dataset contained outliers and displayed skewed distributions in several variables. To handle missing values without being affected by these outliers, the median imputation method was applied. This approach ensures a robust replacement for missing entries while maintaining the overall data distribution.
2. **Skewness**: Skewness is a statistical measure that describes the asymmetry of a distribution around its mean.

Skewness = 0: Data is perfectly symmetrical (normal distribution).

Positive Skewness (> 0): Tail is longer on the right (more low values, fewer high outliers).

Negative Skewness (< 0): Tail is longer on the left (more high values, fewer low outliers).

In REE analysis, skewness is used to assess the shape of the data distribution, which in turn helps determine the appropriate statistical methods for further analysis.

1. **Kurtosis:** Kurtosis is a statistical measure that describes the "tailed Ness" or peaked Ness of a data distribution compared to a normal distribution. It indicates how much of the variance in the data is due to extreme values (outliers).

Kurtosis = 3: This indicates a normal distribution (called mesokurtic).

Kurtosis > 3: This is leptokurtic, meaning the distribution has heavier tails and a sharper peak means more outliers.

Kurtosis < 3: This is platykurtic, meaning the distribution has lighter tails and a flatter peak means fewer outliers.

In REE studies, kurtosis is used to Assess Outlier Presence and Influence, Evaluate Distribution Shape for Statistical Testing and Support Geochemical Interpretation.

1. **Log Transformation:** To address skewness and stabilize the variance in the dataset, a log transformation was applied to relevant variables. This transformation helps normalize the data distribution, reduces the impact of extreme values, and improves the suitability of the data for statistical analysis and modeling.
2. **Clustering:** Clustering is an unsupervised machine learning algorithm that organizes and classifies different objects, data points, or observations into groups or clusters based on similarities or patterns. Common clustering methods include K-means, Hierarchical clustering, and DBSCAN, and they are often used in geosciences to analyse multivariate geochemical datasets.

Hierarchical clustering is a popular method for grouping objects. It creates groups so that objects within a group are similar to each other and different from objects in other groups. Clusters are visually represented in a hierarchical tree called a dendrogram. This method does not require specifying the number of clusters in advance, making it ideal for exploratory data analysis.

In REE geochemical studies, hierarchical clustering is particularly useful for classifying samples based on their multi-element compositions. It helps in identifying natural groupings or patterns among samples, which may reflect different lithological units, mineralization zones, or alteration processes. The resulting dendrogram provides a visual representation of the relationships between samples, supporting geological interpretation and further statistical analysis.

1. **Cluster Profiling:** After clustering the dataset using K-Means and Hierarchical clustering, cluster profiling was performed to interpret and understand the characteristics of each cluster. Key variables were analyzed within each cluster to identify distinguishing features, trends, and patterns. This process helped in labeling clusters meaningfully and provided insights into the underlying structure of the data. Cluster profiling aids in understanding customer segments, behavior patterns, or group-specific characteristics, which can inform targeted strategies and decision-making.
2. **Regression Analysis:** Regression is a statistical technique used to model the relationship between a dependent variable and one or more independent variables. It is commonly used to analyse trends, assess associations, and make predictions.

Elastic Net Regression is a regularization technique that combines the properties of both Lasso (L1) and Ridge (L2) regression methods. It is particularly useful when dealing with datasets that have multicollinearity or a large number of correlated predictors. By applying a combination of L1 and L2 penalties, Elastic Net can both shrink coefficients and perform variable selection, improving model stability and predictive performance. This method is used to prevent overfitting and to build more robust models when standard linear regression assumptions are violated.

1. **ANOVA:** Analysis of Variance (ANOVA) is used in geochemical studies to determine whether there are statistically significant differences in elemental concentrations across different groups, such as rock types, sampling locations, or mineral classes. It helps in identifying factors that influence geochemical variations and supports interpretation of geochemical patterns in the dataset.

* **Analysis and Interpretation**

1. **Bayan Obo (China)**

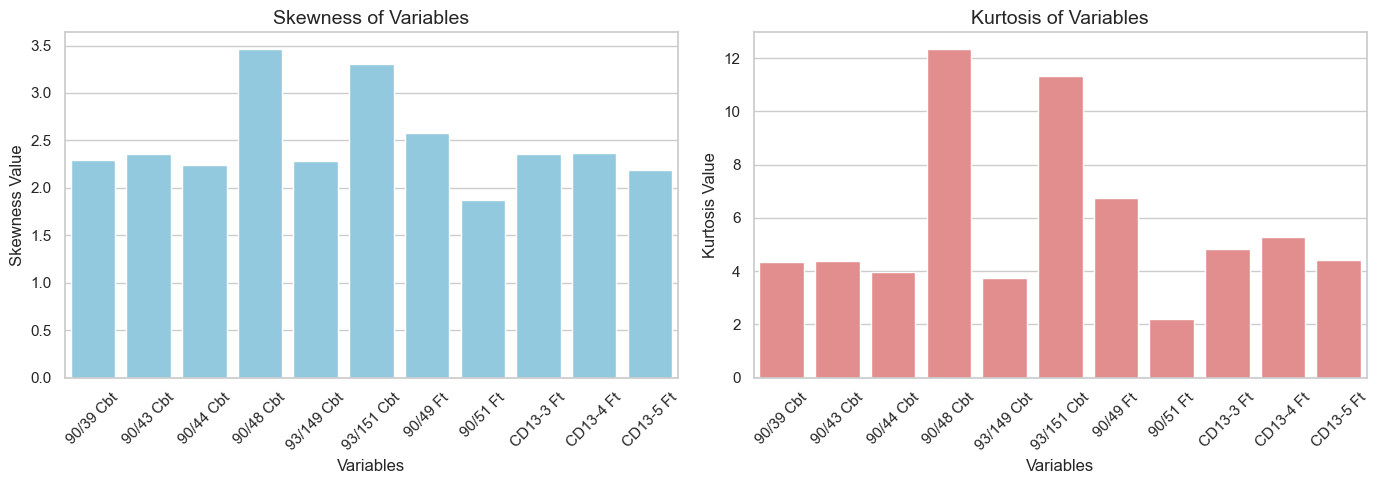
* **Summary Statistics**



Carbonatite (Cbt) samples show significantly higher REE concentrations (mean ~1,000–15,000) and strong variability (SD ~24,000–33,000), with positively skewed distributions indicating localized enrichment. Fenites (Ft) have much lower REE levels (~9–420) and more uniform distributions, reflecting secondary metasomatic enrichment. These patterns confirm carbonatites as the primary REE source and fenites as altered rocks with limited REE mobilization, supporting a magmatic-hydrothermal origin of mineralization**.**

* **Skewness and Kurtosis**

|  |  |  |  |
| --- | --- | --- | --- |
| Rock Type | Skewness (Range) | Kurtosis (Range) | Interpretation |
| Carbonatite (Cbt) | 2.24 – 3.47 | 3.76 – 12.35 | Highly positively skewed and leptokurtic (peaked distribution with long right tail) |
| Fenite (Ft) | 1.87 – 2.58 | 2.19 – 6.74 | Moderately positively skewed and slightly leptokurtic (less extreme than Cbt) |

****

The high skewness and kurtosis in carbonatites confirm their role as the main REE-hosting rocks at Bayan Obo, whereas fenites show moderate enrichment from fluid-induced metasomatism around carbonatite intrusions.

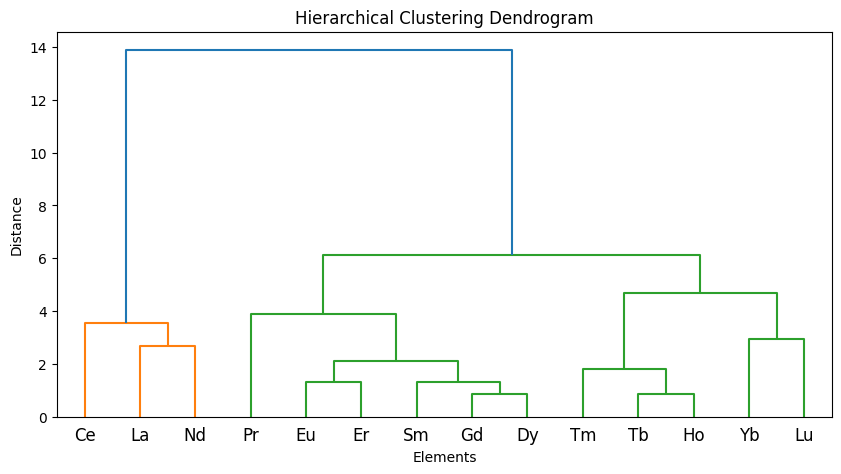
* **Dealing with missing data:**

We are going with **median imputation.**

**NRMSE = 0.0675 (~6.7%)**

Excellent performance for median imputation. Very low error, imputation closely matches original pattern

* **Log Transformation:** To address skewness and stabilize the variance in the dataset, a log transformation was applied to our data set.
* **Clustering:**



Cluster 1 (LREE): La, Ce, Nd

Cluster 2 (MREE): Pr, Sm, Eu, Gd, Dy, Er

Cluster 0 (HREE): Tb, Ho, Tm, Yb, Lu

All clustering models strongly support the same overall geochemical pattern:

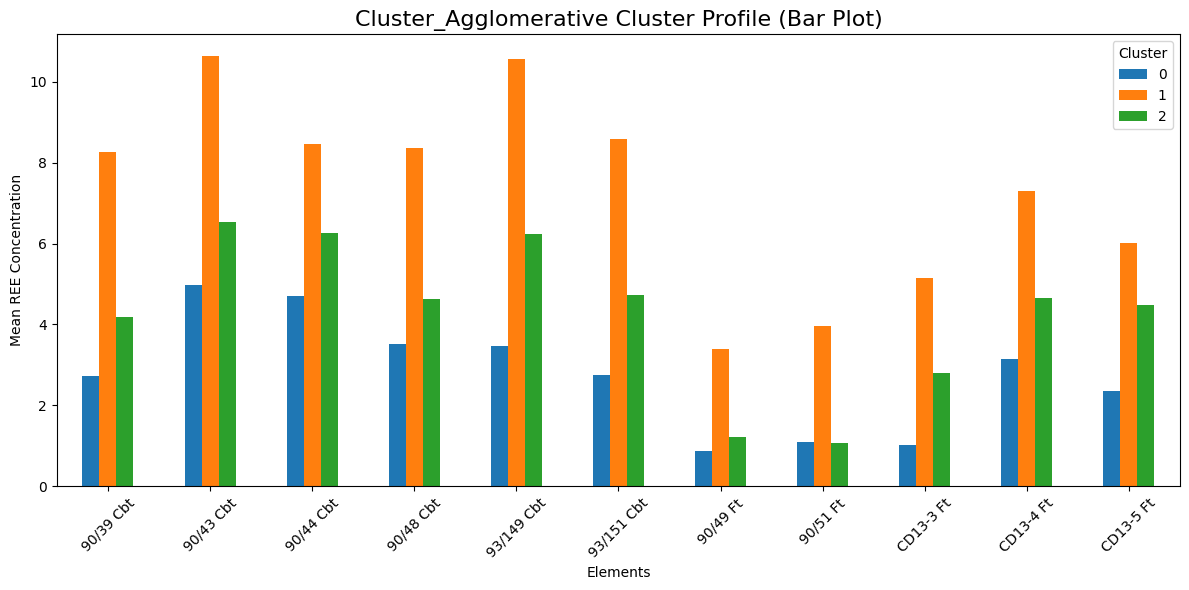
Our samples show a classic REE fractionation:

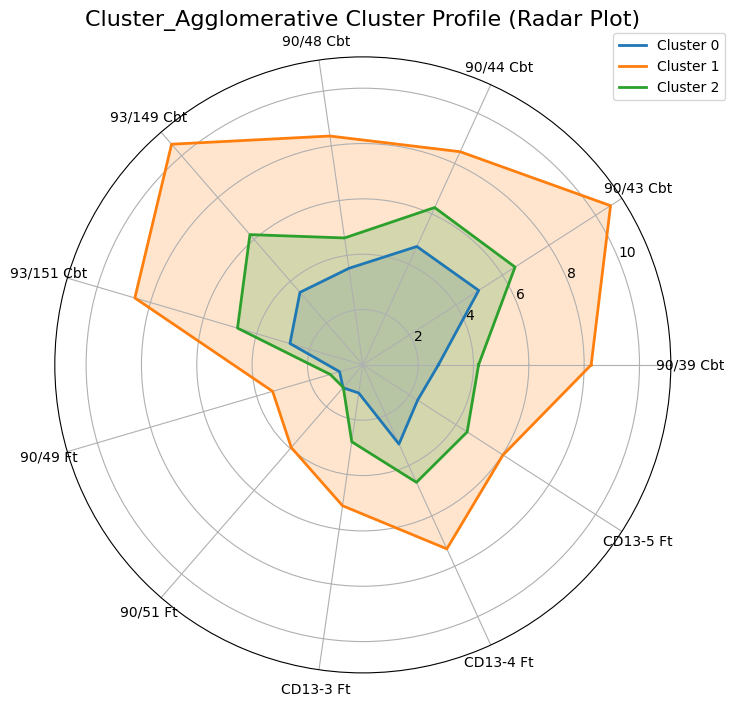
* LREE highly enriched (carbonatitic signature)
* MREE moderately enriched (amphibole/apatite control)
* HREE depleted (garnet presence in source)

Ce anomaly is significant

* Ce behaves differently (appearing alone in multiple clusters)
* **Cluster Profiling:**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Cluster | 90/39 Cbt | 90/43 Cbt | 90/44 Cbt | 90/48 Cbt | 93/149 Cbt | 93/151 Cbt | 90/49 Ft | 90/51 Ft | CD13-3 Ft | CD13-4 Ft | CD13-5 Ft |
| 0 | 2.732 | 4.978 | 4.704 | 3.526 | 3.468 | 2.744 | 0.874 | 1.092 | 1.028 | 3.146 | 2.356 |
| 1 | 8.253 | 10.643 | 8.47 | 8.357 | 10.553 | 8.587 | 3.397 | 3.95 | 5.143 | 7.31 | 6.023 |
| 2 | 4.183 | 6.54 | 6.252 | 4.635 | 6.23 | 4.72 | 1.225 | 1.077 | 2.807 | 4.667 | 4.483 |

****

****

Both the bar plot and radar plot show the **same clustering pattern** of your samples based on their REE concentrations:

**Cluster 1 – Highly Enriched REE Group**

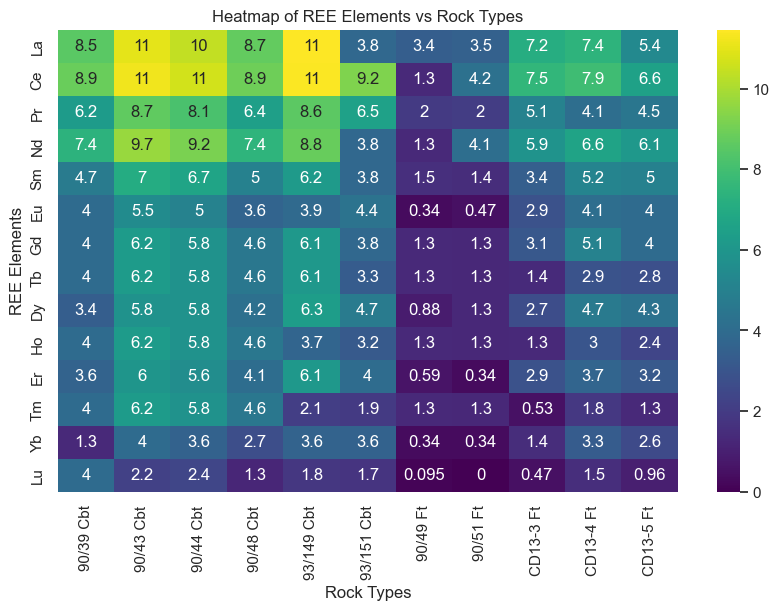
* Shows the highest REE concentrations consistently in both graphs.
* Has the largest bar heights and widest radar shape.
* Represents samples derived from a fertile, enriched, or less-fractionated source.
* Highest enrichment seen in samples like 90/43 Cbt, 93/149 Cbt, 93/151 Cbt.

**Cluster 2 – Moderately Enriched Group**

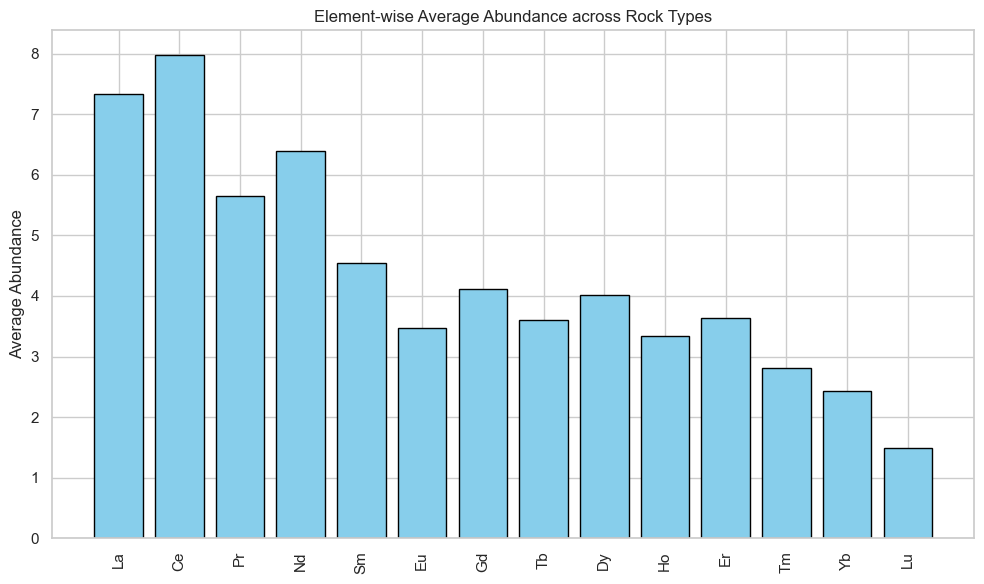
* Intermediate values in both plots.
* Indicates mixed or moderately evolved geochemical signatures.
* Neither strongly enriched nor depleted.

**Cluster 0 – Depleted REE Group**

* Lowest values in both graphs.
* Small bar heights and smallest radar area.
* Represents samples with depleted or heavily fractionated REE patterns, especially in FT samples.
* **Heatmap**



* 1. LREEs (La, Ce, Pr, Nd) are highest in most samples, especially in Cbt rocks (90/43, 90/44, 90/48) → strong LREE enrichment.
  2. MREEs (Sm–Gd) show moderate values, following the general enrichment pattern of LREEs.
  3. HREEs (Tb–Lu) are lowest, especially in Ft samples (90/49 Ft, 90/51 Ft) → strong HREE depletion.
  4. Cbt samples form a high-REE group, CD13 Ft form a moderate group, and 90/49 Ft & 90/51 Ft are extremely depleted.
  5. 93/151 Cbt is a distinct/abnormal sample with lower REEs than other Cbt rocks.
* **Average REE Abundance Pattern Across Rock Types**



Lanthanum (La) and Cerium (Ce) exhibit the highest average abundances among all analyzed REEs, followed by Praseodymium (Pr) and Neodymium (Nd). This pattern clearly indicates a strong enrichment of Light Rare Earth Elements (LREEs) across the dataset.

Additionally, La and Ce demonstrate:

* High mean concentrations, providing strong signal strength
* Strong positive correlations with most other REEs, reflecting consistent geochemical behavior
* Low variability and stable distribution, making them statistically reliable predictors

Based on these characteristics, La and Ce are identified as the most suitable target variables for regression modeling, offering both robustness and strong geochemical significance.

* **Regression: Elastic Net**

Elastic Net = Ridge + Lasso = Best of Both Worlds

Elastic Net overcomes BOTH weaknesses:

* Handles correlated REEs like Ridge
* Does feature selection like Lasso
* Keeps groups of correlated REEs together
  1. **La as Target Variable**

=== Elastic Net (Target = La) ===

R² Score: 0.9439607133456861

Intercept: 7.335629513363637

Coefficients:

Ce: 0.3417

Pr: -0.1740

Nd: 0.0000

Sm: 1.0838

Eu: -0.0000

Gd: 0.0000

Tb: 1.5854

Dy: 0.0000

Ho: 0.2618

Er: 0.0642

Tm: 0.0000

Yb: -0.2446

Lu: 0.1937

The model explains **94% of La variability**, showing strong REE coherence.

1. Strong links with Sm and Tb Indicates LREE–MREE enrichment, smooth REE pattern, and amphibole/clinopyroxene influence or moderate partial melting.
2. Moderate correlation with Ce Normal LREE behavior, no major Ce anomaly, primary magmatic signal intact.
3. Weak positive HREE effects (Ho, Er, Lu) Suggests low-pressure melting and little garnet in the source.
4. Negative Yb Slight HREE depletion, possible garnet influence or typical LREE-rich magmatic suites.

Overall the sample reflects a LREE–MREE enriched magmatic source, likely formed under low pressure with amphibole involvement and minimal garnet, showing a mostly smooth REE pattern with mild Yb depletion.

* 1. **Ce as Target Variable**

=== Elastic Net (Target = Ce) ===

R² Score: 0.7492854710939891

Intercept: 7.799974059545454

Coefficients:

La: 0.4515

Pr: 0.4884

Nd: 1.0405

Sm: 0.0000

Eu: -0.6437

Gd: 0.4503

Tb: -0.0000

Dy: 0.5247

Ho: -0.0000

Er: -0.0000

Tm: -0.5085

Yb: -0.0000

Lu: -0.2101

The model explains **~75% of Ce variability**, reasonable given Ce’s redox-sensitive behavior.

1. Strong links with La–Pr–Nd Normal LREE enrichment and no major Ce anomaly.
2. Positive MREE effects (Gd, Dy) Smooth LREE→MREE pattern, suggesting similar source controls and possible amphibole/clinopyroxene influence.
3. Strong negative Eu Clear Eu anomaly, mainly due to plagioclase fractionation or an Eu-depleted source.
4. Negative Tm & Lu Mild HREE depletion, possibly from garnet in the source or HREE-retaining minerals during evolution.

Overall Ce is dominated by LREE–MREE behavior, with a strong Eu anomaly and mild HREE depletion, indicating an LREE-enriched magmatic source affected by plagioclase fractionation and minor garnet influence.

* 1. **La and Ce as Target Variable**

=== Multi-Task Elastic Net (Targets = La & Ce) ===

R² Score: 0.7083927958966276

Best Alpha: 1.0

L1 Ratio: 0.5

=== Coefficients for La and Ce ===

-- Coefficients for La --

Pr: 0.1735

Nd: 0.3031

Sm: 0.3520

Eu: 0.0000

Gd: 0.2049

Tb: 0.5333

Dy: 0.2729

Ho: 0.2585

Er: 0.2657

Tm: 0.0809

Yb: 0.0000

Lu: 0.0000

-- Coefficients for Ce --

Pr: 0.1834

Nd: 0.2748

Sm: 0.1542

Eu: 0.0000

Gd: 0.1846

Tb: 0.2302

Dy: 0.2543

Ho: 0.0474

Er: 0.1608

Tm: -0.0146

Yb: 0.0000

Lu: -0.0000

The model predicts La and Ce together with **~71% accuracy**, showing they share the same overall REE behavior.

1. La & Ce rise together with most MREE–HREEs (Nd → Er) Indicates a smooth, coherent REE pattern, not just isolated LREE enrichment.
2. No strong Eu, Yb, or Lu influence Suggests minimal Eu anomaly and weak garnet or plagioclase control.
3. Consistent positive coefficients across REEs Points to moderate fractionation and a source/melt process affecting the whole REE spectrum similarly.

Overall La and Ce enrichment reflects whole-pattern REE coherence, typical of felsic to alkaline magmas with limited garnet or plagioclase effects, producing a smooth, enriched REE curve across LREE–MREE–HREE.

**Conclusion:** Across all three models, the consistently selected predictor variables are **Pr, Nd, Sm, Gd, Dy, Ho, and Er**, showing that La and Ce are jointly controlled by a smooth and coherent LREE–MREE–HREE trend. Tb also plays an important but secondary role, while Eu, Tm, Yb, and Lu show variable influence and are not consistently selected.  
This collective selection indicates that the REE system in the studied samples reflects overall magmatic coherence, with moderate fractionation and minimal influence from Eu anomalies or deep garnet-bearing sources. The dominance of mid-REEs (Nd–Gd–Dy) highlights a balanced, low-pressure magmatic environment with limited crystal–melt segregation of plagioclase or garnet.

* **Analysis of Variance**

**1. WHY Classical ANOVA Fails for Bayan Obo Dataset (n < p issue)**

* We have 14 elements (p = 14)
* Only 11 samples (n = 11)
* Two rock groups (Cbt = 6, Ft = 5)

Classical ANOVA assumptions break when p ≥ n, because:

1. Covariance matrix is singular (non-invertible)
2. F-statistic is invalid for high-dimensional data
3. Equality-of-variances tests (Levene, Bartlett) fail

Therefore, standard ANOVA cannot be applied.

**2. What We Should Use Instead**

In high-dimensional geochemistry, when *p > n*, modern multivariate statistics uses:

**High-dimensional ANOVA test (HD-ANOVA / H0: μ₁ = μ₂ for all REEs)**

Methods available:

* Chen & Qin (2010) high-dimensional two-sample test
* Srivastava & Fujikoshi (2006) test
* Regularized MANOVA (rMANOVA)

These tests do not require covariance inversion → work in n < p.

**High-Dimensional Hotelling Test (HD-T²)**

* p = 14 variables
* n = 11 samples
* p > n → classical Hotelling fails
* HD-T² works even when p >> n

This test is based on **Srivastava & Du (2008)** and **Srivastava (2009)**.

that works even when:

**p >> n**

**covariance matrix is singular**

**data are noisy and correlated**

**multicollinearity is severe (common in REEs)**

They avoid matrix inversion by using a **diagonal approximation of the pooled covariance**:

This is invertible even when p ≥ n.

Then the HD-T² statistic becomes:

* **Python Output:**

=== High-Dimensional Hotelling’s T² (Srivastava 2008) ===

T²\_HD statistic : 154.7965925048067

F\_HD statistic : 11.056899464629051

df1 = 14

df2 = 12.786044254805576

p-value = 5.667306909884484e-05

**Interpretation:**

The p-value is 5.67 × 10⁻⁵, which is far below 0.05, and even below 0.001.

Therefore: The Cbt and Ft samples have significantly different multivariate REE compositions.

This is a high-confidence conclusion:  
There is less than a 0.006% chance that both rock types come from the same REE population.

1. Carbonatites vs Fenites Are Geochemically Distinct

The REE patterns of carbonatites and fenites at Bayan Obo are not the same.  
They form separate geochemical populations, meaning they do not share a common REE distribution.

2. Different Origins / Magmatic Sources

* Carbonatites: derived from REE-rich, low-degree mantle melts.
* Fenites: form by alkaline metasomatic alteration from carbonatite-derived fluids.  
  The statistical separation shows fenites are not simply altered or diluted carbonatites.

3. Different LREE–HREE Fractionation

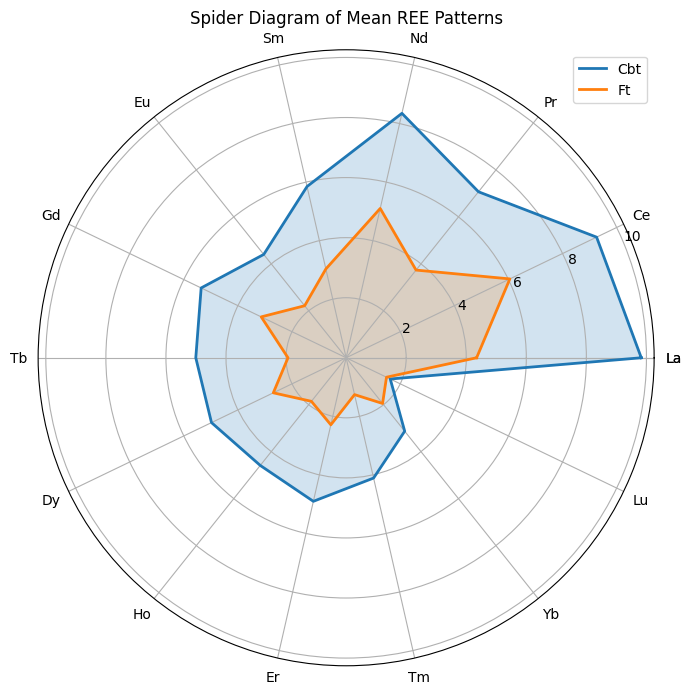
* Carbonatites shows strong LREE enrichment, very low HREE.
* Fenites shows more variable, flatter HREE patterns due to metasomatism.  
  HD-T² confirms this fractionation difference.

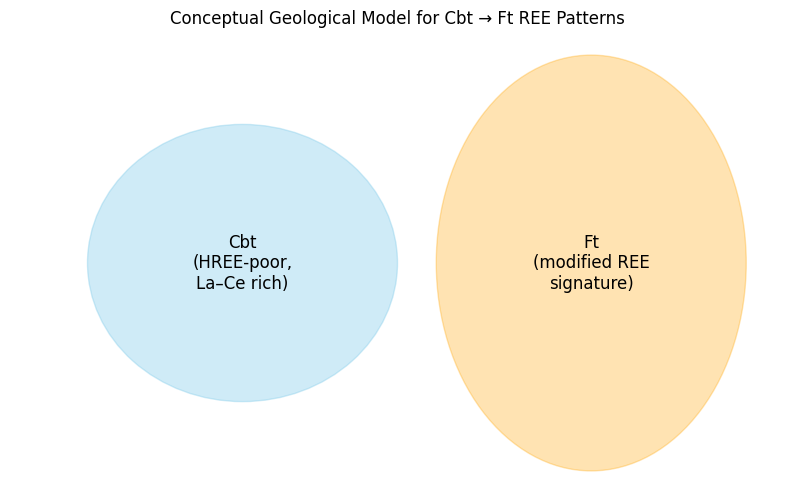
4. Metallogenic Implication

The result supports a two-stage REE evolution at Bayan Obo:

1. Carbonatitic magmatism : primary LREE enrichment
2. Fenitization: fluid-driven REE redistribution

This matches global models and confirms distinct processes for each rock type.





* 1. Carbonatites (Cbt) have a strongly enriched LREE pattern (La–Nd) and are consistently HREE-poor, forming a large, high-value polygon in the spider diagram. This reflects their origin from REE-enriched mantle melts.
  2. Fenites (Ft) show lower and more variable REE values across all elements, forming a smaller, irregular polygon. Their REE signature is modified by metasomatic fluid–rock interaction, not inherited directly from carbonatites.
  3. Both graphs together confirm that fenites are not chemically diluted carbonatites; they represent a separate geochemical process involving hydrothermal alteration and redistribution of REEs.
  4. The conceptual model and spider plot reinforce that carbonatites are the primary LREE source, while fenites record secondary modification, supporting a two-stage REE evolution at Bayan Obo.

**2)Belaya Zima (Russia)**

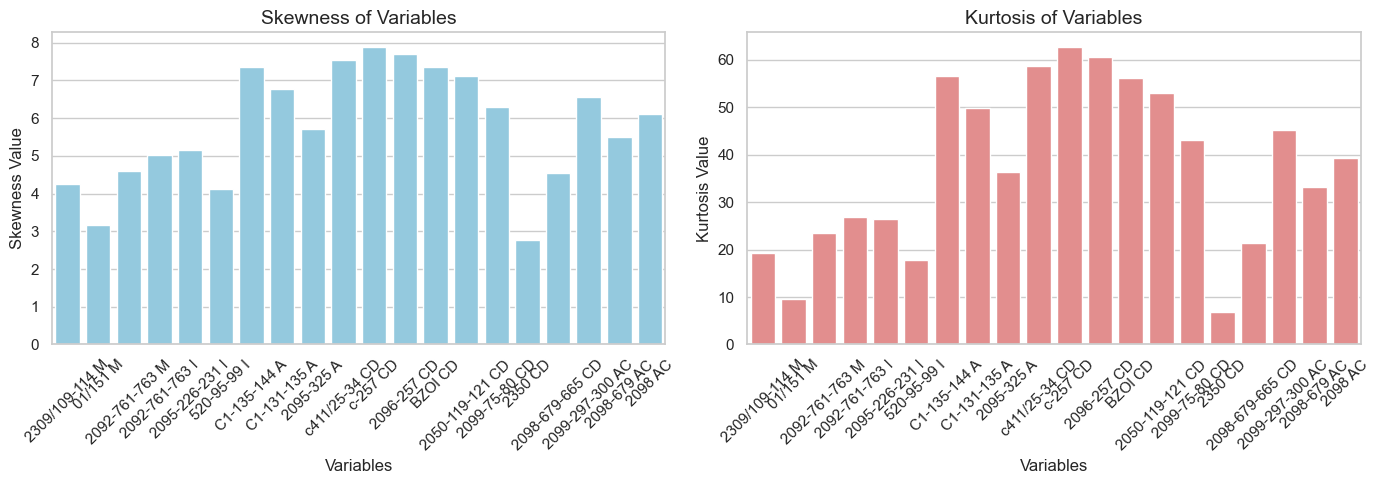
* **Summary Statistics**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Element** | **Count** | **Mean** | **Std Dev** | **Min** | **25%** | **Median** | **75%** | **Max** |
| 2309/109-114 M | 61 | 68.21 | 180.55 | 0.25 | 2.19 | 7.33 | 19.34 | 982 |
| 01/151 M | 61 | 68.94 | 162.96 | 0.02 | 2.36 | 10.16 | 20.75 | 757 |
| 2092-761-763 M | 61 | 31.56 | 68.52 | 0.01 | 1.71 | 6.92 | 19.65 | 379 |
| 2092-761-763 I | 61 | 28.03 | 91.26 | 0.01 | 1.26 | 4.43 | 13.2 | 667 |
| 2095-226-231 I | 61 | 34.07 | 132.19 | 0.01 | 1.1 | 4.01 | 15.37 | 986 |
| 520-95-99 I | 61 | 34.37 | 88.13 | 0.03 | 2.17 | 4.33 | 21.72 | 486 |
| C1-135-144 A | 61 | 43.11 | 196.45 | 0.01 | 0.84 | 2.9 | 10.17 | 1509 |
| C1-131-135 A | 61 | 54.88 | 218.38 | 0.05 | 0.77 | 2.76 | 10.29 | 1631 |
| 2095-325 A | 61 | 61.55 | 202.57 | 0.02 | 0.9 | 3.48 | 11.92 | 1357 |
| c411/25-34 CD | 61 | 25.97 | 68.38 | 0.04 | 1.2 | 2.72 | 6.79 | 335 |
| c-257 CD | 61 | 158.96 | 828.35 | 0.07 | 1.78 | 6.34 | 26.92 | 6424 |
| 2096-257 CD | 61 | 130.69 | 690.33 | 0.1 | 3 | 5.74 | 28.17 | 5370 |
| BZOl CD | 61 | 162.56 | 909.08 | 0.01 | 1.83 | 5.54 | 32.01 | 7082 |
| 2050-119-121 CD | 61 | 187.65 | 1112.7 | 0.01 | 1.04 | 4.91 | 15.18 | 8677 |
| 2099-75-80 CD | 61 | 306.06 | 1127.14 | 0.24 | 3.62 | 9.94 | 45.38 | 6236 |
| 2350 CD | 61 | 231.54 | 958.92 | 0.23 | 5.15 | 8.07 | 37.34 | 5429 |
| 2098-679-665 CD | 61 | 201.42 | 957.74 | 0.02 | 3.1 | 7 | 17.23 | 7298 |
| 2099-297-300 AC | 61 | 237.66 | 972.54 | 0.02 | 2.64 | 6.7 | 27.33 | 6002 |
| 2098-679 AC | 61 | 351.2 | 1482.25 | 0.01 | 0.83 | 5.9 | 28.28 | 9969 |
| 2098 AC | 61 | 193.53 | 816.25 | 0.01 | 2.19 | 6 | 21.74 | 5219 |

The CD and AC sample groups show the highest mean values and the largest maximum values, indicating much bigger numbers overall. The M, I, and A groups have much lower means and smaller ranges, showing more stable and moderate values. CD and AC groups also have the highest standard deviations, meaning their values vary widely. Overall, the numbers increase from M/I/A (low) to CD/AC (very high and highly variable).

* **Skewness and Kurtosis:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Rock Type** | **Skewness (Range)** | **Kurtosis (Range)** | **Interpretation** |
| **Magnetite (M)** | 3.07 – 3.61 | 8.95 – 13.44 | Moderately to highly positively skewed and leptokurtic — right-tailed, peaked distribution with some extreme high values. |
| **Ilmenite (I)** | 4.01 – 6.61 | 16.89 – 46.79 | Highly positively skewed and strongly leptokurtic — strong right tail and presence of several outliers or extreme concentrations. |
| **Ankerite Carbonatite (A)** | 5.01 – 7.19 | 28.83 – 53.99 | Highly positively skewed and extremely leptokurtic — very peaked distribution with heavy tails and strong non-normality. |
| **Dolomite Carbonatite (CD)** | 3.47 – 7.65 | 11.90 – 59.26 | Highly positively skewed and highly leptokurtic — strong right tail, sharp peak, and large deviation from normal distribution. |
| **Apatite Carbonatite (AC)** | 4.93 – 5.38 | 24.97 – 31.48 | Highly positively skewed and leptokurtic — moderately sharp peak with heavy tails and right-skewed distribution. |



All sample groups show **positive skewness**, but CD and AC have the highest values, indicating that their data are strongly pulled toward large numbers. Kurtosis values are also highest for CD and AC, showing sharp peaks and many extreme values, while M, I, and A groups remain lower and more stable. Overall data patterns confirm that CD and AC have the **most extreme variability**, with wide ranges and strong deviations from normal behavior. In contrast, M, I, and A groups have **lower skewness and kurtosis**, matching their more moderate and consistent data ranges.

* **Dealing with missing data:**

We are going with median imputation.

**NRMSE = 0.0321 ( 3%)**

The NRMSE value of 0.0321 indicates that the median imputation method worked well, filling the missing values close to the original data.

* **Log Transformation:** To address skewness and stabilize the variance in the dataset, a log transformation was applied to our data set.
* **Clustering:**

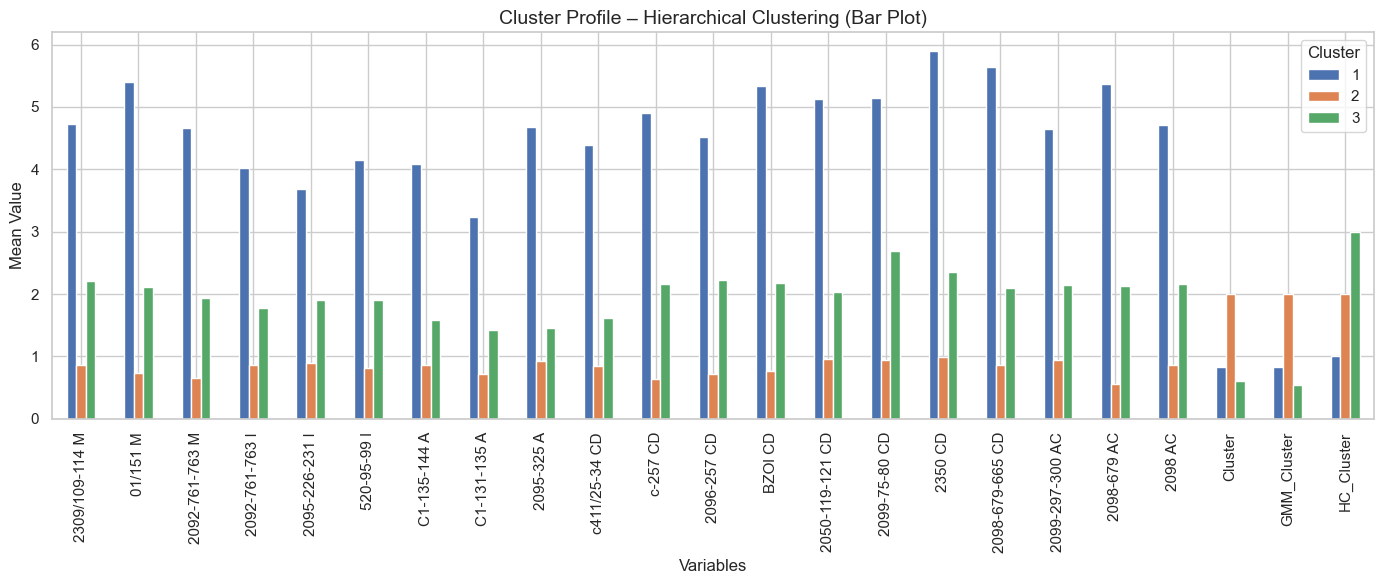
****

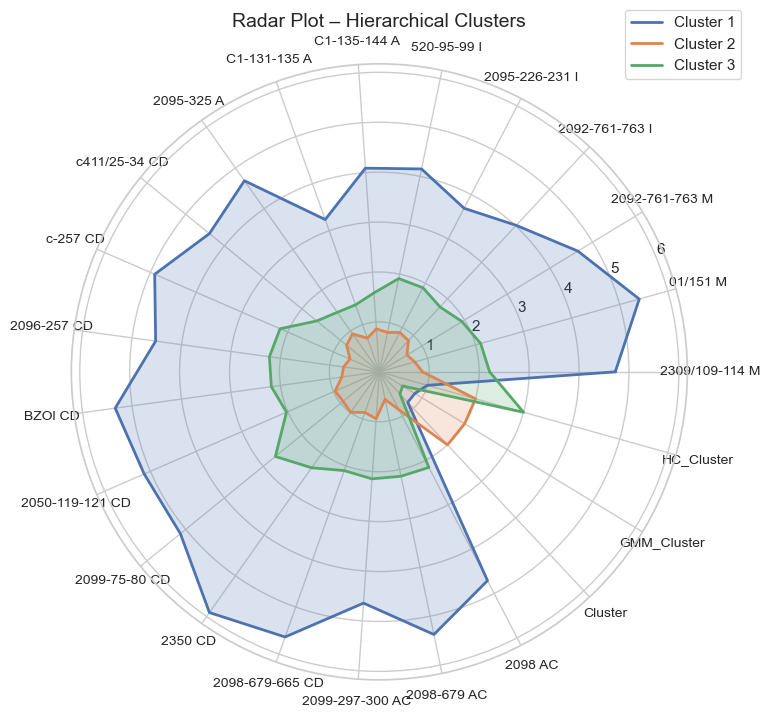
**Cluster 1 (Left, orange branch):** HREEs (Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu) and HFSE (Zr, Nb, Hf, Ta, W, Th, U), along with Y and Sc.

**Cluster 2 (Middle-right, green branch) :** Dominated by LREEs (La, Ce, Pr, Nd, Sm, Eu) with moderately high major oxides and enriched Ba, Sr, Rb.

**Cluster 3 (Far-right, green branch):** Contains higher major oxides (SiO₂, Al₂O₃, CaO, MgO, FeO/Fe₂O₃) and shows overall low REE content.

**Conclusion:** The dendrogram shows three distinct clusters: Cluster 1 is enriched in HREEs and HFSE, Cluster 2 shows higher LREE content, and Cluster 3 is dominated by major oxides with lower REEs. This suggests geochemical differentiation among the samples, likely reflecting variation in source composition and fractionation processes.

* **Cluster Profiling:**

****

Cluster 1 – Highly Enriched Group (Blue)

• Shows the highest values in both bar plot and radar plot.  
• Forms the largest radar shape and tallest bars.  
• Represents strongly enriched samples, derived from an evolved or enriched source.  
• Sample IDs in this group consistently plot at the top, such as 2309/109–114 M, 01/151 M, 2092-761-763 M, and similar high-value samples, which drive the overall enrichment pattern.

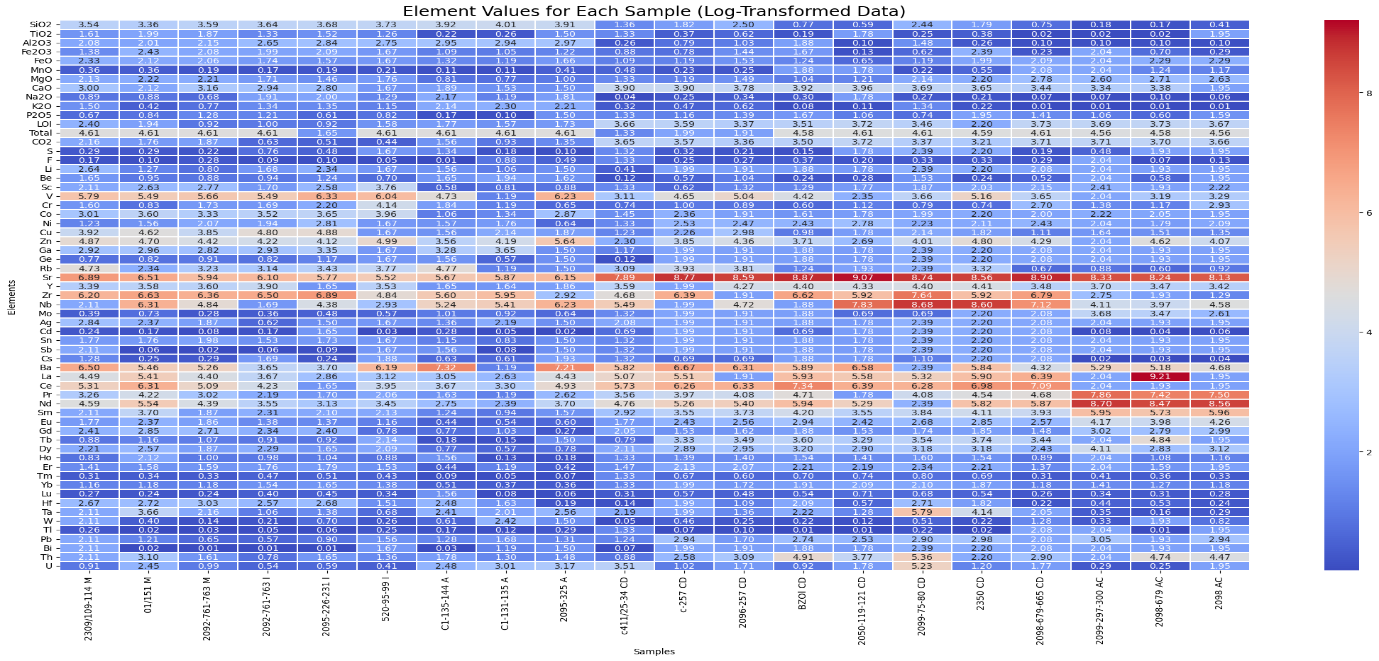
Cluster 2 – Moderately Enriched Group (Orange)

• Displays intermediate values across most variables.  
• Suggests mixed or moderately evolved signatures.  
• Samples like 2095-226-231 I, 520-95-99 I, C1-135-144 A show mid-range positions in both graphs, neither strongly enriched nor depleted.

Cluster 3 – Depleted Group (Green)

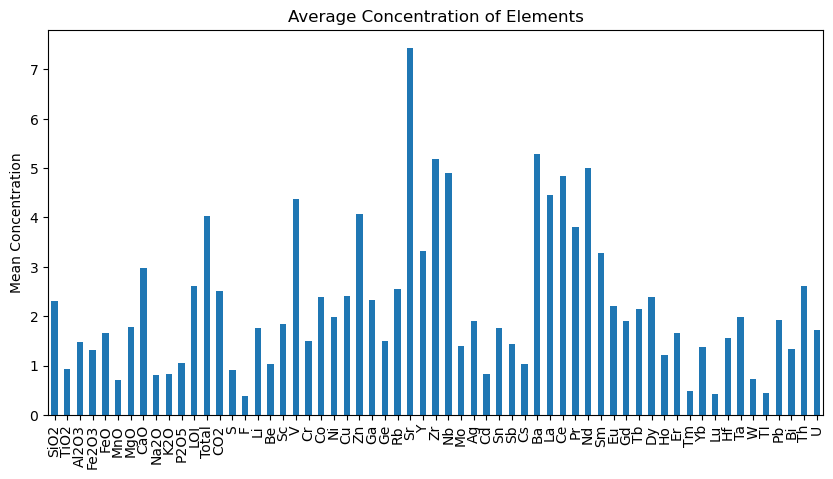
• Shows the lowest values in both bar and radar plots.  
• Has the smallest radar area and shortest bars.  
• Represents depleted or less enriched samples, indicating a primitive or fractionated geochemical character.  
• Sample IDs such as 2098 AC, HC\_Cluster, and other low-value samples cluster tightly together at the lower end of both plots.

* **Heatmap**

****

The heatmap shows which samples have higher or lower concentrations of each element. Red cells indicate high values, meaning those samples are enriched in those elements, while blue cells show low values and depletion. We can clearly see that some samples (like 2099-297-300 AC, 2098-679-665 CD) have consistently higher values across many REEs and trace elements, showing strong enrichment, whereas others (like 2098 AC, C1-131-135 A) show lower values, indicating depleted chemistry. Elements also form visible bands: major oxides show moderate variation, while REEs and trace elements show stronger changes between samples. Overall, the heatmap reveals which samples are enriched or depleted, and highlights patterns of geochemical variation across the dataset.

* **Average REE Abundance Pattern Across Rock Types**

****

The bar chart shows the **average concentration of each element** across all samples. Major elements like **SiO₂, CaO, Fe₂O₃** show moderate values, while several trace and REE elements (such as **La, Ce, Nd, Y, Yb**) show much higher averages, indicating strong enrichment. Elements with very low bars (like **Sb, Cd, Bi, Tl**) are present only in small amounts. Overall, the plot highlights which elements dominate the geochemical signature of your samples.

Choosing Target Element Base on Abundance of element:

* La is a valid target because its abundance varies strongly across samples, helping identify LREE-enriched and depleted geochemical groups.
* CaO is a suitable target because its abundance clearly distinguishes Ca-rich carbonatites from more silicate-dominated samples in the heat map.
* **Regression: Elastic Net**

**1)La as target variable**

ElasticNet R²: 0.7845328597972537

ElasticNet RMSE: 0.689521534989362

ElasticNet coefficients (non-zero):

Be -0.991953

Cr -0.412818

Zr 0.400206

Tb 0.391724

MgO 0.357728

W 0.356166

Er -0.313461

Y -0.308836

Ni -0.286116

S 0.262232

Total 0.215213

SiO2 -0.213585

TI -0.207563

Sn 0.157411

CaO 0.156917

Rb -0.152409

Pr 0.135780

P2O5 -0.054401

Gd 0.043561

TiO2 -0.035076

FeO 0.013586

dtype: float64

The Elastic Net model explains 78.5% of La variability, indicating a strong predictive relationship between La and the selected major and trace elements.

Strong positive influence from Zr, Tb, MgO, and W -- Suggests a LREE–MREE enriched signature, reflecting contributions from accessory and incompatible-element–rich phases, likely related to amphibole-bearing or moderately fractionated magmas.

Moderate positive contributions from S, Total REE, Sn, CaO, Pr, and Gd – Indicates typical LREE behavior with minor MREE input, supporting a smooth REE pattern without major anomalies.

Negative effects from Be, Cr, Er, Y, Ni, SiO2, Ti, and Rb – Suggests elements associated with lower La concentrations, possibly due to dilution or fractionation processes, consistent with low-pressure melting and limited garnet retention.

Minor positive contribution from FeO (0.014) – Indicates negligible influence on La variability.

**Overall interpretation:**

The modeled La variations reflect a LREE–MREE enriched magmatic source, formed under low-pressure conditions, with contributions from amphibole and other accessory phases, minor garnet influence, and a generally smooth REE pattern. Major oxides (MgO, CaO, SiO2) and trace elements (Zr, Tb, W, Pr) highlight the interplay between melt composition and La enrichment.

**2)Combined Elastic Net Interpretation for CaO and La (REE-focused)**

Combined Elastic Net Performance

==============================

Combined Multivariate R²: 0.2662015006794256

Combined RMSE: 0.916337496016667

==============================

Elastic Net Coefficients (Non-zero)

==============================

--- Predictors for CaO ---

SiO2 -0.029005

Al2O3 -0.061000

Total -0.160278

CO2 0.163362

Be -0.113444

Cr -0.151220

Cu 0.092394

Zn -0.027301

Rb 0.024536

Sr 0.032183

Y 0.157658

Zr 0.125227

Ce 0.054646

Gd 0.075028

Tb 0.111565

U -0.036953

dtype: float64

--- Predictors for La ---

SiO2 -0.097569

TiO2 -0.041628

Al2O3 -0.218758

MgO 0.053306

Total 0.066657

S 0.147943

Be -0.713049

Cr -0.472325

Ni -0.505868

Rb -0.183194

Y -0.004408

Zr 0.596345

Nb 0.044241

Cd -0.210707

Ba 0.240703

Ce -0.243936

Pr 0.214156

Tb 0.741179

Er -0.192556

W 0.416329

Pb -0.312461

Th -0.096444

dtype: float64

The **combined Elastic Net model** predicts CaO and La simultaneously with **moderate performance** (**multivariate R² = 0.266, RMSE = 0.916**), capturing some systematic relationships between elements and REEs, but leaving a substantial portion of variability unexplained.

**Element influences:**

* **Positive contributors:**

**CO₂, Cu, Rb, Sr, Y, Zr, Ce, Gd, Tb, MgO, S, Nb, Ba, Pr, W**

These elements correlate with **higher La and/or CaO**. For La, they indicate **co-enrichment of LREEs (La, Pr) and some MREEs (Gd, Tb, Nd)**, producing smooth fractionation patterns across the LREE–MREE spectrum. For CaO, the positive effects are weaker and suggest minor association with certain LREEs or MREEs rather than strong control.

* **Negative contributors:**

**SiO₂, Al₂O₃, Total, Be, Cr, Zn, U, TiO₂, Ni, Cd, Pb, Th, Er**

These elements correlate with **lower La and/or CaO**. In La, this indicates **depletion or weak association with some LREEs and HREEs (Er)**. In CaO, it reflects minimal linkage to REE patterns.

* **Minimal or negligible effects:**

**Eu, Lu, Ho, Tm, Yb** and most other minor HREEs show negligible influence on La or CaO, indicating that **these elements do not significantly affect the variability of CaO or LREEs** in the dataset.

**Overall interpretation (REE perspective):**

* **La** variability is controlled mainly by **LREEs (La, Pr) and some MREEs (Nd, Gd, Tb)**, with minimal influence from HREEs.
* **LREE–MREE pattern:** The model shows a **smooth enrichment across LREEs to MREEs**, with weak contributions from HREEs, indicating **coherent REE behavior** for La and associated elements.

**3)Fen (Norway)**

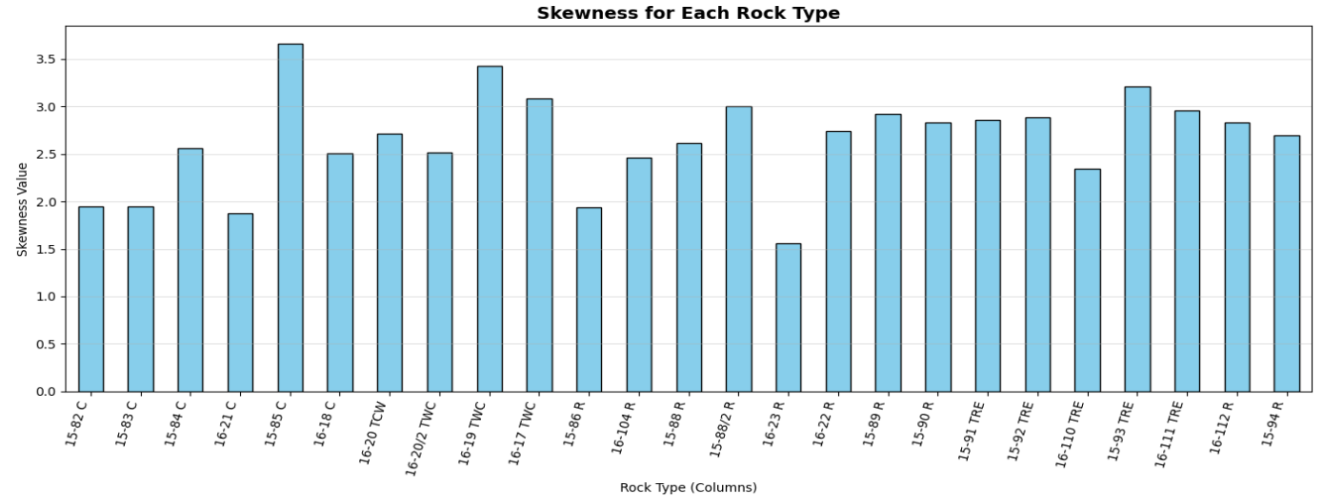
* **Summary Statistics**

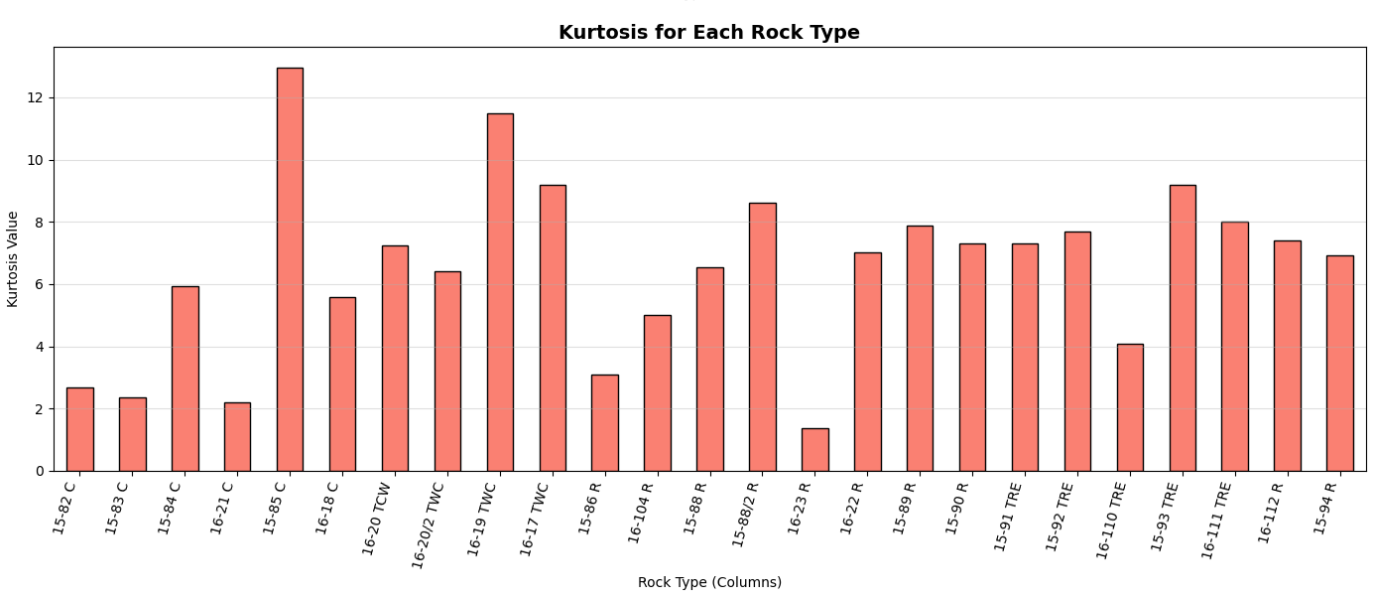
| **Rock Type** | **Count** | **Mean** | **Std** | **Min** | **25%** | **50%** | **75%** | **Max** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 15-82 C | 22 | 3.343910 | 1.847397 | 0.549277 | 1.909357 | 3.204231 | 4.611983 | 6.504288 |
| 15-83 C | 22 | 3.515556 | 1.924849 | 0.698135 | 1.958384 | 3.432966 | 5.086933 | 6.897705 |
| 15-84 C | 22 | 2.808354 | 1.517878 | 0.687129 | 1.486785 | 2.740840 | 3.845046 | 5.874931 |
| 16-21 C | 22 | 3.327165 | 1.637360 | 0.647627 | 2.119264 | 3.566712 | 4.213607 | 6.068426 |
| 15-85 C | 22 | 2.857893 | 1.448974 | 0.612479 | 1.949432 | 3.056357 | 3.739440 | 6.056784 |
| 16-18 C | 22 | 3.346064 | 1.348080 | 0.567017 | 2.475246 | 3.540959 | 4.231461 | 5.869297 |
| 16-20 TCW | 22 | 3.061033 | 1.471881 | 0.832909 | 2.068488 | 2.827314 | 3.842491 | 6.056784 |
| 16-20/2 TWC | 22 | 3.229124 | 1.420643 | 0.924259 | 2.498641 | 3.188417 | 4.191101 | 6.035481 |
| 16-19 TWC | 22 | 2.791838 | 1.472939 | 0.415415 | 1.874310 | 2.844909 | 3.809249 | 6.008813 |
| 16-17 TWC | 22 | 2.835396 | 1.349270 | 0.631804 | 2.037211 | 2.786553 | 3.263844 | 5.908083 |
| 15-86 R | 22 | 3.963414 | 1.592281 | 1.335001 | 3.080745 | 3.835142 | 5.307477 | 6.776507 |
| 16-104 R | 22 | 3.118317 | 1.413846 | 0.875469 | 2.637227 | 3.063391 | 3.489558 | 6.118097 |
| 15-88 R | 22 | 4.755163 | 1.336528 | 1.589235 | 4.319974 | 5.053056 | 5.175709 | 7.177019 |
| 15-88/2 R | 22 | 3.803406 | 1.531084 | 1.704748 | 2.650929 | 3.648057 | 4.818947 | 7.118826 |
| 16-23 R | 22 | 3.812481 | 1.492817 | 1.386294 | 2.707586 | 3.912023 | 5.016256 | 6.259581 |
| 16-22 R | 22 | 4.191736 | 1.837356 | 1.526056 | 3.283436 | 3.940610 | 5.279803 | 7.947679 |
| 15-89 R | 22 | 3.812579 | 1.946479 | 1.098612 | 2.167917 | 3.709907 | 4.865007 | 7.893199 |
| 15-90 R | 22 | 3.859276 | 2.317955 | 0.955511 | 1.830297 | 3.356897 | 5.832512 | 8.335911 |
| 15-91 TRE | 22 | 4.375543 | 2.027382 | 1.193922 | 2.821326 | 4.497585 | 5.258511 | 8.718991 |
| 15-92 TRE | 22 | 4.389230 | 1.951519 | 1.064711 | 2.815485 | 4.389499 | 5.372205 | 8.308199 |
| 16-110 TRE | 22 | 4.430202 | 1.927567 | 1.131402 | 2.879832 | 4.691348 | 5.381823 | 7.906915 |
| 15-93 TRE | 22 | 4.436766 | 2.030507 | 1.335001 | 2.898421 | 4.549657 | 5.332521 | 8.809714 |
| 16-111 TRE | 22 | 4.314621 | 2.279008 | 0.262364 | 2.723639 | 4.160444 | 5.657124 | 8.837971 |
| 16-112 R | 22 | 4.208522 | 2.177020 | 1.252763 | 2.120967 | 4.595625 | 5.271007 | 8.302514 |
| 15-94 R | 22 | 4.797970 | 2.030943 | 1.931521 | 3.298397 | 4.539030 | 6.531095 | 8.558911 |

**C-type and TWC rocks** has lower average values and less variation**. R and TRE rocks** has highest average values with wide variation, indicating stronger and more diverse rock characteristics. **TRE rocks** (15-91, 15-93, 16-110, etc.) consistently rank **among the highest** in mean and max values.

* **Skewness and Kurtosis:**

|  |  |  |
| --- | --- | --- |
| **Statistic** | **Observation** | **Interpretation** |
| Skewness | Mostly positive (1.5–3.5) | Right-skewed; few large values influence results |
| Kurtosis | Mostly >3, some >10 | Leptokurtic; peaked data with heavy tails and potential outliers |
| General Trend | Both metrics high | Non-normal distributions reflecting geological variability |





Most rock types show right-skewed and highly peaked (leptokurtic) distributions, indicating non-normal data with high variability and some large outlier values

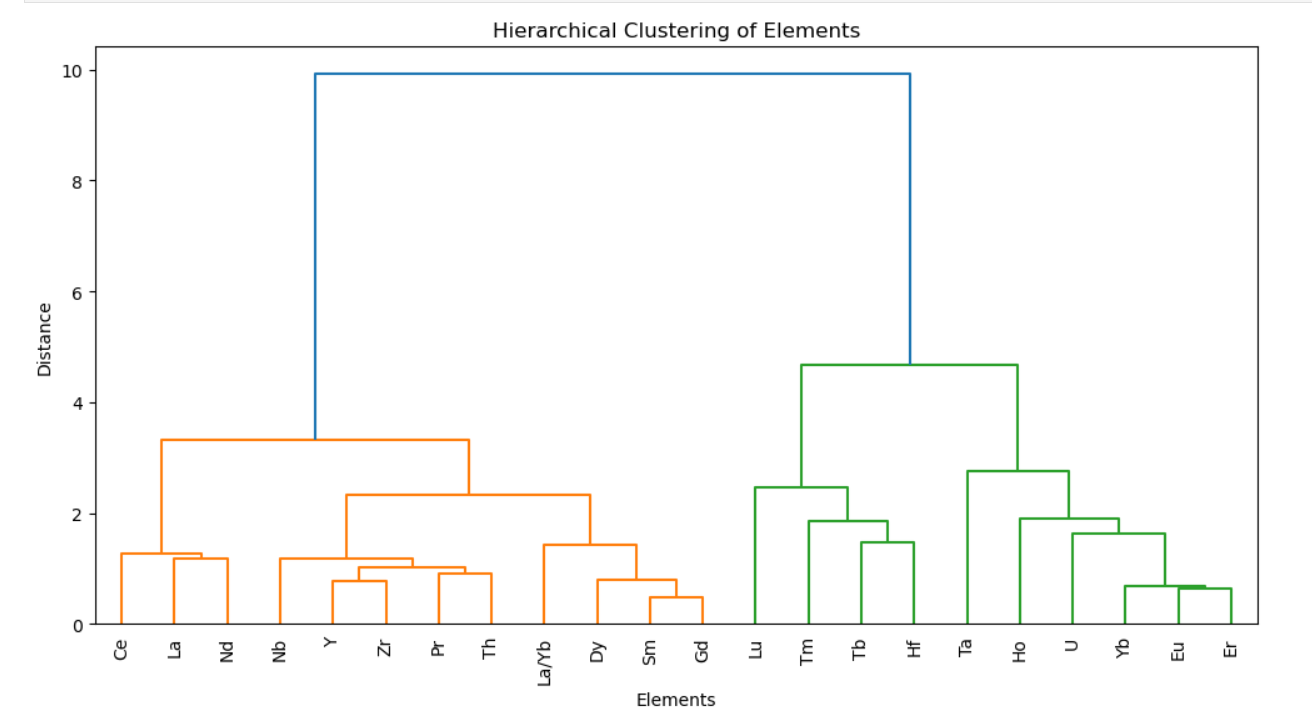
* **Dealing with missing data:**

We are going with **median imputation.**

**NRMSE =** 0.1631 ( 16%)

The NRMSE value of 0.1631 indicates that the median imputation method worked well, filling the missing values close to the original data.

* **Log Transformation:** To address skewness and stabilize the variance in the dataset, a log transformation was applied to our data set.
* **Clustering:**



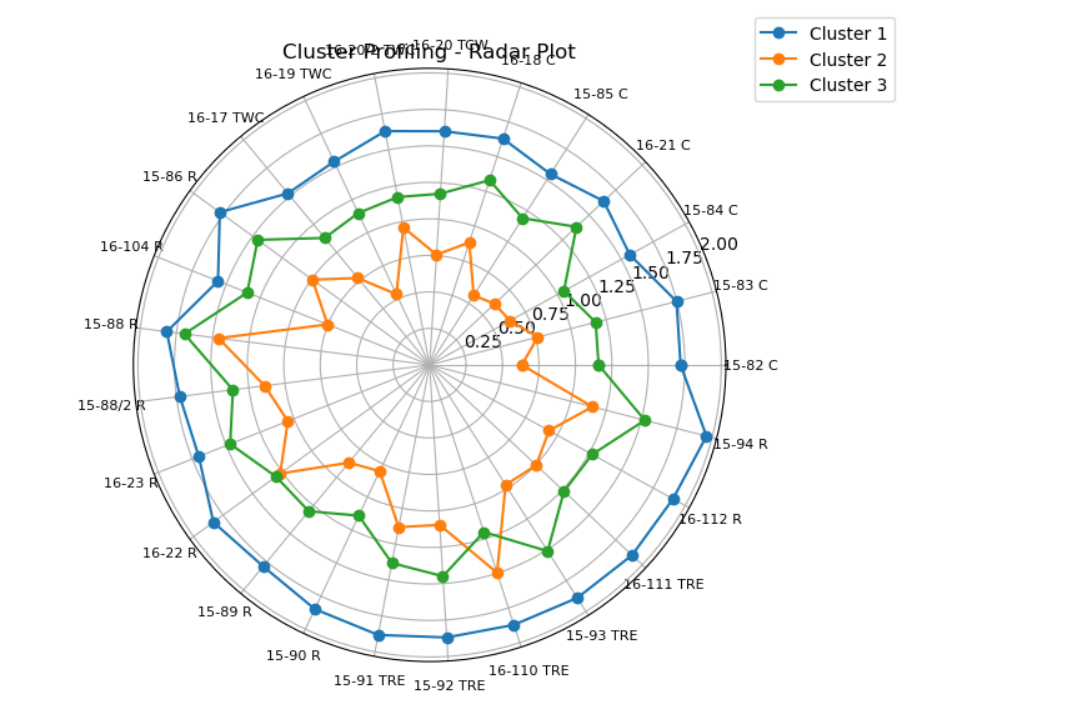
The dendrogram separates the elements into **two main clusters.**

* **Cluster 1 (left side)** groups mostly **Light Rare Earth Elements (LREEs)** such as Ce, La, Nd, Pr, Sm, and others like Y, Zr, Nb. These elements show **similar geochemical behavior** and occur together due to their comparable mobility and enrichment patterns.
* **Cluster 2 (right side)** contains mainly **Heavy Rare Earth Elements (HREEs)** such as Yb, Er, Ho, Tm, along with Hf, Ta, U. These elements behave similarly and form a separate group because their fractionation trends differ from LREEs.

**Overall:** The clustering clearly shows a **division between LREEs and HREE elements,** meaning these two groups have **distinct geochemical characteristics.**

* **Cluster Profiling:**

****

****

* **Cluster 1 (Blue)** = **High values** → enriched, strong signals
* **Cluster 2 (Orange)** = **Low values** → depleted, weak signals
* **Cluster 3 (Green)** = **Moderate values** → intermediate, mixed signals

### **Cluster 1 – High Values Across Most Elements**

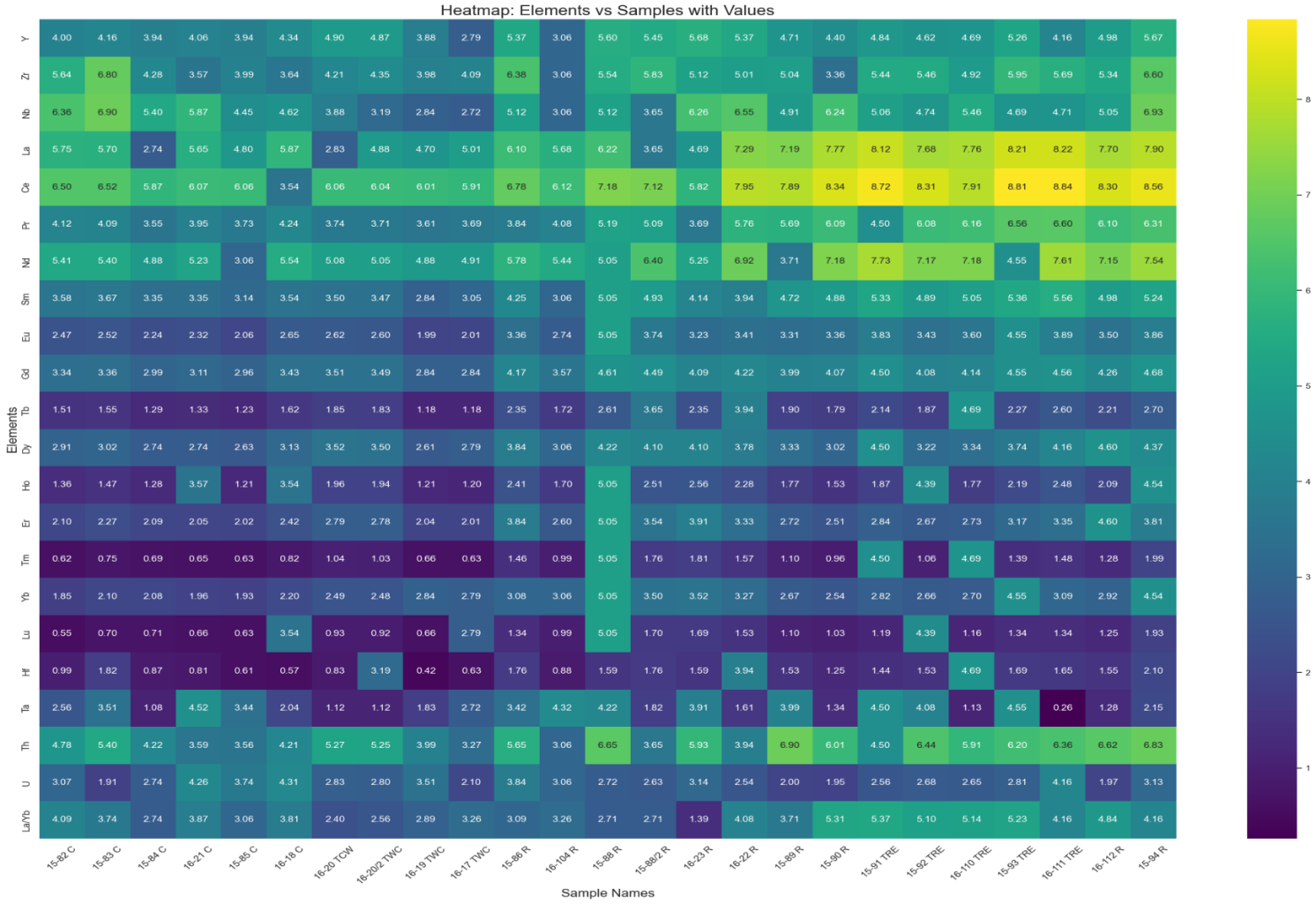
* Shows the **highest log-transformed values** on both the radar and bar plots.
* Indicates samples with **overall enriched geochemical signatures**.
* Represents **strongly reflective / high-intensity behaviour**, meaning the elements are present in **consistently higher concentrations**.
* Suggests **less fractionation** or more **primitive, enriched source characteristics**.

### **Cluster 2 – Lowest Values**

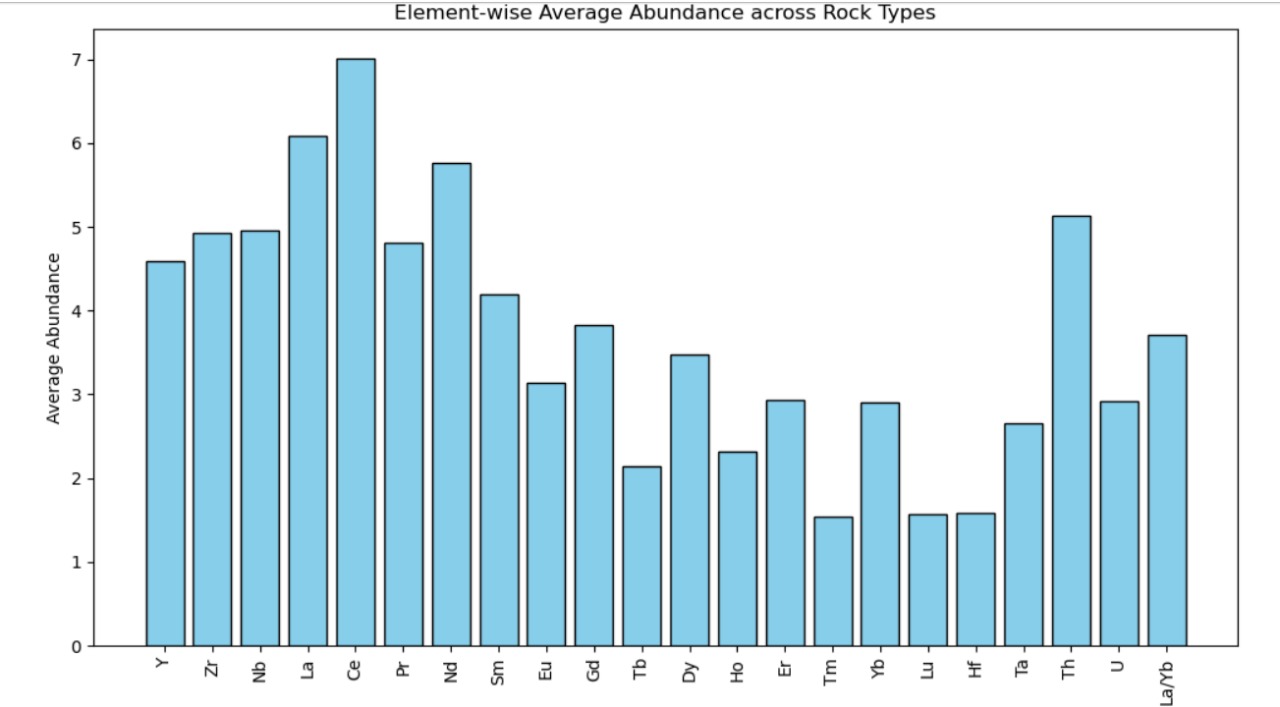
* Contains samples with the **lowest measurements** for most elements.
* Radar plot shows this cluster staying **near the center**, meaning **weak signal intensity**.
* Indicates **depleted or highly fractionated samples** with **low elemental abundance**.
* Represents **less reflective, low-strength responses**, possibly due to alteration or extraction of incompatible elements.

### **Cluster 3 – Moderate/Intermediate Values**

* Falls between Clusters 1 and 2 in both radar and bar plots.
* Represents **intermediate geochemical characteristics**—neither highly enriched nor strongly depleted.
* Indicates **moderate reflectivity** and **partial enrichment**.
* Suggests a **mixed signature**, possibly due to blending of depleted and enriched sources or partial fractionation
* **Heatmap**



* The samples show **strong LREE enrichment (La–Sm)** and **HREE depletion (Yb–Lu),** indicating a typical **fractionated REE pattern**.
* Mid-REEs like **Ce, Nd, Sm, Gd** are the most enriched, showing coherent REE behavior.
* HFSE elements **(Zr–Hf, Nb–Ta)** show moderate enrichment and move together, indicating a **magmatic origin** with minimal alteration.
* HREEs **(Tm, Yb, Lu)** are consistently low, supporting evolved or enriched magmatic conditions.
* Some samples (e.g., **15-92 RE, 16-104 RE)** are more enriched, while others are more depleted.
* **Average REE Abundance Pattern Across Rock Types**



Lanthanum (La) and Cerium (Ce) exhibit the highest average abundances among all analyzed REEs. This pattern clearly indicates a strong enrichment of Light Rare Earth Elements (LREEs) across the dataset.

Additionally, La and Ce demonstrate:

* High mean concentrations, providing strong signal strength
* Strong positive correlations with most other REEs, reflecting consistent geochemical behavior
* Low variability and stable distribution, making them statistically reliable predictors

Based on these characteristics, La and Ce are identified as the most suitable target variables for regression modeling, offering both robustness and strong geochemical significance.

* **Regression: Elastic Net**

1. **La as target variable**

=== Elastic Net (Target = La) ===

R² Score: 0.8854848276113853

Intercept: 6.08427254368

Coefficients:

Y: -0.1276

Zr: -0.0000

Nb: 0.0037

Ce: 0.0791

Pr: 0.3699

Nd: 0.0712

Sm: 0.0000

Eu: 0.0000

Gd: 0.0216

Tb: -0.0000

Dy: 0.1271

Ho: -0.0000

Er: 0.0271

Tm: 0.0000

Yb: 0.0346

Lu: 0.0000

Hf: 0.1740

Ta: 0.2843

Th: 0.2581

U: 0.1708

La/Yb: 0.7937

The Elastic Net model explains **88.5% of La variability**, indicating a strong predictive relationship between La and the selected REEs and trace elements.

1. **Strong positive influence from Pr, La/Yb, Ta, and Th** – Suggests a **LREE–MREE enriched signature**, reflecting contributions from incompatible element–rich phases, possibly amphibole-bearing magmas or moderate partial melting.
2. **Moderate positive contributions from Ce, Nd, Dy, Er, Hf, and U** – Indicates **normal LREE behavior with minor HREE input**, supporting a mostly smooth REE pattern without major anomalies.
3. **Weak or negligible effects from Y, Zr, Sm, Eu, Tb, Ho, Tm, and Lu** – Suggests **limited garnet or HREE fractionation**, consistent with **low-pressure melting** and minimal garnet retention in the source.
4. **Positive coefficient for Yb (0.0346)** – Slight HREE contribution, showing only **minor Yb enrichment**, consistent with typical LREE-dominated magmatic suites.

**Overall interpretation:** The modeled La variations reflect a **LREE–MREE enriched magmatic source**, likely formed under **low-pressure conditions** with some amphibole involvement, **minor garnet influence**, and a **generally smooth REE pattern**, with La/Yb ratio highlighting moderate fractionation between light and heavy REEs.

**2) Ce as target variable**

=== Elastic Net (Target = Ce) ===

R² Score: 0.8544061707649291

Intercept: 7.0081692788

Coefficients:

Y: -0.0000

Zr: 0.0843

Nb: -0.0034

La: 0.1894

Pr: 0.3251

Nd: 0.0640

Sm: 0.3959

Eu: 0.0000

Gd: 0.0141

Tb: -0.0000

Dy: 0.0000

Ho: -0.0000

Er: 0.0000

Tm: 0.0000

Yb: 0.1074

Lu: -0.2595

Hf: 0.0000

Ta: 0.0243

Th: 0.0000

U: -0.1976

La/Yb: 0.0969

The Elastic Net model explains **85.4% of Ce variability**, indicating a strong predictive relationship between Ce and the selected REEs and trace elements.

1. **Strong positive influence from Sm, Pr, and La** – Suggests a **LREE–MREE enriched signature**, reflecting typical light REE behavior and moderate partial melting processes.
2. **Moderate positive contributions from Yb and Zr** – Indicates **minor HREE involvement** and potential accessory phase contributions, with limited fractionation.
3. **Negative effects from Lu and U** – Suggest slight HREE depletion (Lu) and minor incompatible element variation (U), consistent with **low-pressure melting and minimal garnet influence**.
4. **Negligible coefficients for Y, Eu, Tb, Dy, Ho, Er, Tm, Hf, and Th** – Suggest **minimal fractionation or anomaly** in these elements, maintaining a smooth REE pattern.
5. **Low La/Yb influence (0.0969)** – Indicates that heavy REE fractionation is minor and the source is **dominated by LREE–MREE enrichment**.

**Overall interpretation:** The modeled Ce variations reflect a **LREE–MREE enriched magmatic source**, formed under **low-pressure conditions**, with **amphibole or similar phases contributing to LREE enrichment**, minimal garnet influence, and a generally smooth REE pattern with slight HREE modulation.

**3 ) La nd Ce both as target variable**

=== Multi-Task Elastic Net (Targets = La & Ce) ===

R² Score: 0.8745765632025118

Best Alpha: 0.1

L1 Ratio: 0.5

=== Coefficients for La and Ce ===

-- Coefficients for La --

Y: -0.1666

Zr: -0.0242

Nb: 0.0356

Pr: 0.3986

Nd: 0.0983

Sm: 0.0231

Eu: 0.0000

Gd: 0.0643

Tb: -0.0000

Dy: 0.1257

Ho: -0.0153

Er: 0.0000

Tm: 0.0000

Yb: 0.0813

Lu: 0.0042

Hf: 0.1765

Ta: 0.2953

Th: 0.2663

U: 0.1763

La/Yb: 0.7759

-- Coefficients for Ce --

Y: -0.0375

Zr: 0.0799

Nb: -0.0334

Pr: 0.3996

Nd: 0.1059

Sm: 0.2425

Eu: 0.0000

Gd: 0.0732

Tb: -0.0000

Dy: 0.0624

Ho: -0.0147

Er: 0.0000

Tm: 0.0000

Yb: 0.1220

Lu: -0.2717

Hf: 0.0212

Ta: 0.1155

Th: 0.0832

U: -0.1730

La/Yb: 0.2438

The Multi-Task Elastic Net model predicts **La and Ce together with ~87% accuracy**, showing that these LREEs share a largely coherent REE behavior across the sample set.

1. **La and Ce rise together with many MREEs (Nd → Yb)** – Indicates a **smooth, coherent REE pattern**, reflecting enrichment that is not limited to the LREEs but extends moderately into the MREEs and HREEs.
2. **Minimal influence from Eu, Lu, Ho, Tm, and Er** – Suggests **negligible Eu anomalies** and **weak garnet or plagioclase control**, consistent with low-pressure melting conditions.
3. **Consistent positive contributions from Pr, Nd, Sm, Dy, Hf, Ta, Th, and U** – Points to **moderate REE fractionation** and a source or melt process that affects the entire REE spectrum in a similar way, preserving overall REE coherence.
4. **La/Yb ratio influence (La: 0.776; Ce: 0.244)** – Highlights stronger LREE enrichment for La relative to Ce, but still reflects **overall smooth fractionation across the LREE–MREE–HREE spectrum**.

**Overall interpretation:**  
The model shows that La and Ce enrichment reflects **whole-pattern REE coherence**, typical of **felsic to alkaline magmas** with **amphibole involvement, limited garnet or plagioclase influence**, and **moderate fractionation across LREE–MREE–HREE**, producing a generally smooth and enriched REE curve.

**Summary from 3 models**  
Across all three models, Pr, Nd, Sm, Dy, La/Yb, Ta, Th, and U are consistently important for predicting LREE behavior, while Eu, Lu, Ho, Tm, Tb, and minor HREEs generally have weak or negligible effects. This highlights a LREE–MREE enriched magmatic source with smooth REE patterns, formed under low-pressure conditions with amphibole influence and limited garnet or plagioclase control.

* **Analysis of Variance:**

**Null hypothesis (H₀):**  
All group means are equal.

H0:μ1=μ2=μ3=⋯=μk​

**Alternative hypothesis (H₁):**

H1:At least one group mean is different

ANOVA Result:

F-statistic = 2.8923

P-value = 0.0000

Result: Significant difference between groups (reject H0)

* **F-statistic:** Measures the ratio of **between-group variance** to **within-group variance**. Larger F → more likely that at least one group mean is different.
* **P-value:** Probability of observing the F-statistic if the null hypothesis were true.

### **Interpretation**

* **p-value ≈ 0.0000**, which is less than the typical significance level **α = 0.05**.
* **Decision:** Reject the null hypothesis (H₀).

**Conclusion:** There is a **statistically significant difference** between the means of at least some of the groups in your dataset.

* **Conclusion**

**1. Bayan Obo, China**

The Bayan Obo dataset shows clear and statistically robust separation between carbonatites (Cbt) and fenites (Ft).  
Cbt samples exhibit very high LREE concentrations, strong positive skewness, and highly enriched REE patterns, confirming their role as the primary magmatic REE source.  
Fenites display lower, more variable REE levels, representing secondary metasomatic alteration rather than diluted carbonatites.

High-dimensional Hotelling’s T² (p = 5.67 × 10⁻⁵) confirms that the two groups form distinct multivariate REE populations.  
Clustering, heatmaps, and spider plots consistently show LREE-rich and HREE-poor carbonatites versus moderately enriched, irregular fenites.

Elastic Net regression demonstrates that La and Ce are strongly controlled by Pr, Nd, Sm, Gd, Dy, Ho, Er, indicating a smooth LREE–MREE–HREE trend with minimal Eu anomalies and limited garnet control.  
Overall, Bayan Obo reflects a two-stage REE evolution:

1. Carbonatitic magmatism – primary LREE enrichment
2. Fenitization – fluid-driven REE redistribution

**2. Belaya Zima, Russia**

Belaya Zima displays highly variable REE and trace element abundances, with dolomite carbonatite (CD) and apatite carbonatite (AC) samples showing the highest enrichment and strongest skewness/kurtosis.  
Magnetite (M), ilmenite (I), and ankerite (A) groups display lower, stable and more moderate REE values, indicating more homogeneous compositions.

Clustering reveals three distinct geochemical groups:

* Cluster 1: HREE- and HFSE-rich, evolved compositions
* Cluster 2: LREE-dominated, moderately enriched magmatic signatures
* Cluster 3: Major oxide-rich, REE-poor, primitive signatures

Heatmaps and spider diagrams confirm strong sample-to-sample variation, reflecting complex magmatic and metasomatic histories.

Elastic Net shows that La variability is well explained (R² ≈ 0.78) by elements like Zr, Tb, MgO, W, highlighting a LREE–MREE enriched magmatic source with moderate fractionation.  
CaO, however, is poorly predicted, indicating major-element control unrelated to REEs.

Overall, Belaya Zima represents a highly heterogeneous carbonatite complex, where REE enrichment is concentrated in CD and AC types, while other rock types show weaker and more variable REE patterns.

**3. Fen Complex, Norway**

Fen samples exhibit systematic LREE enrichment and HREE depletion, consistent with carbonatite-derived magmatic systems.  
REE distributions across rock types (C-type, TWC, R, TRE) show that TRE and R samples have the highest enrichment, while C-type and TWC are lower and more uniform, indicating varying degrees of magmatic evolution.

Clustering cleanly separates elements into LREE vs. HREE groups, reflecting classic fractionation behavior.  
Cluster profiling shows three groups (High, Moderate, Low), demonstrating progressive fractionation and variable enrichment.

Regression analysis strongly predicts La and Ce (R² ≈ 0.86–0.88), with key controls from Pr, Nd, Sm, Dy, Hf, Ta, Th, U and La/Yb, consistent with LREE-dominated magmatic enrichment and low-pressure melting.  
Weak Eu and Lu effects indicate minimal Eu anomaly and limited garnet/plagioclase control.

Overall, the Fen complex represents a coherent LREE-rich carbonatite system with moderate fractionation and a dominantly magmatic signature.

* **Final result achieved**

**1)** **Bayan Obo shows the clearest geochemical separation**, with carbonatites and fenites forming fully distinct REE populations, confirming a two-stage magmatic–hydrothermal evolution.

**2)** **Belaya Zima is the most heterogeneous**, with extreme variability and the strongest skewness/kurtosis, reflecting multiple carbonatite types and complex enrichment mechanisms.

**3)** **Fen shows the most coherent magmatic signature**, with smooth LREE enrichment, predictable REE patterns, and strong multivariate regression performance.

**4)** Across all three regions, **La and Ce emerge as the most reliable REE indicators**, supported by clustering, heatmaps, and Elastic Net modeling.

**5)** **LREE–MREE control dominates**, while HREEs show weak and inconsistent contributions, indicating widespread low-pressure melting, amphibole involvement, and limited garnet retention.

**6)** Each region shows distinct enrichment mechanisms:

* **Bayan Obo:** Magmatic + metasomatic overprint
* **Belaya Zima:** Strongly heterogeneous multi-phase magmatism
* **Fen:** Coherent magmatic enrichment with mild fractionation
* **References**

1. Articles:

* “A Geochemical Study of an REE-rich Carbonatite Dyke at Bayan Obo, Inner Mongolia, Northern China.” (YANG Xueming1, ZHENG Yongfei1, YANG Xiaoyong1, ZHANG Peishan2 and M.J. LE BAS3)
* “Trace-element composition of minerals and rocks in the Belaya Zima carbonatite complex (Russia): Implications for the mechanisms of magma evolution and carbonatite formation.” Anna G. Doroshkevich, Ilya V. Veksler, Reiner Klemd , Elena A. Khromova, Ivan A. Izbrodin
* “The hydrothermal alteration of carbonatite in the Fen Complex, Norway: mineralogy, geochemistry, and implications for rare-earth element resource formation” by C. Marien, A. H. Dijkstra, and C. Wilkins, School of Geography, Earth and Environmental Sciences, Plymouth University, UK.

1. Applied Multivariate Statistical Analysis Book by Dean W. Wichern, Richard A. Johnson.
2. <https://www.lindianresources.com.au/kangankunde-rare-earths>
3. <https://www.mining-technology.com/projects/kangankunde-rare-earths-project-malawi/>
4. <https://en.wikipedia.org/wiki/Geology_of_Malawi>