**Data processing and outlier removal**

* Extract relevant features:
  + Max\_Diameter\_µm , Elongation , Min\_Diameter\_µm , Shape Factor and Sphericity.
* **The interquartile range (IQR) method** was used to eliminate outliers and ensure data quality.
* StandardScaler is used for **standardization** to keep the data at a uniform scale when clustering.

**Data preprocessing and grouping**

* **Divide the data into 10 groups** according to Age\_Ma and create the Age\_Group variable.
* Use pd.qcut() to ensure that the data are evenly distributed across groups and print the range for each age group.

**Gaussian Mixture Model (GMM) Clustering**

* **A Gaussian mixture model (GMM)** is used to classify data from different time periods .
* We also tried using a time **sliding window to generate different time intervals** , and then used **a Gaussian mixture model (GMM)** to classify data in different time periods .
* We also tried the K-mean model, but the clustering effect was poor and the changing trend of foraminifera species could not be discovered.
* **The Bayesian Information Criterion (BIC)** was used to evaluate different numbers of clusters and select the optimal number of clusters (best\_parameter).
* Within **each time window** or **age group** , the number of clusters **with the lowest BIC value is automatically selected** as the optimal number of clusters.
* **the best model and BIC value** for each age group .

**Clustering effect evaluation**

* **The Silhouette Score** is calculated to evaluate the clustering quality.
* Ensure that clusters are **well separated** and consistent within clusters.

**Visualization and Analysis**

* **BIC evaluation curve** : Displays the BIC values for different numbers of clusters to help determine the optimal number of clusters.
* **Silhouette coefficient over time** : Tracks the trend of clustering quality changes in different time windows.
* **Species quantity change graph : shows the evolution of** the optimal cluster number over geological time and analyzes possible biological evolution trends.
* **Species population change over time : Calculate the average age** of each age group and store the optimal species population to observe the trend over time.