



# An integrated approach for understanding global earthquake patterns and enhancing seismic risk assessment

Mariam Ibrahim<sup>1</sup> · Baidaa Al-Bander<sup>1</sup>

Received: 2 November 2023 / Accepted: 2 February 2024  
© The Author(s) 2024

**Abstract** Earthquakes, as intricate natural phenomena, profoundly impact lives, infrastructure, and the environment. While previous research has explored earthquake patterns through data analysis methods, there has been a gap in examining the time intervals between consecutive earthquakes across various magnitude categories. Given the complexity and vastness of seismic data, this study aims to provide comprehensive insights into global seismic activity by employing sophisticated data analysis methodologies on a century-long dataset of seismic events. The four-phase methodology encompasses exploratory data analysis (EDA), temporal dynamics exploration, spatial pattern analysis, and cluster analysis. The EDA serves as the foundational step, providing fundamental insights into the dataset's attributes and laying the groundwork for subsequent analyses. Temporal dynamics exploration focuses on discerning variations in earthquake occurrences over time. Spatial analysis identifies geographic regions with heightened earthquake activity and uncovers patterns of seismic clustering. K-means clustering is employed to delineate distinct earthquake occurrence clusters or hotspots based on geographical coordinates. The study's findings reveal a notable increase in recorded earthquakes since the 1960s, peaking in 2018. Distinct patterns in seismic activity are linked to factors such as time, human activities, and plate boundaries. The integrated approach enriches understanding of global earthquake trends and patterns, contributing to improved seismic hazard assessments, early warning systems, and risk mitigation efforts.

**Keywords** Earthquakes · Seismic data · Data science analysis · Spatiotemporal analysis · Clustering

## 1 Introduction

Earthquakes result from sudden rock fractures in the Earth's crust, releasing significant energy. They can be categorized as major earthquakes, foreshocks, and aftershocks. Major earthquakes are highly destructive and can predict future major events. Foreshocks occur before major earthquakes, aiding prediction and preventive measures. Aftershocks follow the mainshock, gradually decreasing in frequency and magnitude. Earthquake data represent time series data with location, time, depth, and magnitude information [1–4]. Throughout history, several major earthquakes have left a lasting impact on the affected regions, including the 1960 Great Chilean Earthquake, the 1964 Prince William Sound Earthquake in Alaska, the 2004 Sumatra—Andaman Islands Earthquake, the 2011 Great Tohoku Earthquake in Japan, and the 1952 earthquake near Petropavlovsk-Kamchatsky, Russia. These devastating events had magnitudes high magnitudes and caused widespread destruction, loss of lives, and significant damage to infrastructure. In recent times, there has been a notable increase in the occurrence of both minor and major earthquakes across different regions of the globe and magnitudes, exemplified by recent events in Turkey and Syria [5–7]. These seismic events have garnered significant attention due to their impact on human lives, infrastructure, and the environment. These seismic events are among the most catastrophic natural disasters, resulting in significant casualties and imposing substantial economic burdens on affected communities. The impact of earthquakes extends beyond human lives and infrastructure,

✉ Baidaa Al-Bander  
b.al-bander@keele.ac.uk

<sup>1</sup> School of Computing, Keele University, Keele, UK

often causing secondary environmental repercussions like surface ruptures, soil liquefaction, tsunamis, landslides, and fires [4, 8–17]. The researchers emphasized the devastating consequences of earthquakes, including loss of life, injuries, displacement, and structural damage [18]

By forecasting earthquakes, individuals can take timely actions to protect themselves and reduce damage and economic losses. Researchers tirelessly seek methods to predict earthquakes due to their destructive potential and far-reaching consequences. Accurate forecasts could mitigate impact through preventive measures and public preparedness, focusing on location, timing, and magnitude. Historical trends inform predictions, but the frequent and unpredictable nature of earthquakes presents ongoing challenges. Preparedness, research, and collaboration remain crucial in minimizing their devastating effects [19, 20]. The availability of extensive seismic data and data science tools presents an unprecedented opportunity to gain deeper insights into earthquake dynamics. Traditional methods were limited by manual interpretation and small datasets, hindering accuracy. Data science techniques like machine learning enable researchers to analyse large-scale seismic data for valuable insights, leading to improved forecasting models, hazard assessments, and early warning systems, enhancing our preparedness and response strategies [4, 13, 19, 21–23]. Clustering, a core technique in data mining, emerges as a crucial aspect of this research, aiming to group similar seismic events for deeper insights into earthquake dynamics [24, 25]. K-means, a widely employed algorithm, exhibits speed and efficiency, although its drawbacks include the need for predefined centroids and sensitivity to initial choices [25]. Data mining, emphasizing convenience and completeness, plays a pivotal role, with clustering as a fundamental operation, contributing to tasks such as image processing, sequence analysis, and pattern recognition [26]. As the volume of data increases, the necessity for data mining tools becomes paramount, and classification emerges as a vital technique for knowledge discovery [27].

The significance of the presented study lies in its pioneering integration of data science methodologies, spatial analysis, and comprehensive interoccurrence time analysis, which collectively provide unprecedented insights into global seismic trends and earthquake behaviour for enhanced earthquake prediction, hazard assessment, and risk mitigation. The remainder of this paper is structured as follows: Sect. 2 provides a brief review of the previous work related to the analysis of global earthquake trends and patterns; Sect. 3 describes the dataset used in this study, along with an explanation of the proposed method; Sect. 4 presents the results obtained, while Sect. 5 presents the results and discussion of these results; Sect. 6 compares our findings with existing methods; Finally, Sect. 6 serves as the conclusion of the paper.

## 2 Literature survey

Several articles have explored the analysis of global earthquake trends and patterns using data science techniques. The authors of [28] conducted a literature review on earthquake prediction and prevention, categorizing methods into machine learning, data mining, and seismic feature extraction. They highlighted the importance of reducing prediction errors for accurate earthquake predictions and identifying high-risk areas. In recent studies, researchers have explored various aspects of seismic activities and earthquake distributions using data analysis techniques. In [29], the authors focused on analysing the spatial distribution of seismic activities in China by utilizing provincial seismic data. Through spatial autocorrelation analysis, they identified significant global autocorrelation characteristics, revealing a spatial agglomeration pattern of earthquakes in mainland China. Moreover, they observed a decreasing trend in the disparities of seismic activity among different regions over time. This suggests a potential convergence in seismic activity across China.

Another study presented in [30] delved into the spatial–temporal characteristics of seismicity clusters, aiming to understand their distribution and heterogeneity. By categorizing seismic clusters into persistent clusters and burst clusters based on duration, they analysed their spatial distributions. The findings indicated that plate interaction played a substantial role in shaping the distribution of persistent clusters, while the burst clusters displayed less spatial heterogeneity. This suggests that different mechanisms may govern the formation and behaviour of these distinct types of seismic clusters. Additionally, [12] conducted spatiotemporal analyses to gain insights into earthquake distributions. Their investigation revealed intriguing findings related to the behaviour of earthquakes. Applying scaling relationships resulted in data collapses, indicating critical behaviour within the seismological phenomenon. Furthermore, the presence of long-range spatiotemporal correlations between earthquakes and q-exponential distributions suggested the existence of self-organized criticality. These observations contribute to the understanding of the underlying dynamics and mechanisms involved in seismic events. While the study conducted in [29] shed light on the spatial patterns and regional convergence of seismic activity in China, [30] explored the spatial–temporal characteristics of seismicity clusters, providing insights into their distribution and heterogeneity [12]. Moreover, understanding of the seismic nature is expanded by uncovering critical behaviour and long-range correlations in earthquake distributions [29].

The authors of [31] introduced a method for analysing earthquake time-series data, distinguishing clustered aftershock sequences from regular background events. It utilized inter-event time statistics and coefficient of variation

(COV), employing a sliding temporal window to filter out time-correlated events and model background events as a Poisson process. The research showed the approach's competitive performance in seismicity declustering, emphasizing the usefulness of inter-event time statistics and COV in assessing seismic risk. Shan et al. proposed a method to analyse temporal and spatial evolution trends in earthquakes in California and Nevada. The study finds a regular cycle of decreasing-rising frequency for earthquakes of magnitude 4.5 or above. The spatial concentration of earthquakes exhibits a conch movement pattern, indicating the epicentre moving closer to the study area's centre. The spatial distribution pattern aligns with the direction of the San Andreas Fault Zone [8]. Furthermore, [32] emphasized the significance of understanding spatial distribution patterns (SDPs) of natural disasters for effective risk mitigation. Their study analysed global disasters from 1980 to 2016 using biclustering techniques, providing insights into different disaster types and their impact on fatality rates across regions. The findings revealed uneven SDPs of fatality rates compared to occurrence rates, classifying selected countries into four classes based on the occurrence of major disasters like storms, floods, epidemics, droughts, and earthquakes in specific regions.

Yousefzadeh et al. [19] demonstrated the effectiveness of Support Vector Machine (SVM) and Deep Neural Network (DNN) models in predicting high-magnitude earthquakes by introducing the novel parameter Fault Density. Moreover, [33] emphasized the importance of appropriate model selection and data preprocessing in leveraging time series data for earthquake risk analysis. Their findings highlighted the potential of advanced deep learning methods in enhancing understanding of earthquakes and improving prediction capabilities. Other researchers have also employed data mining and statistical techniques to analyse earthquake patterns in different regions. [23] applied K-means neutrosophic clustering to Ecuador earthquake data, identifying patterns for predicting future earthquake behaviour and preventive measures. Similarly, [33] analysed the spatial distribution pattern of earthquakes in Iraq using statistical and data mining techniques.

Despite these efforts, a detailed analysis of time intervals between successive earthquakes of different magnitudes is lacking. The present research aims to fill this gap by comprehensively analysing these time differences, leveraging data science tools and geospatial analyses to gain insights into the earthquake frequency, regularity, spatial distribution, and behaviour of seismic clusters. The conducted study aims to enhance seismic risk assessment and disaster preparedness, providing valuable insights for policymakers, researchers, and stakeholders involved in earthquake monitoring and mitigation efforts. By strengthening earthquake forecasting capabilities, this study contributes to the scientific community and ensures

the protection of lives and infrastructure. The main contributions and advantages of this work are summarized as follows:

1. It contributes significantly to several areas by analysing historical earthquake data to discern temporal fluctuations and enduring patterns in seismic activity, it is imperative to conduct an analysis.
2. It attempts to contribute to the literature by describing the earthquake magnitude scale.
3. Mapping the spatial distribution helps identify regions with higher seismic activity, aiding disaster preparedness and early warning systems.
4. Assessing earthquake frequency, magnitude, and intensity informs resource allocation and risk reduction strategies.
5. The application of cluster analysis identifies earthquake hotspots and potential future seismic events in clustered regions.
6. Overall, the presented study enhances earthquake forecasting, hazard assessments, and disaster management efforts.

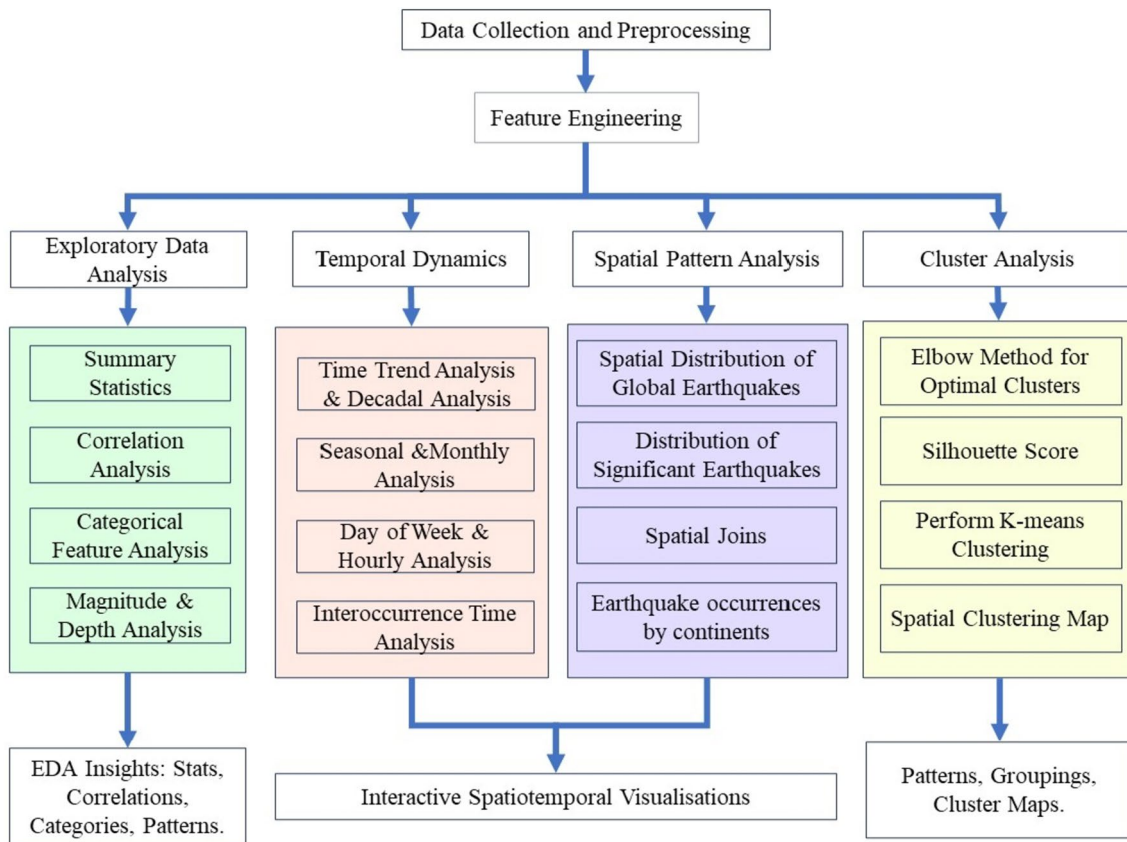
### 3 Materials and method

The developed method integrates diverse data analytical techniques to explore global earthquake patterns comprehensively. Prior to data analysis, the data underwent meticulous data pre-processing, involving cleaning procedures to remove duplicates, errors, and inconsistencies, ensuring the dataset's reliability. The Exploratory Data Analysis phase provides foundational insights into the dataset's characteristics, laying the groundwork for subsequent analyses. The exploration of temporal dynamics focused on understanding how earthquake occurrences varied over time. Spatial analysis aimed to identify geographic hotspots of earthquake occurrences and reveal patterns of seismic clustering. The study applied K-means clustering to identify distinct clusters or hotspots of earthquake occurrences based on geographical coordinates. K-means is a clustering technique utilized to group data points into distinct clusters based on similarity, with the aim of identifying patterns and relationships within the dataset [34]. Figure 1 illustrates the composite workflow diagram of the adopted methodology in this research.

#### 3.1 Data Pre-processing and exploratory analysis of earthquake dataset

To conduct the data analysis, a rich publicly available dataset was sourced from the United States Geological Survey<sup>1</sup> (USGS). The USGS dataset is a comprehensive collection of earthquake data provided by the United States Geological

<sup>1</sup> <https://earthquake.usgs.gov/data/comcat/>



**Fig. 1** Block diagram of the developed method

Survey (USGS). It includes information about earthquakes that have occurred worldwide and provides valuable insights into the characteristics and patterns of seismic activity. The USGS collects earthquake data from various sources, including seismographs, seismic networks, and earthquake monitoring stations around the world. These instruments record ground motion and other seismic parameters during an earthquake event. The dataset has various features that provide a detailed and comprehensive description of each earthquake event recorded in the USGS dataset. It includes earthquake events recorded from 1904 to 2023, with 4,036,902 unique entries across 22 columns.

Data pre-processing involves systematically cleaning, transforming, and refining raw data to enhance its quality and suitability for analysis [35]. This process is fundamental for refining the seismic data, enhancing its quality, and preparing it for advanced data science methodologies, spatial analysis, and interoccurrence time analysis, ultimately contributing to a more accurate understanding of global seismic trends and behaviours. The analysis of the earthquake dataset began with a crucial data pre-processing step. By combining meticulous data pre-processing with insightful feature engineering, the analysis established a robust foundation for further exploration and interpretation of earthquake

occurrences, fostering a deeper understanding of the dataset's distribution, patterns, and relationships. Upon establishing the consolidated data frame, an in-depth analysis of missing values has been conducted. The information about the number of missing values in each column of the dataset has been determined. The columns are listed along with the count of missing values for each column:

- 'time', 'date', 'event\_time', 'latitude', 'longitude', 'depth', 'mag', 'net', 'id', 'updated', 'place', 'type', 'status', 'locationSource', 'magSource': These columns were the key columns used in the study and they have zero missing values, indicated by a count of 0. This means that there are no missing values in these columns.
- 'magType': This column has 11,074 missing values. This suggests that some earthquakes may not have a recorded magnitude type. These columns were not included as it had no major contribution to the analysis carried out in the research. 'nst', 'gap', 'dmin', 'rms', 'horizontalError', 'depthError', 'magError', 'magNst': These columns have a significant number of missing values. 'nst' has 1,190,635 missing values, 'gap' has 1,077,681 missing values, 'dmin' has 1,719,084 missing values, 'rms' has 202,742 missing values, 'horizon-

**Table 1** Earthquake magnitude classification and effects

Magnitude	Classification	Earthquake effects	Estimated number each year
2.5 or less	Very minor	Usually not felt but can be recorded by seismograph	Millions
2.5 to 5.4	Minor	Often felt, but only causes minor damage	500,000
5.5 to 6.0	Moderate	Slight damage to buildings and other structures	350
6.1 to 6.9	Strong	May cause a lot of damage in very populated areas	100
7.0 to 7.9	Major	Major earthquake. Serious damage	10–15
8.0 or greater	Great	Great earthquake. Can totally destroy communities near the epicentre	One every year or two

<https://www.mtu.edu/geo/community/seismology/learn/earthquake-measure/magnitude/>

talError' has 1,820,242 missing values, 'depthError' has 499,182 missing values, 'magError' has 1,937,267 missing values, and 'magNst' has 1,134,114 missing values. These missing values signify potential inconsistencies or incomplete information in this dataset and had no major contribution to the analysis carried out in this study.

Feature engineering played a pivotal role in enhancing the dataset. The feature engineering method employed in the research on analysing global earthquake trends involves the extraction and transformation of relevant data attributes to enhance the analytical accuracy of earthquake patterns. This process includes selecting significant features from the raw earthquake data, such as magnitude, location coordinates, depth, and time of occurrence. Additionally, derived features like interoccurrence time intervals between successive earthquakes of different magnitudes are calculated, providing insights into the temporal dynamics of seismic activity. These engineered features are then used as input variables for spatiotemporal, interoccurrence time and statistical analyses, enabling a more comprehensive understanding of global seismic trends and patterns. Table 1: The Earthquake Magnitude Classification and Effects table categorizes the strength of earthquakes based on their magnitude and describes the typical effects experienced at each level and the estimated number each year.

Descriptive statistics were computed to summarize the magnitude and depth columns, providing measures like mean, standard deviation, minimum, maximum, and quartiles. Value counts determined earthquake classification frequencies, highlighting the most common categories based on magnitude. Correlation analysis and statistical was conducted to explore relationships between variables.

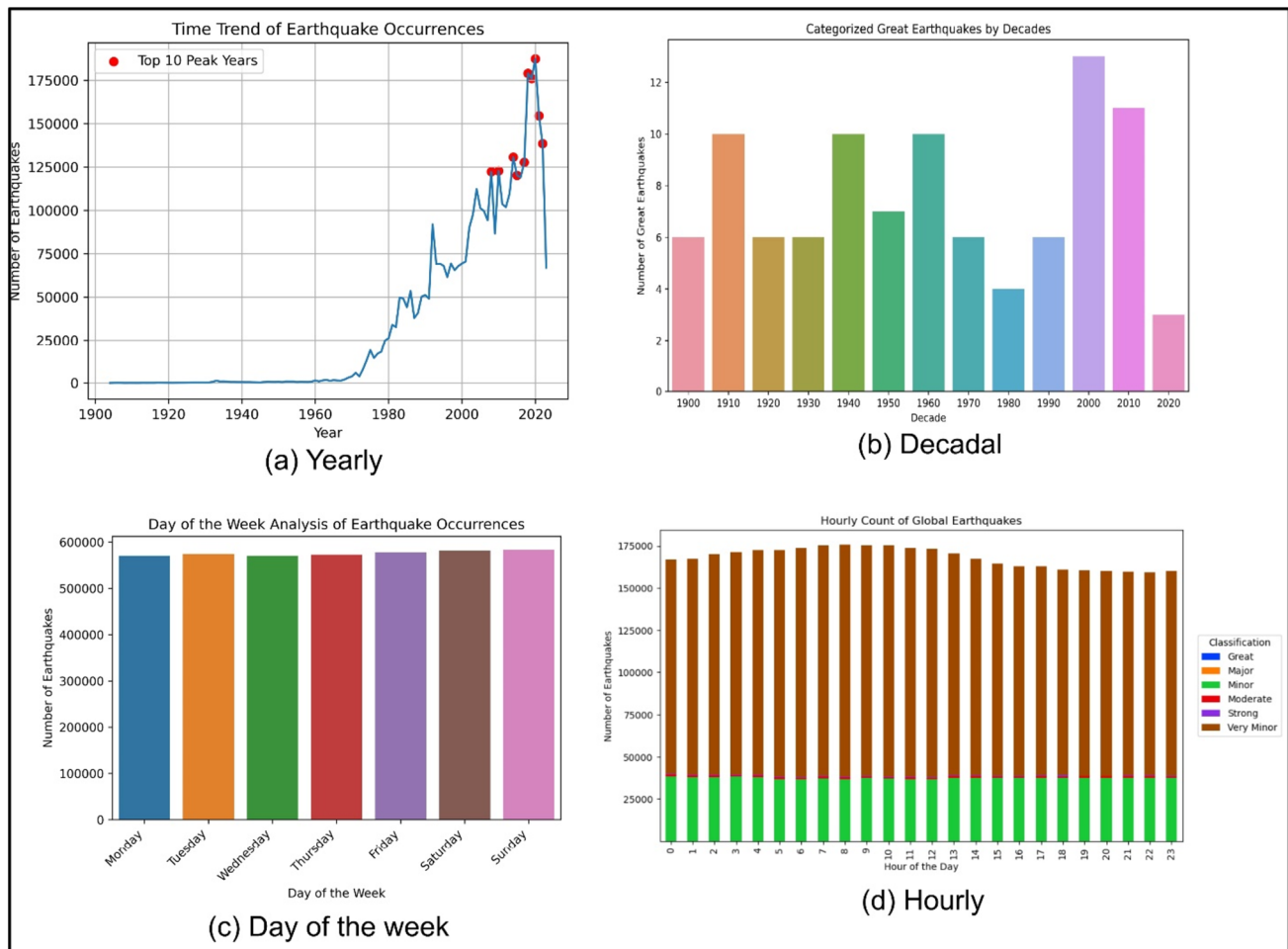
### 3.2 Temporal and spatial analyses of global earthquakes

Long-term trends are identified through time-trend analysis, shedding light on seismic occurrences over the years. Seasonal analysis unearths recurring patterns linked to specific times of the year, while monthly analysis delves into shorter-term temporal variations. Examining seismic activity by the day of the week provides insights into weekly patterns, and hourly analysis probes correlations with specific time periods. The interoccurrence time analysis, calculating intervals between consecutive earthquakes, offers valuable insights into temporal behaviors, forming a comprehensive understanding of the temporal intricacies of global earthquake events.

Simultaneously, the spatial aspect is meticulously analyzed, focusing on two key components: global spatial distribution and significant earthquakes within continents. Employing geospatial analysis techniques, including spatial joins and map visualizations, uncovers intricate spatiotemporal earthquake occurrences. The integration of advanced data science techniques and geospatial tools reveals meaningful patterns and trends in global earthquake activity. This holistic approach leads to the identification of geographical hotspots, providing crucial insights into seismic events' spatial distribution and intensity. Furthermore, a detailed analysis of significant earthquakes within continents is conducted. By spatially joining earthquake data with continent boundaries, each earthquake point is associated with its corresponding continent. This process yields a dataset rich in information, offering valuable insights into the frequency and spatial patterns of significant earthquakes in different regions of the world.

### 3.3 Cluster analysis

The primary objective here is to determine the optimal number of clusters for spatially grouping earthquakes and visualise these clusters on a map. According to [23], two



**Fig. 2** Temporal trend of global earthquake occurrences

well-known methods with a decent performance is used to determine the optimal number of clusters: the elbow method and silhouette score analysis. The elbow method involves iterating over a range of  $k$  values (number of clusters) from 2 to 10. Adopting the technique proposed by Ricardo et al. [23], For each  $k$  value, the K-means clustering algorithm is applied to the dataset, and the inertia (within-cluster sum of squared distances) is computed. The inertia measures how well data points are grouped within their clusters. The inertias for different  $k$  values are stored in the "inertias" list. The elbow method plot is used to identify the "elbow point," where the inertia starts decreasing at a slower rate. This point is indicative of the optimal number of clusters. Based on the plot, a  $k$  value of 5 seems reasonable for this dataset. Additionally, the silhouette score is calculated for each  $k$  value. The silhouette score assesses the similarity of a data point to its own cluster compared to other clusters. A higher silhouette score indicates well-separated and well-defined clusters. The silhouette scores are stored in the "silhouette\_scores" list. The elbow method and silhouette

score plots were visualised to help determine the optimal number of clusters. In this case, both analyses suggest that five clusters would be appropriate for spatially grouping the earthquakes. With the optimal number of clusters identified (optimal\_  $k=5$ ), the K-means clustering algorithm is applied again to the dataset. The "cluster" column in the DataFrame is updated with the cluster labels assigned to each earthquake based on the optimal  $k$  value. Finally, the clustered earthquakes are visualised on a map using scatter points.

## 4 Results and discussion

### 4.1 Temporal dynamics of global earthquake occurrences

Examining the overall trend of earthquake occurrences over time helps to identify any long-term patterns or changes. Figure 2 displays the temporal trend of global earthquake occurrences at different resolutions.

The analysis reveals fluctuations in earthquake occurrences over time, with a relatively low number of recorded events before 1960. This can be attributed to limited monitoring stations, incomplete historical data, and lower population density. However, from the 1960s onwards, there is a significant increase in recorded earthquakes, indicating improvements in monitoring networks and technology, and increased human presence in earthquake-prone regions. The dataset emphasizes the importance of continuous advancements in seismic monitoring capabilities to accurately capture and document earthquake events.

The analysis reveals that the year 2018 had the highest number of recorded earthquakes, indicating a significant increase in seismic activity compared to previous years. The rise in earthquakes may be influenced by various factors, such as geological characteristics, tectonic plate movements, or improved monitoring capabilities. This highlights the complex nature of earthquake occurrences, driven by geological, tectonic, and environmental factors. Understanding these patterns is essential for effective earthquake monitoring and response strategies. The findings emphasize the importance of robust seismic monitoring systems and preparedness measures in earthquake-prone regions, as advancements in data collection and global collaboration have contributed to the increase in recorded earthquakes. Continued investment in monitoring infrastructure and comprehensive mitigation strategies are crucial to minimize the impact of earthquakes on affected populations.

In the early twentieth century, from the 1900s to the 1910s, there was a notable cluster of "Great" earthquakes, with a total of 6 and 10 occurrences, respectively. This suggests a period of heightened seismic activity during this time. The subsequent decades, from the 1920s to the 1940s, exhibited a relatively consistent level of "Great" earthquake occurrences, ranging from 6 to 10 events. This stability in seismic activity implies a relatively stable tectonic environment during this period. The analysis shows fluctuations in the occurrence of "Great" earthquakes over the decades. The 1950s had 7 occurrences, followed by a resurgence in the 1960s (10 events) and 1970s (6 events). The 1980s saw a decline (4 events), while the 1990s had a slight increase (6 events). The twenty-first century witnessed a significant rise, with 13 occurrences in the 2000s, 11 in the 2010s, and 3 recorded in the 2020s so far. These variations indicate the dynamic nature of seismic activity and suggest periods of heightened tectonic activity.

Further analysis revealed distinct patterns across different months of the year, categorized into three groups based on mean earthquake counts. High seismic activity is notably observed during the summer months, particularly in July, aligning with the prevalent summer season in the Northern Hemisphere. These months exhibit a heightened seismic occurrence. Months including January, March, May, June,

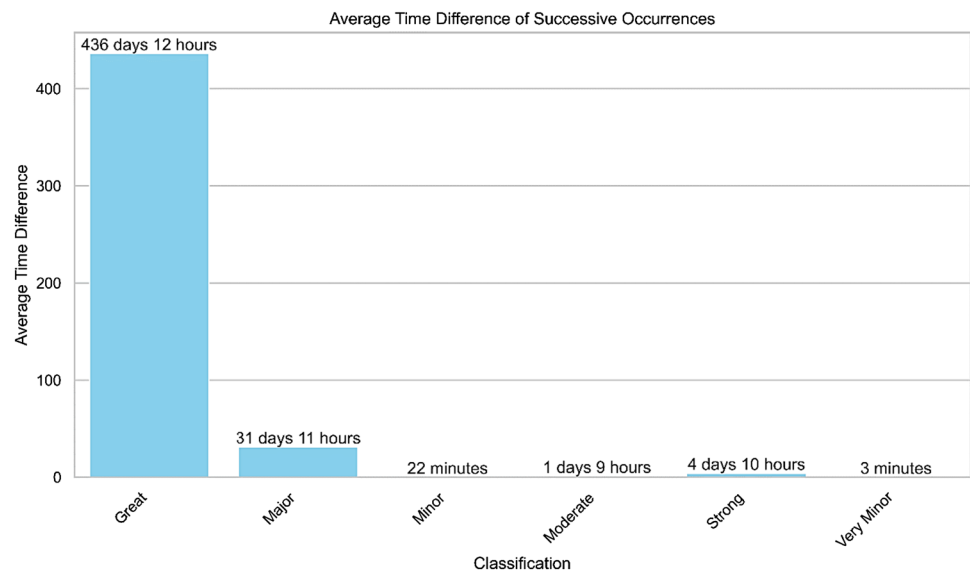
August, and December demonstrate a moderate level of seismic activity, indicating relative stability without significant peaks or declines. Conversely, lower seismic activity characterizes months like February, April, September, October, and November, suggesting a decrease in seismic occurrences compared to other periods. It's important to note that while these months show comparatively fewer earthquakes, seismic activity persists to some extent even during these times.

The analysis of the day of the week variation in earthquake occurrences reveals interesting insights into the distribution and patterns of seismic activity throughout the week. The analysis reveals the distribution of earthquakes throughout the week, with each day accounting for around 14% of the total count. Weekdays (Monday to Thursday) show consistent earthquake counts, while Friday and Saturday exhibit slightly higher counts. Sunday stands out with the highest count, suggesting a peak in seismic activity at the end of the weekend. The uniform pattern of seismic activity during weekdays underscores the importance of considering seismic risks in everyday activities and urban planning. The higher counts on weekends may indicate a potential correlation with human activities and energy consumption during these periods, emphasizing the need for increased awareness and preparedness.

Hourly analysis of earthquake occurrences provides valuable insights into the distribution and frequency of earthquakes throughout the day. The data reveals interesting patterns and variations in the number of earthquakes across different magnitudes throughout the 24-h period. In general, the analysis highlights the temporal aspect of seismic activity and helps us understand the dynamics of earthquake occurrences during different times of the day. One notable finding is the variation in the magnitude distribution of earthquakes throughout the day. When examining the hourly patterns of earthquake occurrences, distinct fluctuations are observed. The frequency of earthquakes tends to be higher during the early morning hours, typically between 2 and 6 AM. Interestingly, there is a decrease in the occurrence of earthquakes during the daytime, with a relatively lower frequency observed in the afternoon hours. However, the occurrence of earthquakes rises again during the evening hours, roughly between 5 and 9 PM. This observation suggests that seismic activity may exhibit diurnal patterns, influenced by factors such as temperature changes, stress accumulation, or human-induced activities during different times of the day.

Comparing the hourly occurrence of earthquakes across different magnitudes, we notice consistent patterns. "Very Minor" earthquakes consistently have the highest frequency throughout the day, indicating their prevalence in seismic records. The occurrence of "Minor," "Moderate," and "Strong" earthquakes follows a similar pattern, with relatively lower frequencies compared to "Very Minor" earthquakes. On the other hand, "Major" earthquakes show a

**Fig. 3** Average time difference of successive earthquake occurrences



consistent and relatively low occurrence throughout the day. The occurrence of "Great" earthquakes appears sporadic and less frequent during most hours. These findings emphasize the importance of considering the temporal aspect when analysing earthquake data. The observed patterns suggest that seismic activity exhibits temporal fluctuations and highlights the need to study the underlying causes and mechanisms driving these variations.

#### 4.2 Interoccurrence time analysis

The Interoccurrence Time Analysis (ITA) involves examining the time intervals or time gaps between consecutive earthquake events. This analysis is useful for understanding the patterns and behaviours of earthquake occurrences over time. Figure 3 shows the average time difference of successive earthquake occurrences for each class of earthquake.

Exploring the earthquake interoccurrence times reveals distinct patterns for different magnitudes. The analysis reveals that on average, there is an interval of approximately 436 days and 12 h between successive Great earthquakes. These seismic events are of significant magnitude and tend to occur at relatively infrequent intervals, separated by several months. Such extended periods between Great earthquakes indicate their potential to cause substantial damage and impact regions with prolonged seismic activity. For Major earthquakes, the average interoccurrence time is approximately 31 days and 11 h. This suggests a relatively shorter duration between successive Major earthquakes compared to Great earthquakes. Major earthquakes are powerful and can cause significant damage to buildings and structures, making their more frequent occurrence a concern for seismic hazard assessment and preparedness efforts. For Minor earthquakes, on average, have an interoccurrence time

of approximately 22 min. This remarkably short duration indicates that Minor earthquakes occur in rapid succession, with very little time between individual events. While Minor earthquakes may not cause significant damage, their frequent occurrences contribute valuable seismic data for monitoring and research.

The moderate earthquakes have an average interoccurrence time of approximately 1 day and 9 h. This duration signifies a longer interval compared to Minor earthquakes but still suggests a relatively frequent occurrence. Moderate earthquakes can cause slight damage to buildings and structures, making their interoccurrence pattern crucial for seismic risk assessment and mitigation planning. The analysis reveals that Strong earthquakes have an average interoccurrence time of approximately 4 days and 10 h. This duration indicates a less frequent occurrence compared to Minor and Moderate earthquakes. Strong earthquakes have the potential to cause significant damage in highly populated areas, and their interoccurrence patterns contribute to understanding regional seismic activity. Very Minor earthquakes have an average interoccurrence time of approximately 3 min. This finding indicates that Very Minor earthquakes occur in rapid succession, with almost no time between individual events. While they may not be felt by humans, their frequent occurrences provide valuable data for seismic monitoring and research.

The earthquake interoccurrence time analysis by decade for different earthquake classifications reveals intriguing patterns and trends in seismic activity over the past 13 decades (from 1900 to 2020s). The analysis of interoccurrence time between consecutive earthquake events reveals insightful patterns for each magnitude category. For Great earthquakes, the average time difference varied significantly over the decades. The 1920s witnessed longer intervals, around 661 days



and 11 h, while the 2020s had shorter intervals of approximately 80 days and 11 h, indicating a higher frequency in recent years. Major earthquakes generally showed shorter interoccurrence times compared to Great earthquakes. The 1910s had the longest average time difference, about 41 days and 11 h, while the 1990s recorded the shortest, approximately 23 days and 12 h, indicating increased activity during that decade.

Furthermore, variations were also observed for Minor earthquakes. The 1920s had the longest average time difference, around 75 days and 13 h, whereas the 2020s saw a significant decrease to as short as 4 min, indicating a substantial increase in frequency. Moderate earthquakes showed a consistent average time difference, with the 1900s having the longest, around 35 days and 10 h, and later decades averaging around 9 h, suggesting a consistent level of seismic activity. For Strong earthquakes, the interoccurrence time remained stable across decades, averaging approximately 3 to 4 days and 10 h, indicating a consistent occurrence rate throughout the past 13 decades.

In general, the analysis of earthquake interoccurrence times indicates that earthquakes of different magnitudes exhibit distinct patterns in their occurrences. Great earthquakes, being of significant magnitude, tend to occur at relatively infrequent intervals, separated by several months. On the other hand, Minor earthquakes occur rapidly and successively, with very little time between individual events.

### 4.3 Spatial patterns, hotspots and clusters in global earthquakes

The spatial distribution of global earthquakes from 1900 to 2023 provides valuable insights into the occurrence and geographical patterns of seismic activity around the world over the years. Figure 4 reflects the distribution for all classes of earthquakes for significant earthquakes (magnitude  $\geq 5.5$ ). By visualizing the spatial distribution, we can identify regions that are more prone to seismic events and observe any potential trends or clustering of earthquakes in specific areas.

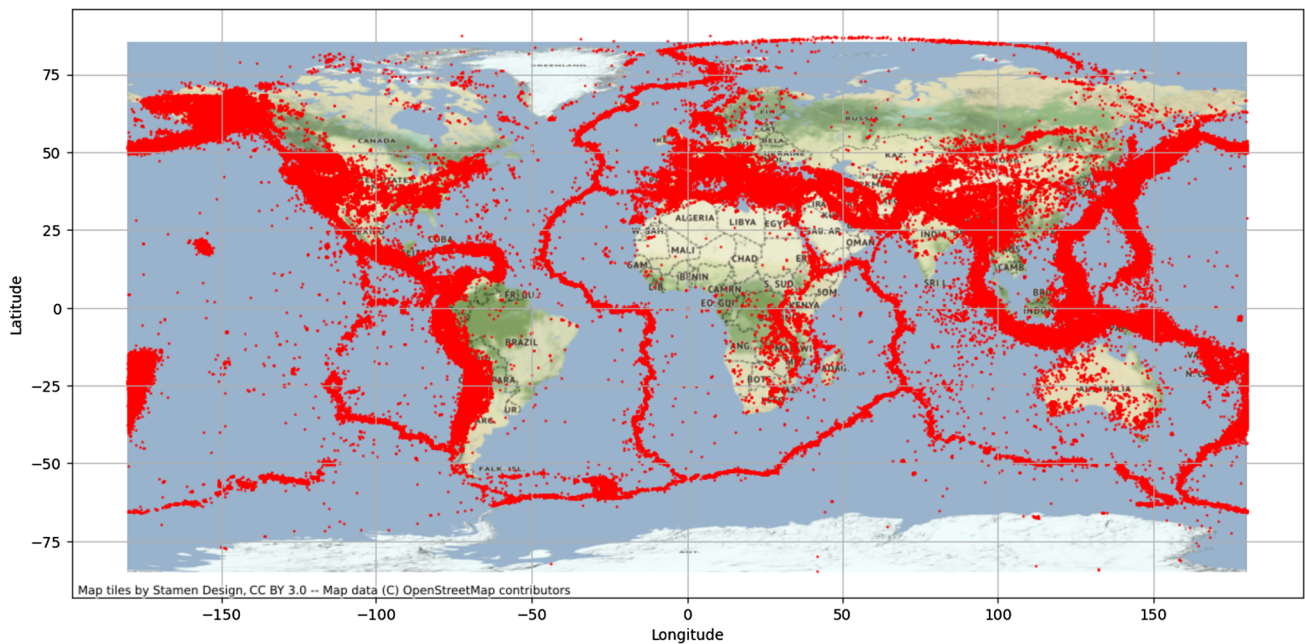
The global earthquake distribution reveals certain hotspots, with regions along the western coasts of North and South America, the central Atlantic Ocean, the Himalayan region, and Eastern Asian countries like Indonesia, Japan, and Korea being more susceptible to seismic activity. California and Alaska record the highest earthquake counts across various magnitudes, with California showing significant 'Minor', 'Moderate', and 'Strong' earthquakes, while Alaska leads in 'Great' and 'Major' earthquakes. 'Very Minor' and 'Minor' earthquakes dominate in many regions, providing valuable data for seismic monitoring. Meanwhile, Indonesia and Japan experience more significant seismic events, including 'Moderate', 'Strong', and

'Major' earthquakes. Regions like Chile, Indonesia, Japan, and the USA also encounter 'Great' earthquakes, emphasizing the need for monitoring and preparedness in high-risk areas. Certain regions, including Greece, Turkey, Iran, and Chile, exhibit a higher frequency of 'Moderate' and 'Strong' earthquakes, emphasizing the significance of seismic activity and the need for risk assessment and mitigation. The distribution of earthquake classes varies across regions, with some experiencing predominantly 'Very Minor' and 'Minor' earthquakes, while others face more significant seismic events. Prioritizing regions for earthquake preparedness and risk mitigation based on their seismic potential is crucial. Continuous monitoring and planning in seismically active regions like California and Alaska enhance community safety and resilience.

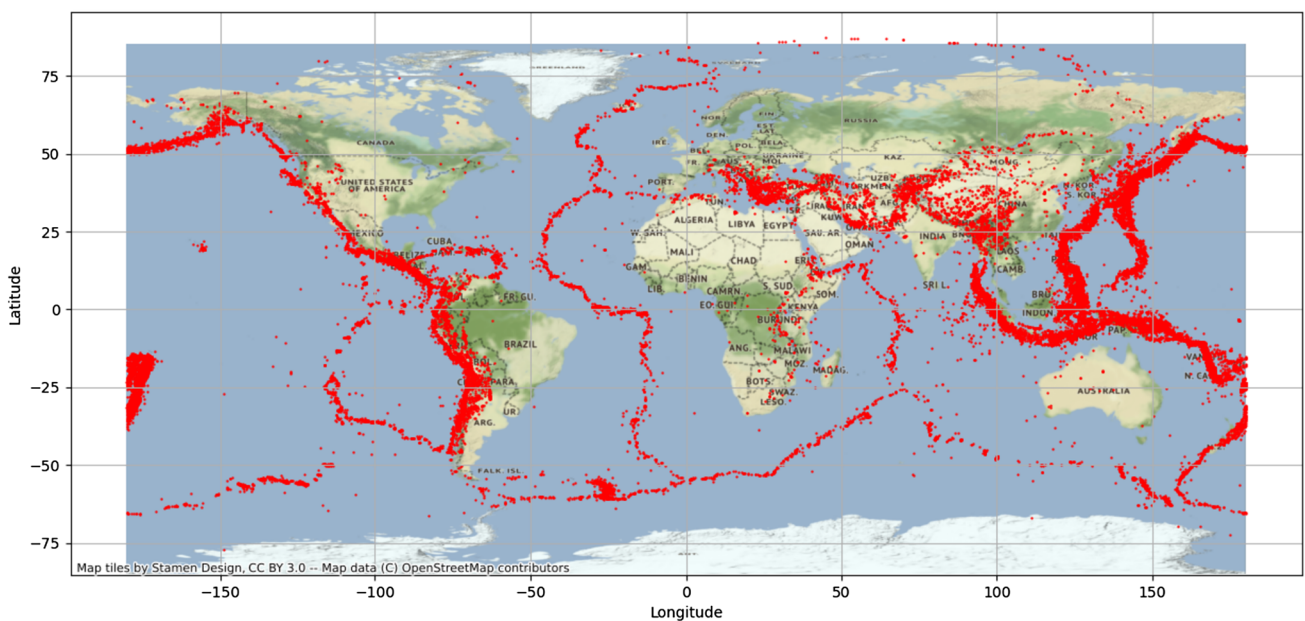
The significant earthquake counts by continents provide valuable insights into the distribution of seismic activity across different regions of the world. The results reveal valuable insights into the distribution and relative seismicity of earthquakes across different continents. Asia emerges as the most seismically active continent with a substantial count of 3971 earthquakes, representing approximately 47.81%. This high earthquake count in Asia is primarily influenced by the presence of multiple tectonic plate boundaries, including the collision of the Indian Plate with the Eurasian Plate and subduction zones in the Pacific Ring of Fire. The collision and subduction processes lead to frequent earthquakes in countries like India, China, Nepal, Japan, Indonesia, and the Philippines, making Asia a hotspot for seismic activity.

South America follows closely with 1548 significant earthquakes, constituting around 18.64% of the total. The western coast of South America, along the Peru–Chile Trench, experiences powerful earthquakes due to the subduction of the Pacific Plate beneath the South American Plate. This subduction zone has historically produced devastating earthquakes, such as the 1960 Valdivia earthquake. Additionally, the collision of the South American Plate with the Nazca Plate contributes to seismic activity in the Andes mountains. North America, with 1015 earthquakes, accounting for approximately 12.22% of the total, exhibits seismic activity along the western part, primarily along the San Andreas Fault system. The interaction of the Pacific Plate with the North American Plate creates a seismically active region, affecting areas like California. While not as active as Asia or South America, North America experiences moderate to strong earthquakes due to these tectonic interactions.

Oceania, which includes Australia, New Zealand, and the Pacific islands, accounts for 925 significant earthquakes, representing about 11.14%. This region is situated along the Pacific Ring of Fire, characterized by numerous tectonic plate boundaries and subduction zones. The Tonga Trench and the Kermadec Trench are some of the subduction zones contributing to seismic activity in Oceania. In contrast,



(a) Spatial distribution of global earthquakes from 1900 to 2023

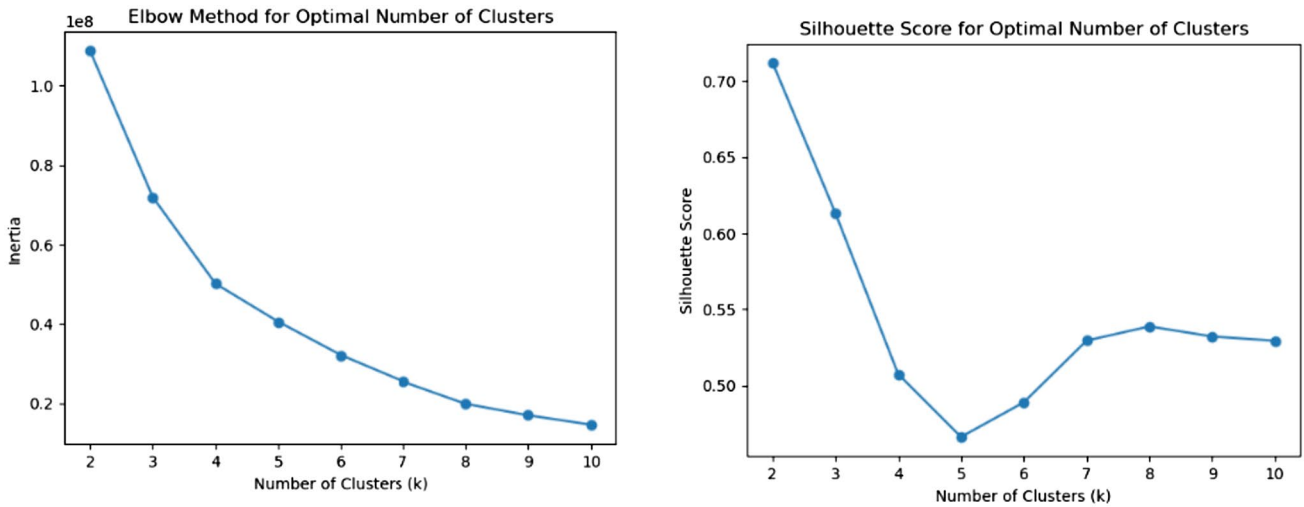


(b) Spatial distribution of significant global earthquakes from 1900 to 2023

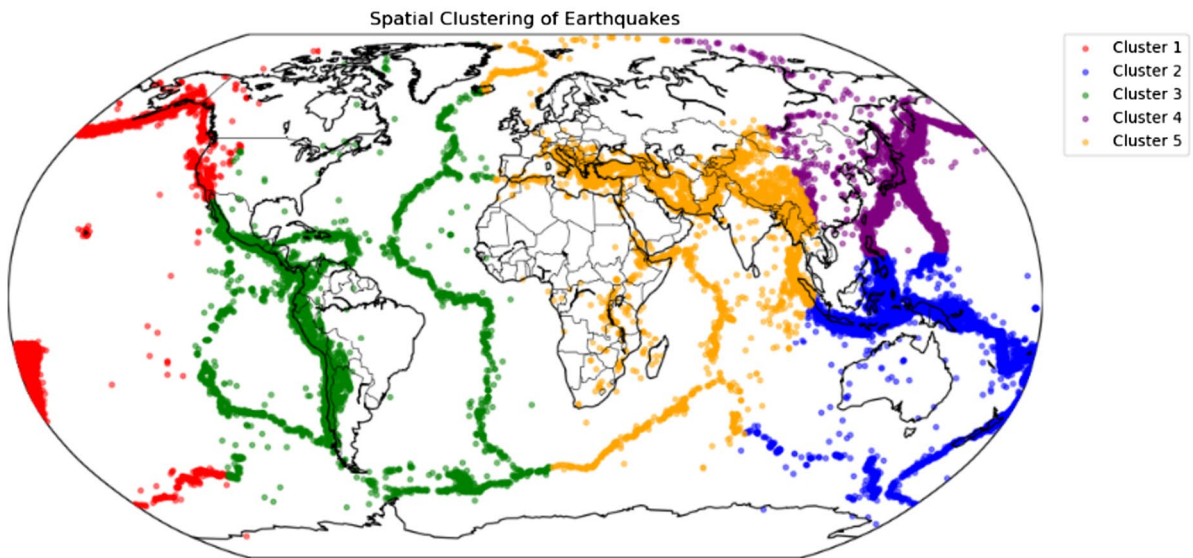
**Fig. 4** Global earthquake distribution (1900–2023)

Europe accounts for 613 earthquakes, making up approximately 7.38% of the total. While Europe is generally considered a seismically less active region compared to others, it still experiences notable seismicity, particularly around the Mediterranean region and the Alpine-Himalayan belt. The Mediterranean region, influenced by the interaction of the African Plate, Eurasian Plate, and Anatolian Plate, witness earthquake occurrences in countries like Turkey and Greece.

Africa, with 230 significant earthquakes, which constitutes around 2.77% of the total, has relatively fewer seismic events compared to other continents. Most of Africa is located on the stable African Plate, with fewer active plate boundaries. However, regions along the eastern edge of Africa, such as the East African Rift, experience seismic activity due to tectonic movement and rifting. Antarctica, with only 3 significant earthquakes, making up a mere



(a) Elbow method silhouette score plots



(b) Spatial clustering map of significant global earthquakes

**Fig. 5** Clustering Analysis of Significant Global Earthquakes

0.04% of the total, has the lowest seismic activity among the continents. This is not unexpected, given Antarctica’s ice-covered and relatively isolated nature, with minimal tectonic activity. Inferences drawn from the significant earthquake counts by continents reveal that regions situated near active plate boundaries, such as Asia and South America, exhibit higher earthquake counts. These areas are prone to powerful and potentially destructive seismic events. On the other hand, regions with fewer active plate boundaries, like Africa and Antarctica, have lower earthquake counts, indicating a lower frequency of seismic activity.

The cluster analysis revealed that the optimum number of clusters for the earthquake dataset was determined to be 5 based on the elbow method and silhouette score analysis. Figure 5 shows the clustering analysis of significant global earthquakes. The figure presents the Elbow method silhouette score plots and the corresponding spatial clustering map.

Hence, earthquakes in the dataset can be effectively grouped into 5 distinct clusters, each representing a specific pattern or characteristic in seismic activity. Upon examining the spatial distribution of the clustered earthquakes, a significant finding emerged: the clusters were predominantly centered around the boundaries of various major tectonic

plates. The presence of earthquake clusters aligned with tectonic plate boundaries further suggests a strong correlation between seismic activity and plate interactions. Earthquakes tend to be more frequent and intense in regions where tectonic plates converge, diverge, or slide past one another. These interactions generate stress and strain in the Earth's crust, leading to the release of energy in the form of seismic events. The distribution of earthquake clusters along tectonic plate boundaries is a significant confirmation of the tectonic theory of earthquakes, which posits that plate movements are a primary driving force behind seismic activity [36].

Inferences can be drawn from these results to enhance our understanding of earthquake patterns and their association with tectonic plate movements. By identifying and characterizing these clusters, seismologists and geologists can gain valuable insights into the geological processes driving seismicity. This information can be utilized to improve earthquake monitoring and hazard assessment, which is crucial for enhancing disaster preparedness and response. Moreover, the identification of seismic hotspots around tectonic plate boundaries can aid in the assessment of seismic risk in regions prone to large earthquakes. Understanding the distribution and behaviour of earthquake clusters helps to prioritize resources and implement effective mitigation strategies in areas with higher seismic activity.

## 5 Comparative with existing methods

The presented study employed a methodology consisting of exploratory data analysis, temporal dynamics exploration, spatial pattern analysis, and K-means clustering. This approach allowed to gain insights into seismic trends and patterns at a global scale. In comparison to [23] while their focus was on Neutrosophic K-means clustering for prediction, our study encompassed a broader analysis pipeline, including temporal and spatial exploration, offering comprehensive insights into earthquake trends and patterns. Similarly, compared to [33] who concentrated on spatial clustering techniques within a specific region, our research aimed to understand global earthquake patterns and trends.

The present research's use of K-means clustering provided additional insights into distinct earthquake hotspots along tectonic plate boundaries, reinforcing the importance of plate interactions. In line with [32], our research also delved into spatial pattern analysis, identifying seismic clusters. However, the focus of the present study on analysing global earthquake trends further contributes to the understanding of seismic activity on a worldwide scale. The authors of [33] focused on earthquake patterns within Iraq and utilized various statistical techniques. Our study, on the other hand, employed a combination of data science techniques to analyse a century-long dataset of global earthquake

occurrences, encompassing temporal dynamics, spatial patterns, and clustering behaviour worldwide.

From [29], the analysis of seismic spatial distribution in China underlines the importance of understanding regional seismic patterns, as demonstrated by the identification of positive spatial autocorrelation and agglomeration patterns in specific time intervals. This aligns with the present research, which aims to analyse global earthquake trends and patterns, and emphasizes the significance of spatial patterns and agglomeration at both regional and global scales. Comparing with the study by [30] their classification of seismic clusters into persistent and burst types and the consideration of multiple spatial factors highlights the complexity of earthquake clustering mechanisms.

This aligns with the present research's approach of utilizing K-means clustering to reveal distinct earthquake hotspots and trends, albeit on a broader global scale. Their insights into spatial factors contributing to cluster characteristics complement our findings by offering a deeper understanding of how seismic activity varies across different regions. The study presented by [12] introduced a modified model to study earthquake distributions, demonstrating the presence of self-organized criticality and long-range spatiotemporal correlations in seismic events. This complements the present research by showcasing different methodologies to analyse seismic patterns, and further strengthens the idea that critical behaviour and correlations play a significant role in earthquake occurrences.

In general, the research presented in this study expands beyond the scope of individual methodologies employed in these studies, presenting a holistic approach to understanding global earthquake trends and patterns. The findings of the present study contribute to the broader understanding of seismic activity, reinforcing the importance of spatial patterns, and temporal dynamics in earthquake occurrences.

## 6 Conclusions

The present research has significantly advanced our understanding of global earthquake patterns through the integration of diverse data science methodologies and spatial analyses. By meticulously exploring a century-long dataset from the USGS, we uncovered intricate spatiotemporal relationships, identified seismic hotspots, and delved into the temporal dynamics of earthquake occurrences through innovative Interoccurrence Time Analysis. These insights have profound implications for earthquake prediction, hazard assessment, and disaster mitigation efforts worldwide. The findings underscore the importance of continuous monitoring and research to enhance our understanding of seismic activity and inform robust disaster preparedness strategies. Moreover, this study highlights the need for further

investigations focusing on the interplay between seismic events and external factors such as climate change, volcanic activity, and human-induced activities, to provide a more comprehensive understanding of earthquake dynamics. The future scope of this research lies in further advancing data science methodologies for a more nuanced analysis of global earthquake trends and patterns. Specifically, there is potential for refining existing spatial analyses, exploring advanced statistical approaches, and integrating emerging technologies to enhance the interpretability and accessibility of seismic data. Future investigations could delve into the development of more sophisticated real time visualization tools and interactive platforms, providing researchers and stakeholders with comprehensive and user-friendly insights.

### Declarations

**Conflicts of interest** Authors have no conflict of interest to declare.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

### References

- Mousavi SM, Ellsworth WL, Zhu W, Chuang LY, Beroza GC (2020) Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat Commun*. <https://doi.org/10.1038/s41467-020-17591-w>
- Kanber M, Santur Y (2023) Time series and data science pre-processing approaches for earthquake analysis. *Eur J Sci Technol*. <https://doi.org/10.31590/ejosat.1265261>
- Bantidi TM, Nishimura T (2022) Spatio-temporal clustering of successive earthquakes as inferred from analyses of global CMT and NIED F-net catalogs. *Earth, Planets Sp*. <https://doi.org/10.1186/s40623-022-01677-4>
- Mondol M (2021) “analysis and prediction of earthquakes using different machine learning techniques,” [Online]. Available: [www.kaggle.com/usgs/earthquake-database](http://www.kaggle.com/usgs/earthquake-database)
- Dal Zilio L, Ampuero J-P (2023) Earthquake doublet in Turkey and Syria. *Commun Earth Environ* 4(1):1–4. <https://doi.org/10.1038/s43247-023-00747-z>
- Naddaf M (2023) Turkey–Syria earthquake: what scientists know. *Nature* 614(7948):398–399. <https://doi.org/10.1038/D41586-023-00364-Y>
- Toda S et al (2023) Stress change calculations provide clues to aftershocks in 2023 Türkiye earthquakes. *Temblor*. <https://doi.org/10.32858/TEMBLOR.295>
- Shan W, Wang Z, Teng Y, Wang M (2021) Temporal and spatial evolution analysis of earthquake events in California and Nevada based on spatial statistics. *ISPRS Int J Geoinf*. <https://doi.org/10.3390/ijgi10070465>
- Rahadi I et al (2022) Applying K-means algorithm for clustering analysis earthquakes data in West Nusa Tenggara. *Indones Phys Rev* 5(3):197–207
- Bao Z, Zhao J, Huang P, Yong S, Wang X (2021) A deep learning-based electromagnetic signal for earthquake magnitude prediction. *Sensors (Basel)*. <https://doi.org/10.3390/S21134434>
- T. J. Hsieh, G. K. Chen, and K. L. Ma (2010) “Visualizing field-measured seismic data,” *IEEE Pacific Visualization Symposium 2010, PacificVis 2010—Proceedings*, pp. 65–72. Doi: <https://doi.org/10.1109/PACIFICVIS.2010.5429610>.
- Ferreira DSR et al (2022) Spatiotemporal analysis of earthquake occurrence in synthetic and worldwide data. *Chaos Solitons Fractals*. <https://doi.org/10.1016/j.chaos.2022.112814>
- Zhang B, Hu Z, Wu P, Huang H, Xiang J (2023) EPT: a data-driven transformer model for earthquake prediction. *Eng Appl Artif Intell* 123:106176. <https://doi.org/10.1016/J.ENGAPPAL.2023.106176>
- Tehseen R, Farooq MS, Abid A (2020) Earthquake prediction using expert systems: a systematic mapping study. *Sustainability (Switzerland)*. <https://doi.org/10.3390/su12062420>
- Zheng Z, Shi HZ, Zhou YC, Lu XZ, Lin JR (2022) “Earthquake impact analysis based on text mining and social media analytics
- Saba S, Ahsan F, Mohsin S (2017) BAT-ANN based earthquake prediction for Pakistan region. *Soft Comput* 21(19):5805–5813. <https://doi.org/10.1007/s00500-016-2158-2>
- Weber GH, Schneider M, Wilson DW, Hagen H, Hamann B, Kutter BL (2003) “Visualization of experimental earthquake data
- Banagr P, Gupta D, Gaikwad S, Marekar B, Patil J (2020) Earthquake prediction using machine learning algorithm. *Int J Recent Technol Eng (IJRTE)* 8(6):4684–4688. <https://doi.org/10.35940/ijrte.E9110.018620>
- Yousefzadeh M, Hosseini SA, Farnaghi M (2021) Spatiotemporally explicit earthquake prediction using deep neural network. *Soil Dyn Earthq Eng*. <https://doi.org/10.1016/j.soildyn.2021.106663>
- Sruthi AAVL, Bhargavi R, Gospati VR (2021) Analysis of seismic data using machine learning algorithms. *IOP Conf Ser Mater Sci Eng* 1070(1):012042. <https://doi.org/10.1088/1757-899x/1070/1/012042>
- Al Banna MH et al (2020) Application of artificial intelligence in predicting earthquakes: state-of-the-art and future challenges. *IEEE Access* 8:192880–192923. <https://doi.org/10.1109/ACCESS.2020.3029859>
- Haridas M (2015) “Big data applications for earthquake predictions Haridas M
- Je. Ricardo et al., (2021) “Neutrosophic sets and systems {Special issue: impact of neutrosophic scientific publication in latin neutrosophic k-means for the analysis of earthquake data in Ecuador
- Gupta P, Sharma AK (2011) A framework for hierarchical clustering based indexing in search engines. *Int J Inf Technol* 3(2):329–334
- Kumar TV, Guruprasad HS (2015) Clustering of web usage data using hybrid k-means and PACT algorithms. *Int J Inf Technol* 7(2):871–876
- Rastogi R, Mondal P, Agarwal K, Gupta R, Jain S (2015) GA based clustering of mixed data type of attributes (numeric, categorical, ordinal, binary and ratio-scaled). *Int J Inf Technol* 7(2):861–866
- Lamba A, Kumar D (2016) Optimization of KNN with firefly algorithm. *Int J Inf Technol* 8(2):997–1003

28. Soleimani NS, Pashazadeh ZS (2014) Classification of databases and methods for seismic data analysis and earthquake prediction". *Int J "Tech Phys Probl Eng IJTPE Issue 18:18–26*
29. Cao Z et al (2022) Spatial distribution analysis of seismic activity based on GMI, LMI, and LISA in China. *Open Geosci 14(1):89–97*. <https://doi.org/10.1515/geo-2020-0332>
30. Yang J, Cheng C, Song C, Shen S, Zhang T, Ning L (2019) Spatial-temporal distribution characteristics of global seismic clusters and associated spatial factors. *Chin Geogr Sci 29(4):614–625*. <https://doi.org/10.1007/s11769-019-1059-6>
31. Vijay RK, Nanda SJ (2023) Earthquake pattern analysis using sub-sequence time series clustering. *Pattern Anal Appl 26(1):19–37*. <https://doi.org/10.1007/s10044-022-01092-1>
32. Shen S et al (2018) Spatial distribution patterns of global natural disasters based on biclustering. *Nat Hazards 92(3):1809–1820*. <https://doi.org/10.1007/s11069-018-3279-y>
33. Al-Heety EAMS (2016) Spatial analysis of earthquakes in Iraq using statistical and data mining techniques. *Iraqi Geol J Al-Heety 39:1–15*
34. Jamsandekar SS, Mudholkar RR (2014) Fuzzy classification system by self generated membership function using clustering technique. *Int J Inf Technol 6(1):697–704*
35. Khan M, Quadri SMK (2013) Effects of using filter based feature selection on the performance of machine learners using different datasets. *Int J Inf Technol 5(2):597–603*
36. Ryan HF, Ross SL, Graymer RW "Earthquakes, faults, and tectonics summary and introduction