

DataSources:**Website:**

Table lists the largest earthquakes by year.

https://en.wikipedia.org/wiki/Lists_of_earthquakes

Columns listed in table are Year,Magnitude,Location Depth (km),MMI,Notes,Deaths,Injuries,Event,Date

API:This API allows custom searches for earthquake information using a variety of parameters.Earthquake Data provided by USGS.gov is very neat and complete. You can arrange your query based on parameters like coordinates, dates, alert level, tsunami causing or not, duration,Magnitude,latitude and logitude.

<https://earthquake.usgs.gov/fdsnws/event/1/>

FlatFile:

The Earthquakes dataset is a very good source to study the patterns and trends in seismic (earthquake) activity at global and regional levels.

Datasets contain records of 782 earthquakes from 1/1/2001 to 1/1/2023. The meaning of all columns is as follows:

title: title name given to the earthquake

magnitude: The magnitude of the earthquake

date_time: date and time

cdi: The maximum reported intensity for the event range

mmi: The maximum estimated instrumental intensity for the event

alert: The alert level - "green", "yellow", "orange", and "red"

tsunami: "1" for events in oceanic regions and "0" otherwise

sig: A number describing how significant the event is. Larger numbers indicate a more significant event. This value is determined on a number of factors, including: magnitude,maximum MMI, felt reports, and estimated impact

net: The ID of a data contributor. Identifies the network considered to be the preferred source of information for this event.

nst: The total number of seismic stations used to determine earthquake location.

dmin: Horizontal distance from the epicenter to the nearest station

gap: The largest azimuthal gap between azimuthally adjacent stations (in degrees). In general, the smaller this number, the more reliable is the calculated horizontal position of the earthquake. Earthquake locations in which the azimuthal gap exceeds 180 degrees typically have large location and depth uncertainties

magType: The method or algorithm used to calculate the preferred magnitude for the event

depth: The depth where the earthquake begins to rupture

latitude / longitude: coordinate system by means of which the position or location of any place on Earth's surface can be determined and described

location: location within the country

continent: continent of the earthquake hit country

country: affected country

Based on the datasets and API selected,Key focus areas include identifying the trends and patterns across various event parameters including magnitude, depth and significance of the events, type of magnitude measurement used, and tsunami occurrence etc.

Some of the common columns are listed out as magnitude, longitude, latitude, place happened and year. By these columns I need to derive some pattern among the occurrence of the earthquake happened. I want to focus on identifying the event patterns by year, by month, by day of the month, and by hour of the day and will try to identify and visualise the quake events by country and analyse the top 3 quake prone countries in detail.

Planning to leverage bar diagrams, pie charts, histograms, boxplots, scatterplots to visualise and decide on the pattern. Number missing values in the dataset are 1246. Major Challenge for this dataset is filling large number of missing values by a logical approach.