UNIVERSITÄT DUISBURG ESSEN

Offen im Denken

Digibau 3: Introduction to Data Science for Engineers

Assignment 2

Mana Mohajer ▪ 24.11.2021

DIGIBAU

Institut für Digitalisierung im Bauwesen

Assignment 2

DIGIBAU
Institut für Digitalisierung im Bauwesen

UNIVERSITÄT DUISBURG ESSEN
Offen im Denken

In this assignment you experience working with real-time data of earthquakes.

"The number of earthquakes with small magnitudes are higher compared to the number of earthquakes with large magnitudes". This is a qualifying expression which we can agree on even without any seismological knowledge.

In order to quantify the seismicity, Gutenberg-Richter recurrence relationship is defined as following and can be used for any region and any one-year time interval:

$$\log(N) = a - b\,M$$

In which N is the **cumulative number** of earthquakes with magnitude higher than M. Two constants, a and b, derived from regression analysis. Our goal is to derive a and b through performing a linear regression on our desired real-time data of earthquakes.

Assignment 2

DIGIBAU
Institut für Digitalisierung im Bauwesen

UNIVERSITÄT DUISBURG ESSEN
Offen im Denken

Your tasks:

1. The first step is to get the data from the proper platform. Most of the countries have own earthquake catalogues. We use the earthquake catalogue provided by 'GeoForschungsZentrum (GFZ)'.

If you open the webpage of GEOFON – GFZ , you can search in the earthquake catalogue and get your data. It is required to customize the following filters:

Search Criteria: Select any one-year time interval you would like and a minimum magnitude of 2. It is not necessary to insert the specific latitude and longitude ranges. Leave it empty. Your search will be on data for the whole world.

Output Format: Select HTML. The default of 40 for the number of events which have your criteria is enough but you are free to increase it to have access to more events.

Assignment 2

DIGIBAU
Institut für Digitalisierung im Bauwesen

UNIVERSITÄT DUISBURG ESSEN
Offen im Denken

2. After clicking on **search,** you see a list of earthquakes with your criteria. Select one of them as you like. You will see a part of the map where the epicenter of this earthquake in shown by red circles. On the bottom of this image, go to **Nearby events** and select **Text(CSV)**. You are finished with getting the data related to all occurred earthquakes near the selected epicenter. Download the data and save it. Now it is time to read this data in your Jupyter notebook and derive a and b constants!

3. The first row of your data looks like as following:

#EventID|Time|Latitude|Longitude|Depth/km|Author|Catalog|Contributor|ContributorID|MagType|Magnitude|MagAuthor|EventLocationName|EventType

Extract the available values in the ‘MagType’ and ‘Magnitude’ columns. The magnitude of earthquake can be reported in different types. Therefore, you see different notations in ‘MagType’ column. For example, M_L is used for Richter but rarely used in scientific content.

Assignment 2

DIGIBAU
Institut für Digitalisierung im Bauwesen

UNIVERSITÄT DUISBURG ESSEN
Offen im Denken

Count the repetition of each notation.
Extract and report the ‘MagType’ with the maximum repetition. Call it MagType_max.
Extract all the corresponding ‘Magnitude’ values to MagType_max.

4. Define the different intervals for you extracted ‘Magnitude’ values in step 3 based on the minimum and maximum values. For example, if maximum magnitude value in your data is 7:

(2-2.9) , (3-3.9), ..., (6-7)

Count the number of occurred earthquakes in each magnitude interval. Accordingly, calculate N as the **cumulative number** of earthquakes with magnitude higher than M.

5. Plot your scattered data containing the points of (M, log(N)). Put labels for your axes.

Assignment 2

DIGIBAU
Institut für Digitalisierung im Bauwesen

UNIVERSITÄT DUISBURG ESSEN
Offen im Denken

6. Use the available function in SciPy library to fit a line to your data. Practically, the function performs a linear regression analysis.

7. Plot a figure containing the fitted line and the scattered data. Use legend for both. Report the equation of the fitted line representing the Gutenberg-Richter relationship:

$$\log(N) = a - b\,M$$

The deadline to work on this assignment is 07.12.2021. Name your Jupyter notebook as Assignment02_NAME_FAMILYNAME **and upload it in Moodle. Upload the file of your downloaded data as well.**