

Matière : DEVELOPPEMENT PERSONNEL

Enseignant : KAZARIAN Avétis

Echéance : dimanche 19 mars 2023 à 23h59

Description

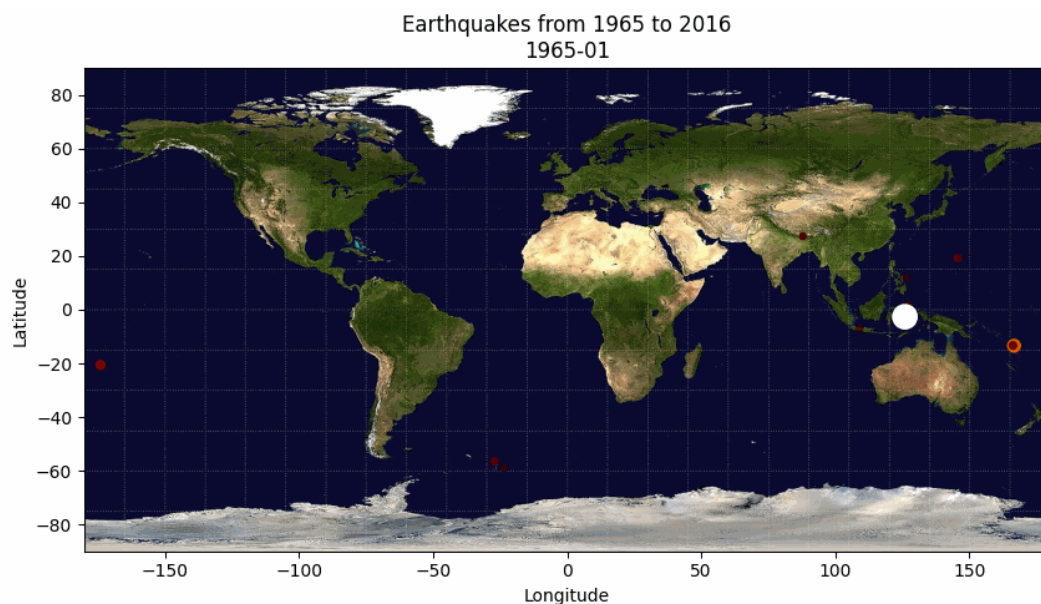
Write a post intended for a social media account about a BigData/AI subject showing your proficiency and skillset around that very subject with the following constraints: - it should contain at least one image, either purely decorative or illustrative/explanatory; - it must be at least 2 paragraphs; - it must be at least 500 words; - you have 3 weeks. Alternatively it can be a short introduction to an online example of running app/code showcasing your skills with a open source project that YOU have made (link to the repository is mandatory). The format must be a PDF file with the following name: firstName_lastName.pdf

Rendu : Rami EL-WAZZI

Hello network, I wanted to share this project that I am quite proud of. I have been working on it for the past 3 weeks, to challenge myself and improve my skills in Data Science, specifically improving my data Scientist skills. To do that I chose a dataset on Kaggle that includes data about significant earthquakes that have happened since 1st Jan 1965 and up to 31st Dec 2016. Here is the link to get the dataset : <https://www.kaggle.com/usgs/earthquake-database>.

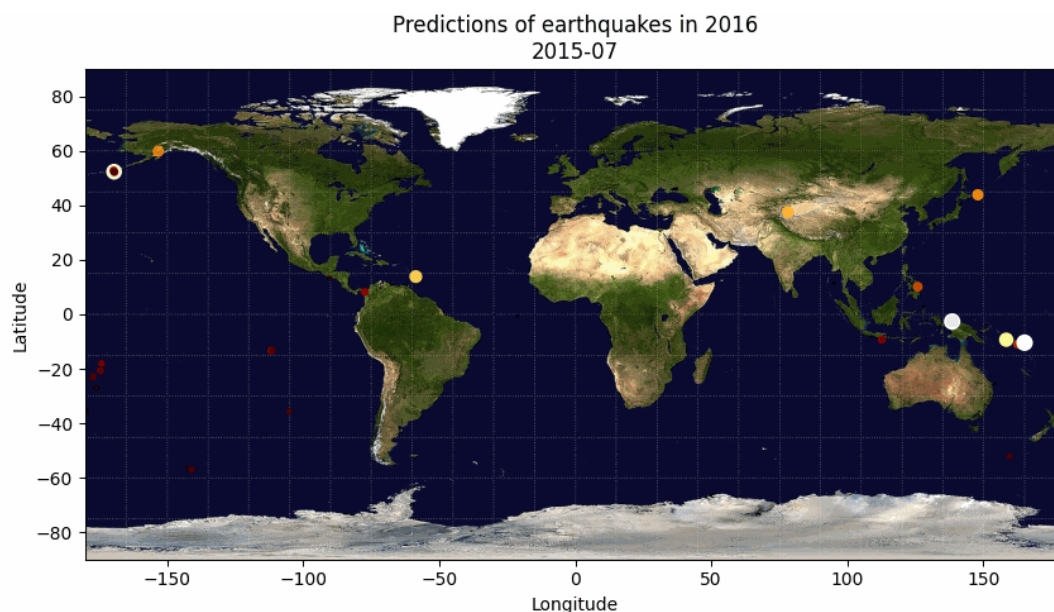
It contains 23K rows which isn't all that much considering the timespan, but it is what I worked on, as the result isn't as ant to me as the challenge. With this data frame I was going to create a visualization of earthquakes.

23K rows spread over 21 columns, including the magnitude, latitude, longitude, depth, and the type of earthquake (including nuclear explosions which are considered a type of earthquake).



Then I wanted to see if I could predict the magnitude of an earthquake based on the location, depth, and the type of earthquake. I used sklearn to train a model of Linear Regression, I used the data before 2016 to train the model, and the data from 2016 to test the model. Using the model, I was able to create a dataset of predicted longitude, latitude, depth, and magnitude. I then used this dataset to create a visualization of the predicted earthquakes over the visualization of the actual earthquakes that have happened in 2016, to see if my model was accurate. You can see here a gif representing all earthquakes that happened every month before 1st of January 2016 in each frame.

Then starting January 2016, the added circles (in blue and green) represent the predicted earthquakes. The legend is the same as the previous gif, but instead of brown -> white representing magnitudes that range from 4.5 to 9, for the predicted. earthquakes the colors are blue -> green.



As you can see, the model is not very accurate, but it is still interesting to see the difference between the actual earthquakes and the predicted earthquakes, as the model, while it hasn't accurately predicted earthquakes, is still following the general shape of the tectonic plates. The model is not very accurate because the dataset is not very big, and also, because the data used include 5 columns to be predicted for more than 10 columns of input. I had to cut corners to make the model trainable in a reasonable amount of time. I linked the code i used to create the model and the visualization.

I have ideas to improve the model.

Some of them include, crossing the data with other datasets, to get more information about the earthquakes, also instead of using the data to calculate the mean Longitude, Latitude, Magnitude, and Depth, of ALL earthquakes that happened up to only 3 days before the predicted value. I could use the data to calculate those values by regions, and for a larger span of time.


those ideas will have to wait until I have more time to work on it. or a better computer :D.

Anyway I hope you enjoyed this post, thanks for your interest if you have read it, and if you have any questions or suggestions, feel free to contact me.

Link to my github repo: https://github.com/Rami-lwz/earthquake_predictions

Voici le post, j'ai inclus le lien du repo, je vous invite à y jeter un œil.

Voici un tutoriel pour mieux voir mon travail si vous en avez l'envie :

	Rami-lwz Create README.md	c-fac40a 28 minutes ago	🕒 2 commits
📁 csv	first commit	32 minutes ago	
📁 gifs_final	first commit	32 minutes ago	
📁 html	first commit	32 minutes ago	
📁 pythton	first commit	32 minutes ago	
📄 README.md	Create README.md	28 minutes ago	

Dans « csv » vous avez le dataset, ainsi que les versions après avoir clean les données.

Dans « gifs_final » vous avez les gifs des visualisations.

Dans « html » vous avez un petit fichier html crée par chatgpt, qui ressemble a un post linkedin.
(Puisque je ne posterai pas ça, je compte améliorer mon travail avant de le poster si je le fais)

Dans « pythton » vous avez « earthquakes.py » qui permet de clean le dataset pour êtres utilisés pour la visualisation. Et « earthquake_animation.ipynb » qui permet de créer les gifs. Si vous voulez essayer de lancer les fichiers vous aurez besoin de python 3.10 et des librairies suivantes numpy, pandas, matplotlib, Seaborn, time, Datetime, sklearn.

LinkedIn Post Example

 Rami EL-WAZZI

314159 followers

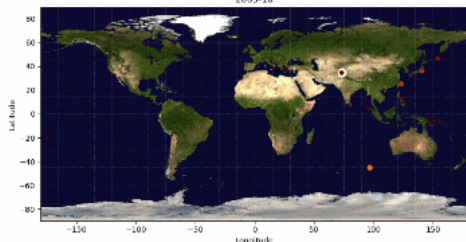
2 days ago · 🌐

Hello network, I wanted to share this project that I am quite proud of. I have been working on it for the past 3 weeks, to challenge myself and improve my skills in DataScience, specifically improving my data Scientist skills. To do that I chose a dataset on Kaggle that includes data about significant earthquakes that have happened since 1st Jan 1965 and up to 31st Dec 2016. Here is the link to get the dataset : <https://www.kaggle.com/usgs/earthquake-database>

It contains 23K rows which isn't all that much considering the timespan, but it is what I worked on, as the result isn't as important to me as the challenge. With this dataframe I was going to create a visualization of earthquakes. 23K rows spread over 21 columns, including the magnitude, latitude, longitude, depth, and the type of earthquake (including nuclear explosions which are considered a type of earthquake).

You can see here a gif representing all earthquakes that happened every month in each frame. The circles get bigger and whiter with the magnitude. The more a circle is transparent, the deeper it is (under the surface of the sea).

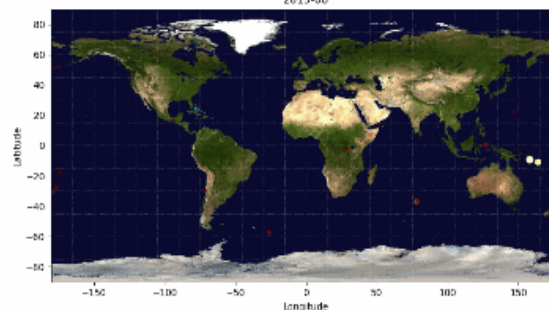
Earthquakes from 1965 to 2016
2005-10



Then I wanted to see if I could predict the magnitude of an earthquake based on the location, depth, and the type of earthquake. I used sklearn to train a model of Linear Regression. I used the data before 2016 to train the model, and the data from 2016 to test the model. Using the model I was able to create a dataset of predicted longitude, latitude, depth, and magnitude. I then used this dataset to create a visualization of the predicted earthquakes over the visualization of the actual earthquakes that have happened in 2016, to see if my model was accurate.

You can see here a gif representing all earthquakes that happened every month before 1st of January 2016 in each frame. Then starting January 2016, the added circles (in blue and green) represent the predicted earthquakes. The legend is the same as the previous gif, but instead of brown -> white representing magnitudes that range from 4.5 to 9, for the predicted earthquakes the colors are blue -> green.

Predictions of earthquakes in 2016
2015-08



As you can see, the model is not very accurate, but it is still interesting to see the difference between the actual earthquakes and the predicted earthquakes, as the model, while it hasn't accurately predicted earthquakes, is still following the general shape of the tectonic plates. The model is not very accurate because the dataset is not very big, and also because the data used include 5 columns to be predicted for more than 10 columns of input. I had to cut corners to make the model trainable in a reasonable amount of time. I linked the code I used to create the model and the visualization. I have ideas to improve the model.

Some of them include, crossing the data with other datasets, to get more information about the earthquakes, also instead of using the data to calculate the mean Longitude, Latitude, Magnitude, and Depth, of ALL earthquakes that happened up to only 3 days before the predicted value. I could use the data to calculate those values by regions, and for a larger span of time.

those ideas will have to wait until I have more time to work on it, or a better computer :D.

Anyway I hope you enjoyed this post, thanks for your interest if you have read it, and if you have any questions or suggestions, feel free to contact me.