Dear Data Analysts :hugging_face:**Stage 2 Task: Flood Prediction**Using all the data you can find, predict the likely date of the next flood in Lagos. ***Provide a detailed justification for your answer.***Submit your answer in a ***beautiful*** PDF using this [link](https://docs.google.com/forms/d/e/1FAIpQLSdwMnIFqc9pMIMRzLwfweRzZKNxHth3FZYzoMpmGNcFv7c0iA/viewform?usp=sf_link).

Maybe this data set can help:thinking_face:As a data analyst, you can use Power BI's built-in features to analyze and predict flood occurrences. Here's a simple approach:1. \*Create a calendar table\*: Create a table with dates for the year, including the date column from your dataset.  
2. \*Calculate precipitation totals\*: Use the "Measure" feature to calculate the total precipitation for each day.  
3. \*Create a flood indicator\*: Use a conditional statement (e.g., IF) to create a column that indicates if a flood occurred (Yes/No) based on a threshold (e.g., precipitation > 10 mm).  
4. \*Visualize flood history\*: Use a bar chart or line chart to show the flood indicator over time.  
5. \*Use the "Forecast" feature\*: Select the precipitation column and use the "Forecast" feature to predict future precipitation values.  
6. \*Create a flood prediction measure\*: Use a conditional statement to predict if a flood will occur (Yes/No) based on the forecasted precipitation values.  
7. \*Visualize flood predictions\*: Use a bar chart or line chart to show the predicted flood occurrences.Power BI's built-in features make it easy to perform these steps without requiring advanced data science knowledge. You can also use the "Quick Insights" feature to automatically generate predictions based on your data.Remember to adjust the threshold value and forecasting parameters to suit your specific needs. Current precipitation threshold for Lagos this year is 1936.2mm

the columns that are most relevant to flood prediction are:- Precipitation  
- Precipitation probability  
- Precipitation cover  
- Precipitation type  
- Sea level pressure

You can consider deleting the following columns:- Temp max  
- Temp min  
- Feels like Max  
- Feels like min  
- Dew  
- Humidity  
- Wind gust  
- Wind speed  
- Wind direction  
- Cloud cover  
- Visibility  
- Solar radiation  
- Solar energy  
- UV index  
- Severe risk  
- Sun rise  
- Sunset  
- Moon phaseThese columns are not directly related to flood prediction, and deleting them will simplify your dataset and reduce the risk of overfitting your model.The columns you should keep are:- Date  
- Precipitation  
- Precipitation probability  
- Precipitation cover  
- Precipitation type  
- Sea level pressureYou can use these columns to train a model that predicts the likelihood of flooding based on historical data. Additionally, you may want to consider creating new features that are derived from these columns, such as:- Total precipitation over a certain period (e.g. 24 hours)  
- Average precipitation probability over a certain period  
- Maximum precipitation intensity over a certain periodThese derived features can help improve the accuracy of your model by capturing more nuanced patterns in the data.



TosensenTosensen:two:  [9:37 PM](https://hng11.slack.com/archives/C07AJDXN4AG/p1720125470530109)

* get incidents of flood in the past using newspaper or any structured data. Note the month, day, year. Do a graph of this
* get average precipitation for lagos for each day of the year
* get moon cycles
* get ocean levels through each part of the year for that area
* find out the winds that carry moisture and graph them
* look at coorrelated rainfall areas (e.g india) and see if there is structured data
* graph all this to look for peaks, and compare to incidents of flood in the past
* based on that, project into the future and figure out when all variables will converge again to create flood
* look for other variables I may not have mentioned that better correlate with flood reports
* e.g size of lake chad, size of the niger river, cloud coveer, etc