Welcome to Factored 2024 Datathon

August 12th - August 25th 2024

Factored

Agenda

01

Keynote

Israel Niezen
CEO

02
Welcome Aboard

Timeline

08/12/2024, 6:00 PM EST - **Datathon Starts**

08/25/2024, 23:30 EST - **Last Submission deadline**

Technical - Business Review phase

09/06/2024 - Closing Event - **Winners announcement**

We're here to support you!

Andrés Rojas
Senior Data Analyst

Héctor Mejía
Senior Data Engineer

Jose Baene
Senior Geospatial Analyst

Diego Medina
Marketing Manager

John Santa
Event Coordinator

03
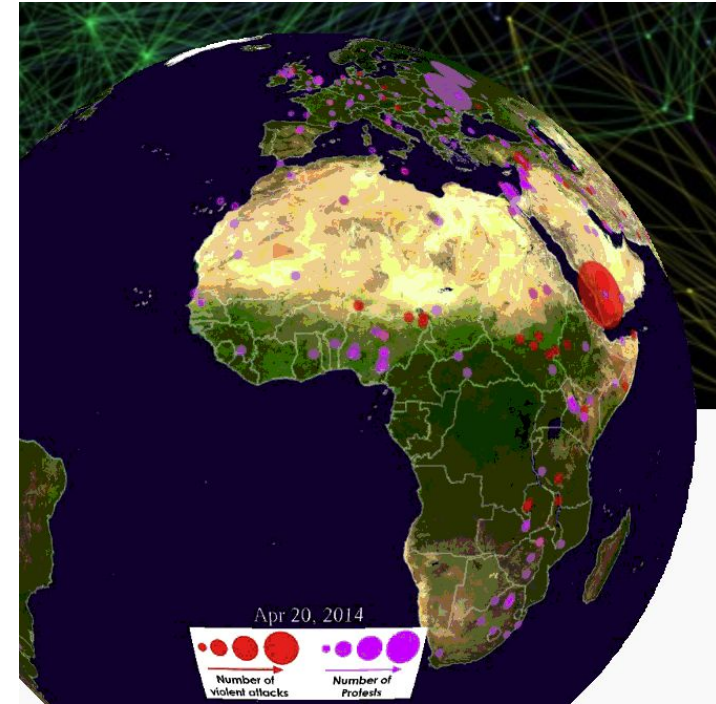# Quick Overview

352

Participants

102*

Teams

* Initial count of teams, subject to change when new teams are organized in our #2024-looking-for-a-team

The GDELT - "The global database of society"

Monitor Print, Broadcast and web news from almost everywhere in the world! (Global Database of Events, Language, and Tone)

We want you to use it to uncover critical insights from global events that:

- Enable early intervention and strategic planning
- Helps businesses and organizations navigate complex and volatile environments.

The Challenges - News Analysis

Create a summary of news for a specific period picking topics such as Social, Political, and Economical.You should use Natural Language Processing (NLP) techniques to summarize the most relevant news for a topic of your interest in a specific timeframe. General guidelines that will help with the analysis are:

Your task is to synthesize vast amounts of news data into insightful summaries that capture the essence of current events, helping stakeholders make informed decisions based on reliable and unbiased information.

Minimize Political Polarization, it should be more focused on facts than opinions.

Include the geographical and time dimensions within the analysis.

Avoid using fake news as input, so you should be able to identify them.

The Challenges - Risk Assessment

Explore Global Trends and Conduct Risk Assessments across Social, Political, and Economic Domains.You should search and uncover anomalies that enable you to detect early signs of political unrest, armed conflicts, economic crises, or disease outbreaks. The insights enable proactive decision-making and risk mitigation strategies. General guidelines that will help with the analysis are:

Provide early warning insights and predictions to aid in risk mitigation for businesses and organizations facing social, political, and economic uncertainties.

Assessing how Social and Political Risks (News) influence Economic events (Flagged Events).

Predict negative sentiment in Economic events based on Social and Political news.

Choose one topic—Politics, Social, or Economics—to understand and predict patterns for deteriorating situations in different geographies.

The Challenges

Overall, this challenge is evaluated based on two angles:

Technical

- Overall quality of your solution. Does your Data Pipeline works correctly, regarding ingestion, data quality, delivery?
- How reliable are your model results?
- Does your solution works as expected?

Business

- Does your solution answers a business question?
- Are the insights from the workflow be used to take actions?
- Think that you're part of a company data team or consultants required to provide value to a client!

The Challenge

Details on the Data Connection

Considerations to have when accessing the data:

- Data should be accessed through the RAW CSV files loaded in the GDELT webpage
  - **GDELT 1.0 GKG Data**
  - **GDELT 1.0 Events Data**
- In addition to the static data, participants must ingest the live streamed data that is updated daily.

We will share with you a document detailing more information about the dataset. Please be on the lookout for this document on the Slack Workspace!

04

Details

Event Logistics

Communication!
We will add you to the Datathon community in **Slack**:

Recommendations:
- Keep in touch always with your team!!
- We have some preset Channels, join and explore them!
- We might create additional ones to share specific topics that may help you going through with the challenge
- We´ve set a technical discussions channel to get you in touch with mentors! #technical-discussions-2024

If you don't have a team you can participate by yourself but bear in mind that having a team will greatly improve your results

Event Logistics

Document your code

All code should be stored in a Github Repository specific to your project

1. Set up the repo with the following formatting
   factored-datathon-2024-[your team's name]
2. Add **anferojasr**, **hector6298**, **jnbaene** as collaborators on your repository
3. We will share a Forms to collect each team link to the repo, only the team leader should answer this form
4. Further instructions will be provided during the datathon

Remember we have commit windows that are every day between 19:30 EST and 21:30 EST. (Not mandatory but you will get extra points)

Event Logistics

You're free to use any language and/or set of tools you deem necessary!

Some (maybe) valuable resources:

Microsoft Azure

Databricks

Snowflake

AWS

Submission Details

For your submission to be considered successful you should:

1. Share the link to the GitHub Repository
2. Share a link to where your tool is deployed
3. Share a 4 - 6 Slides presentations with details on your tool
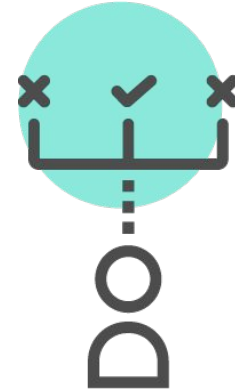
Submit your tool no matter what!!!

Submit all of this to **datathon.admin@factored.ai**

Evaluation Criteria

First and foremost our solution should work

1. Overall project rationale and documentation
2. Data Engineering: How are you dealing with the Extraction and Transformation of the data
3. Machine Learning: Model selection, optimization, implementation and tracking
4. Data Analytics: Data Quality and Providing relevant insights from the solution
5. Software Engineering: Backend, Frontend and Deployment.

Code of Conduct

We are an inclusive community.

Violence and discrimination will not be tolerated.

Remember we do have a [code of conduct](#).

05
Prizes

Prizes

US$ 3,000
2nd

US$ 6,000
1st

US$ 1,000
3rd

Be INTERVIEWED by our engineering & talent team at FACTORED!

Thank you!
Happy Coding!