Globant S.A. is a software solutions company. We are a company that provides engineering, design, and innovation services for clients. We are a digitally native corporation that helps organizations reinvent themselves to create a way forward and unleash their potential, Operating across multiple countries located in America, Europe and Asia.
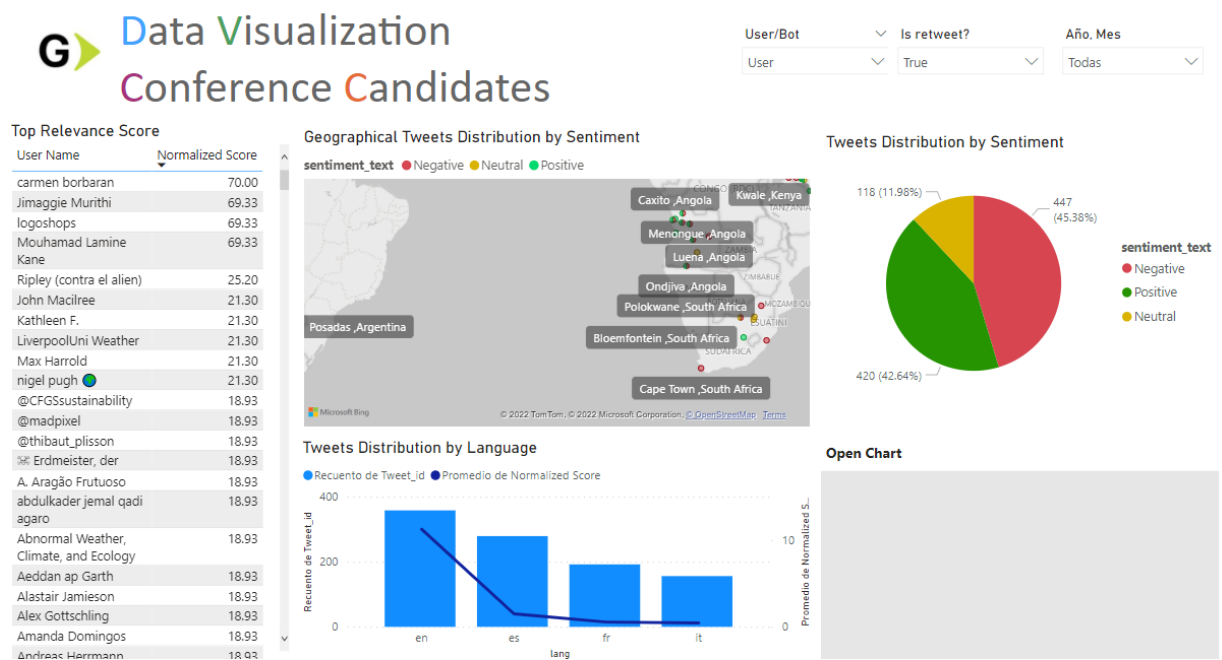
In Globant, The **Data & AI** studio is responsible for projects related to the management of data for all uses and the analysis of data to drive business processes and improve business outcomes through the more effective decision making and enhanced customer experiences.

Globant is a Pioneer of innovation. **This time, the company is organizing a new Data Visualization Best Practices conference.** The Data & AI will look for the most relevant people in the "Data Visualization" topic on social networks. For this purpose Twitter was chosen as the social network with a more impressive impact on media and people.

THE DA & AI Studio has collected Twitter Data with the hashtag #DATAVIZ and now we are requesting you to organize, clean and provide an analysis to review which candidates are the most relevant option to reach them.

D & A Studio need you to:

- Analyze and determine which twitter user could be the best candidate creating the **relevance** KPI from the "retweets" and "favorites" measures.
- Replicate the following dashboard where you can find 5 visualizations (3 graphs, 1 map and 1 table) and 3 filters .

- 1 Table with the top 20 Relevant candidates.
- 1 bar chart with the tweets distribution by language.
- 1 Map with the geographical tweets distribution by sentiment.
- 1 Pie chart of tweets distribution by sentiment.
- 1 open chart (as per your consideration) regarding tweets distribution by day on a weekly basis.

- (Optional) Apply any additional features to enhance the user experience in the dashboard, such as drill-down capabilities, enabled search bar in filters, navigation buttons, custom tooltips or any other elements you want to propose.

This must be done in PowerBI or Tableau. No other tool is accepted.

Assumptions:

1. "Source" column is a reference from who created the tweet. You need to identify whether the tweet was created by a bot (if the source column contains the word "bot") or a real user and group them, classify as User or Bot in a new column. **This will be used as a filter**.

2. Valid location will have the below format:

   [City] , [Country]

   Location descriptions such as 'mx', 'world', etc. are considered as Invalid. Once invalid locations are identified, assign them to Greenland (65.430427, -46.023625) as the new location , so these are going to be grouped in the map visualization

3. Group the sentimient_value in a new column accordingly to:
   - Sentiment values from -5 to -1 are considered negative tweets
   - Sentiment value with 0 are considered as neutral
   - Sentiment values from 1 to 5 are considered as positive tweets

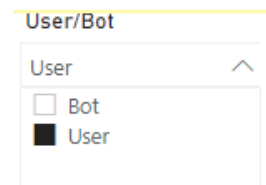   **This will be used as a filter**.

4. The entire Dashboard should be able to be filtered by :
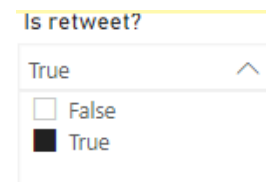   - Year-Month, from the creation date of tweet

   - User/Bot Group column, in order to differentiate between users and bots

   - Is retweet? column

5. In order to standardize values, score should be calculated for Retweet and Favorite :
   - Score Retweet: ([retweet_count]-[Retweet_MIN])/([Retweet_MAX]-[Retweet_MIN])
   - Score Favorite: ([favorite_count]-[Favorite_MIN])/([Favorite_MAX]-[Favorite_MIN])

   Note:

   Retweet_MIN: Minimum value of retweet_count on the entire dataset.

   Retweet_MAX: Maximum value of retweet_count on the entire dataset.

   Favorite_MIN: Minimum value of favorite_count on the entire dataset.

   Favorite_MAX: Maximum value of favorite_count on the entire dataset.

6. Once scores are calculated, you will be asked to ponder these for the relevance calculation. D & A Studio considers that Retweets has a weight of 70% and Favorites 30% :
   - Ponderation : ([Score Retweet]*100*70 + [Score Favorite]*100*30)/100
   - Considering the sentiment, the relevance could be negative or positive
   - Sentiment consideration: [sentiment_value]>=0 then [Ponderation]*1 else [Ponderation]*-1)