# Big Data for Public Policy and Decision Making

Can big data assess the effectiveness of Government Crises Responses?

Expert Forum for Producers and Users of Disaster -related Statistics , 10 June 2021

Beirut, Lebanon

# Project Platform

## Why and What

- In crisis and disaster settings, there is a need for up-to-date real-time data & quick evaluation of the effectiveness of response policies being implemented in achieving their objective and target.
- Dashboard with real-time updates on a number of selected topics and indicators using Big Data sources
- An online platform that generates near real-time data and insights into the indicators of effectiveness of policies being implemented after a disaster by using non-traditional open big data sources.

## Objectives

- Provide the statistics offices a way to generate faster insights on the trends and key socio-economic indicators requested by the ministries and DRM.
- Aims to help policymakers develop responses and monitor adopted policies and trends

# Target users

1. Civil servants dealing with assessing socio-economic risks or shocks
2. National Statistics offices
3. Ministries of planning and other line ministries concerned with national risk mitigation
4. Disaster management unit within General Security
5. Office of the UN Resident Coordinator
6. Development practitioners in the multilateral organizations (working with line ministries)

# Platform user categories

1. *Regular users*
   a. Can view and use the homepage, Twitter Statistics, Google Search Trends, and Google Nowcasting pages.
   b. The users can use the platform for the countries assigned to them by the admin.
2. *Administrator*
   a. Regular user functionality + user management
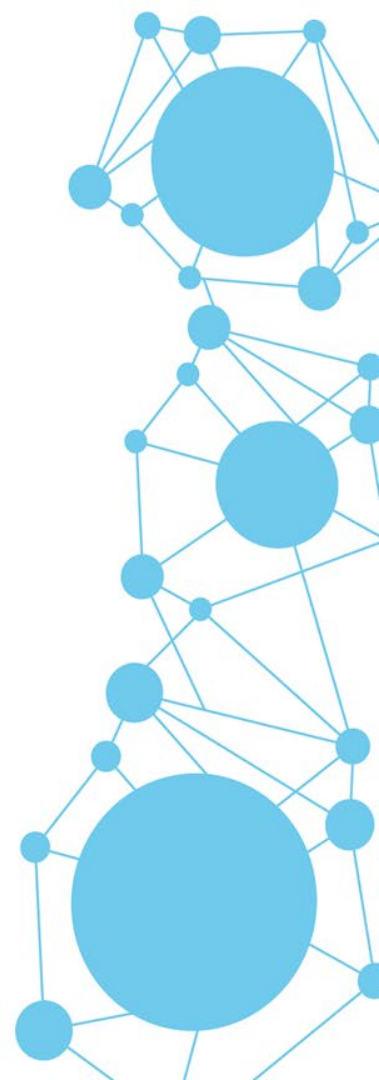   b. Can create and add users to the platform.

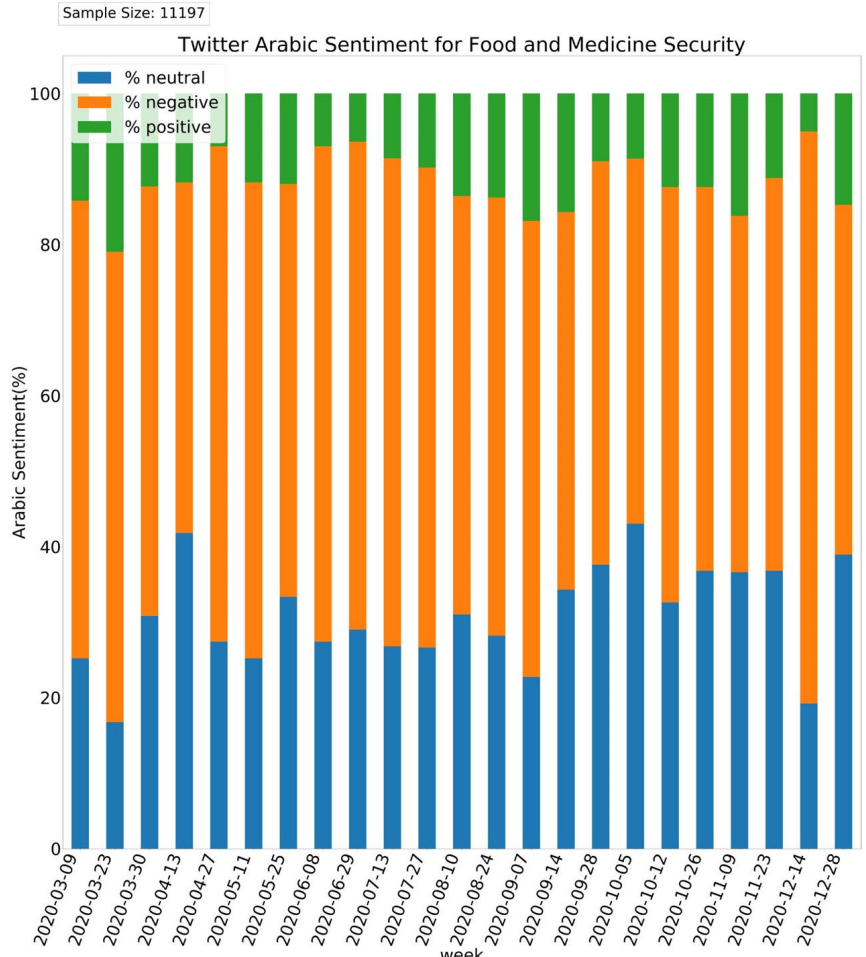# Data used in the platform?

## Twitter
Sentimental analysis and Word Clouds

## Google Search Trend (GST)
Trends and Nowcasting

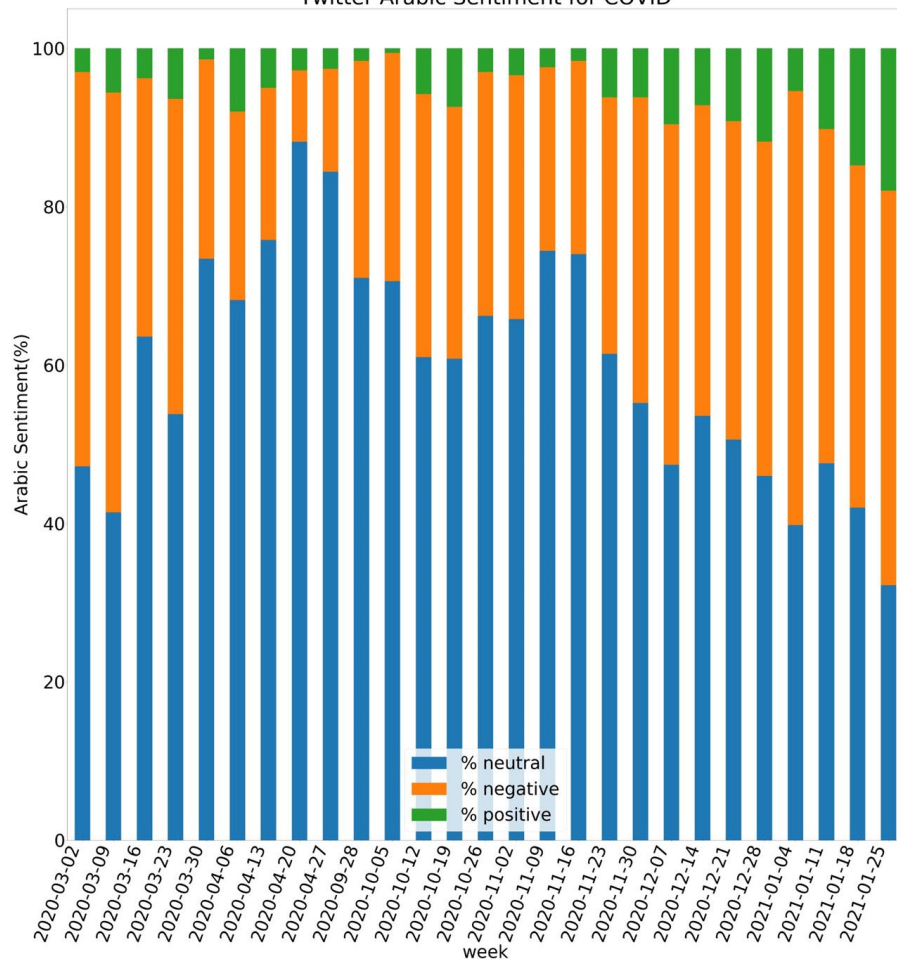# Examples – sentiment analysis and word clouds



Food and medicine security in
Lebanon

Sample Size: 13499

**Twitter Arabic Sentiment for COVID**

Arabic weekly sentiment analysis for _Covid -19 related equipment & supply shortages_

March 2020 - Jan 2021
Sample size: 13,499

P.S. to avoid bias by small sample size, we drop weeks that have sample size less than 100

March 2020



Jan 2021

- Mainly mentioning hariri hospital in the beginning of the pandemic v.s. more hospitals in Jan 2021
- Stay at home hashtag trending in March, as opposed to different words representing requesting blood for covid patient (بحاجة-ماسة-دم-فئة) in addition to extensions hospitals are making to fit covid patients (تستوعب-ملحق) and خطر → much more serious words

P.S. to avoid bias by small sample size, we drop months that have sample size less than 500

# Examples – Google Search Trends (GST)



Google Search Trends for oxygen concentrator

Google Search Trends for oxygen machine
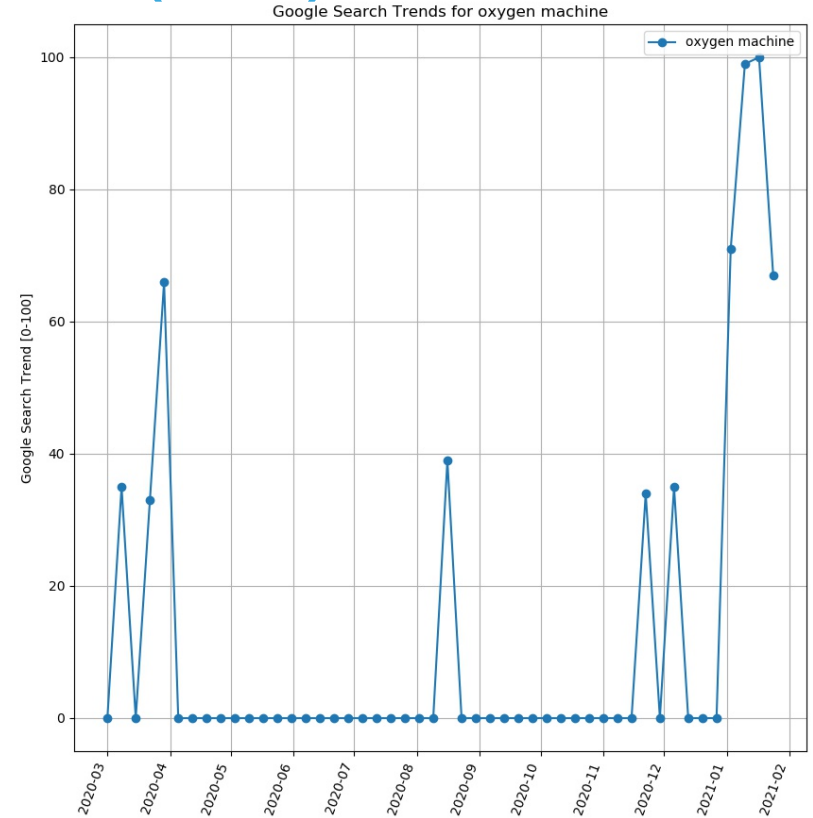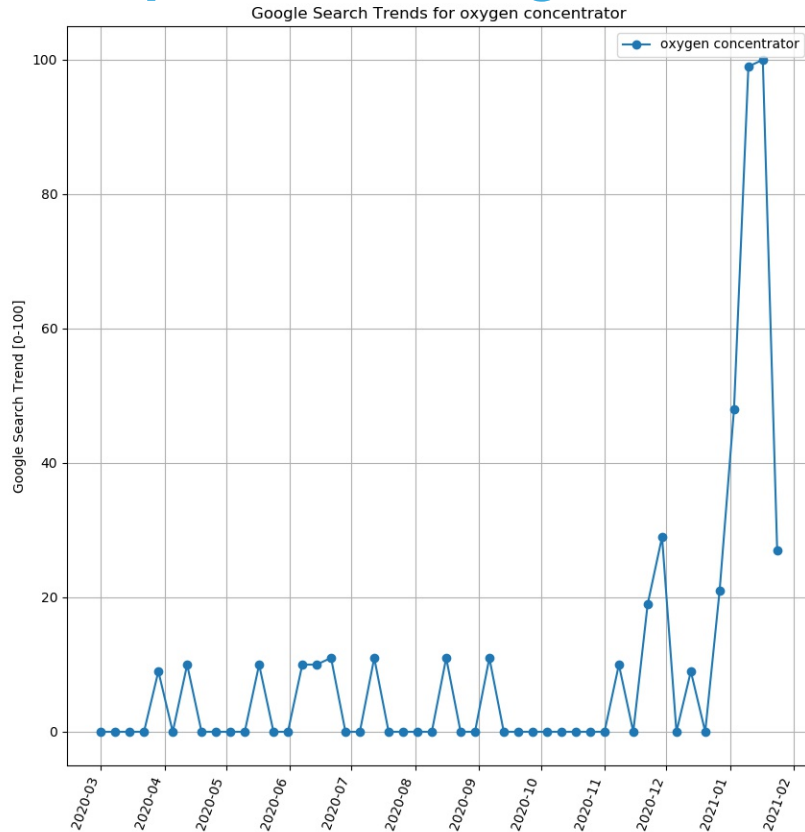
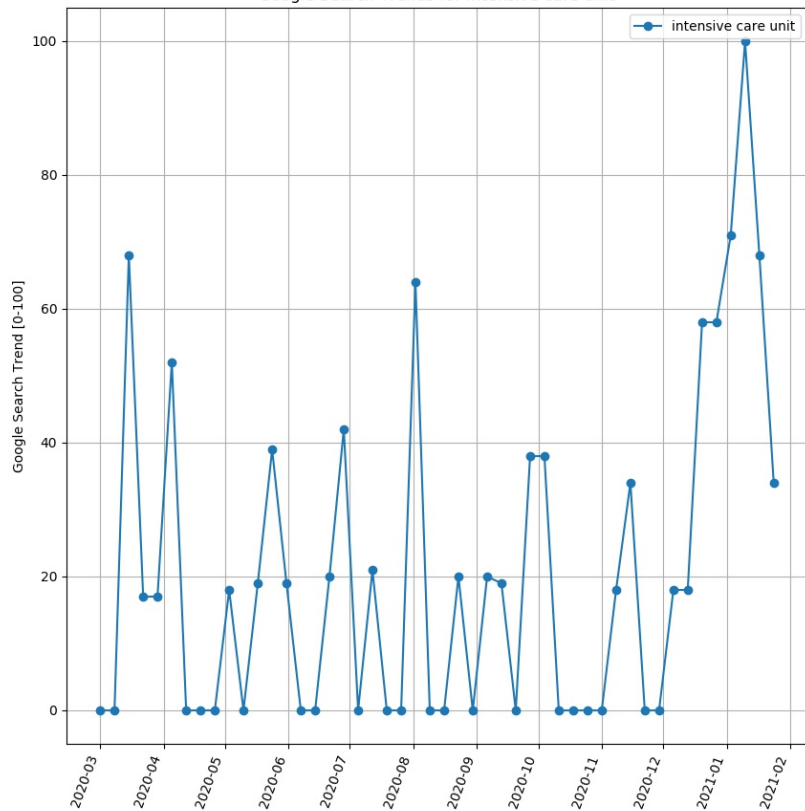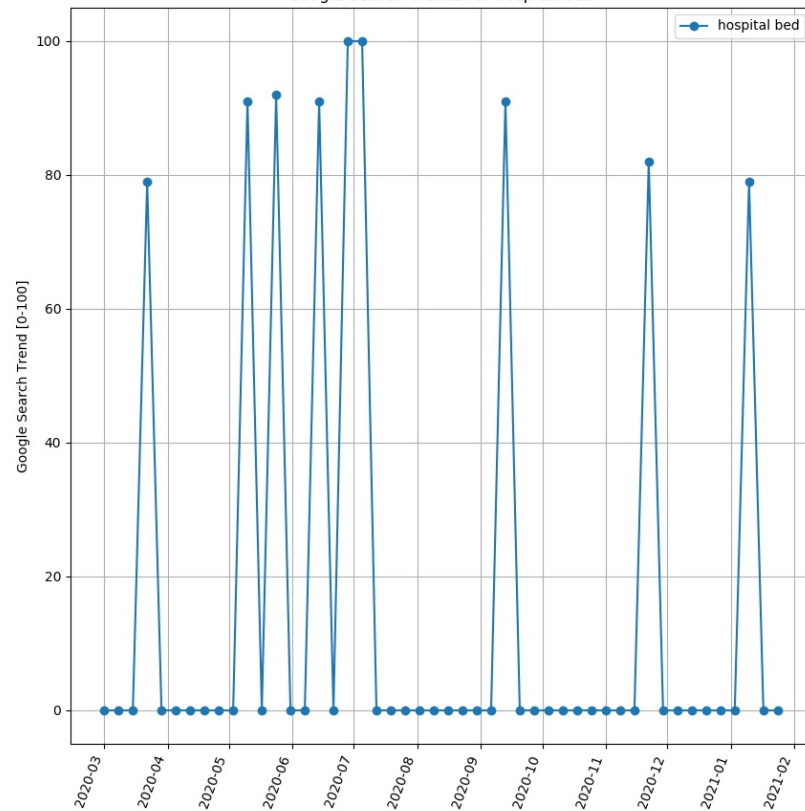O2 concentrator and O2 machines searches on Google peaked in January 2021, when people were purchasing them to treat their loved ones at home

ICU and hospital beds search trends

# Nowcasting

- Comes from "Now" and "Forecasting"

  - Has been widely used in meteorology since 1860

- **Nowcasting is predicting the present, the very near future, and the recent past.**

  - Mostly used as part of predicting the current value of observable economic indicators.

- Model that mimics real-life behaviour and reads data in real-time to obtain **timely estimates of the current state** of the indicators.

  - Model needs to be updated as frequently as possible as new data becomes available

# Why nowcasting?

- Sometimes we need to forecast the present and even the near past because we do not have enough information.

- This can be due to many reasons, such as delay and difficulties in data-collection.
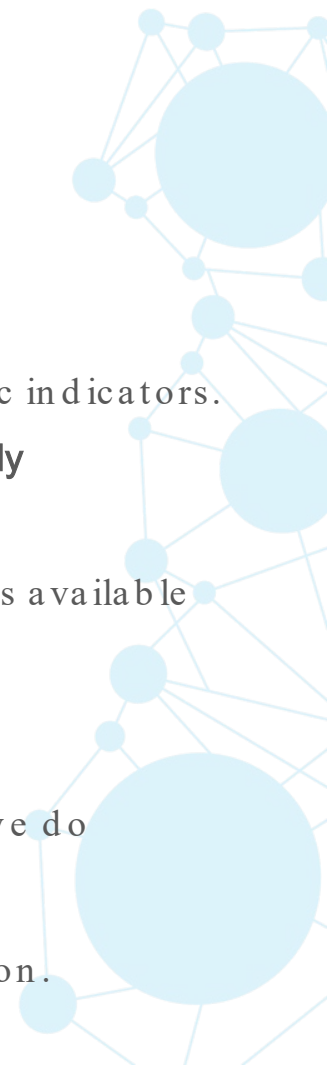
# Time-series Analysis

- Analysis of time series data presents some of the most difficult analytical challenges:
  - Typically, there is the least amount of data to work with, while needing to inform some of the most important decisions.

Usage scenarios

# Platform usage

## Twitter

- Depending on the nature of the crisis, the user is able to select a category of interest (ex. health, education, economy, employment, food security etc.)
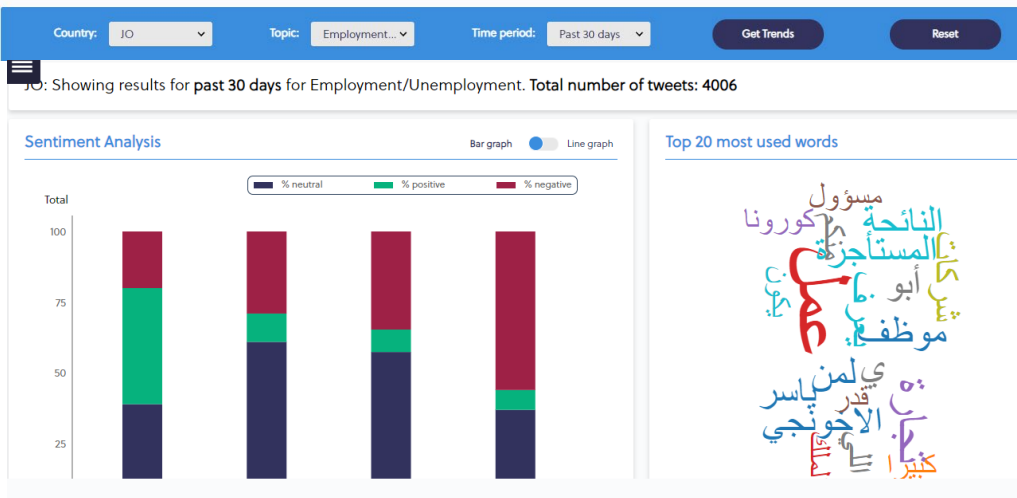- The user can also select the time period of interest (ex. past week, past month)

### Twitter Statistics

| Country: | JO ⌄ | Topic: | Employment... ⌄ | Time period: | Past 30 days ⌄ | Get Trends | Reset |

JO: Showing results for **past 30 days** for Employment/Unemployment. **Total number of tweets: 4006**

**Sentiment Analysis**   Bar graph ⬤ Line graph

Legend: ■ % neutral   ■ % positive   ■ % negative

Total
100
75
50
25

**Top 20 most used words**

مسؤول
كورونا
النائحة
المستأجرة
أبو
موظف
يلمن ياسر
قدر
الاحونجي
كثيرا

## What will the user know?

- What has been the sentiment around the key topic selected in the time period selected.
- What has been the trend over the selected time period.
- By knowing the key dates for implementation of certain policies or measures, the user is able to see if the implementation of the policy has affected the sentiment expressed - did it worsen or improve?
- Detecting key events: by observing the sentiment trends over time, if a sharp difference is detected this could be indicative of an important event or occurrence that has triggered a change in sentiment.
  - *Some examples: Monitoring social media for natural  -hazard disasters, for example, floods, fires.*
  - *Localized economic crises, food security issues.*

# Platform usage

## GST

- Similarly to Twitter, the user selects the topic and the time period of interest
- The selectable results are also displayed as a line graph and geospatially on a map by governorate



| Country: | JO | Topic: | Employment... | Time period: | Past 30 days | Get Trends | Reset |

: Showing results for **Past 30 days** for Employment and Unemployment.

### Interest over time

- [x] job [ ] salary [x] employment [ ] minimum wage [ ] unemployment [ ] employee benefits

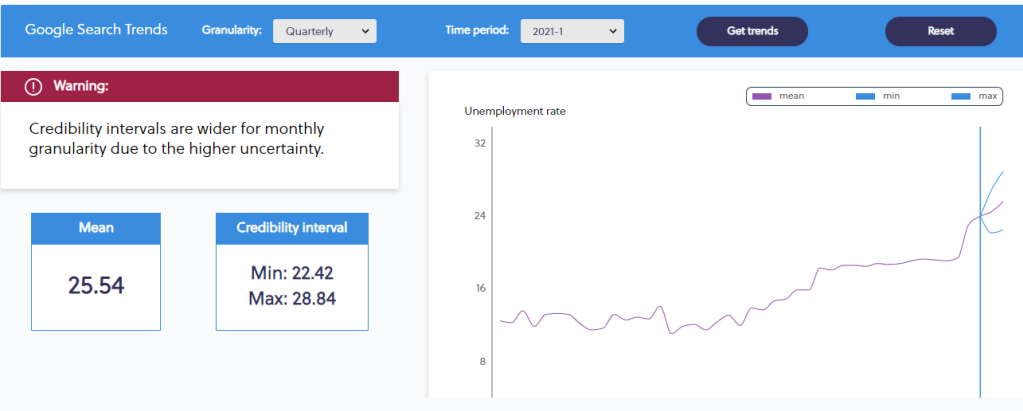### Interest by governorate

job

## What will the user know?

- What the **trend** of interest over the selected time period, and where is this interest most concentrated
- By comparing with key dates of policy adoption and implementation, the user is able to see if the **implementation of the policy has affected the interest in this topic** .
    - This could help understand if people are aware of the existence of the policy, if they are looking for information about it and related terms.
    - Whether the policy has affected the spatial distribution of the interest in the given topic, helping with targeting of communication campaigns.
    - For example, in a natural-hazard disaster by viewing the interest in various topics by geographical location, it can be inferred where food security or financial needs are greatest.

# Platform usage

## GST for Nowcasting

- The user is able to select the temporal granularity of the nowcasting predictions: monthly or quarterly.
- After, the user selects the month from 1 to 12.
- The platform returns the predicted mean value, and the credibility interval for this value and displays the trend over time as well.



| Google Search Trends | Granularity: Quarterly | Time period: 2021-1 | Get trends | Reset |

**Warning:**
Credibility intervals are wider for monthly granularity due to the higher uncertainty.

Unemployment rate — mean — min — max

| Mean | Credibility interval |
|------|---------------------|
| 25.54 | Min: 22.42 Max: 28.84 |

## What will the user know?

- Crises today are complex and interconnected with various socio-economic aspects
- By using GST nowcasting the users will be able to predict and foresee how the current situation is affecting the crisis.
- This in turn provides the crucial information on the impact of the crises and data to develop a response measure.
- By utilizing the nowcasting tool before and after a certain period (ex. 1 month) of policy implementation, the user is able to compare the effects of the specific policy on the nowcasted unemployment rate.

# Platform usage scenarios

Example: Natural hazard disaster
(flooding, earthquake, drought)

Example to give some pointers on how the platform can be utilized in different types of disasters/crises.

Assumptions:

➔ Certain locality is affected more than others and requires humanitarian assistance.
➔ The road access to this locality might be compromised due to the disaster and people could be isolated.
➔ Some relevant topics from the platform would be *"humanitarian aid, food and medicine security, health "*

## Twitter

● By choosing the topic, the policymakers can detect what is the sentiment around the current humanitarian aid provided, food and/or medicine security

## GST

● Users can monitor what has been the interest per topic to detect any peaks that could indicate that there's an increased need.
● Spatial geolocation to detect any area that might in greater need.

## GST for Nowcasting

● To predict and understand how the unemployment rate could be affected by the ongoing disaster.

*However, this version of the platform is a prototype that needs to be further tested with the users and modified accordingly.

# Platform usage scenarios

## Example: Economic crisis

Example to give some pointers on how the platform can be utilized in different types of disasters/crises.

### Assumptions:
➔ A protracted interconnected economic crisis that is affecting multiple sectors of the economy and the society.
➔ Certain localities can be affected more than others given their economic and geographical integration.
➔ Some relevant topics from the platform would be *"economy/SME, poverty, employment, humanitarian aid, education, health"*

## Twitter
● By choosing the topic, the policymakers can detect what is the sentiment around the current humanitarian aid provided, food and/or medicine security.
● Monitoring of sentiment trends and response to a given policy.

## GST
● Users can monitor what has been the interest per topic to detect any peaks that could indicate that there's an increased need.
● Spatial geolocation to detect any area that might be in greater need.

## GST for Nowcasting
● To predict and understand how the unemployment rate could be affected by the ongoing crisis.

# Issues related to the sources used

- Some Data sources (SM) are made available only during Crisis !
- In cyclic crisis like Covid-19, possible to "near real time" correction
- Bias due to technology dependence is a calculated risk
- Balanced between data richness (accuracy) and the protection of privacy
- CODE - Council for the Orientation for Development and Ethics