



SESRIC WEBINAR SERIES ON STATISTICAL EXPERIENCE SHARING



دائرة الإحصاءات العامة
Department of Statistics

Big Data Applications and Utilising Non-Traditional Data Sources and Methods for Official Statistics, 10 June 2021



**Department of Statistics (DOS)
Jordan**

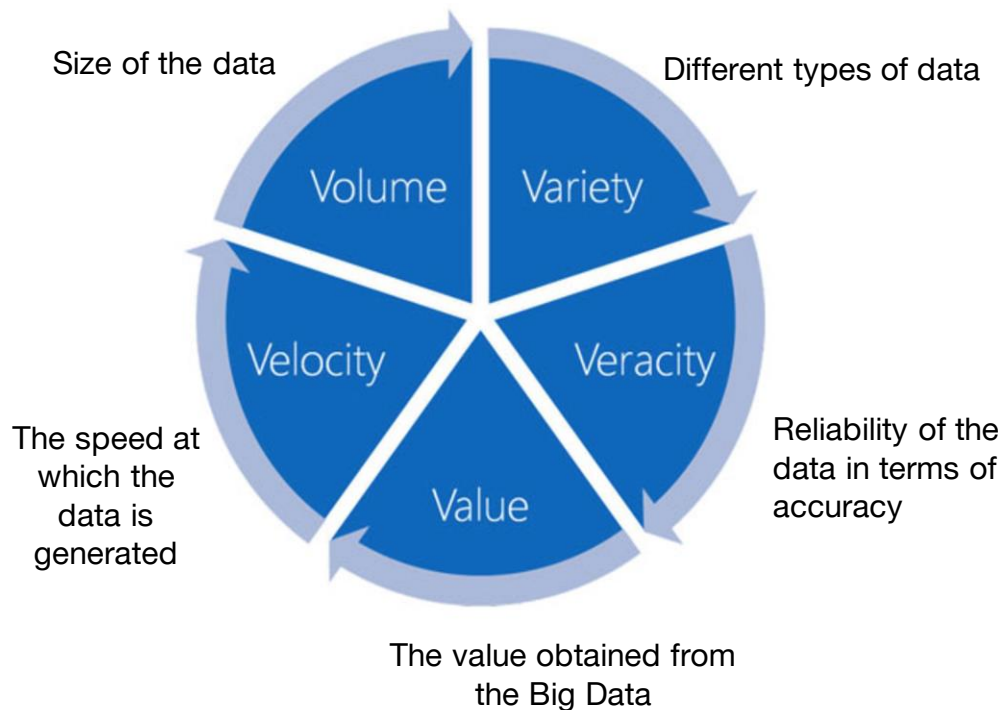
Ahlam Alrousan
IT director



Leveraging Big Data and nontraditional data for Policy Effectiveness Evaluation During Covid19



What is Big Data? What the challenges





Project Objectives



دائرة الإحصاء العامة
Department of Statistics

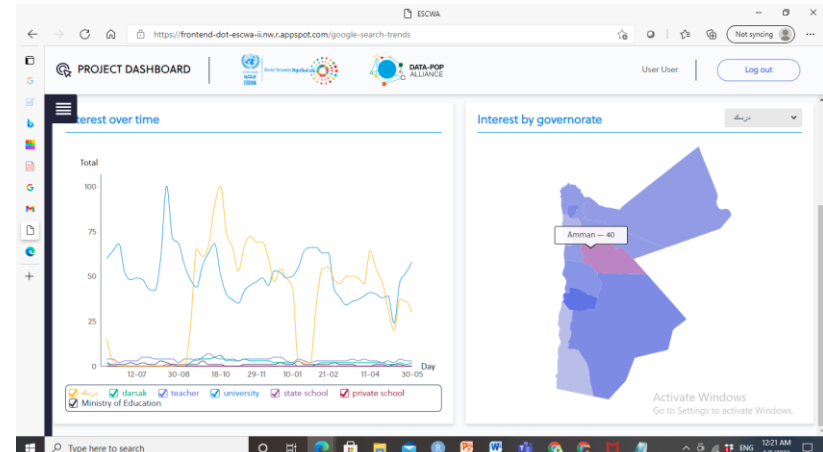


Evaluate the effectiveness of a set number of government policies

adopted to respond to the crisis and mitigate its effects on the populations, with a particular emphasis on using alternative sources of data in evaluation.

Develop a prototype of a platform for policymakers

to help develop responses, implement and monitor adopted policies by leveraging new kinds of data and analytical approaches .





Project Methodology



دائرة الإحصاءات العامة
Department of Statistics

- **Policies selection criteria for each countries (Jordan and Lebanon)**

→ *5 top ranked policies selected for each country*

- **Policy decomposition into indicators for evaluation**
- **Mapping of non-traditional open source data sources**

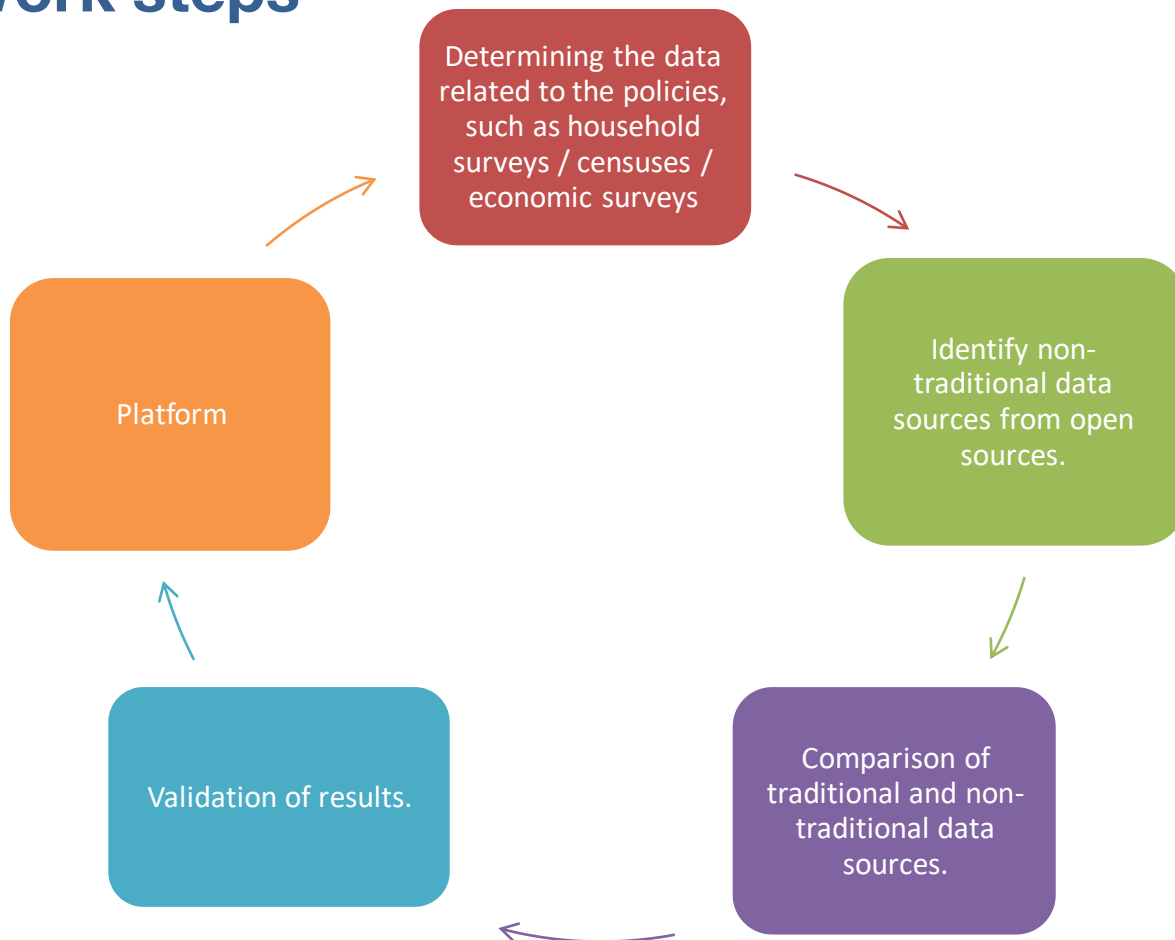




Work steps



دائرة الإحصاءات العامة
Department of Statistics



Policy description	Information			
	Policy Owner	Beneficiaries	Stakeholders	Data sources
Tkiyet Um Ali - social protection for those not covered by SSC & NAF - The Ministry of Planning and International Cooperation provided a JD375,000 grant to Tkiyet Um Ali within the Economic and Social Productivity Enhancement Programme with the aim of supporting 4,022 families of daily wage workers across the Kingdom	Ministry of Planning and International Cooperation	[daily wage workers, families living below the food poverty line, unemployed] who are listed in Tkiyet Um Ali databases, and not benefiting from National Aid Fund or the Social Security Institution	Tkiyet Um Ali (NPO), Ministry of Planning and International Cooperation	https://www.tua.jo https://www.facebook.com/TkiyetUmAli/

Policy logframe/ ToC				
	Description	Indicators	Traditional Data	Non traditional
Activities	The ministry funds Tkiyet Um Ali with JD375,000	NA	Tkiyet Um Ali database - number of households applied v.s. official statistics of households leaving in	Was their targeting effective? comparing between the database of Tkiyet Um Ali and the poverty mappings and results of the mobility
	Tkiyet Um Ali decides the families to receive the aid based on their database			
Outputs	Target 25,425 individuals and 4000 families	Number of households reached;	Tkiyet Um Ali database - number of families who received the aid	Google trends; Twitter; GDELT; News media; Sentiment analyses on: - for income, food security & consumption & if the media have been sounding alerts/alarms
	Reach the families most affected by the lockdown	Characteristics and stats of households that received the aid	Tkiyet Um Ali database -	
Outcomes	Income security (maintain incomes) for the most vulnerable	Average income levels of beneficiaries	Tkiyet Um Ali database - average income of the HH that received aid	
Impact	Reduced hunger impact of	% families who received the aid	Tkiyet Um Ali	Mobility data, Open Street Map & satellite data to

Selected policies for Jordan

N°	Category	Policy description and its significance	Initiating ministry/ Policy owner
JO1	Cash transfer / Food distribution	Tkiyet Um Ali - social protection for those not covered by SSC & NAF	Ministry of Planning and International Cooperation Tkiyet Um Ali
JO2	Education	'Darsak' official e-learning portal (https://darsak.gov.jo/)	Ministry of Education
JO3	Cash transfer / Food distribution	Jordan Emergency Cash Transfer COVID-19 Response Project	National Aid Fund
JO4	Disability	HCD: Protection and Response Measures for Persons with Disabilities	The Higher Council for the Rights of the People with Disabilities
JO5	Support to Small companies	Defence order no.9: Tadamun 1 & Tadamun 2	Social Security Corporation

Non traditional Data source

Social media and news analysis

Data Sources

Twitter

GDELT

Availability

+

From April 1st to Oct 9th
Keywords: poverty, economy/SMEs, food/medicine security, policy, related to the policy (humanitarian assistance, financial support, etc).
Total number of tweets analyzed: 7937

From April 1st to Oct 9th
Total number of articles analyzed 61,610 (18,958 in Arabic and 42,652 in English)

Methodology

Sentimental analysis: negative vs positive sentiment

Word Clouds: most used words across tweets

Google trends

Google trends data

Google trends daily index from January 2020 until now

Descriptive analysis across the time: to identify behaviour of search in google on time

Human mobility

+

Facebook Movement
Range Maps
Google Mobility Reports

From February until now
The daily change in movement of Facebook users in Jordan. Facebook divides the geographical space in $600 \times 600 \text{m}^2$ tiles and provides the following information.

Analysis of mobility and other socio - economic indicators

Satellite images

Copernicus

Sentinel satellites

Landsat 8

MODIS Rapid Response System

Night-time lights in satellite images could be used as a proxy for poverty.
Another interesting application that would be explored in this project is measurement of air pollution via satellite images

Non-traditional data sources - to measure sentiment and level of satisfaction

Policy name and description	What are we measuring?	JO1	JO2	JO3	JO4	JO5
Twitter Data Wordcloud analysis Sentiment analysis	To identify what are the topics that people are preoccupied or are talking about the most.	X	X	X	X	X
	To understand what is the general sentiment expressed about the topics of interest.	X	X	X	X	X
Google search trends	To gauge whether the policy announcement or other major events have generated interest among the population. Juxtaposing these results with the major events could show what people are searching for and what are their needs and interests.	X	X	X	X	X
GDELT analysis	To gauge the sentiment expressed about certain topics as well as the policy being studied.	X	X	X	X	X

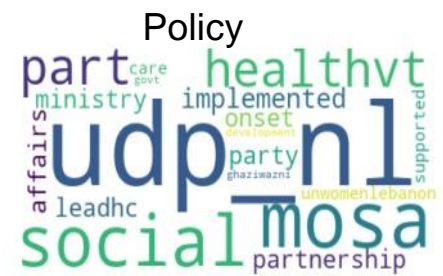
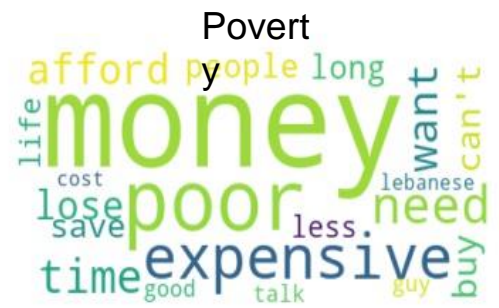
Non-traditional data sources—to measure mobility and movement

Policy name and description	What are we measuring?	JO1	JO2	JO3	JO4	JO5
Mobility analysis Facebook Data4Good Google mobility	Changes in mobility patterns since the start of the pandemic per district/region.	X	X	X		X
Mobility analysis coupled with traditional socio-economic indicators (poverty, income levels, unemployment, etc).	Mobility patterns have been shown to correlate with socio-economic status in several countries. Combining mobility analysis data with other available socio-economic indicators could shed light on and suggest certain correlations or potential relationships that need to be taken into account when designing a policy or measuring its effectiveness.	X	X	X		X
Satellite images	Night-time lights in satellite images could be used as a proxy for poverty. Another interesting application that would be explored in this project is measurement of air pollution via satellite images.	X		X		
Open street map	Mapping of educational institutions and/or medical institutions as a complement to the analysis of policies LB2 and JO2 on education and LB5 on hospitals.		X			

Define topics and keywords of interest

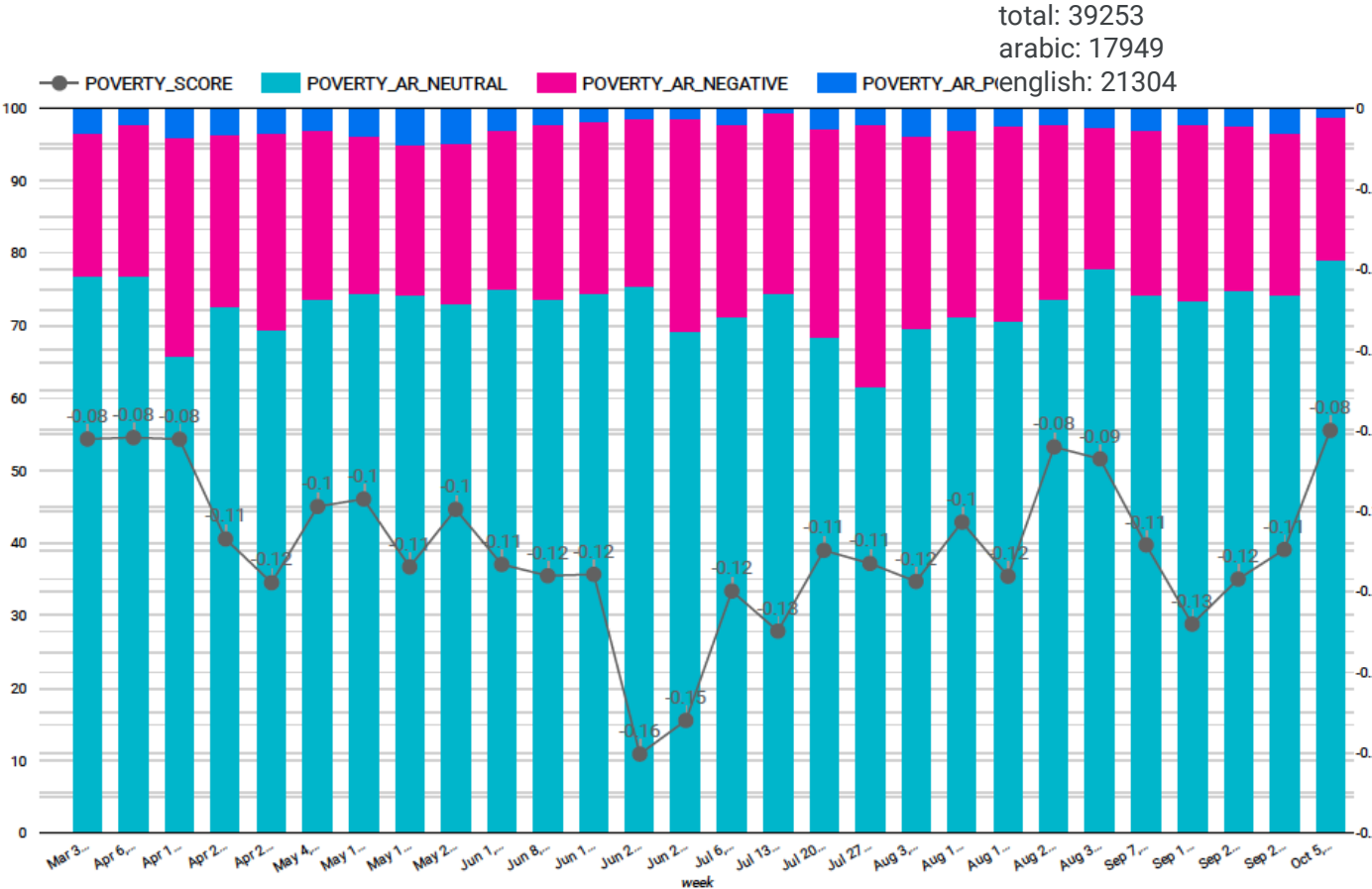
Topic	English keywords	Arabic keywords
Policy	Darsak, distance learning, online class	تعليم عن بعد ,دراسة عن بعد ,منصة درسك
Ministry	Ministry of education, minister of education	وزارة التربية,وزير التربية
Education	education, private school, public school, students, teachers, parents	تعليم ,مدارس رسمية ,مدارس خاصة,الطلاب أهالي الطلاب,أساتذة

Define topics and keywords of interest



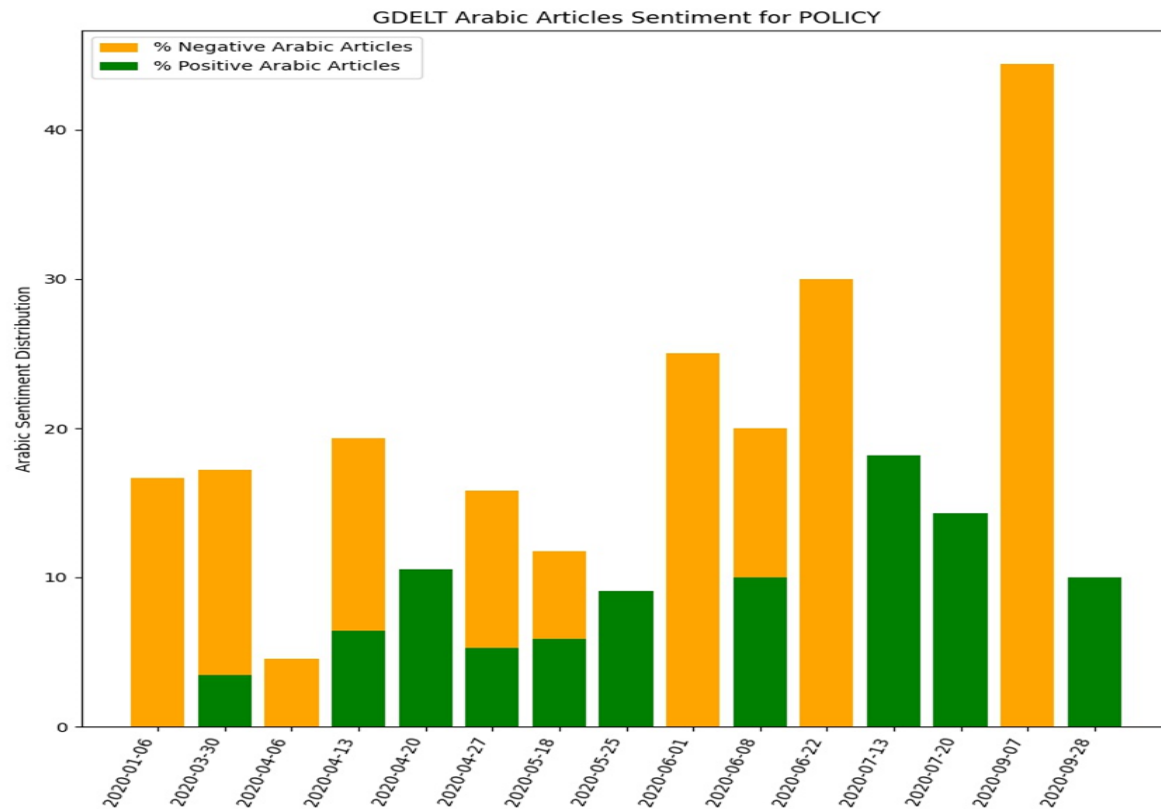
GDELT arabic and english sentiment: POVERTY

The results from social media and news confirm the population's concern about the current economic situation, as well as the need to provide adequate material support to the entire population in general. Goals .



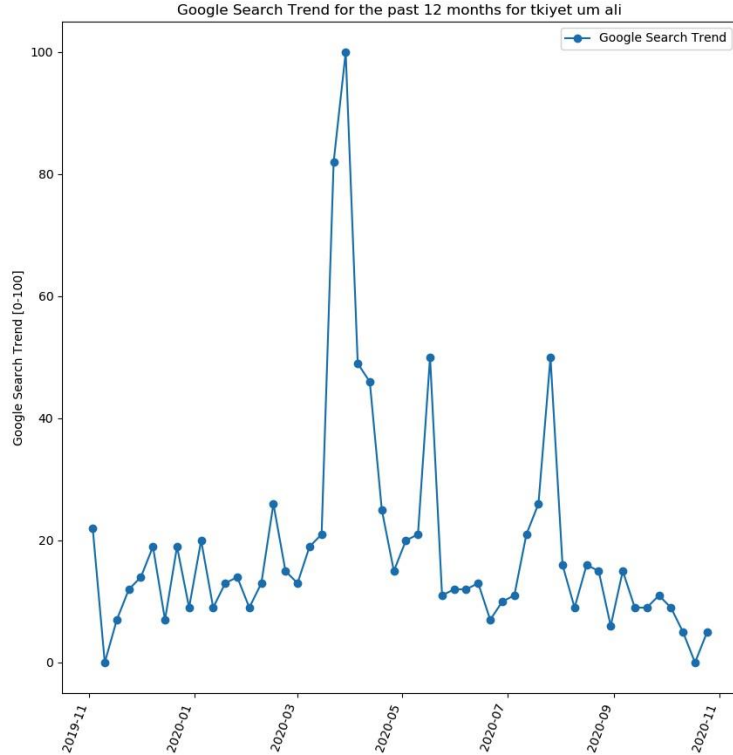
Analysis of negative and positive emotions on social media

GDLT



Negative feelings expressed in all categories are high with a peak in June. A rise in the positive sentiment index in mid-July.

GOOGLE Trend Search Results

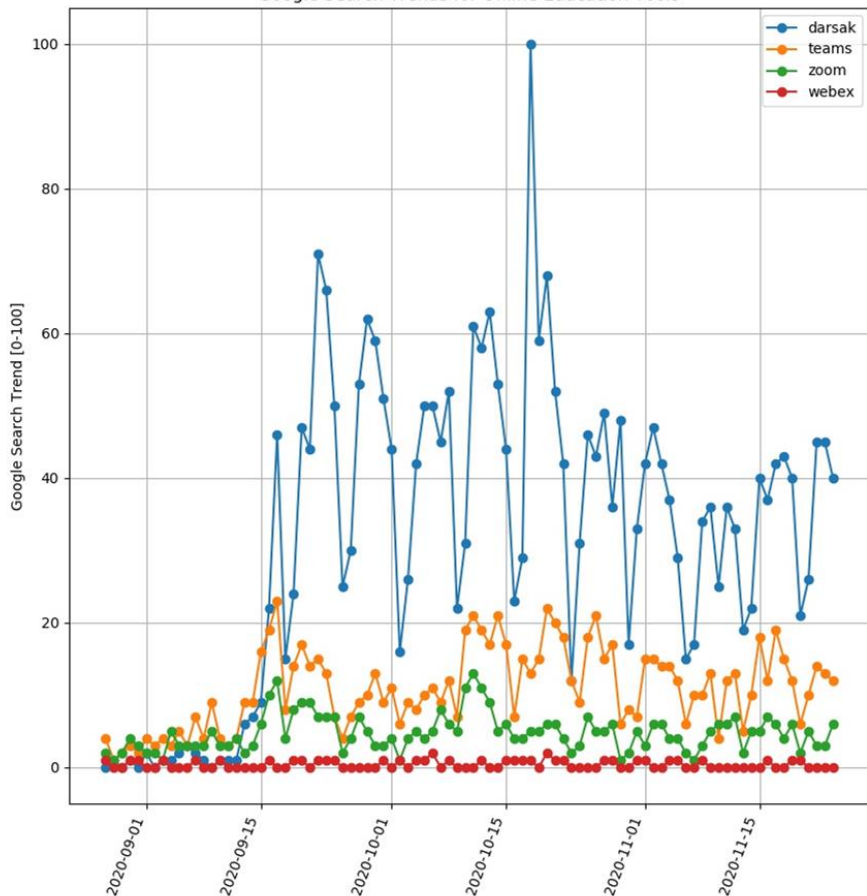


Tkiyet Um Ali: major peak in April (100%) and decline again in May and August by 50%

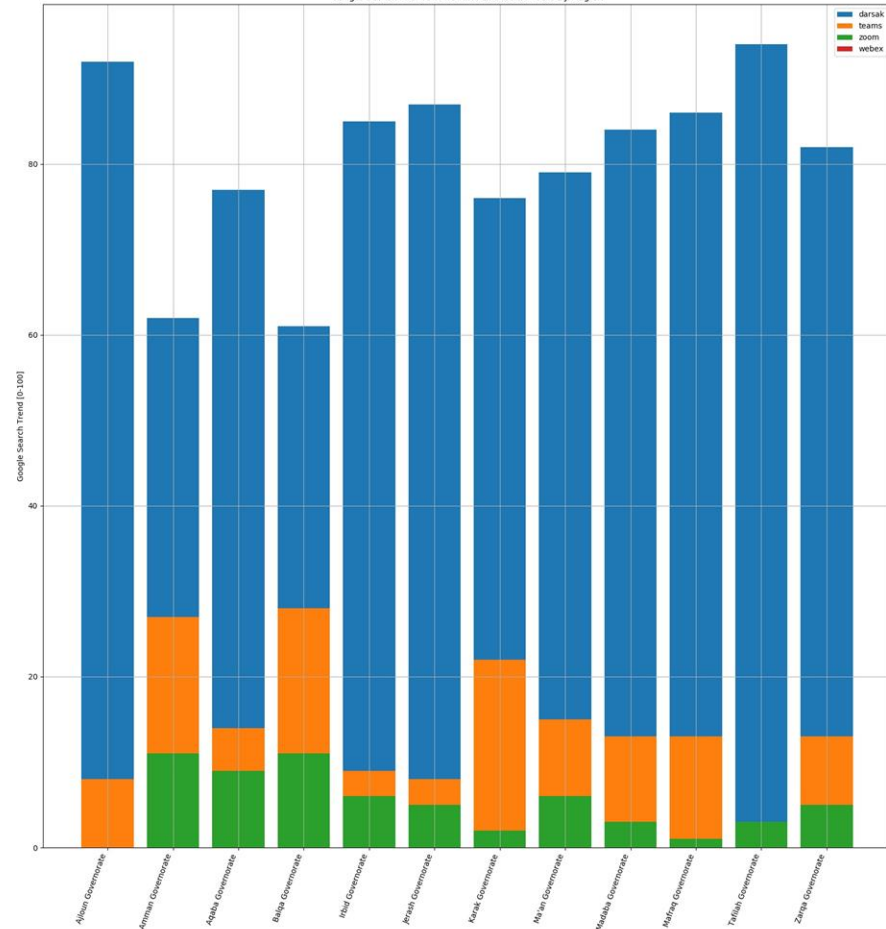
Recommendation to compare available data with the aid distribution in terms of individuals and targeted locations



Google Search Trends for Online Education Tools

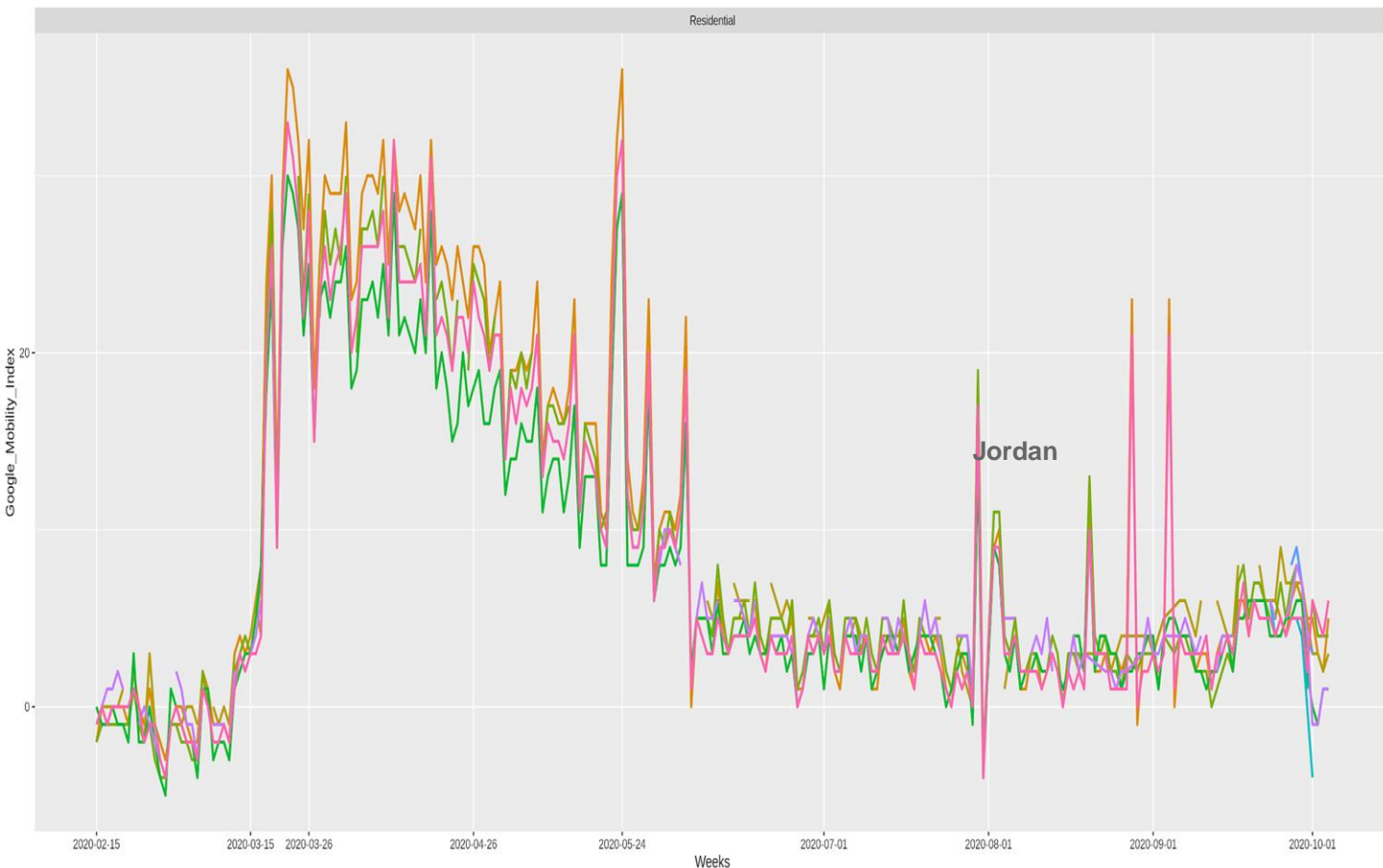


Google Search Trends for Online Education Tools By Region



Darsak v.s. Other tools for online teaching. (Sep - Nov) across time and region

Mobility analysis during a pandemic



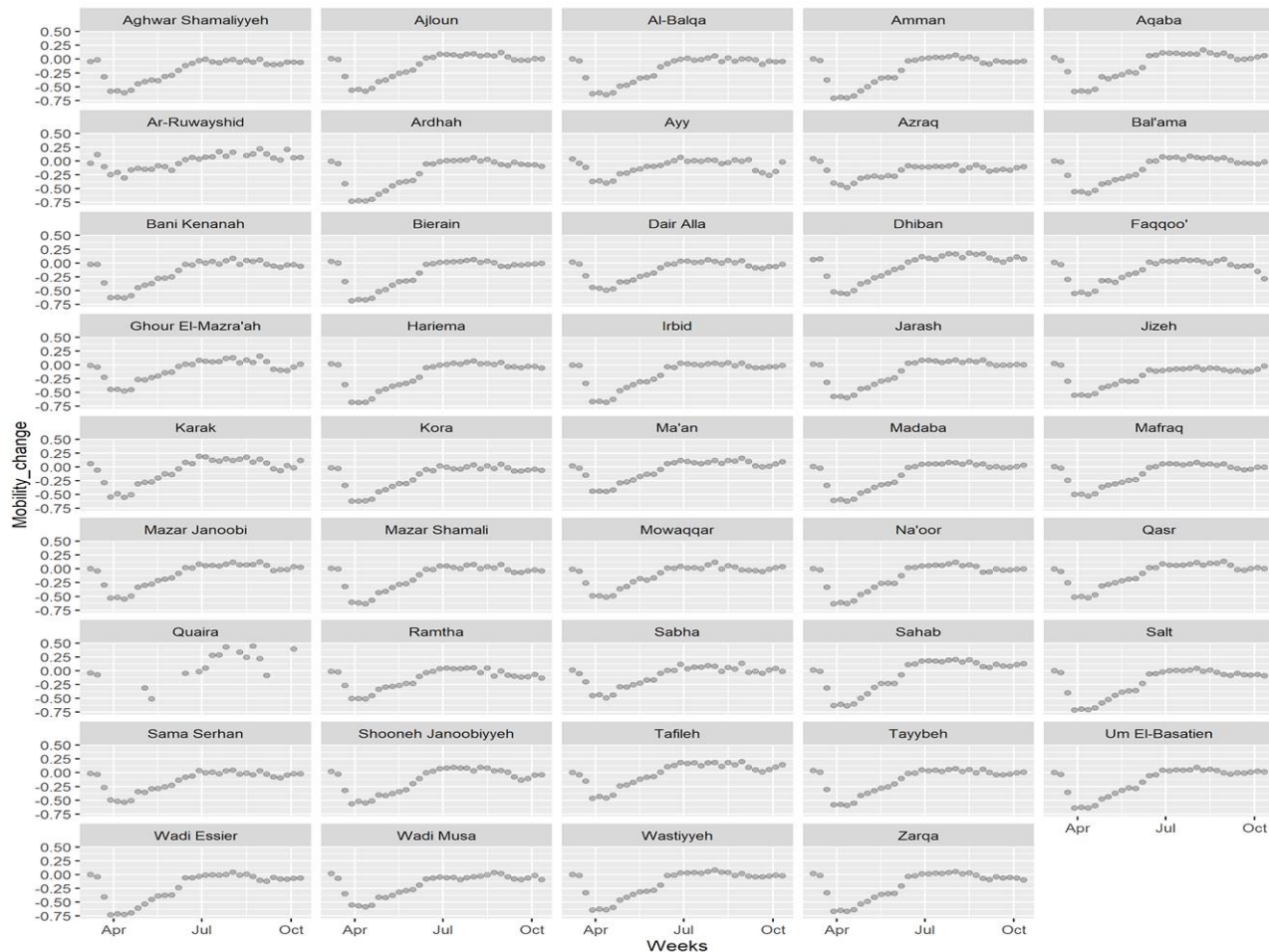
Mobility
decreased
sharply during
lockdown

FACEBOOK Mobility Result by Areas

→ The results are presenting need of people to go out to work.

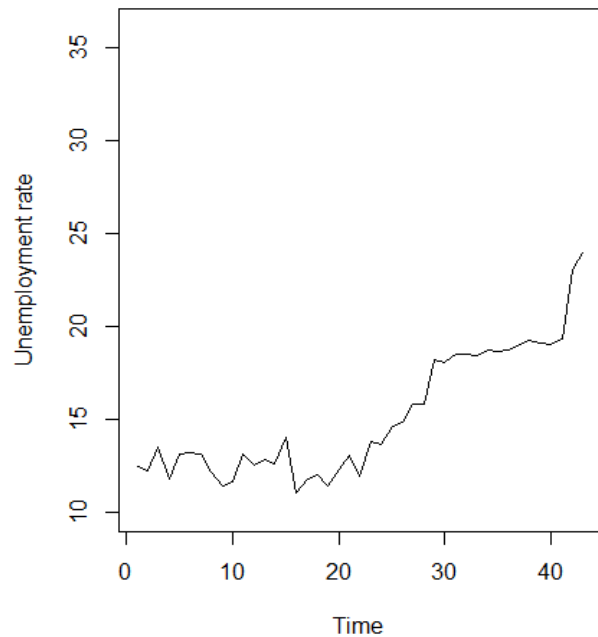
→ It is recommended to Supporting economically affected individuals is critical.

→ identify high-traffic populations to provide support and advices to reduce the potential risks.

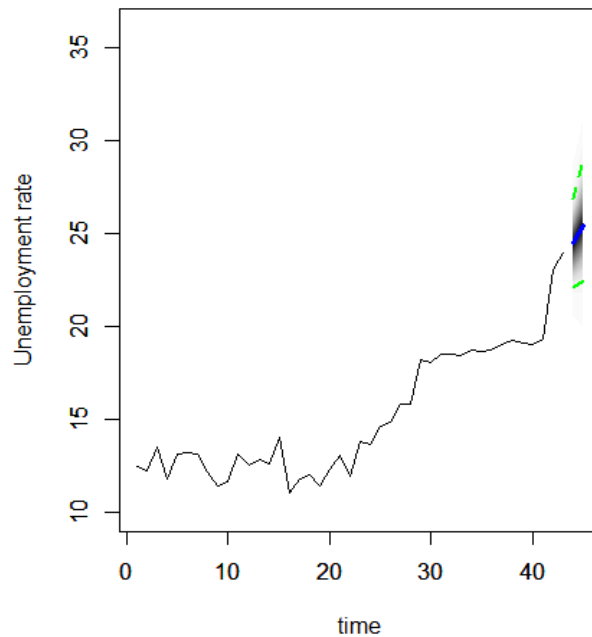


Prediction (Employment & Nonemployment)

Actual Unemployment rate



**Predicted Unemployment rates
for the next two Quarters**



Platform

SPENDING

Consumer Spending

BUSINESSES

Small Business Revenue

Small Businesses Open

Job Postings

EMPLOYMENT

Employment

Unemployment Claims

EDUCATION

Online Math Participation

Student Progress in Math

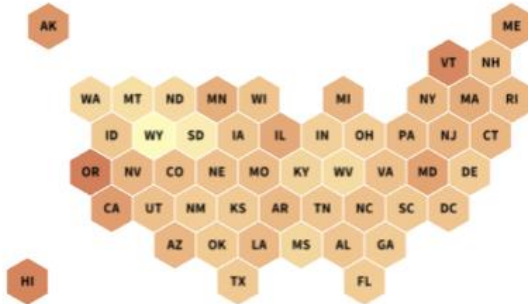
PUBLIC HEALTH

COVID-19

Time Outside Home

view: **National** States Counties Metros | compare: **Subcategories** Geographies

Percent Change in Employment ⓘ



Total

☐ Total

Income

☒ Low (<\$27K)

☒ Middle (\$27K-\$60K)

☒ High (>\$60K)

Industry

☐ Professional & Business Ser...

☐ Education & Health Services

☐ Retail & Transportation

☐ Leisure & Hospitality

PROJECT DASHBOARD



User User

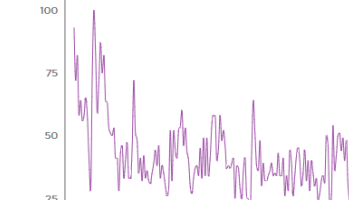
Log out

Country: **JO** Topic: **Employment...** Time period: **Past 7 days** **Get Trends** **Reset**

JO: Showing results for **Past 7 days** for Employment and Unemployment.

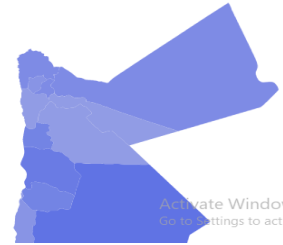
Interest over time

☒ job ☒ salary ☐ minimum wage ☐ unemployment
☐ employee benefits



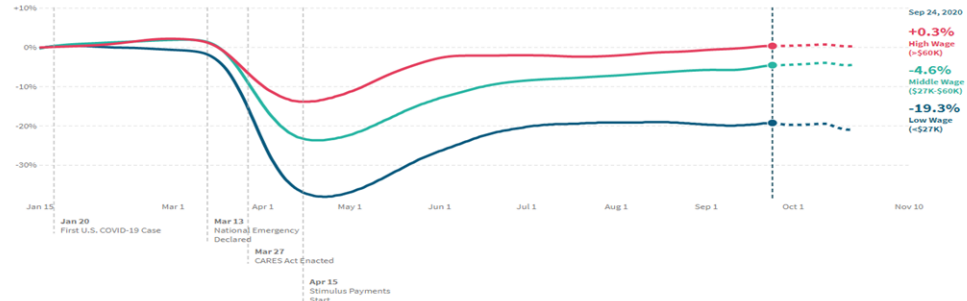
Interest by governorate

salary



In the United States, as of September 24, 2020, employment rates among workers in the bottom wage quartile decreased by **19.3%** compared to January 2020 (not seasonally adjusted).

DOWNLOAD CHART





The importance of Big Data and Data Science to support decision makers

- Availability of real-time data.
- Quick access to results.
- Supporting statistical data with unconventional data.
- Reduce the cost of data collection.
- deep analysis.
- Forecasting.
- Keep up with technology.





- Access to data in real time and in a timely manner and during the period required for action and decision-making.
- Analysing Non traditional data sources makes it easier to get to more accurate details.
- Take advantage of big data sources to respond to all challenges and crises such as
 - ✓ Large numbers of refugees in the country.
 - ✓ Evaluation/analysis of policies adopted during the pandemic.
- A good option while traditional data collection is difficult.



- Unconventional data for the most important sources is only available during the crisis.
 - *Who determines the domestic/international crisis.*
 - *What are the criteria.*
- Suspected concerns/trends have been confirmed through social media such as Twitter.
- Data bias due to technology dependence is a calculated risk



Ethical Concerns



دائرة الإحصاءات العامة
Department of Statistics

(Potential) concerns

- **Potential bias**
- **Data gaps** (e.g. data on people with disabilities)
- **Exclusion**: People who do not have access to digital media.
- **Data Protection and Privacy Measures**

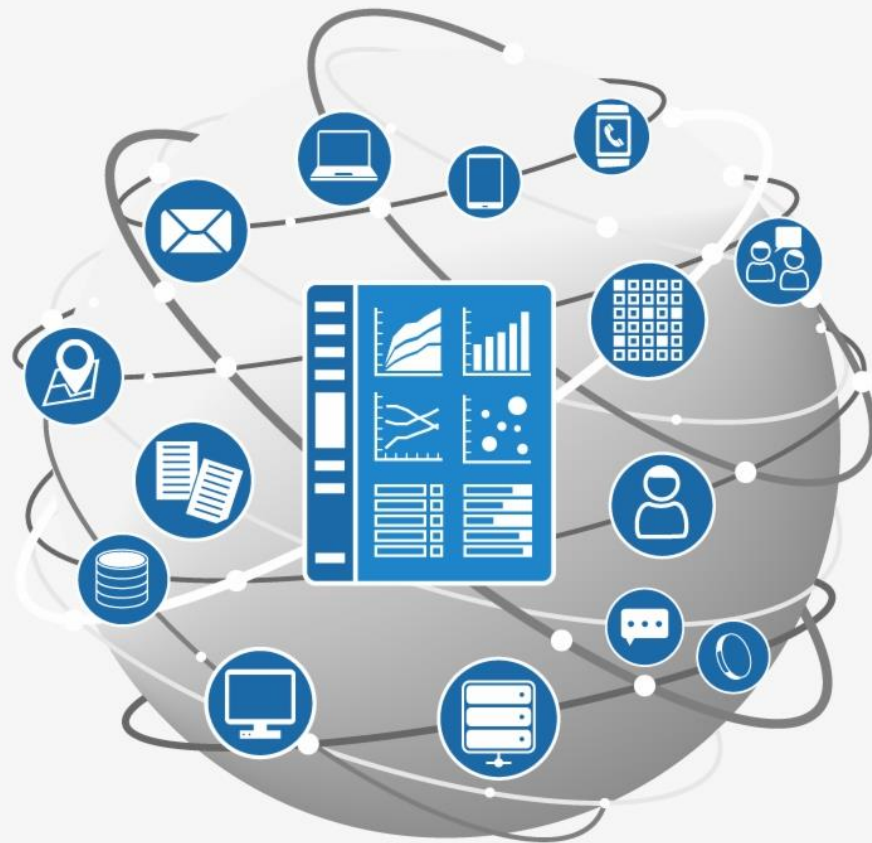




دائرة الإحصاءات العامة
Department of Statistics

THANK YOU

Ahlam ALrousan



Big Data