BIG DATA ANALYTICS FOR OFFICIAL STATISTICS

Dr. Abbas Maaroof
aimaaroof@gmail.com
29-03-2021

WORKSHOP OUTLINE

1. Big Data Analytics-Day #1
2. Machine Learning Tools: R Programming language and Graphical Interfaces -Day#1
3. Data Mining Analytical Tools-Day#2
4. Data Visualization and Using Big Data For SDGs-Day#2
5. Digital Data Collection-Day#3
6. Drone Technology and Multispectral Sensor, Geospatial and Remote sensing Technology-Day#3
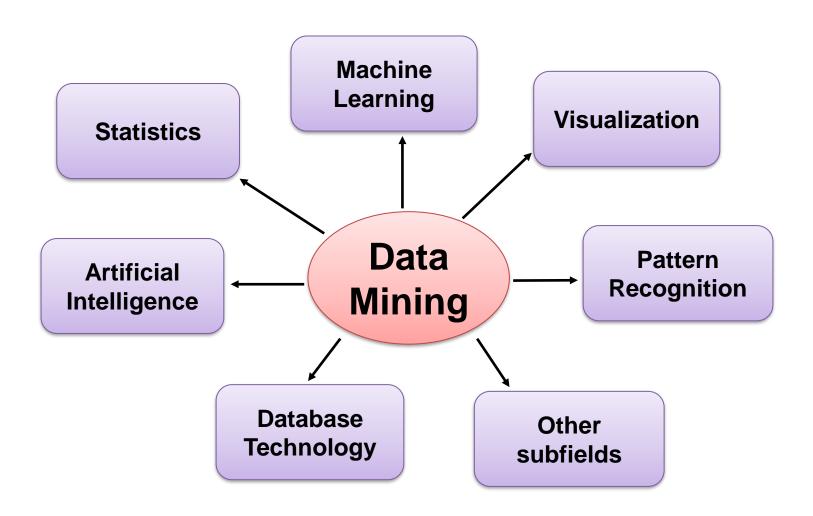
Part#1: Data Mining analytical tools

What is Data Mining?

Data mining or Natural Language Processing (NLP): is the process of extracting useful information or knowledge from a large volume of structured/ unstructured data and databases (relational and non-relational), to discover interesting patterns from large amounts of data.

Text mining is an automatic process that uses natural language processing to extract valuable insights from unstructured text. By transforming data into information that machines can understand, text mining automates the process of classifying texts by sentiment, topic, and intent. Social media(Facebook, Twitter, Instagram, etc.), text, email , and Internet (Google)…etc. (It can be done at a fraction of the cost and in real time).

➢ Traditional statistics, household surveys and census data have been effective in tracking medium to long-term development trends, but are less effective in generating a real-time snapshot for policymakers , while Data Mining method can be done at a fraction of the cost and in real time.

➢ There is an ocean of data (Big Data)— generated by citizens in both developed and developing countries— that did not exist even a few years ago. Videos, Audios, Mobile phones, social media and Internet searches all leave digital traces that, when anonymized, aggregated and analyzed, can reveal significant insights that help governments make faster and more informed decisions.

Data mining confluence of multi-disciplinary subfields

Data mining confluence of multi-disciplinary subfields
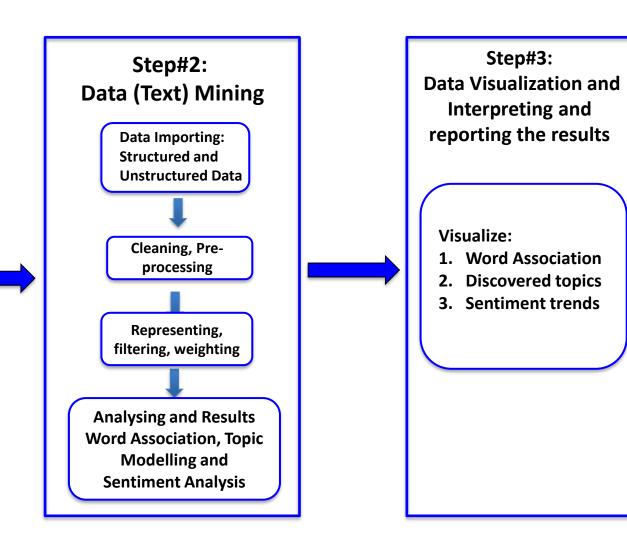
Why is Text Mining Important?

Individuals and organizations generate tons of data every day. Stats claim that almost [80% of the existing text data is unstructured](#), meaning it's not organized in a predefined way, it's not searchable, and it's almost impossible to manage. In other words, it's just not useful.

1. Scalability: with text mining it's possible to analyze large volumes of data in just seconds.

2.Real-time analysis: thanks to text mining, companies can prioritize urgent matters accordingly including, detecting a potential crisis, and discovering product flaws or negative reviews in real time. Why is this so important? Because it allows companies to take quick action.

3. Consistent Criteria: when working on repetitive, manual tasks people are more likely to make mistakes

Data Mining Analytical Steps

Step#1:
Data Collection

1. **Identifying** the Data source information: Such as Social media(Facebook, Twitter, Instagram, etc.), text, email , and Internet (Google)…etc.

2. **Extracting** the relevant information from the data

Step#2:
Data (Text) Mining

Data Importing: Structured and Unstructured Data

Cleaning, Pre-processing

Representing, filtering, weighting

Analysing and Results Word Association, Topic Modelling and Sentiment Analysis

Step#3:
Data Visualization and Interpreting and reporting the results

Visualize:
1. Word Association
2. Discovered topics
3. Sentiment trends

Best Data Mining Techniques with Examples

1. Classification

Example-Data mining can help classify customers in the "low," "medium," or "high" credit risks category by analyzing their purchase history and financial background.

2. Prediction

Example-It allows you to predict the sale of a customer based on their previous purchases, credit histories, and financial status.

3. Cluster Analysis

Example-Market Segmentation

4. Tracking Patterns

Example, a business can notice that a certain product is sold more before a certain festival.

5. Outlier Analysis

Example-Fraud Detection

6. Association Analysis

Example-Market Basket Analysis

5. Evolution Analysis

Example-Forecasting stock market index using Time Series Analysis

Data Mining Prerequisites and Types

The software for data mining falls into two very general categories.

Text Interfaces

These are generally programming languages that use written commands:

1. One of the most fundamental tools in data mining is the statistical **programming language R-Free-Open Source**
2. **Programming language Python 3-Free- Open Source**

Graphical Interfaces

These include specialized applications that use menus, widgets, virtual connections, and it's really easy to see the process:

1. **RapidMiner**- There is a free version and there is a paid version- You can download the free version
2. **KNIME**- You can download it for free
3. **Orange**-You can download it for free
4. **BigML** on Server free for small task but they charge with big data analytics-Nice way to work with.
5. **Apache Spark**
6. **Rattle**

How is Text Analytics Used?

Text Analytics Use Cases

Manufacturers
- Identify root causes of product issues quicker
- Identify trends in market segments
- Understand competitors' products

Government
- Identify fraud
- Understand public sentiments about unmet needs
- Find emerging concerns that can shape policy

Financial Institutions
- Use contact center transcriptions understand customers
- Identify money laundering or other fraudulent situations

Retail
- Identify profitable customers and understand the reasons for their loyalty
- Manage the brand on social media

Legal
- Identify topics and keywords in discovery documents
- Find patterns in defendant's communications

Healthcare
- Find similar patterns in doctor's reports
- Use social media to detect disease outbreaks earlier
- Identify patterns in patient claims data

Telecommunications
- Prevent customer churn
- Suggest up-sell/cross-sell opportunities by understanding customer comments

Life Sciences
- Identify adverse events in medicines or vaccines
- Recommend appropriate research materials
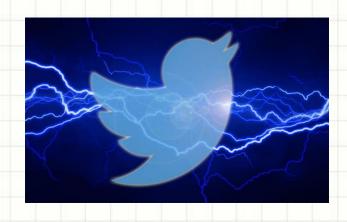
Insurance
- Identify fraudulent claims
- Track competitive intelligence
- Manage the brand on social media

zencos

Applying a hybrid of powerful analytical capabilities Text analysis, combined with predictive modelling and data visualization

https://www.zencos.com/blog/text-mining-examples-advanced-analytics/

Text Mining:
Twitter Data Analysis

Case Study #1: Rohingya Refugees

Cases Study #1: Extract Tweets using a search Term for **Rohingya Refugees**
(Real time analytics)

Number of searched Tweets = 2197

Example of Tweets for Rohingya

[[1]]
[1] "SangwanShilpa: RT @ashokepandit: Go and ask those women constables who were molested by your Rohingya friends on this very day when they destroyed the Ama
[[2]]
[1] "BarmsBham: RT @IMIX_UK: 'Nothing is impossible if you put your mind to it. I was born in a refugee camp, I had no right to education... no right to a…"
[[3]]
[1] "manir_rhg: RT @nslwin: Aung San Suu Kyi and NLD have won a 2nd term. \n\nCan a leopard change its spots?\n\nMy prediction: she will defend genocide, delay…"
[[4]]
[1] "JamalArkani8: They took away all our belongings they denied our experience they tried to exterminate us but we hold our own exper… https://t.co/3MVJPmZc28"
[[5]]
[1] "Switzer45773858: RT @RoMohammadRohi1: Just evil. #Facebook is obstructing the investigation of the Burmese #Genocide of #Rohingya Muslims. Refusing to hand…"
[6]]
[1] "nigib1: RT @jeremycorbyn: \"Solidarity around the world for people who are victims of political or human rights abuses should not know any boundarie…"
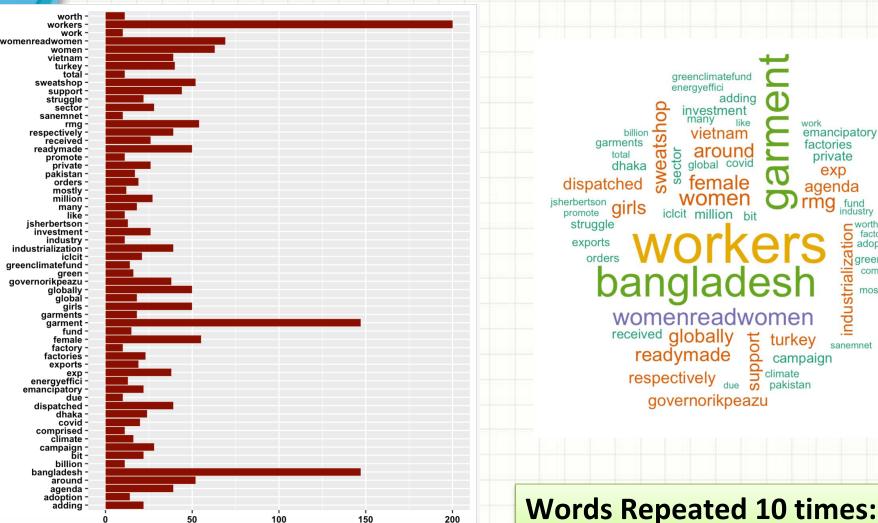
Viewing some of the tweets and retweets

Inspect 10 most frequently used words

Repeated 10 times:

```
> (freq.terms <- findFreqTerms(tdm, lowfreq = 10))
 [1] "adding"           "adoption"          "agenda"            "around"
 [5] "bangladesh"       "billion"           "bit"               "campaign"
 [9] "climate"          "comprised"         "covid"             "dhaka"
[13] "dispatched"       "due"               "emancipatory"      "energyeffici"
[17] "exp"              "exports"           "factories"         "factory"
[21] "female"           "fund"              "garment"           "garments"
[25] "girls"            "global"            "globally"          "governorikpeazu"
[29] "green"            "greenclimatefund" "iclcit"             "industrialization"
[33] "industry"         "investment"        "jsherbertson"      "like"
[37] "many"             "million"           "mostly"            "orders"
[41] "pakistan"         "private"           "promote"           "readymade"
[45] "received"         "respectively"      "rmg"               "sanemnet"
[49] "sector"           "struggle"          "support"           "sweatshop"
[53] "total"            "turkey"            "vietnam"           "women"
[57] "womenreadwomen"   "work"              "workers"           "worth"
>
```

Words Repeated 10 times:

Word Association

Explore frequent terms and their associations

Word Association

Now that we have the most frequently occurring terms in hand, we can check for correlations between some of these words.

In this context, correlation is a quantitative measure of the co-occurrence of words in multiple tweets.

Any specific word which can useful for analysis and help us identify the highly correlate words with that term. If word always appears together then correlation=1.0.

Specifying a correlation limit of 0.1
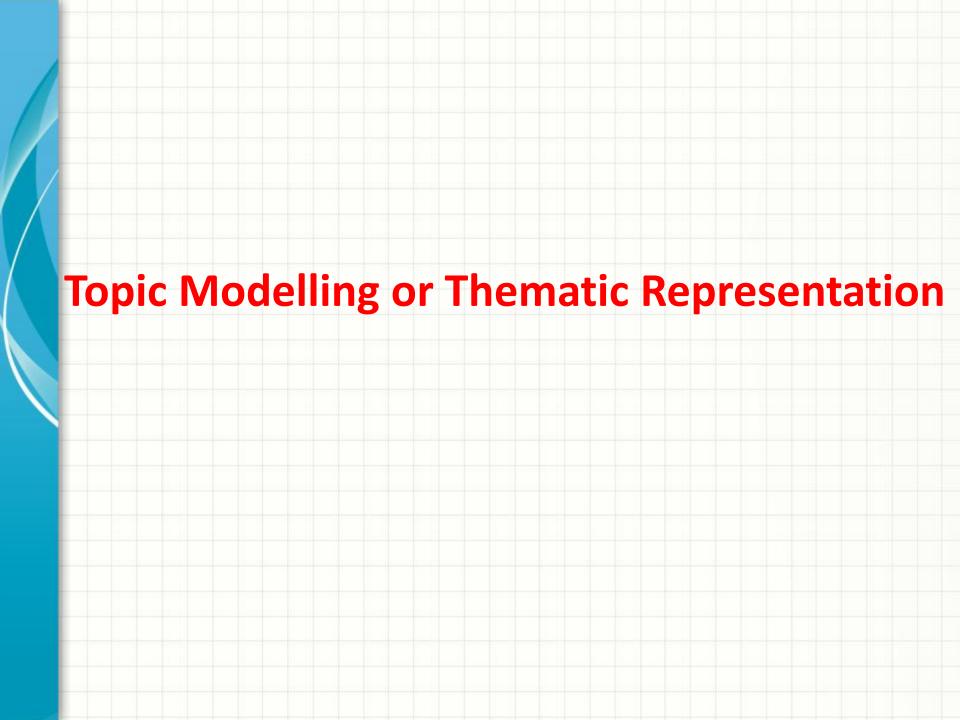
Let us pick seven keywords from slide 3 and correlate with limit of 0.1

findAssocs(tdm, "**covid**", 0.1)
findAssocs(tdm, "**girls**", 0.1)
findAssocs(tdm, "**readymade**", 0.1
findAssocs(tdm, "**garment**", 0.1)
findAssocs(tdm, "**workers**", 0.1)
findAssocs(tdm, "**dispatched**", 0.1)
findAssocs(tdm, "**bangladesh**", 0.1

Which words are associated with "covid" ?

```
> findAssocs(tdm, "covid", 0.1)
$covid
          many     factories     cancelled      clothing        leaving
          0.83          0.69          0.66          0.66           0.66
      liveblog     retailers         today           due   cleanclothes
          0.66          0.⌷Spellcheck 0.66          0.62           0.58
      incurred       julyoct          loss positivenewsuk          stock
          0.58          0.58          0.58          0.58           0.58
        unsold           dec        orders         wages         brands
          0.58          0.53          0.53          0.45           0.31
      continue         fewer           led      pandemic        percent
          0.31          0.31          0.31          0.31           0.31
         piles       placing        second         shops           shut
          0.31          0.31          0.31          0.31           0.31
          wave       exports            bd       limited            one
          0.31          0.26          0.24          0.24           0.24
 asiafloorwage        bangla         expor             h            sto
          0.22          0.22          0.22          0.22           0.22
          vqxs       surplus    bangladesh       garment            wor
          0.22          0.20          0.17          0.14           0.14
       fashion           sit
          0.13          0.13
```

Topic Modelling or Thematic Representation

Topic Modelling or Thematic Representation

Topic modelling is an 'unsupervised' machine learning technique, in other words, one that doesn't require training. Topic models are suitable for the exploration of data.

The calculation of topic models aims to determine the proportionate composition of a fixed number of topics in the documents of a collection. It is useful to experiment with different parameters in order to find the most suitable parameters for your own analysis needs

Top six topics for ten most frequent words

```
> terms(topicModel, 10)
         Topic 1             Topic 2                 Topic 3        Topic 4        Topic 5          Topic 6
 [1,] "womenreadwomen"  "industrialization"  "rmg"          "bangladesh"  "workers"        "garment"
 [2,] "workers"         "respectively"       "female"       "million"     "support"        "bangladesh"
 [3,] "women"           "vietnam"            "sweatshop"    "sector"      "campaign"       "factories"
 [4,] "exports"         "exp"                "readymade"    "received"    "dhaka"          "covid"
 [5,] "garments"        "governorikpeazu"    "around"       "investment"  "bit"            "orders"
 [6,] "mostly"          "turkey"             "globally"     "private"     "emancipatory"   "many"
 [7,] "billion"         "agenda"             "girls"        "global"      "iclcit"         "like"
 [8,] "comprised"       "dispatched"         "workers"      "climate"     "struggle"       "due"
 [9,] "worth"           "bangladesh"         "women"        "green"       "adding"         "dispatched"
[10,] "total"           "factory"            "total"        "fund"        "garment"        "industry"
```
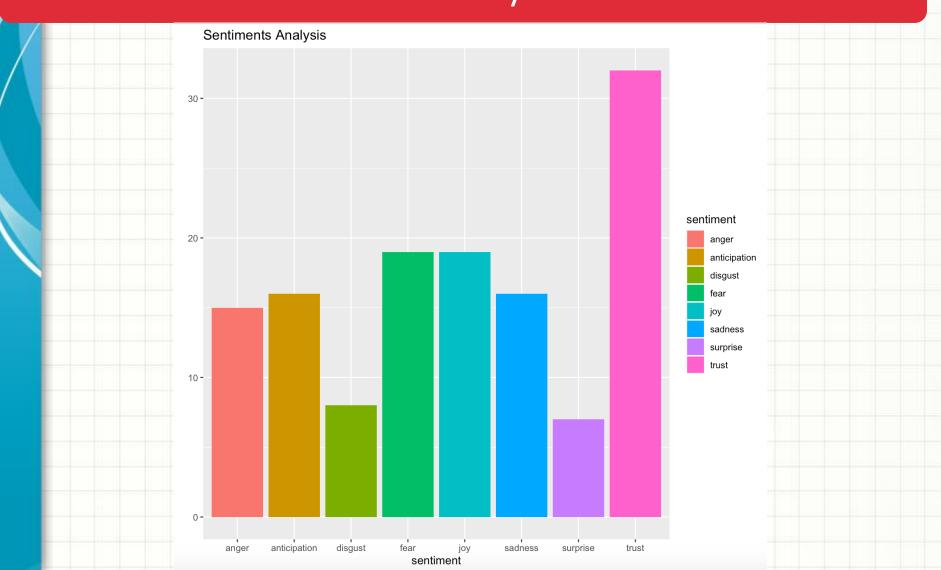
Classification by Emotion and Polarity

My sentiment analysis is based on the NRC Word-Emotion Association Lexicon
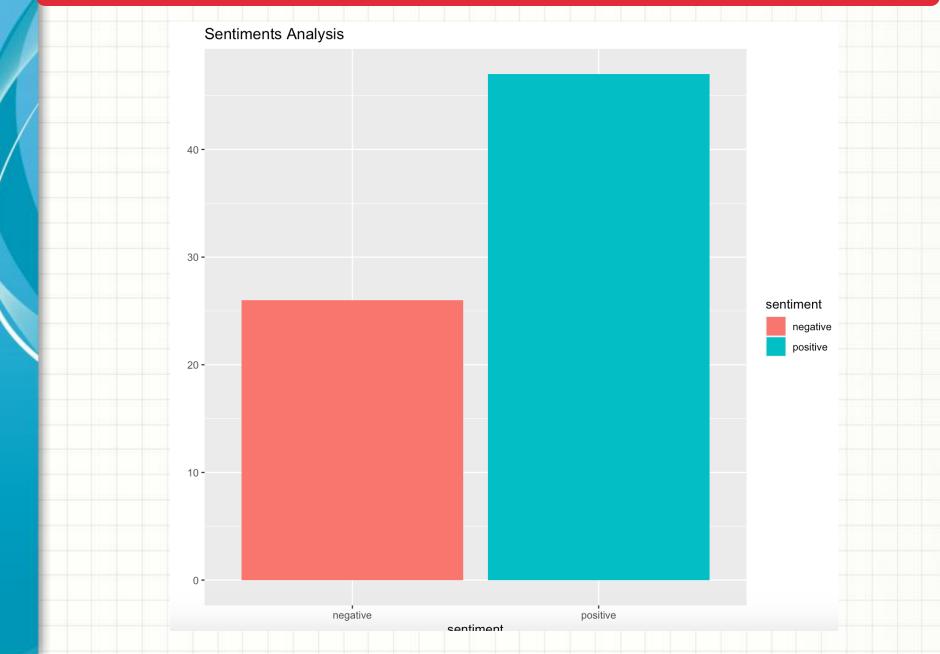
The NRC Emotion Lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The annotations were manually done by crowdsourcing.

More details can be found at this link:
https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

Classification by Emotion

Sentiments Analysis

To give more descriptive analytics to sentiment analysis I have analysed the influence of each sentiment words with related frequent words in the world cloud package

To give more descriptive analytics to sentiment analysis I have analysed the influence of each sentiment words with related frequent words in the world cloud package as shown below

Let us analyse for example "Fear" emotion

The influence of words classified as "fear" **NRC Word-Emotion Association Lexicon** are analyzed in the wordcloud() function, as show below, inside the wordcloud figure shows the words "flood" and "prevent" come up on top; and second "Persecution" and "suffering" and the remaining words are less feared. The people are more feared of flood than persecution and pandemic

Text Mining:
Facebook Data Analysis

Case Study #2: COVID-19 in Philippines

Exploring Public Response to COVID-19 on Philippines Facebook with LDA Topic Modelling and Sentiment Analysis
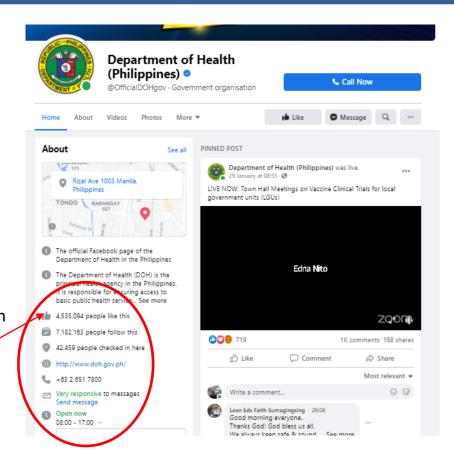
1. Data Collection for Department of Health (Philippines) Facebook page two consecutive years 2019 and 2020
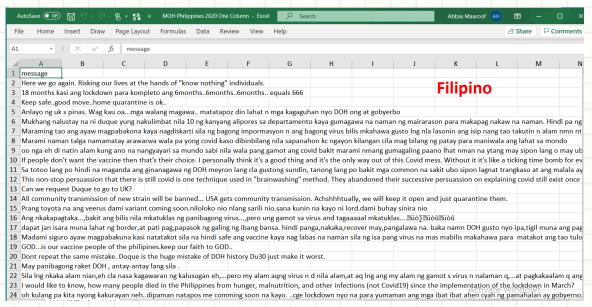
Facebook Page	No. of Like	No. of Followers	Period	No of Posts and comments extraxted
Department of Health (Philippines)	4,440,370	7,084,707	(1/1/2019)-(31/12/2019)	62280
			1/1/2020)-(31/12/2020)	69802
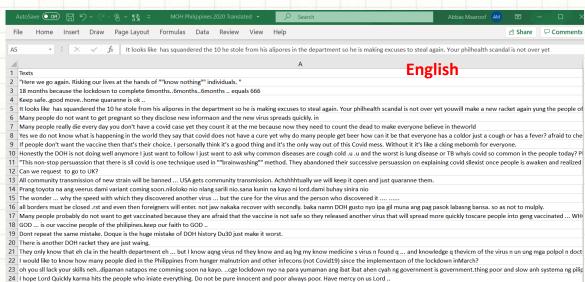
To understand the public perceptions related to COVID- 19 on Department of Health (Philippines) Facebook page, this research collects data from Facebook posts and related comments by using **Facepager**. It then analyses the data with text mining techniques, including Latent Dirichlet Allocation (LDA) topic modelling and sentiment analysis.

This is the most liked and followed Facebook page-The Department of Health. Total posts and comments collected for this evaluation: 425611

Google Translate Example from Filipino to English

Filipino

	A
1	message
2	Here we go again. Risking our lives at the hands of "know nothing" individuals.
3	18 months kasi ang lockdown para kompleto ang 6months..6months..6months.. equals 666
4	Keep safe..good move..home quarantine is ok..
5	Anlayo ng uk s pinas. Wag kau oa...mga walang magawa.. matatapoz din lahat n mga kagaguhan nyo DOH ong at gobyerbo
6	Mukhang nalustay na ni duque yung nakulimbat nila 10 ng kanyang alipores sa departamento kaya gumagawa na naman ng mairarason para makapag nakaw na naman. Hindi pa ng
7	Maraming tao ang ayaw magpabakona kaya nagdiskarti sila ng bagong impormasyon n ang bagong virus bilis mkahawa gusto lng nla lasonin ang isip nang tao takutin n alam nmn nt
8	Marami naman talga namamatay arawarwa wala pa yong covid kaso dibinbilang nila sapanahon kc ngayon kilangan cila mag bilang ng patay para maniwala ang lahat sa mondo
9	oo nga eh di natin alam kung ano na nangyayari sa mundo sabi nila wala pang gamot ang covid bakit marami nmang gumagaling paano lhat nman na ytang may sipon lang o may ub
10	If people don't want the vaccine then that's their choice. I personally think it's a good thing and it's the only way out of this Covid mess. Without it it's like a ticking time bomb for ev
11	Sa totoo lang po hindi na maganda ang ginanagawa ng DOH meyron lang cla gustong sundin, tanong lang po bakit mga common na sakit ubo sipon lagnat trangkaso at ang malala ay
12	This non-stop persuassion that there is still covid is one technique used in "brainwashing" method. They abandoned their successive persuassion on explaining covid still exist once
13	Can we request Duque to go to UK?
14	All community transmission of new strain will be banned... USA gets community transmission. Achshhhtually, we will keep it open and just quarantine them.
15	Prang toyota na ang veerus dami variant coming soon.niloloko nio nlang sarili nio.sana kunin na kayo ni lord.dami buhay sinira nio
16	ang nkakapagtaka...,bakit ang bilis nila mkatuklas ng panibagong virus...,pero ung gamot sa virus and tagaaaaal mkatuklas....☒üò⟩☒üòû☒üòû
17	dapat jan isara muna lahat ng border,at pati pag,papasok ng galing ng ibang bansa. hindi panga,nakaka,recover may,pangalawa na. baka namn DOH gusto nyo ipa,tigil muna ang pag
18	Madami siguro ayaw magpabakuna kasi natatakot sila na hindi safe ang vaccine kaya nag labas na naman sila ng isa pang virus na mas mabilis makahawa para matakot ang tao tulo
19	GOD...is our vaccine people of the philipines.keep our faith to GOD..
20	Dont repeat the same mistake. Doque is the huge mistake of DOH history Du30 just make it worst.
21	May panibagong raket DOH , antay-antay lang sila .
22	Sila lng nkaka alam nian,eh cla nasa kagawaran ng kalusugan eh,...pero my alam aqng virus n d nila alam,at aq lng ang my alam ng gamot s virus n nalaman q,...at pagkakaalam q ang
23	I would like to know, how many people died in the Philippines from hunger, malnutrition, and other infections (not Covid19) since the implementation of the lockdown in March?
24	oh kulang pa kita nyong kakurayan neh..dipaman natapos me comming soon na kayo. ..cge lockdown nyo na para yumaman ang mga ibat ibat ahen cyah ng pamahalan ay gobyerno.

English

	A
1	Texts
2	"Here we go again. Risking our lives at the hands of ""know nothing"" individuals. "
3	18 months because the lockdown to complete 6months..6months..6months .. equals 666
4	Keep safe..good move..home quaranne is ok ..
5	It looks like has squandered the 10 he stole from his alipores in the department so he is making excuses to steal again. Your philhealth scandal is not over yet youwill make a new racket again yung the people of
6	Many people do not want to get pregnant so they disclose new informaon and the new virus spreads quickly. in
7	Many people really die every day you don't have a covid case yet they count it at the me because now they need to count the dead to make everyone believe in theworld
8	Yes we do not know what is happening in the world they say that covid does not have a cure yet why do many people get beer how can it be that everyone has a coldor just a cough or has a fever? afraid to che
9	If people don't want the vaccine then that's their choice. I personally think it's a good thing and it's the only way out of this Covid mess. Without it it's like a cking mebomb for everyone.
10	Honestly the DOH is not doing well anymore I just want to follow I just want to ask why common diseases are cough cold .u .u and the worst is lung disease or TB whyis covid so common in the people today? Pl
11	"This non-stop persuassion that there is sll covid is one technique used in ""brainwashing"" method. They abandoned their successive persuassion on explaining covid sllexist once people is awaken and realized
12	Can we request to go to UK?
13	All community transmission of new strain will be banned ... USA gets community transmission. Achshhhtually we will keep it open and just quaranne them.
14	Prang toyota na ang veerus dami variant coming soon.niloloko nio nlang sarili nio.sana kunin na kayo ni lord.dami buhay sinira nio
15	The wonder ... why the speed with which they discovered another virus ... but the cure for the virus and the person who discovered it ....
16	all borders must be closed .rst and even then foreigners will enter. not jaw nakaka recover with secondly. baka namn DOH gusto nyo ipa gil muna ang pag pasok labasng bansa. so as not to mulply.
17	Many people probably do not want to get vaccinated because they are afraid that the vaccine is not safe so they released another virus that will spread more quickly toscare people into geng vaccinated ... WH
18	GOD ... is our vaccine people of the philipines.keep our faith to GOD ..
19	Dont repeat the same mistake. Doque is the huge mistake of DOH history Du30 just make it worst.
20	There is another DOH racket they are just waing.
21	They only know that eh cla in the health department eh ... but I know aqng virus nd they know and aq lng my know medicine s virus n found q ... and knowledge q thevicm of the virus n un ung mga polpol n doct
22	I would like to know how many people died in the Philippines from hunger malnutrion and other infecons (not Covid19) since the implementaon of the lockdown inMarch?
23	oh you sll lack your skills neh..dipaman natapos me comming soon na kayo. ..cge lockdown nyo na para yumaman ang ibat ibat ahen cyah ng government is government.thing poor and slow anh systema ng pilip
24	I hope Lord Quickly karma hits the people who iniate everything. Do not be pure innocent and poor always poor. Have mercy on us Lord ..

2. Word Frequency and Cloud

2020

Period : (1/1/2020)-(31/12/2020)

Plotting 100 frequently used words

Inspect 100 words frequently used

```
> head(freq, 100)
     covid    vaccine     people       virus       know        new       want        now         ban       like        get
     14494       9578       7560        5780       3856       3784       3692       3646        3164       3050       3020
      days     travel     really philippines       good      every       hope       many  government       even vaccinated
      2600       2574       2460        2442       2354       2340       2270       2220        2120       2060       2010
      make       life      world     believe     please       mary       year   lockdown        link       made      china
      1980       1930       1870        1822       1760       1660       1656       1650        1624       1610       1600
   country     family       help       happy        die       take       need       safe         let     inject      right
      1554       1550       1502        1490       1480       1480       1460       1442        1430       1380       1376
      lord       home       work       start     resign        see      enter    another     already   business   pandemic
      1340       1310       1290        1250       1210       1190       1150       1130        1110       1110       1110
     thank  countries     posive        stop    contact        day       sure       poor      number       true       long
      1100       1080       1080        1080       1050       1040       1040       1030        1010       1010       1000
   trading       face       earn     without       care     person       died investment    vaccines    disease    earning
      1000        990        980         980        964        960        958        950         950        930        930
      keep       test    variant       never      today       case       give      think       cases      trust     always
       910        910        892         880        874        870        870        870         862        860        850
    spread     health     afraid   filipinos       free        pay      bless      hindi     nothing        way       sana
       850        842        840         810        792        790        780        780         780        780        770
    strain
       760
>
```
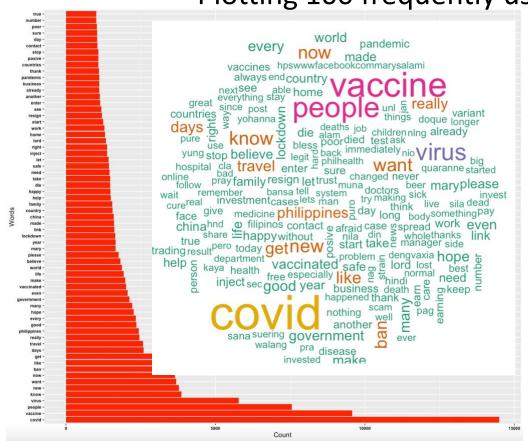
Data processing

Plotting 100 frequently used words

This word cloud shows the frequency of these specific words appearing in the Facebook text data mined. The largest fonts (shown in yellow and purple) are the most frequently used words. The second largest fonts (shown in red and green) indicate the second most frequently used words. The next few slides outline how the reader should interpret this specific word cloud.
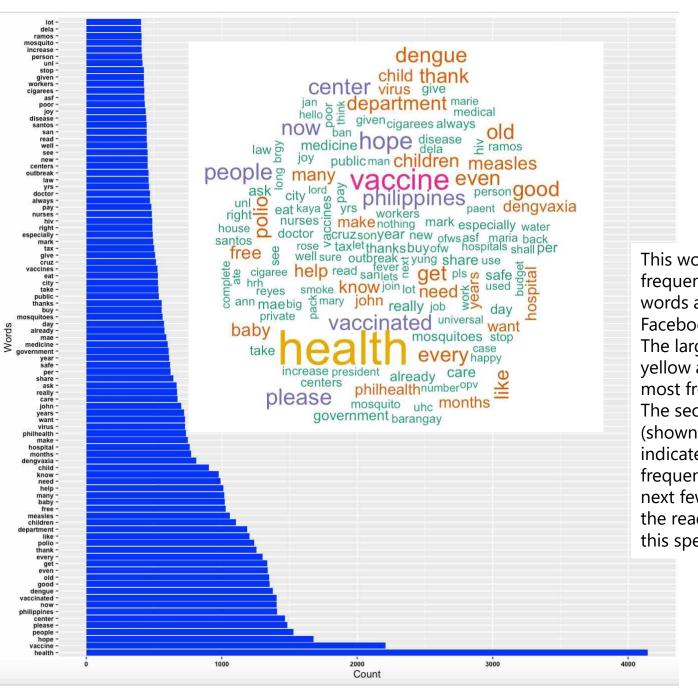
3. Word Frequency and Cloud

2019

Period : (1/1/2019)-(31/12/2019)

Inspect 100 words frequently used
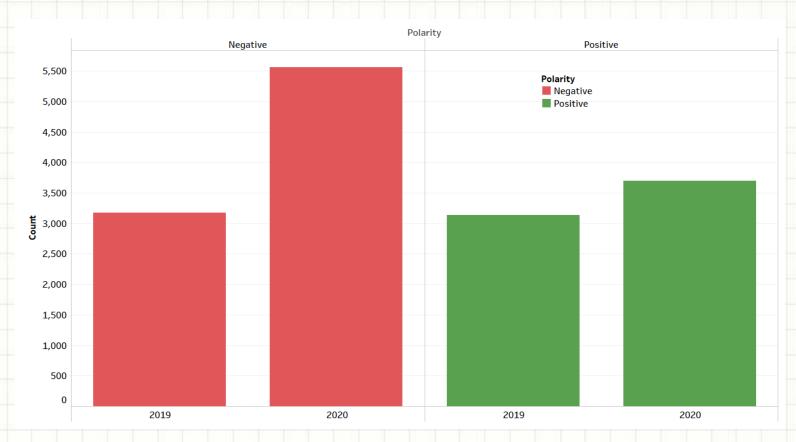
```
> freq <- sort(colSums(as.matrix(dtm)), decreasing=TRUE)
> head(freq, 100)
```

health	vaccine	hope	people	please	center	philippines	now	vaccinated
4144	2210	1676	1528	1484	1468	1408	1404	1404
dengue	good	old	even	get	every	thank	polio	like
1376	1354	1350	1340	1334	1302	1258	1238	1204
department	children	measles	free	baby	many	help	need	know
1188	1104	1060	1028	1020	1018	1010	990	976
child	dengvaxia	months	hospital	make	philhealth	virus	want	years
904	812	774	764	748	734	728	728	722
john	care	really	ask	share	per	safe	year	government
700	672	670	668	642	622	618	612	608
medicine	mae	already	day	mosquitoes	buy	thanks	public	take
602	592	578	572	564	556	556	536	532
city	eat	vaccines	cruz	give	tax	mark	especially	right
530	530	526	514	506	500	494	492	488
hiv	nurses	pay	always	doctor	yrs	law	outbreak	centers
486	484	480	470	468	464	458	458	454
new	see	well	read	san	santos	disease	joy	poor
454	452	448	446	446	444	442	438	432
asf	cigarees	workers	given	stop	unl	person	increase	mosquito
430	428	426	424	424	414	410	408	408
ramos	dela	lot	son	work	medical	ann	yung	barangay
408	404	404	400	394	392	386	384	382
uhc								
376								

This word cloud shows the frequency of these specific words appearing in the Facebook text data mined. The largest fonts (shown in yellow and purple) are the most frequently used words. The second largest fonts (shown in red and green) indicate the second most frequently used words. The next few slides outline how the reader should interpret this specific word cloud.

4. Sentiment Analysis
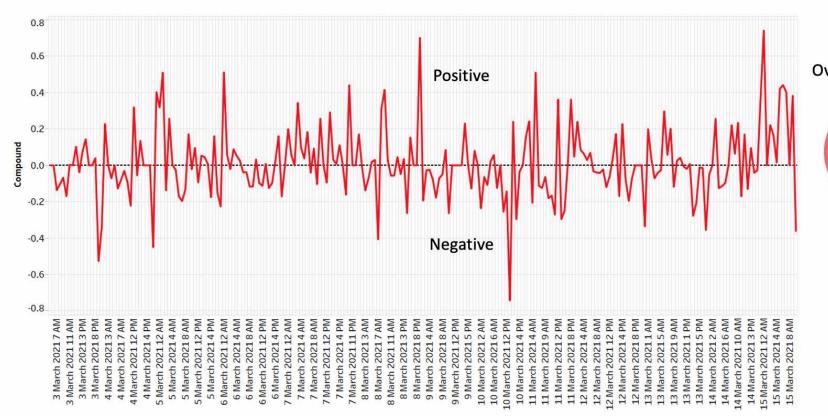
The total sentiment of the **Department of Health (Philippines) Facebook page**
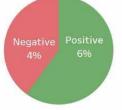
Real Time Monitoring

The *compound* by score
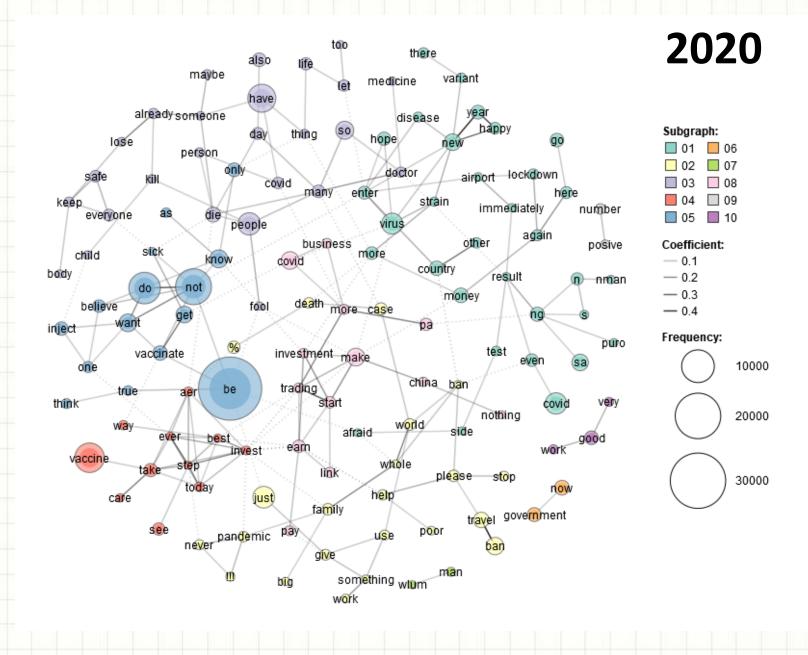
Compound represents the total sentiment of a comment on Vaccine, where -1 is the most negative and 1 the most positive at a real time

3. Co-Occurrence Network of Words

2020

PART#1: DATA VISUALIZATION AND USING BIG DATA FOR SDGS

What is Data visualization?

Data visualization is the act of taking information (data) and placing it into a visual context, such as a map or graph.

Data visualizations make big and small data easier for the human brain to understand, and visualization also makes it easier to detect patterns, trends, and outliers in groups of data.

Good data visualizations should place meaning into complicated datasets so that their message is clear and concise.

Why data visualization is important?

1. Interactive images speak louder than words.
2. Data visualization is especially important when it comes to big data
3. The results from complex algorithms are much easier to understand in a visual format
4. Data visualization can be use in any field such as in finance, marketing, tech, design, or anything else, you need to visualize data
5. Visualized data gives stakeholders, business owners, and decision-makers a better prediction

Google, Facebook, Amazon, Apple, Twitter and Netflix all ask better questions of their data –and make better business decisions –**by using data visualization.**

Tools used to Build Data Visualization

Tools used to Build Data Visualization

Best tools for Big Data visualization can be outlined as follows:

1.Tableau
2.Infogram
3.ChartBlocks
4.Datawrapper
5.Plotly
6.RAW
7.Visual.ly
8.D3.js
9.Ember Charts
10.Google Charts
11.Highcharts
12.Polymaps
13.Modest Maps

https://bigdata-madesimple.com/review-of-20-best-big-data-visualization-tools/

Tool	Learning Curve	Code Sets	Special Features	Premium Level Available?	Tutorial Availability	Import Formats
Charted	Low	No coding required	Simplicity in chart creation	No	None	CSV/TSV/Google and Dropbox
Datawrapper	Low	No coding required	Allows for brand style matching	Yes	None	CSV/Excel/Google Sheets
Leaflet	Moderate	JavaScript	Mobile-friendly mapping	No	Adequate	N/A
RawGraphs	Low to Moderate	No	Extensive Non-Conventional Charting Types	No	Adequate	CSV/TSV/Any Copy/Paste & APIs
Chartist.JS	Moderate	JavaScript CSS/SASS	Uses SVG and quick animation	No	Limited	N/A
D3.JS	Moderate to High	JavaScript	Extensive flexibility in developing visualizations	No	Extensive	N/A
Plotly	Moderate	JavaScript Python R	Provides for collaborative development	Yes	Limited Master classes available ($)	Can pull from SQL (Client available)
Polymaps	Moderate to High	JavaScript	Provides for dynamic map development	No	Adequate	N/A
OpenHeatMap	Low	No coding required	Simplicity in map creation	No	Limited	CSV/Excel/Google
DyGraphs	Moderate to High	JavaScript/CSS	Allows for handling of large data sets	No	Limited	Primarily CSV and URL Assignment
Candela	Moderate to High	JavaScript Python R	Provides complete suite of tools	No	Adequate	N/A

11 Best Open-Source Data Visualization Tools

https://www.kolabtree.com/blog/11-best-open-source-data-visualization-tools/
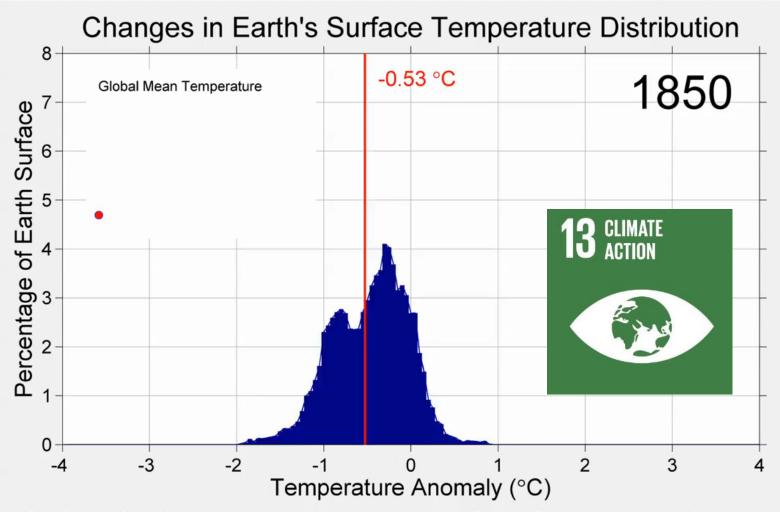
Free Tools! ☺

Global warming is a lie? One thing I can tell you: **Data never lies. Here's a cool visualization showing how the distribution over Earth's surface of annual average temperature anomalies has been shifting due to global warming since 1850**. TLDR? **Earth is heating up! hashtag#datahashtag#visualizationhashtag#globalwarming**
Source:
 https://lnkd.in/efswdQp

Changes in Earth's Surface Temperature Distribution

Data Source: 12-month surface temperature anomaly distributions from Berkeley Earth, relative to 1951-1980 average

Case Study#1: Iraq SDGs Visualized dashboards

https://public.tableau.com/profile/abbas1073#!/

Part#1: Iraq SDGs Data Mapping and Gap Analysis Assessment

Objective of Iraq SDGs Data Mapping and Gap Analysis Assessment

What is the objective of Iraq SDGs Data Mapping and Gap Analysis Assessment?

The <u>objective</u> of this assignment to collect information necessary to map the data that is available for the Sustainable Development Goals (SDGs) Indicators in two tiers Categorization

The information collected will <u>assist</u> in the identification of potential data gaps or <u>areas requiring further development to inform the elaboration of an action plan for improving data quality to monitor SDGs-2030 agenda at national level.</u>

SDGs Indicators Category (Level)

The IAEG has categorized the SDG indicators into two "Tiers":

Tier (1): Ready-Feasible with more effort

Tier (2): Not Ready-Feasible with Strong Effort

As of the 51st Session (New York 3 - 6 March 2020) of the United Nations Statistical Commission, the global indicator framework does not contain any Tier III indicators)

SDGs Five Thematic Areas

Iraq SDGs Data Matrix For the Online Platform

Data Gap in SDGs Data Matrix

Country	Level	Thematical Areas	Goal No.	Goal	Target No.	Target	Indicator No.	Indicator	Year-2020	Year-2019	Year-2018	Year-2017	Year-2016	Year-2015	Year-2014	Year-2013	Year-2012	Year-2011	Data Source	Unit of measurement	Index measurement methodology	Level of detail	Index calculation patrol	Potential specialized agencies	Partner agencies
Iraq	Tier 2	People	Goal 1	Goal 1: No Poverty	Target 1.1.	Target 1.1. By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than $1.25 a day	Indicator 1.1.1.	Indicator 1.1.1. The proportion of the population living below the international poverty line is classified by gender, age, employment status and geographical location (urban/rural)																World Bank	ILO
Iraq	Tier 1	People	Goal 1	Goal 1: No Poverty	Target 1.2.	Target 1.2. By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions	Indicator 1.2.1.	Indicator 1.2.1.The proportion of the population living below the national poverty line is classified by sex and age	31,7		20,05			22,5	22,50				Poverty monitoring and assessment survey in Iraq	%	Number of poor / population	Iraq	As needed	World Bank	UNICEF
Iraq	Tier 2	People	Goal 1	Goal 1: No Poverty	Target 1.2.	Target 1.2. By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions	Indicator 1.2.2.	Indicator 1.2.2. The proportion of men, women and children of all ages living in poverty in all its dimensions in accordance with national definitions														National Gov		UNICEF World Bank UNDP"
Iraq	Tier 1	People	Goal 1	Goal 1: No Poverty	Target 1.3.	Target 1.3. Implement nationally appropriate social protection systems and measures for all, including floors, and by 2030 achieve substantial coverage of the poor and the vulnerable	Indicator 1.3.1.	Indicator 1.3.1. Proportion of population covered by social protection floors/systems, by sex, distinguishing children, unemployed persons, older persons, persons with disabilities, pregnant women, newborns, work-...	(899563) Male (436943) Female (1336506) Total	(703102) Male (438,281) females (1141383) Total	(3848220) Male (2455041) Female (6303261) Total	(492202) Male (413,571) females (905773) Total							Ministry of Labour and Social Affairs	Number	Number of beneficiaries covered Social protection network	Iraq / Province / Sex	Annual	ILO	World Bank
Iraq	Tier 2	People	Goal 1	Goal 1: No Poverty	Target 1.4.	Target 1.4. By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basic services, ownership and control over land and other forms of property, inheritance, natural resources, appropriate new technology and financial services, including microfinance	Indicator 1.4.1.	Indicator 1.4.1. The proportion of the population living in households has access to basic services															UN-Habitat	UNICEF, WHO
Iraq	Tier 2	People	Goal 1	Goal 1: No Poverty	Target 1.4.	Target 1.4. By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basic services, ownership and control over land and other forms of property, inheritance, natural resources, appropriate new technology and financial services, including microfinance	Indicator 1.4.2.	Indicator 1.4.2. Proportion of total adult population with secure tenure rights to land, with legally recognized documentation, and who perceive their rights to land as secure, by sex and by type of tenure															World Bank UN-Habitat"	FAO UNSD UN Women UNEP IFAD

Data Methodology Mapping Results of All SDGs Indicators Relevant for Iraq

| IAEG- Category	National Methodology	Data	Number	Percentage
Tier (1)-Ready-Feasible with more effort	Methodology exists	1.Data only available at national level, not at subnational level (e.g. district, city or village level) 2. Indicators require data that need to be improved, adjusted or modified based on existing national data.	94	38%
Tier (2)-Not Ready-Feasible with strong effort	Methodology does not exist	1. Data is not available 2. The indicators require data collected using a new methodology or approach	153	62%
			247	100%
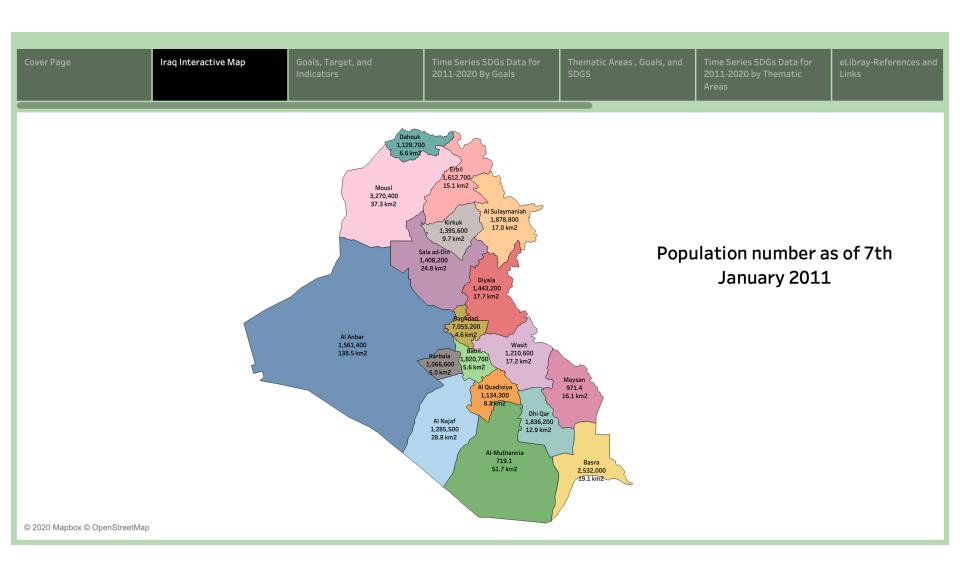
Tier (1)-Ready-Feasible with more effort

Goal 1: No Poverty	Indicator 1.2.1. The proportion of the population living below the national poverty line is classified by sex and age
Global Methodology	$$P_0 = \frac{1}{N} + \sum_{i=1}^{N} I(y_i < z) = \frac{N_p}{N}$$
Global Indicator Variables	1. Np is the total, urban or rural number of poor. 2. N is the total, urban or rural population.
Global Methodology Data Disaggregation	1. Rural area 2. Urban area
Nationally Adapted Methodology	Number of poor / population
Nationally Adapted Variables	Number of poor
Level and frequency of collection	National; as needed
Values per year	(22.50%), 2012 (22.50)%, 2014 (20.5%), 2018 (31.7%), 2020
Missing Variables (Data gaps) against Global Methodology	1. Rural area 2. Urban area
Data Source	1.Central Statistical Organization 2. Poverty monitoring and assessment survey in Iraq 3. Household surveys 4. Study of the impact of the Corona pandemic on poverty and vulnerability in Iraq
Potential specialized agencies	World Bank
Partner agencies	UNICEF
Proposed Data Collection methods and Advanced Analytical Tools	1. Advanced Mobile Based Data Collection tools support capturing data as text, numbers, dates, GPS, photos, video and audio such as Open Data Kit and Kobo Toolbox. They allow for offline data collection with mobile devices in remote areas with advanced data curation and visualization on the device. 2.R and Python are programming languages and free modern software environment for data Science, statistical computing and data visualization. The R and python languages have become very popular among statisticians and data miners (Big Data analytics) and is widely used for advanced data analysis in statistical methodology, survey creation, data cleaning and data analysis 3.Data Visualization tools to visualize large amounts of complex structure and unstructured data and to identify areas that need attention or improvement such as Tableau, Highcharts, Power BI, etc. 4. Graphical User Interface softwares; these include specialized applications that use menus, widgets, virtual connections, no programming skills required Applying a hybrid of powerful analytical capabilities Text analysis, combined with predictive modelling and data visualization such as RapidMiner, KNIME, and Orange 5. Big Data Analytics and Data Mining tools for large volumes of data using Machine Learning modelling
Links	Global Metadata 1.2.1.

Tier (2)-Not Ready-Feasible with Strong Effort

Goal 9: Industry, Innovation and Infrastructure	Indicator 9.5.1. Research and development expenditure as a proportion of GDP
Global Methodology	Research and development expenditure as a proportion of GDP (R&D$_{Intensity}$) is calculated as: $$R\&D_{intensity} = \frac{\textit{The total intramural expenditure on R\&D (GERD)}}{GDP} \times 100$$
Global Indicator Variables	Computation of the indicator Research and development (R&D) expenditure as a proportion of Gross Domestic Product (GDP) is self-explanatory, using readily available GDP data as denominator.
Global Methodology Data Disaggregation	R&D expenditure can be disaggregated by sector of performance, source of funds, field of R&D, type of research and type of cost. The Frascati Manual provides more details related to these breakdowns (what these breakdowns/classifications are, the purposes, including user needs, the main criteria that are applied, etc).
Nationally Adapted Methodology	NIL
Nationally Adapted Variables	NIL
Level and frequency of collection	NIL
Values per year	NIL
Missing Variables (Data gaps) against Global Methodology	1. Total intramural expenditure on R&D 2. GDP
Year data was last collected, frequency of collection	NIL
Proposed Data Source	1. Central Statistical Organization (CSO) 2. Ministry of Higher Education and Scientific Research
Potential specialized agencies	UNESCO-UIS
Partner agencies	NIL
Proposed Data Collection methods and Advanced Analytical Tools	1. Advanced Mobile Based Data Collection tools support capturing data as text, numbers, dates, GPS, photos, video and audio such as Open Data Kit and Kobo Toolbox. They allow for offline data collection with mobile devices in remote areas with advanced data curation and visualization on the device. 2. R and Python are programming languages and free modern software environment for data Science, statistical computing and data visualization. The R and python languages have become very popular among statisticians and data miners (Big Data analytics) and is widely used for advanced data analysis in statistical methodology, survey creation, data cleaning and data analysis. 3. Data Visualization tools to visualize large amounts of complex structure and unstructured data and to identify areas that need attention or improvement such as Tableau, Highcharts, Power BI, etc. 4. Graphical User Interface softwares; these include specialized applications that use menus, widgets, virtual connections, no programming skills required Applying a hybrid of powerful analytical capabilities Text analysis, combined with predictive modelling and data visualization such as RapidMiner, KNIME, and Orange 5. Big Data Analytics and Data Mining tools for large volumes of data using Machine Learning modelling.
Links	**Official SDG Metadata URL** https://unstats.un.org/sdgs/metadata/files/Metadata-09-05-01.pdf

Interactive Iraq Map for Each 18 Provinces Page

Dahouk
1,128,700
6.6 km2

Erbil
1,612,700
15.1 km2

Mousl
3,270,400
37.3 km2

Al Sulaymaniah
1,878,800
17.0 km2

Kirkuk
1,395,600
9.7 km2

Sala ad-Din
1,408,200
24.8 km2

Diyala
1,443,200
17.7 km2

Al Anbar
1,561,400
138.5 km2

Baghdad
7,055,200
4.6 km2

Wasit
1,210,600
17.2 km2

Babil
1,820,700
5.6 km2

Karbala
1,066,600
5.0 km2

Maysan
971.4
16.1 km2

Al Quadisiya
1,134,300
8.2 km2

Al Najaf
1,285,500
28.8 km2

Dhi Qar
1,836,200
12.9 km2

Al-Muthannia
719.1
51.7 km2

Basra
2,532,000
19.1 km2

Population number as of 7th January 2011

Goals, Targets, and Indicators Page

Level
- ☑ (All)
- ☑ Tier 1-Ready-Feasible With More Effort
- ☑ Tier 2-Not Ready-Feasible With Strong Effort

Thematic Areas
- ☑ (All)
- ☑ Partnership
- ☑ Peace
- ☑ People
- ☑ Planet
- ☑ Prosperity

Goal	Target	Indicator		
Goal 1: No Poverty	Target 1.2. By 2030, reduce at least by half the propor..	Indicator 1.2.1.The proportion of the population living below the national poverty line is classified by sex a..		🟥
	Target 1.1. By 2030, eradicate extreme poverty for all ..	Indicator 1.1.1. The proportion of the population living below the international poverty line is classified by ..		🟥
	Target 1.2. By 2030, reduce at least by half the propor..	Indicator 1.2.2. The proportion of men, women and children of all ages living in poverty in all its dimensions..		🟥
	Target 1.3. Implement nationally appropriate social pr..	Indicator 1.3.1. Proportion of population covered by social protection floors/systems, by sex, distinguishin..		🟥
	Target 1.4. By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basi..	Indicator 1.4.1. The proportion of the population living in households has access to basic services		🟥
		Indicator 1.4.2. Proportion of total adult population with secure tenure rights to land, with legally recogniz..		🟥
	Target 1.5. By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters	Indicator 1.5.1.Number of deaths, missing persons and injuries transported or evacuated by disaster supp..		🟥
		Indicator 1.5.2. Direct economic losses of disasters to global GDP		🟥
		Indicator 1.5.3. Number of countries with national and local disaster risk reduction strategies		🟥
		Indicator 1.5.4. Proportion of local governments that adopt and implement local disaster risk reduction str..		🟥
	Target 1.a. Ensure significant mobilization of resources from a variety of sources, including through enhanced development cooperation, in order to provide adequate and predictable means for developing countries, in particular least developed countries, to implement pro..	Indicator 1.a.1. Total official development assistance grants from all donors that focus on poverty reductio..		🟥
		Indicator 1.a.2 Proportion of total government spending on essential services (education, health and socia..		🟥
		Indicator 1.a.3. Sum of total grants and non-debt-creating inflows directly allocated to poverty reduction p..		🟥
	Target 1.b. Create sound policy frameworks at the nati..	Indicator 1.b.1. Pro-poor public social spending		🟥
Goal 2: Zero Hunger	Target 2.c. Adopt measures to ensure the proper funct..	Indicator 2.c.1. Indicator of food price anomalies		🟦

Iraq SDGs Thematic Areas Page

Goal	Target	Indicator	
Goal 1: No Poverty	Target 1.2. By 2030, reduce at least by half the proportion ..	Indicator 1.2.1.The proportion of the population..	◼
	Target 1.1. By 2030, eradicate extreme poverty for all peop..	Indicator 1.1.1. The proportion of the populatio..	◼
	Target 1.2. By 2030, reduce at least by half the proportion ..	Indicator 1.2.2. The proportion of men, women a..	◼
	Target 1.3. Implement nationally appropriate social protec..	Indicator 1.3.1. Proportion of population covere..	◼
	Target 1.4. By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basic services, ow..	Indicator 1.4.1. The proportion of the populatio..	◼
		Indicator 1.4.2. Proportion of total adult popula..	◼
	Target 1.5. By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters	Indicator 1.5.1.Number of deaths, missing perso..	◼
		Indicator 1.5.2. Direct economic losses of disast..	◼
		Indicator 1.5.3. Number of countries with nation..	◼
		Indicator 1.5.4. Proportion of local governments..	◼
	Target 1.a. Ensure significant mobilization of resources from a variety of sources, including through enhanced development cooperation, in order to provide adequate and predictable means for developing countries, in particular least developed countries, to implement programmes and ..	Indicator 1.a.1. Total official development assist..	◼
		Indicator 1.a.2 Proportion of total government ..	◼
		Indicator 1.a.3. Sum of total grants and non-deb..	◼
	Target 1.b. Create sound policy frameworks at the national..	Indicator 1.b.1. Pro-poor public social spending	◼
Goal 2: Zero Hunger	Target 2.c. Adopt measures to ensure the proper functioni..	Indicator 2.c.1. Indicator of food price anomalies	◻
	Target 2.1. By 2030, end hunger and ensure access by all people, in particular the poor and people in vulnerable situations, including infants, to safe, nutritious and suffici..	Indicator 2.1.1. Prevalence of undernutrition	◻
		Indicator 2.1.2. Prevalence of moderate or sever..	◻
	Target 2.2. By 2030, end all forms of malnutrition, including achieving, by 2025, the internationally agreed targets on stunting and wasting in children under 5 years of age, and address the nutritional needs of adolescent girls, pregnant and lactating women and older persons	Indicator 2.2.1. Prevalence of stunting (height f..	◻
		Indicator 2.2.2. Prevalence of malnutrition (wei..	◻
		Indicator 2.2.3. 2.2.3 Prevalence of anaemia in ..	◻
	Target 2.3. By 2030, double the agricultural productivity a..	Indicator 2.3.2. Average income of small-scale fo..	◻
	Target 2.3. Target 2.3. By 2030, double the agricultural pro..	Indicator 2.3.1. Indicator 2.3.1. Volume of produ..	◻
	Target 2.4. By 2030, ensure sustainable food production sy..	Indicator 2.4.1. Proportion of agricultural area u..	◻
	Target 2.5. By 2020, maintain the genetic diversity of seeds, cultivated plants and farmed and domesticated animals and their related wild species, including through soundly mana..	Indicator 2.5.1. Number of plant and animal gen..	◻
		Indicator 2.5.2. Proportion of local breeds classi..	◻
	Target 2.a. Increase investment, including through enhanced international cooperation, in rural infrastructure, agricultural research and extension services, technology d..	Indicator 2.a.1. The agriculture orientation inde..	◻
		Indicator 2.a.2. Total official flows (official deve..	◻
	Target 2.b. Correct and prevent trade restrictions and dist..	Indicator 2.b.1. Agricultural export subsidies	◻

Level

- ☑ (All)
- ☑ Tier 1-Ready-Feasible With More Effort
- ☑ Tier 2-Not Ready-Feasible With Strong Effo...

Thematic Areas

- ☑ (All)
- ☑ Partnership
- ☑ Peace
- ☑ People
- ☑ Planet
- ☑ Prosperity

Partnership
Goal 17: Partnerships For The Goals

People
Goal 1: No Poverty

People
Goal 2: Zero Hunger

Peace
Goal 16: Peace, Justice and Strong Institutions

Planet
Goal 15: Life on Land

People
Goal 3: Good Health and Well-Being

Planet
Goal 14: Life Below Water

People
Goal 4: Quality Education

SDGs Thematic Areas

Planet
Goal 13: Climate Action

People
Goal 5: Gender Equality

Planet
Goal 12: Responsible Consumption and Production

People
Goal 6: Clean Water and Sanitation

Prosperity
Goal 7: Affordable and Clean Energy

Planet
Goal 11: Sustainable Cities and Communities

Prosperity
Goal 8: Decent Work and Economic Growth

Prosperity
Goal 10: Reduce inequality within and among countries

Prosperity
Goal 9: Industry, Innovation and Infrastructure

Time Series SDGs Data for 2011-2020 By Goals Page
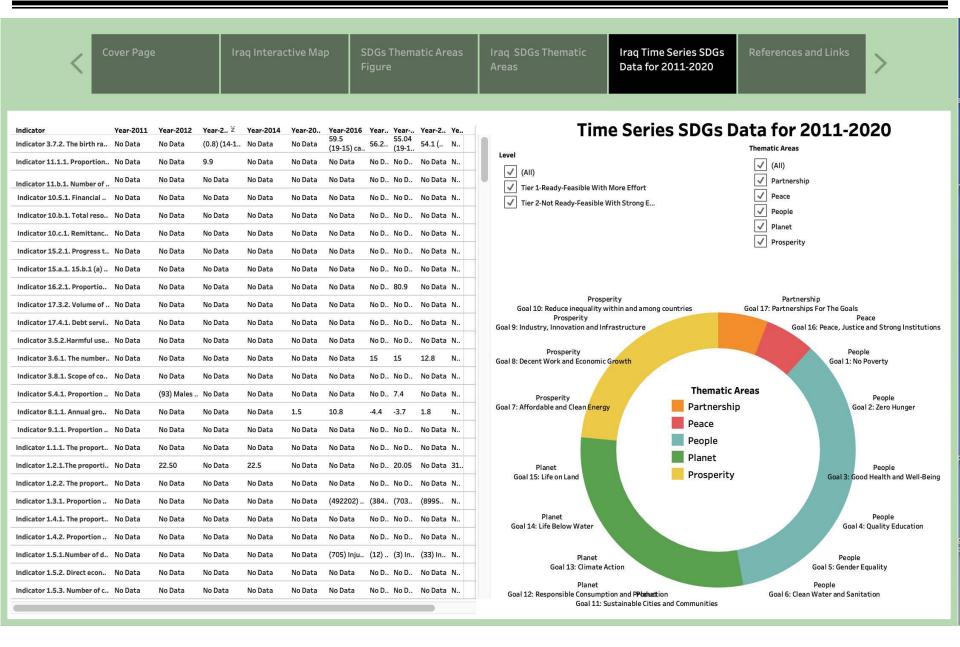
Level

- [✓] (All)
- [✓] Tier 1-Ready-Feasible With More Effort
- [✓] Tier 2-Not Ready-Feasible With Strong Effort

Indicator	Year-2011	Year-2012	Year-2013	Year-2014	Year-2015	Year-2016	Year-2017	Year-2018	Year-2019	Year-20..	
Indicator 3.7.2. The birth rate among adolescent girls (ages 10-14 and 15-19) pe..	No Data	No Data	(0.8) (14-10) for category (59.0) (19-15) ..	No Data	No Data	59.5 (19-15) category	56.2 (19-15) category	55.04 (19-15) category	54.1 (19-15) category	No Data	●
Indicator 11.1.1. Proportion of urban population living in slums, informal settle..	No Data	No Data	9.9	No Data	No Data	No Data	No Data	No Data	No Data	No Data	●
Indicator 11.b.1. Number of countries that ..	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	●
Indicator 10.5.1. Financial Soundness Indicators	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	●
Indicator 10.b.1. Total resource flows for development, by recipient and donor count..	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	●
Indicator 10.c.1. Remittance costs as a proportion of the amount remitted	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	●
Indicator 15.2.1. Progress towards sustainable forest management	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	●
Indicator 15.a.1. 15.b.1 (a) Official development assistance on conservation a..	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	●
Indicator 16.2.1. Proportion of children aged 1–17 years who experienced any phys..	No Data	No Data	No Data	No Data	No Data	No Data	No Data	80.9	No Data	No Data	●
Indicator 17.3.2. Volume of remittances (in United States dollars) as a proportion of to..	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	●
Indicator 17.4.1. Debt service as a proportion of exports of goods and services	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	●
Indicator 3.5.2.Harmful use of alcohol, determined in accordance with national co..	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	●
Indicator 3.6.1. The number of deaths resulting from fatal injuries due to road tra..	No Data	No Data	No Data	No Data	No Data	No Data	15	15	12.8	No Data	●
Indicator 3.8.1. Scope of coverage of	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	No Data	●

Iraq Time Series SDGs Data for 2011-2020 Page

Indicator	Year-2011	Year-2012	Year-2..	Year-2014	Year-20..	Year-2016	Year..	Year-..	Year-2..	Ye..
Indicator 3.7.2. The birth ra..	No Data	No Data	(0.8) (14-1..	No Data	No Data	59.5 (19-15) ca..	56.2..	55.04 (19-1..	54.1 (..	N..
Indicator 11.1.1. Proportion..	No Data	No Data	9.9	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 11.b.1. Number of ..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 10.5.1. Financial ..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 10.b.1. Total reso..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 10.c.1. Remittanc..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 15.2.1. Progress t..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	Progress t..	N..
Indicator 15.a.1. 15.b.1 (a) ..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 16.2.1. Proportio..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	80.9	No Data	N..
Indicator 17.3.2. Volume of ..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 17.4.1. Debt servi..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 3.5.2.Harmful use..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 3.6.1. The number..	No Data	No Data	No Data	No Data	No Data	No Data	15	15	12.8	N..
Indicator 3.8.1. Scope of co..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 5.4.1. Proportion ..	No Data	(93) Males ..	No Data	No Data	No Data	No Data	No D..	7.4	No Data	N..
Indicator 8.1.1. Annual gro..	No Data	No Data	No Data	No Data	1.5	10.8	-4.4	-3.7	1.8	N..
Indicator 9.1.1. Proportion ..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 1.1.1. The proport..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 1.2.1.The proporti..	No Data	22.50	No Data	22.5	No Data	No Data	No D..	20.05	No Data	31..
Indicator 1.2.2. The proport..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 1.3.1. Proportion ..	No Data	No Data	No Data	No Data	No Data	(492202) ..	(384..	(703..	(8995..	N..
Indicator 1.4.1. The proport..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 1.4.2. Proportion ..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 1.5.1.Number of d..	No Data	No Data	No Data	No Data	No Data	(705) Inju..	(12) ..	(3) In..	(33) In..	N..
Indicator 1.5.2. Direct econ..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..
Indicator 1.5.3. Number of c..	No Data	No Data	No Data	No Data	No Data	No Data	No D..	No D..	No Data	N..

Time Series SDGs Data for 2011-2020

Level
- ☑ (All)
- ☑ Tier 1-Ready-Feasible With More Effort
- ☑ Tier 2-Not Ready-Feasible With Strong E...

Thematic Areas
- ☑ (All)
- ☑ Partnership
- ☑ Peace
- ☑ People
- ☑ Planet
- ☑ Prosperity

Prosperity Goal 10: Reduce inequality within and among countries
Prosperity Goal 9: Industry, Innovation and Infrastructure
Prosperity Goal 8: Decent Work and Economic Growth
Prosperity Goal 7: Affordable and Clean Energy
Planet Goal 15: Life on Land
Planet Goal 14: Life Below Water
Planet Goal 13: Climate Action
Planet Goal 12: Responsible Consumption and Production
Planet Goal 11: Sustainable Cities and Communities
Partnership Goal 17: Partnerships For The Goals
Peace Goal 16: Peace, Justice and Strong Institutions
People Goal 1: No Poverty
People Goal 2: Zero Hunger
People Goal 3: Good Health and Well-Being
People Goal 4: Quality Education
People Goal 5: Gender Equality
People Goal 6: Clean Water and Sanitation

Thematic Areas
- Partnership
- Peace
- People
- Planet
- Prosperity

USING BIG DATA FOR SDGS

1 NO POVERTY

North Korea vs. South Korea

Satellite image and machine learning example
Combining satellite imagery and machine learning to predict poverty.

The elimination of poverty worldwide is the first of 17 UN Sustainable Development Goals for the year 2030. To track progress towards this goal, we require more frequent and more reliable data on the distribution of poverty than traditional data collection methods can provide. Sustainability and artificial intelligence lab at Stanford University proposed an approach that combines machine learning with high-resolution satellite imagery to provide new data on socioeconomic indicators of poverty and wealth. More detailed explanation of how it all works can be found at their science article paper [1].

The lack of reliable data in developing countries is a major obstacle to sustainable development, food security, and disaster relief. Poverty data, for example, is typically scarce, sparse in coverage, and labor-intensive to obtain. Remote sensing data such as high-resolution satellite imagery, on the other hand, is becoming increasingly available and inexpensive. Unfortunately, such data is highly unstructured and currently no techniques exist to automatically extract useful insights to inform policy decisions and help direct humanitarian efforts [2].

They have demonstrated an accurate, inexpensive, and scalable method for estimating consumption expenditure and asset wealth from high-resolution satellite imagery. Using survey and satellite data from five African countries--Nigeria, Tanzania, Uganda, Malawi, and Rwanda--we show how a convolutional neural network can be trained to identify image features that can explain up to 75% of the variation in local-level economic outcomes. Their method, which requires only publicly available data, could transform efforts to track and target poverty in developing countries. It also demonstrates how powerful machine learning techniques can be applied in a setting with limited training data, suggesting broad potential application across many scientific domains [1].

Reference:

[1] Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S, "Combining satellite imagery and machine learning to predict poverty", Science, 2016
[2] Michael Xie and Neal Jean and Marshall Burke and David Lobell and Stefano Ermon, "Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping" secarXiv: 1510.00098v2 [cs.CV] 27 Feb 2016.

Example of Social media example: Twitter
Use of Social media Technology for Perception Assessment [9].

This example shows that online social media conversations might represent a new source of information to monitor food security or other SDGs goals.

Pulse Lab Jakarta analyzed Twitter conversations discussing food price increases between March 2011 and April 2013. This research also explores the relations between such conversations, food price inflation and external events.

Since 2010, Indonesia has witnessed substantial increases in food prices: the price of rice increased 51% between December 2009 and February 2012. Taxonomies, that is groups of words and phrases with related meanings, were developed in the Bahasa Indonesia language to identify relevant content. A classification algorithm was trained to categorize the extracted tweets as positive, negative, confused, or neutral to analyze their sentiment. Using simple time-series analysis, the researchers quantified the correlation between the volume of food-related Twitter conversations and official food inflation statistics. A relationship was found between retrospective official food inflation statistics and the number of tweets speaking about food price increases. Moreover, upon analyzing fuel price tweets, it was found that perceptions of food and fuel prices were related.
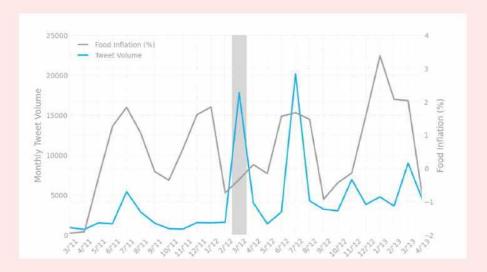
Figure (2). Plot of Monthly Food Price-Related Tweet Volume with Official Food Price Inflation Statistics. The grey highlighted area marks the month of March 2012 during which proposals on fuel subsidies cuts were under consideration by the Government of Indonesia.

Reference: UN Global Pulse, "Mining Indonesian Tweets to Understand Food Price Crises", (2014).

4.1. By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes.

4.1.1. The literacy rate of men and women age group 15-24

Case study:
Can mobile usage predict illiteracy in a developing country?[1]

Abstract: The lack of a standard definition of literacy is a big problem. Two common methods for determining literacy are self-assessment and testing. With self-assessment, the person is asked to rate his or her own literacy. With testing, of course, an actual test is administered and evaluated.

The present study provides the first evidence that illiteracy can be reliably predicted from standard mobile phone logs. By deriving a broad set of mobile phone indicators reflecting users' financial, social and mobility patterns we show how supervised machine learning can be used to predict individual illiteracy in an Asian developing country, externally validated against a large-scale survey. On average the model performs 10 times better than random guessing with a 70% accuracy. Further we show how individual illiteracy can be aggregated and mapped geographically at cell tower resolution. Geographical mapping of illiteracy is crucial to know where the illiterate people are, and where to put in resources. In underdeveloped countries such mappings are often based on out-dated household surveys with low spatial and temporal resolution. One in five people worldwide struggle with illiteracy, and it is estimated that illiteracy costs the global economy more than $1 trillion dollars each year [2]. These results potentially enable cost effective, questionnaire-free investigation of illiteracy-related questions on an unprecedented scale.

[1] Pål Sundsøy, Telenor Group Research, Big Data Analytics Snarøyveien 30,1331 Fornebu, Norway.
[2] World Literacy Foundation: The Economic & Social Cost of Illiteracy. (Accessed 2015) Available at: http://worldliteracyfoundation.org/wp-content/uploads/2015/05/The-EconomicSocial-Cost-of-Illiteracy.pdf

Target: 3.6. By 2030 halve the number of global deaths and injuries from road traffic accidents and, in the interim, by 2020, stabilize and then reduce global deaths and injuries from road traffic accidents.

Indicator: 3.6.1. Number of victims died from traffic per 100,000 population.

Case study:
Big Data Analytics and Visualization with Spatio-Temporal Correlations for Traffic Accidents

Abstract: Big data analytics for traffic accidents is a hot topic and has significant values for a smart and safe traffic in the city. Based on the massive traffic accident data from October 2014 to March 2015 in Xiamen, China, they propose a novel accident occurrences analytics method in both spatial and temporal dimensions to predict when and where an accident with a specific crash type will occur consequentially by whom. Firstly, they analyze and visualize accident occurrences in both temporal and spatial view. Second, they illustrate spatio-temporal visualization results through two case studies in multiple road segments, and the impact of weather on crash types. These findings of accident occurrences analysis and visualization would not only help traffic police department implement instant personnel assignments among simultaneous accidents, but also inform individual drivers about accident-prone sections and the time span which requires their most attention.

Target: 7.3. Double the global rate of improvement in energy efficiency by 2030.

Indicator: 7.3.1. Energy intensity measured in terms of primary energy and GDP.

Case study:
Applying Big Data analytics for energy efficiency.

Abstract: Global energy requirements are continuously increasing. Conventional methods of producing more energy to meet this growth pose a great threat to the environment. CO2 emissions and other bi-products of energy production and distribution processes have dire consequences for the environment. Efficient use of energy is one of the main tools to restrain energy consumption growth without compromising on the customers requirements. Improving energy efficiency requires understanding of the usage patterns and practices. Smart energy grids, pervasive computing, and communication technologies have enabled the stakeholders in the energy industry to collect large amounts of useful and highly granular energy usage data. This data is generated in large volumes and in a variety of different formats depending on its purpose and systems used to collect it. The volume and diversity of data also increase with time. All these data characteristics refer to the application of Big Data.

This work focuses on harnessing the power of Big Data tools and techniques such as MapReduce and Apache Hadoop ecosystem tools to collect, process and analyse energy data and generate insights that can be used to improve energy efficiency. Furthermore, it also includes studying energy efficiency to formulate the use cases, studying Big Data technologies to present a conceptual model for an end-to-end Big Data analytics platform, implementation of a part of the conceptual model with the capacity to handle energy efficiency use cases and performing data analysis to generate useful insights. The analysis was performed on two data sets. The first data set contained hourly consumption of electricity consumed by a set of different buildings. The data was analysed to discover the seasonal and daily usage trends. The analysis also includes the classification of buildings on the basis of energy efficiency while observing the seasonal impacts on this classification. The analysis was used to build a model for segregating the energy inefficient buildings from energy efficient buildings. The second data set contained device level electricity consumption of various home appliances used in an apartment. This data was used to evaluate different prediction models to forecast future consumption on the basis of previous usage. The main purpose of this research is to provide the basis for enabling data driven decision making in organizations working to improve energy efficiency.

Reference: Hussnain Ahmed, "Applying Big Data analytics for energy efficiency", Master's thesis, Aalto University (2014)

Use of satellite images for computing the SDG 11 indicators

Using satellite images, it is possible to obtain historical and updated data on land cover to analyze the expansion of urban agglomerations.

11 SUSTAINABLE CITIES AND COMMUNITIES

Indicator 11.1.1

Proportion of urban population living in slums, informal settlements or inadequate housing.

Use: satellite imagery classification to determine informal settlements.

Indicator 11.2.1

Proportion of population that has convenient access to public transport, by sex, age and persons with disabilities.

Use: classification of satellite images to determine the urban area of cities.

Indicator 11.3.1

Ratio of land consumption rate to population growth rate.
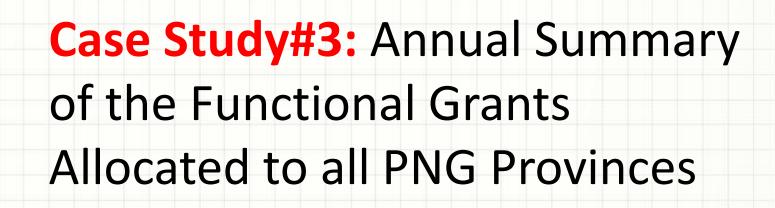
Use: classification of satellite images to determine land consumption of cities.

Indicator 11.7.1

Average share of the built-up area of cities that is open space for public use for all, by sex, age and persons with disabilities.
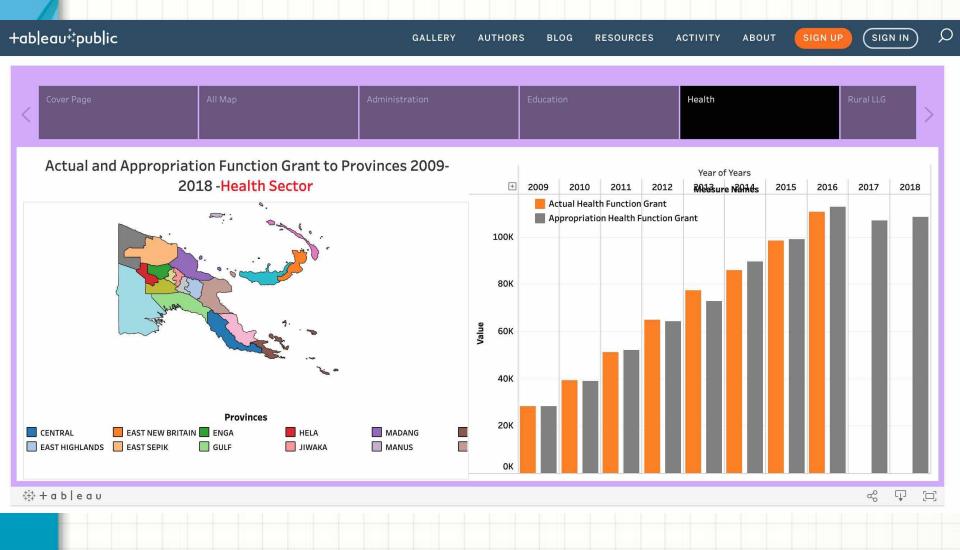
Use: land cover classification using a huge remote sensing imagery collection (petabytes) to identify the urban area selected cities and their green places.

Case Study#2: Papua New Guinea Department of Treasury National Budget-2018

https://public.tableau.com/shared/BYXRCZKGY?:toolbar=n&:display_count=y&:origin=viz_share_link

Case Study#3: Annual Summary of the Functional Grants Allocated to all PNG Provinces

Thank you