

# Getting Started with Data Science

*Parvathy Krishnan*

*CTO, Analytics for a Better World*



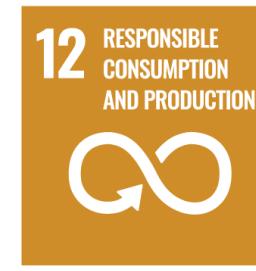
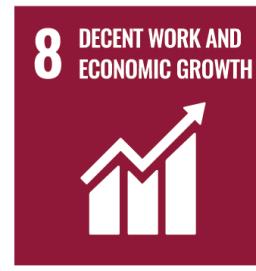
**Analytics for a  
Better World**

# Why become data-driven?

- Increase Impact
- Drive Action
- Enhance Communication
  - Internal Communication
  - External Communication

# Analytics to contribute to the:

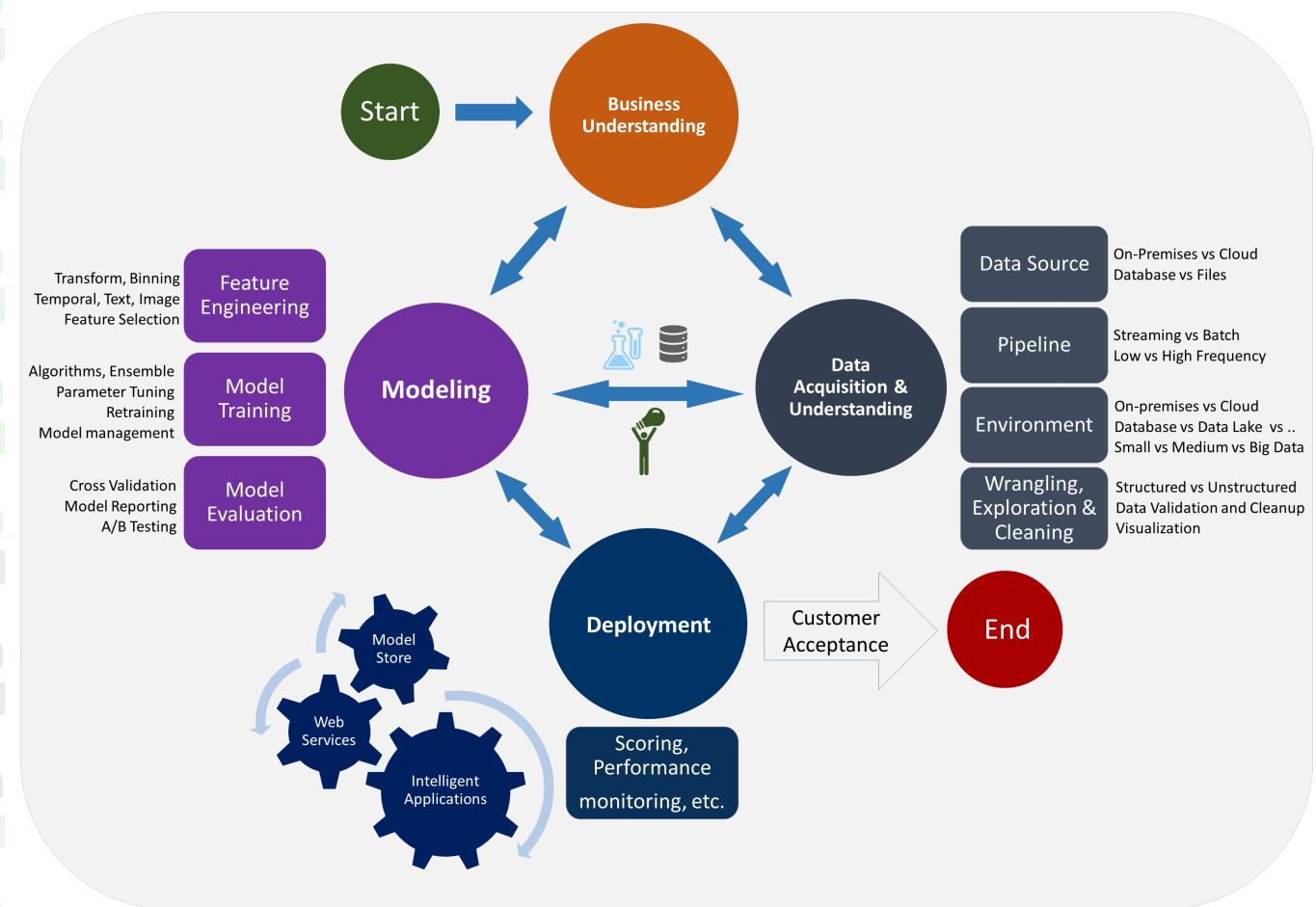
## SUSTAINABLE DEVELOPMENT GOALS



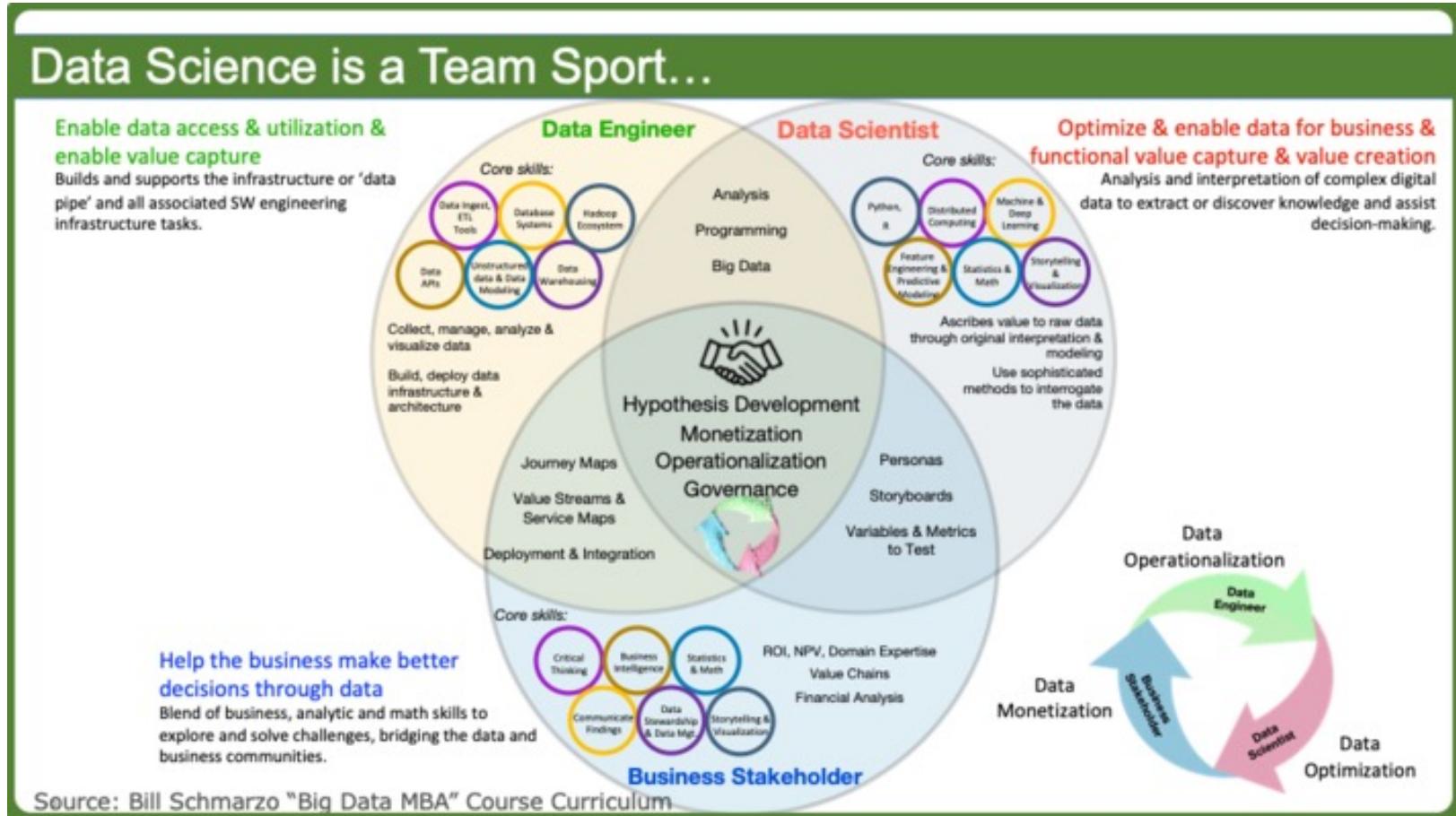
# Agenda

- Data Science Lifecycle
- Skills needed for Data Science
- Types of Data
- 4 stages of Data Maturity
- Open data-Open tools to get started

# Data Science Lifecycle



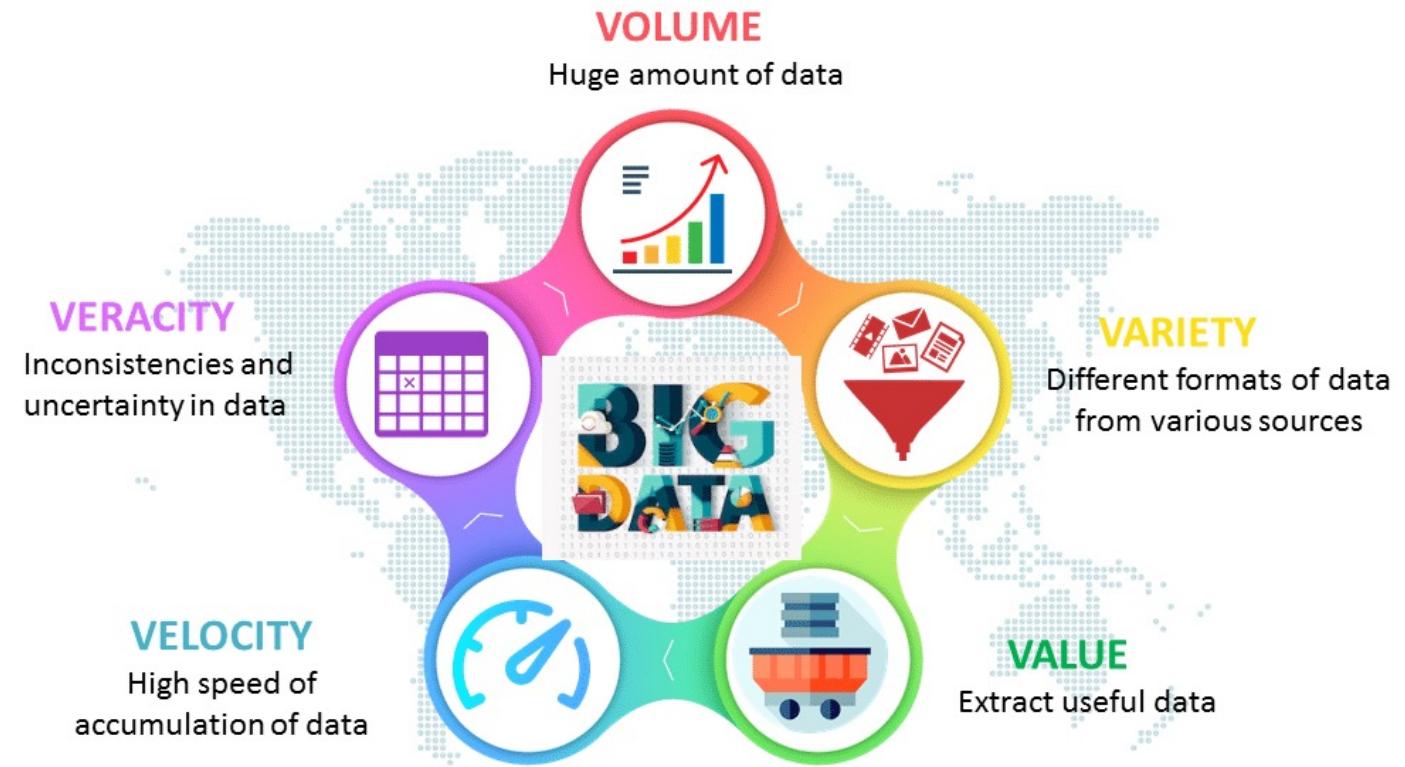
# Data Science is a Team Sport...



# Skills needed for Data Science and Why?

- Programming
- Computer science
- Math and stats
- Machine learning
- Social science
- Problem and Project Scoping
- Project management
- Privacy and security
- Ethics, fairness, bias, and transparency
- Communications
- Social issues

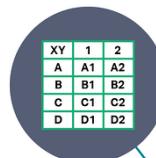
# Types of Data



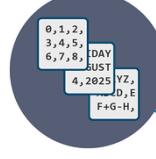
# Types of Data

## Structured Data vs Unstructured Data

Can be displayed in rows, columns and relational databases



Numbers, dates and strings



Estimated 20% of enterprise data (Gartner)



Requires less storage

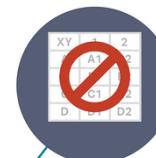


Easier to manage and protect with legacy solutions



## Unstructured Data

Cannot be displayed in rows, columns and relational databases



Images, audio, video, word processing files, e-mails, spreadsheets



Estimated 80% of enterprise data (Gartner)



Requires more storage



More difficult to manage and protect with legacy solutions

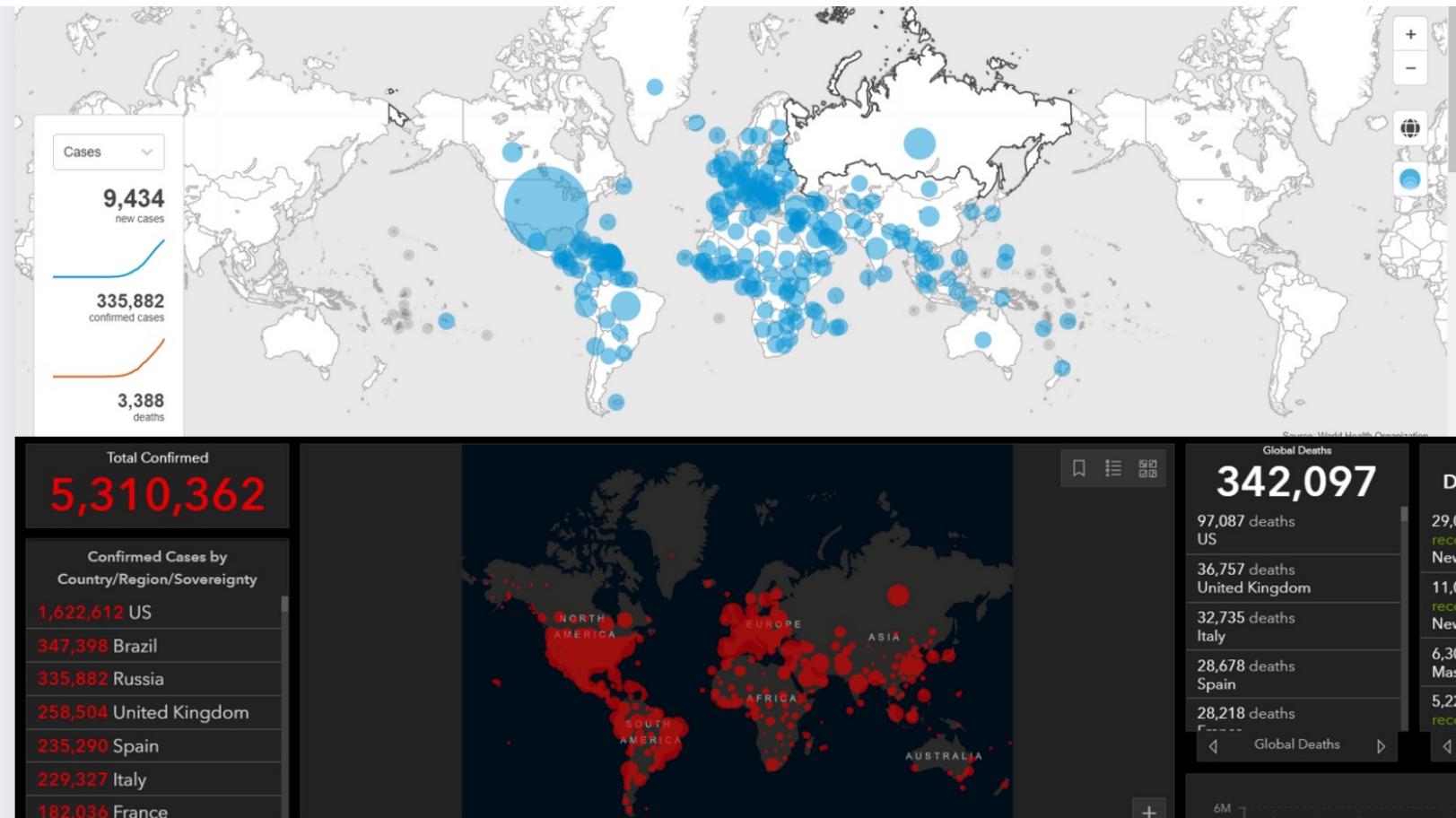


# Case Study - Structured Data

## Reference

[https://www.ted.com/talks/hans\\_rosling\\_the\\_best\\_stats\\_you\\_ve\\_ever\\_seen?language=nl](https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen?language=nl)

# Case Study - Structured Data



# Case Study - Unstructured Data (Text)

 MEDISYS

Search [advanced search](#)

Top Stories  
Event Extraction  
Recent Disease Incidents  
Alert Statistics  
Communicable Diseases  
Symptoms  
Bioterrorism  
Nuclear  
Chemical  
ECDC  
EFSA  
EMCDDA  
ENV\_RISKS  
Food Security  
SAM  
Medical Devices  
Vaccination  
Other  
Continents  
Official Sources  
Sources List

**Most Active Topics**

**Xylella fastidiosa**  
In combination with: Ukraine; Italy; Greece;  
Unaprol lancia l'allarme: 30 milioni di olivi abbandonati da salvare in Italia, un patrimonio unico al mondo di biodiversità e storia [winews](#) Wednesday, October 5, 2022 5:48:00 PM CEST | info [en] [other]  
Top terms: unaprol lancia l'allarme, olivi abbandonati, villa adriana, biodiversità e storia, tavola dell'imperatore adriano

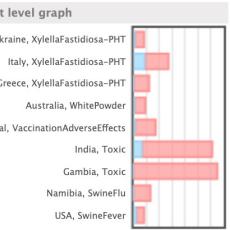
09:39 | economia | Olio: Colidretti-Unaprol, in Italia 30 milioni di ulivi da salvare a causa clima e rincari [sassarinotizie](#) Wednesday, October 5, 2022 9:44:00 AM CEST | info [en] [other]  
Top terms: colidretti e unaprol, cambiamento climatico, causa clima, consorzio olivicolo italiano, 09:39 economia olio

**White Powder**  
In combination with: Australia;  
Melbourne Storm [theherald-au](#) Thursday, October 6, 2022 12:07:00 PM CEST | info [other]  
Top terms: melbourne storm, roy ward, chairman matt tripp, christian nicolussi, pengilly cameron munster

'Really excited' Bartel returns to spotlight after 'white powder' scandal Former WAG Nadia Bartel is officially returning to the spotlight following la [perthnow](#) Thursday, October 6, 2022 10:49:00 AM CEST | info [other]  
Top terms: year's plate scandal, wag nadia bartel, officially returning, snorting white powder, excited'bartel returns

Tools  
Thursday, October 6, 2022  
2:16:00 PM CEST  
[manage](#) [info](#)

Alert level graph



Country	Alert Level
Ukraine, XylellaFastidiosa-PHT	High
Italy, XylellaFastidiosa-PHT	High
Greece, XylellaFastidiosa-PHT	High
Australia, WhitePowder	High
Portugal, VaccinationAdverseEffects	Medium
India, Toxic	High
Gambia, Toxic	High
Namibia, SwineFlu	Medium
USA, SwineFever	Low

Previous 14 days average  
Alert level:  
high medium low  
[Full statistics...](#)

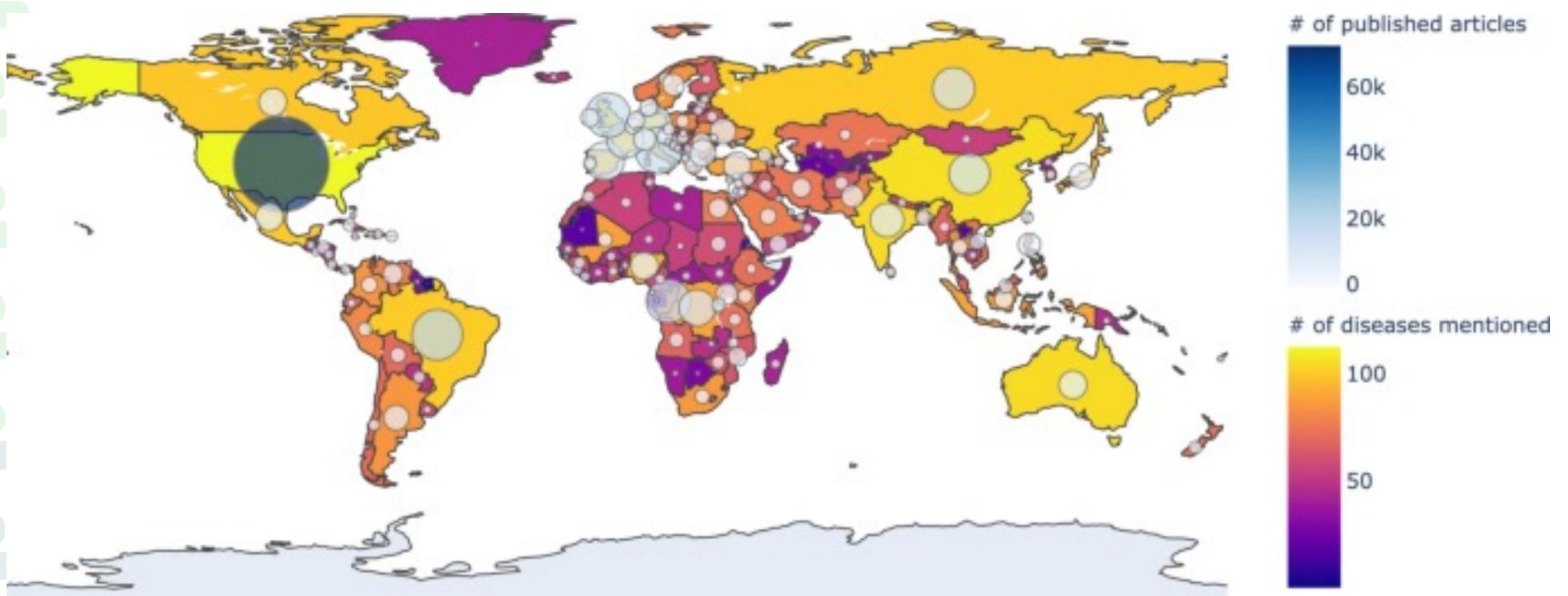
Manuals/Tutorials (access to wiki upon request)  
Contact: [jrc-emmi-info@ec.europa.eu](mailto:jrc-emmi-info@ec.europa.eu)  
[Legal Notice](#)  
[Cookies – Privacy Policy](#)



The Medical Information System (MediSys) is a fully automatic 24/7 public health surveillance system monitoring infectious diseases, bioterrorism, and chemical, biological, radiological and nuclear (CBRN) threats in open-source media. MediSys became online in August 2004 and has been continuously extended.

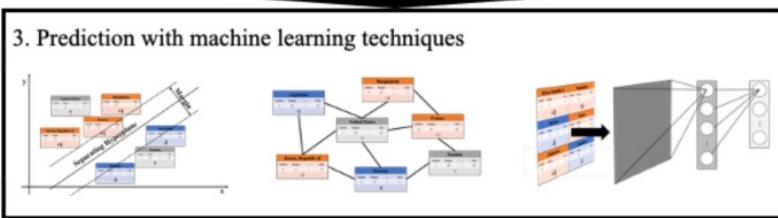
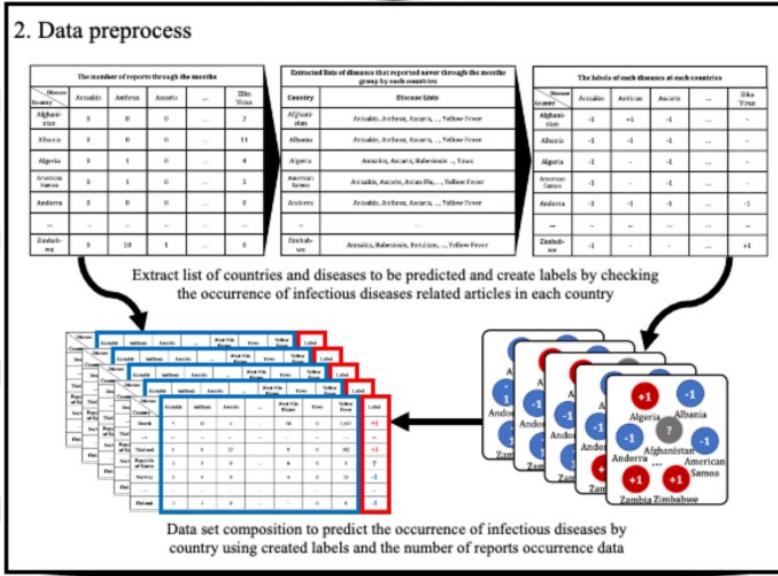
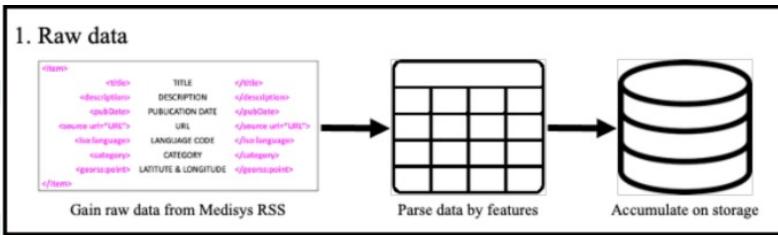


# Case Study - Unstructured Data (Text)

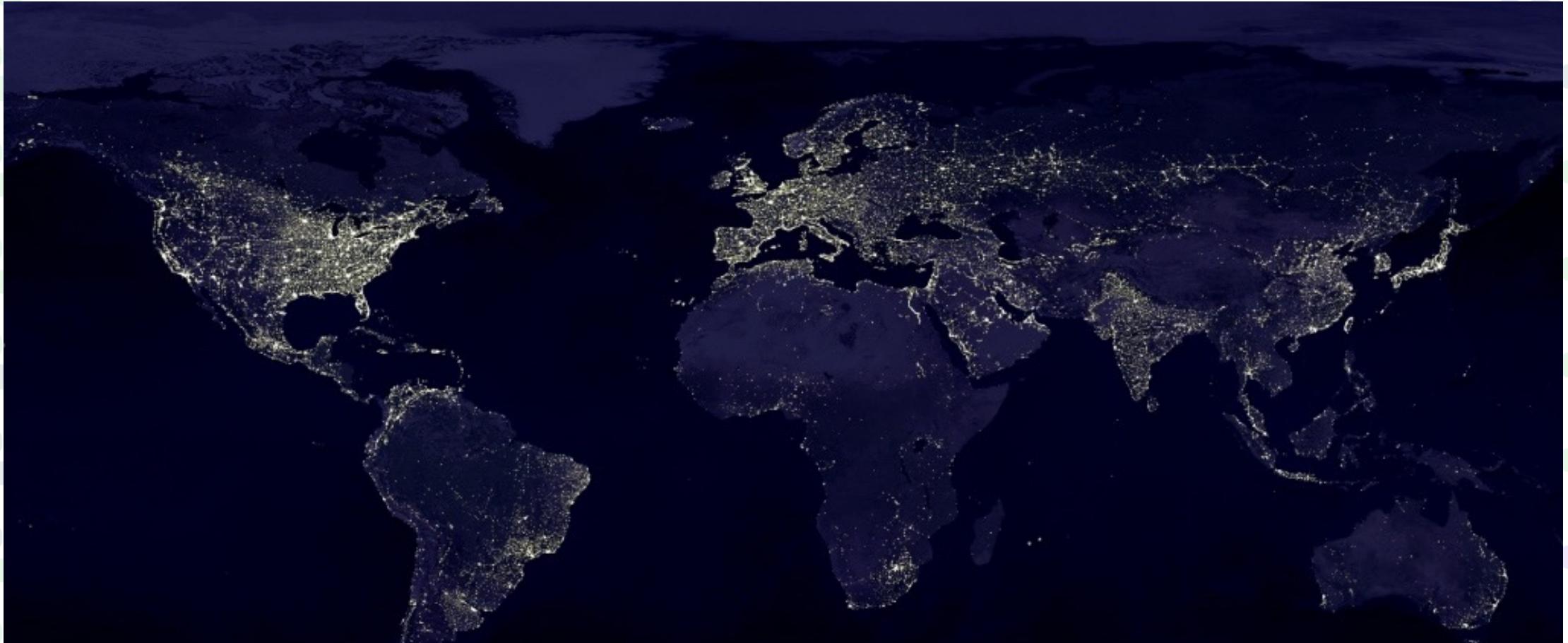


The number of articles published in each country collected from Medisys from January to December 2019: the closer the color of the country to yellow, the more diseases occurred, and the larger the circle, the more articles have occurred.

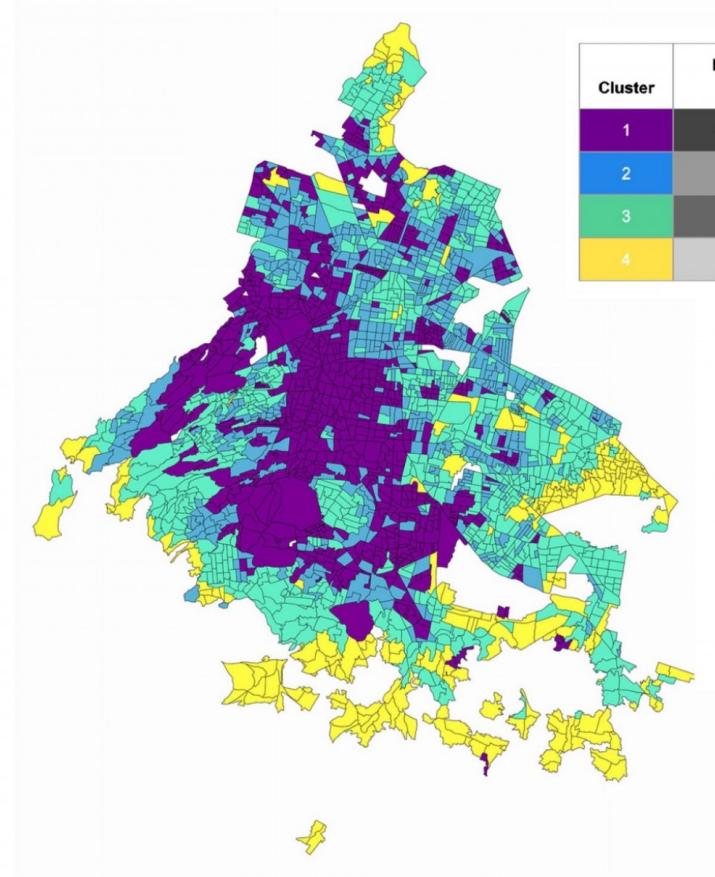
# Case Study - Unstructured Data (Text)



# Case Study - Unstructured Data (Remote Sensing Data)



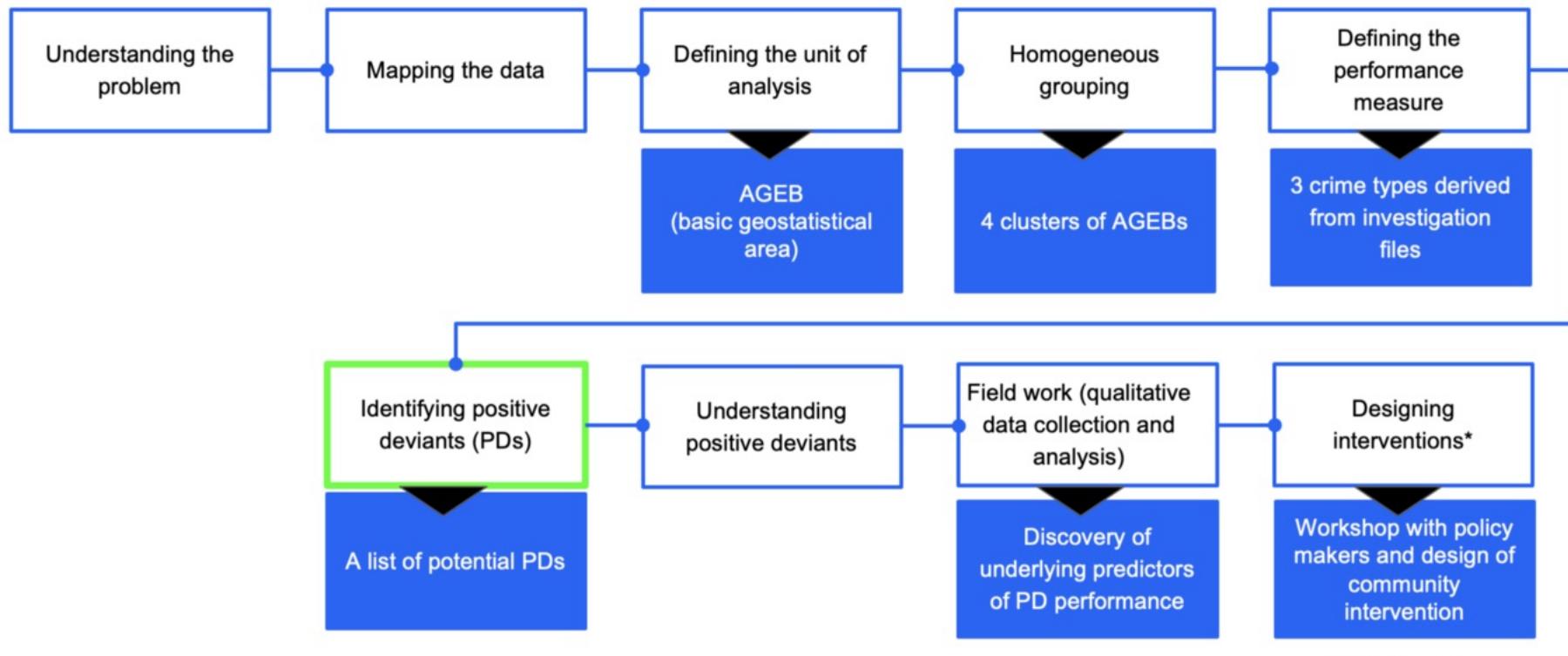
# Case Study – Data Stack



Cluster	Number of AGEBs	Population density	Incoming daily trips	Marginalization index
1	Very high	Medium	Very high	Very low
2	Medium	High	High	Low
3	High	High	Medium	Medium
4	Low	Low	Very low	High

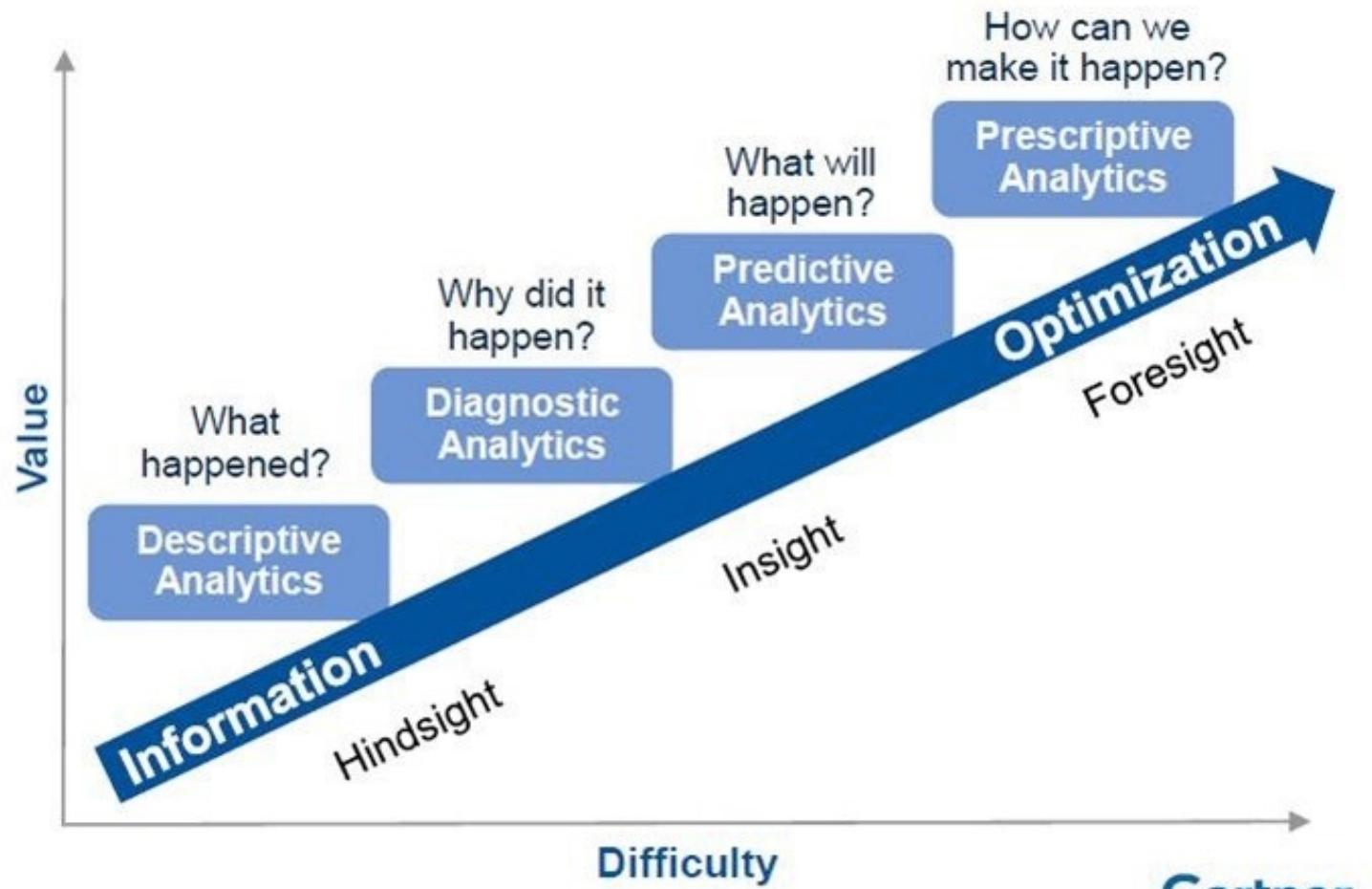
*Identifying Safe(r) Public Spaces for Women in Mexico City*

# Case Study – Data Stack



*Identifying Safe(r) Public Spaces for Women in Mexico City*

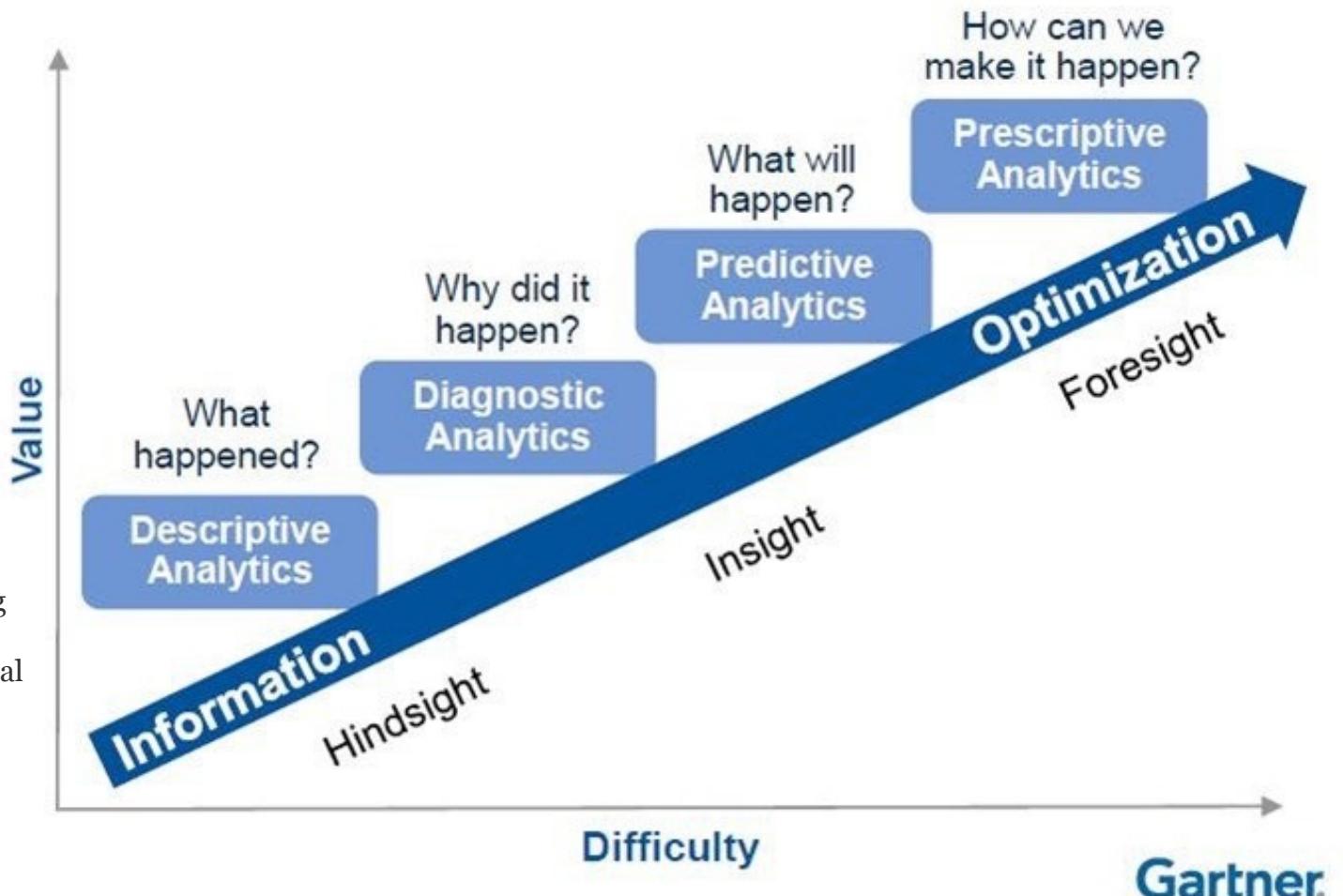
# 4 stages of Data Maturity



Gartner.

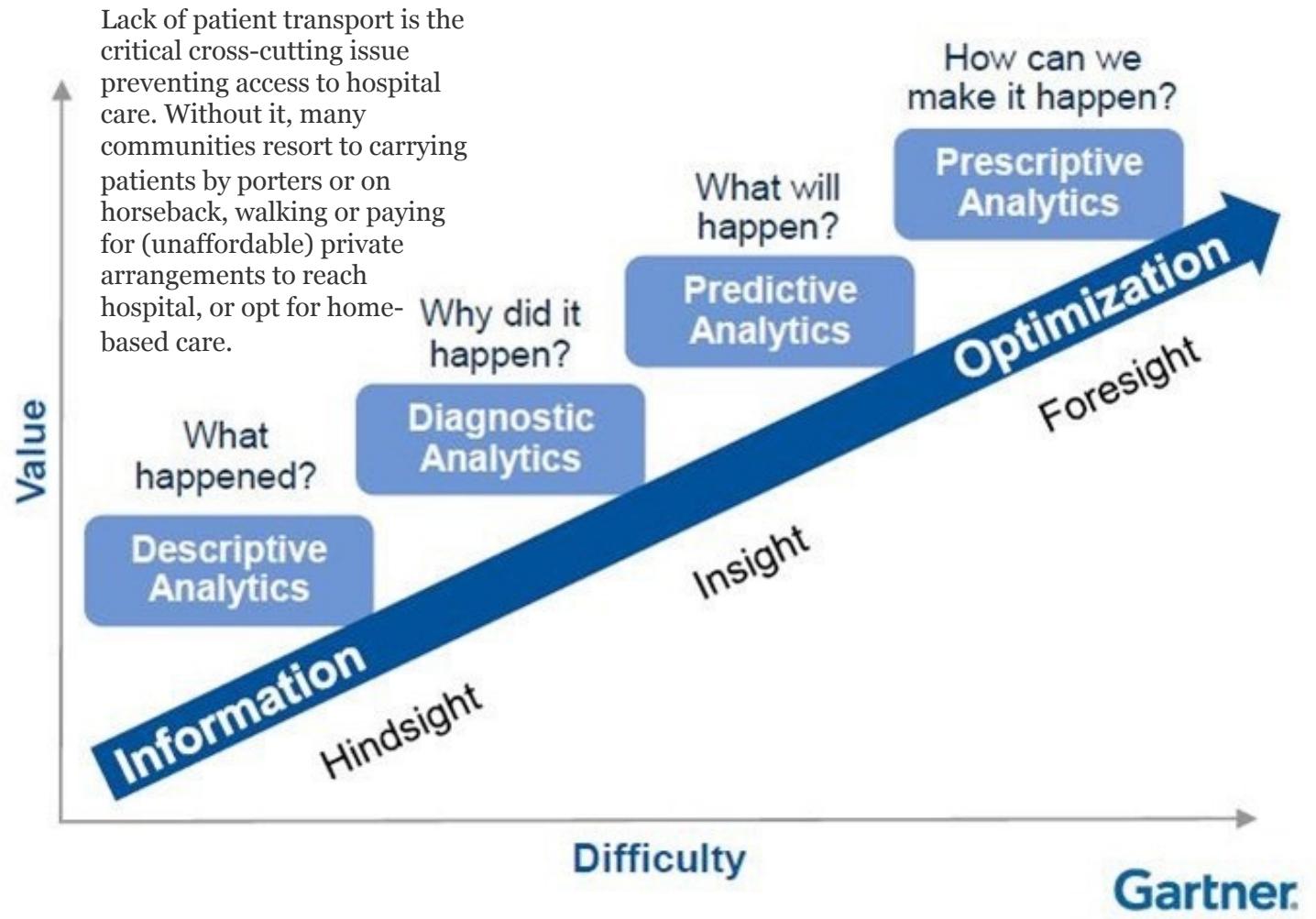
# 4 stages of Data Maturity

Despite public health care being free at the point of delivery, wealthier patients access hospital care at nearly twice the rate of poorer patients.

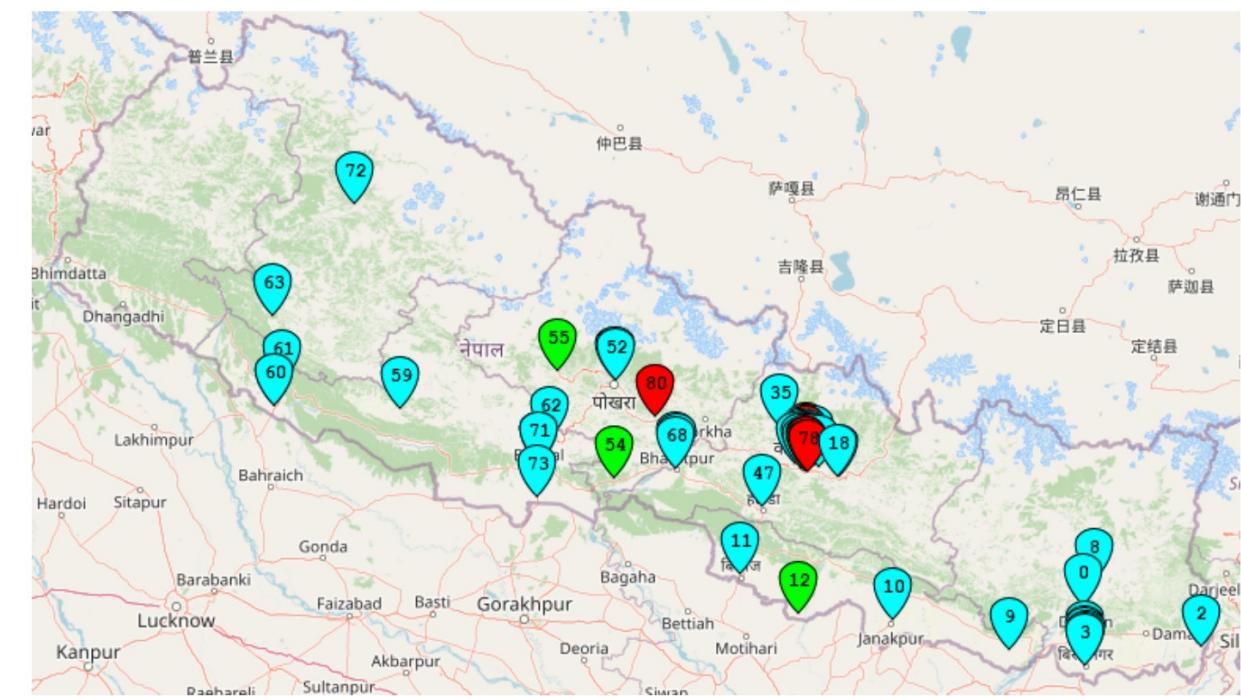
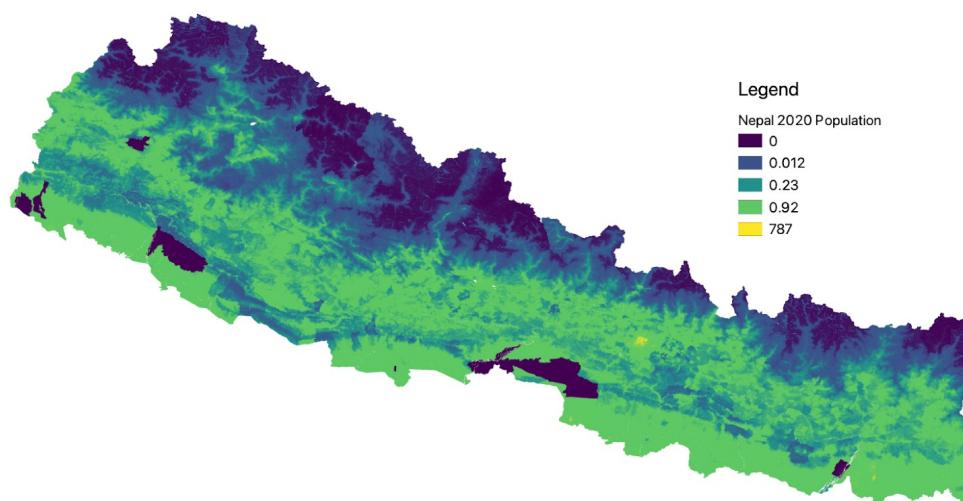


Gartner

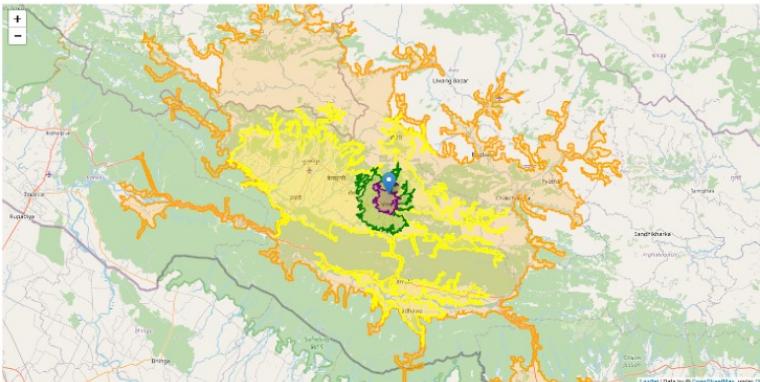
# 4 stages of Data Maturity



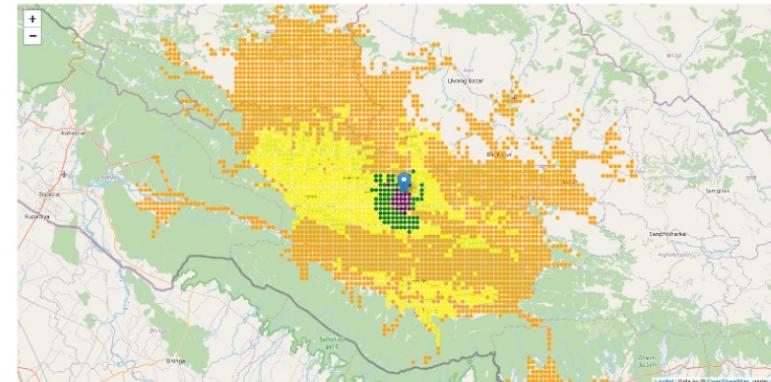
# 4 stages of Data Maturity



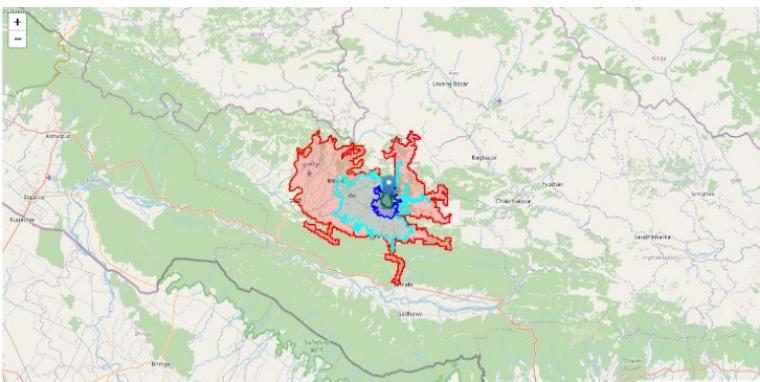
# 4 stages of Data Maturity



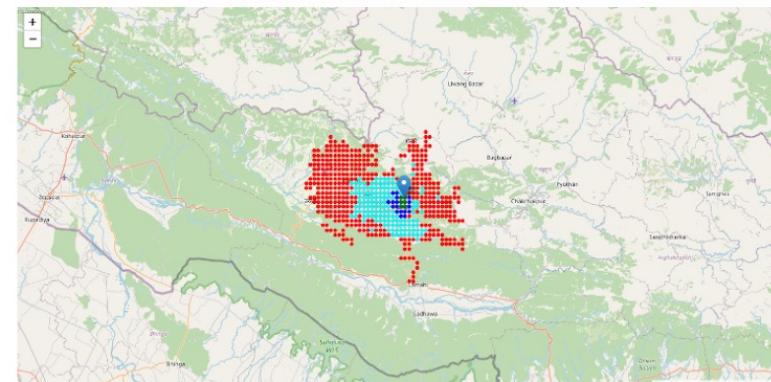
Isodistances



Individual assessment of each household cluster

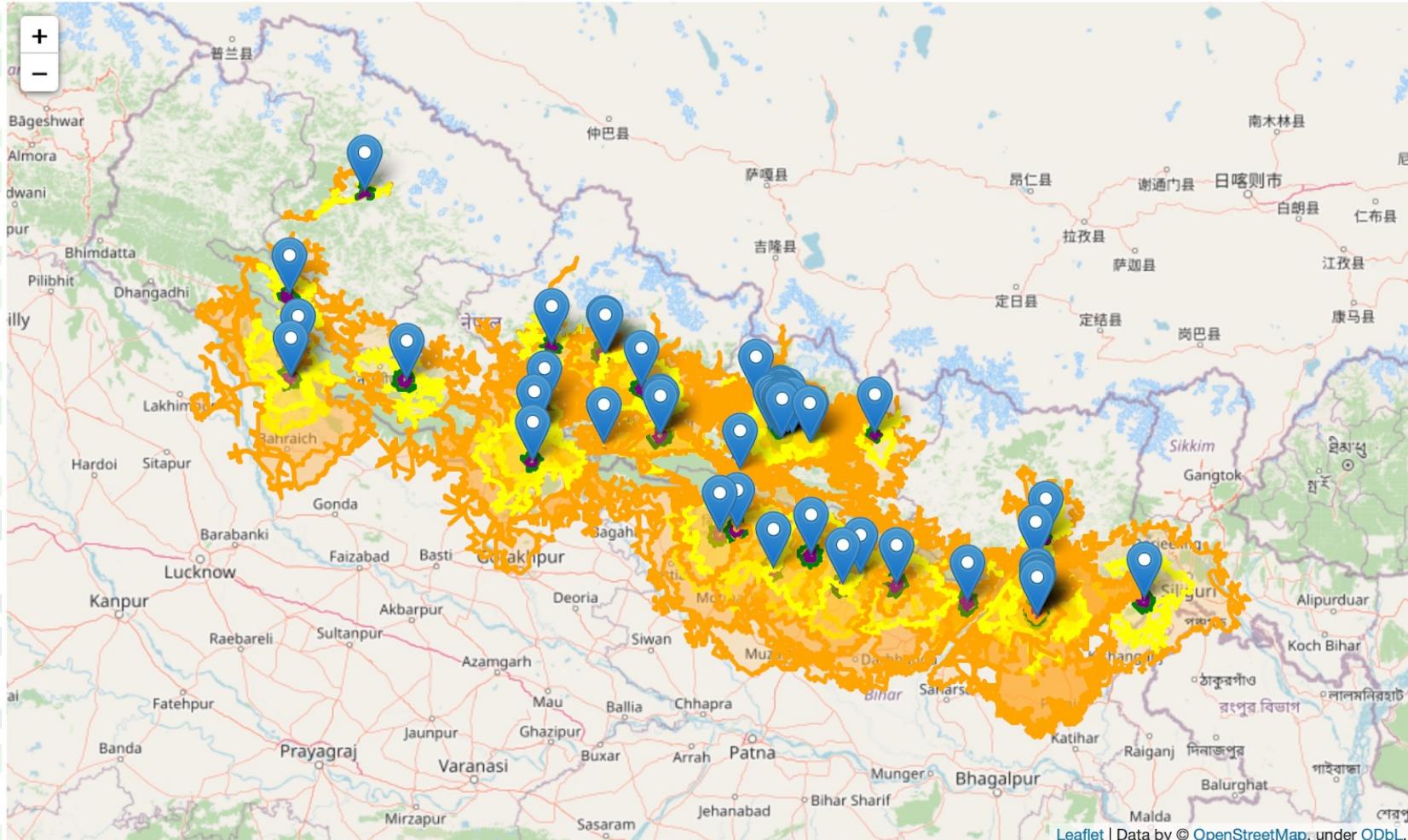


Isochrones

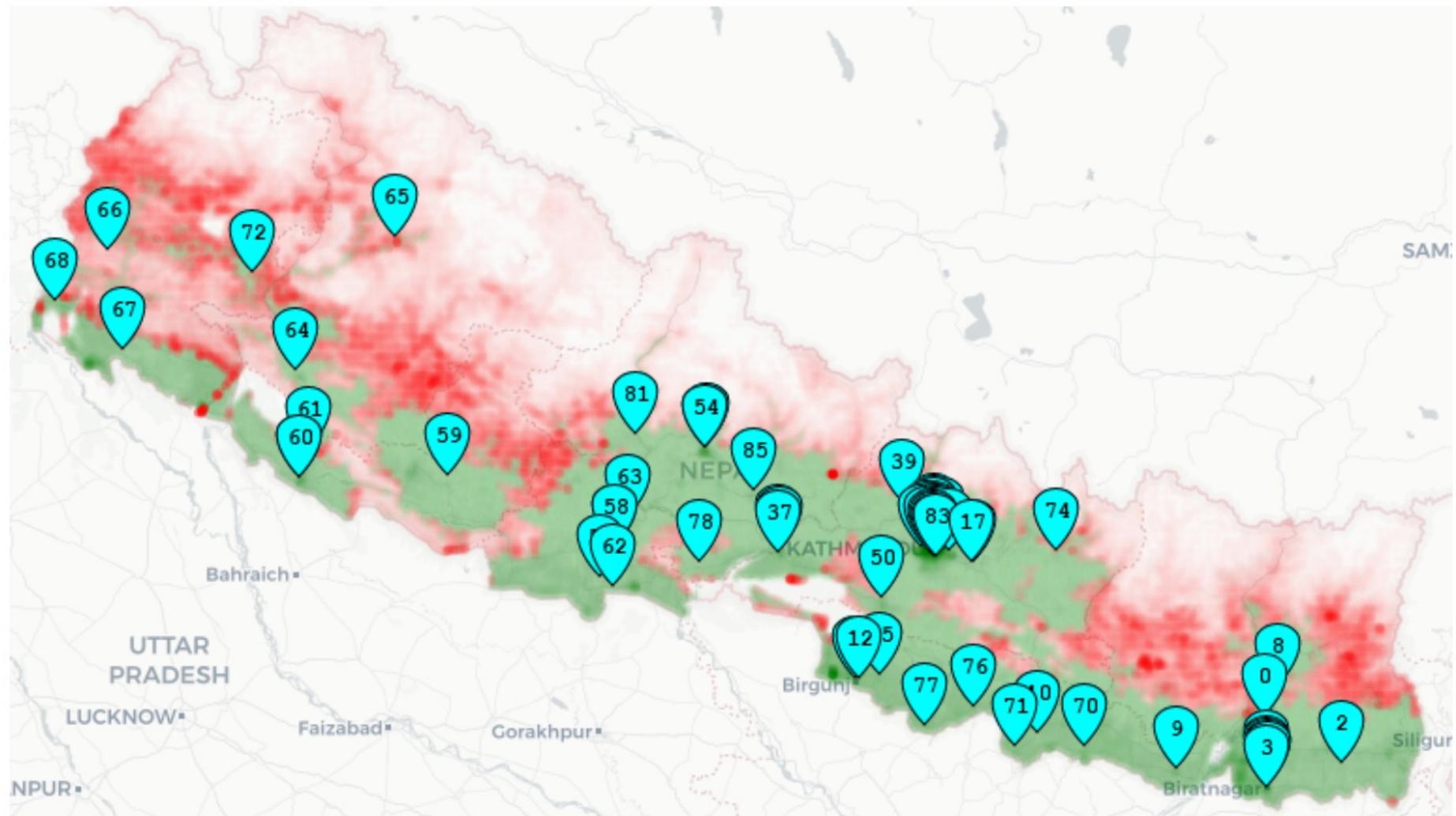


Individual assessment of each household cluster

# 4 stages of Data Maturity



# 4 stages of Data Maturity



# Open Data/Tools to get started

- Humanitarian Data Exchange
- WorldPop
- World Bank Open Data
- Open Data Initiative ESA
- GDELT Project
- Facebook Data for Good



**Thank you  
for your interest.**

**Analytics for a Better World Institute**  
[www.analyticsbw.org](http://www.analyticsbw.org)

**Location(s):**

**Amsterdam Business School**  
University of Amsterdam,  
Plantage Muidergracht 12, 1018 TV  
Amsterdam, The Netherlands.

**ORTEC**  
Houtsingel 5, 2719 EA Zoetermeer,  
The Netherlands

**Contact ABW:**  
[hello@analyticsbw.org](mailto:hello@analyticsbw.org)  
[LinkedIn](#)

**For enquiries / contact:**  
Robert Monné  
[robert.monne@analyticsbw.org](mailto:robert.monne@analyticsbw.org)  
[LinkedIn](#)