

DATAATHON

NOVEMBER 2021 - MARCH 2022

**THE DATA SCIENCE CHALLENGE
BRIDGING INDIAN AGRICULTURAL
PROTECTION GAP**

SCOR
The Art & Science of Risk

BEREXIA

Global Advisor for Change





Please note all questions you might have during the presentation.

You will have the opportunity to ask them to the team at the end of the presentation.

DATAATHON

NOVEMBER 2021 - MARCH 2022

BRUNO MARTINAUD

Responsable Académique Entrepreneuriat
Ecole Polytechnique

SCOR
The Art & Science of Risk

BEREXIA

GLOBAL ADVISOR FOR CHANGE



DATAATHON

NOVEMBER 2021 - MARCH 2022

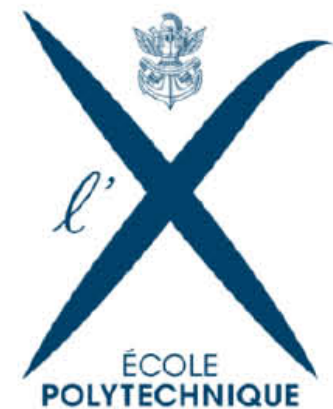
FRANÇOIS DE VARENNE

Member of the SCOR's Executive Committee
Investments, Transformation and Group Corporate Finance

SCOR
The Art & Science of Risk

BEREXIA

GLOBAL ADVISOR FOR CHANGE



DATAATHON

NOVEMBER 2021 - MARCH 2022

KATI NISIPASU

Head of P&C Reserving, Capital Modeling & Risk Management
SCOR Global P&C

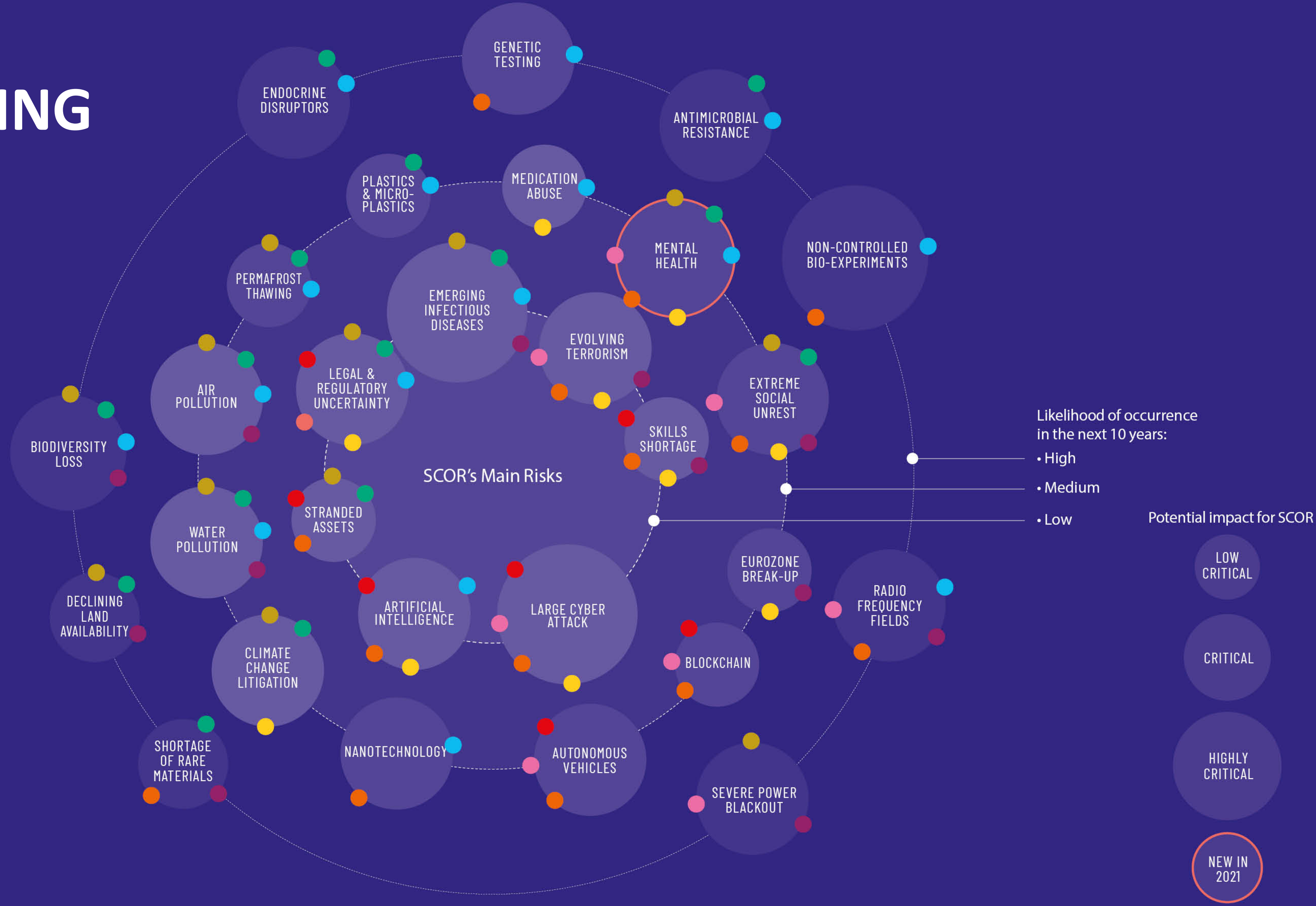
SCOR
The Art & Science of Risk

BEREXIA

GLOBAL ADVISOR FOR CHANGE



EMERGING RISKS RADAR



DATAATHON

NOVEMBER 2021 - MARCH 2022

SARAH CONRADT

Senior Risk Analyst, Agri & Specialty Risk Modelling
SCOR Global P&C

SCOR
The Art & Science of Risk

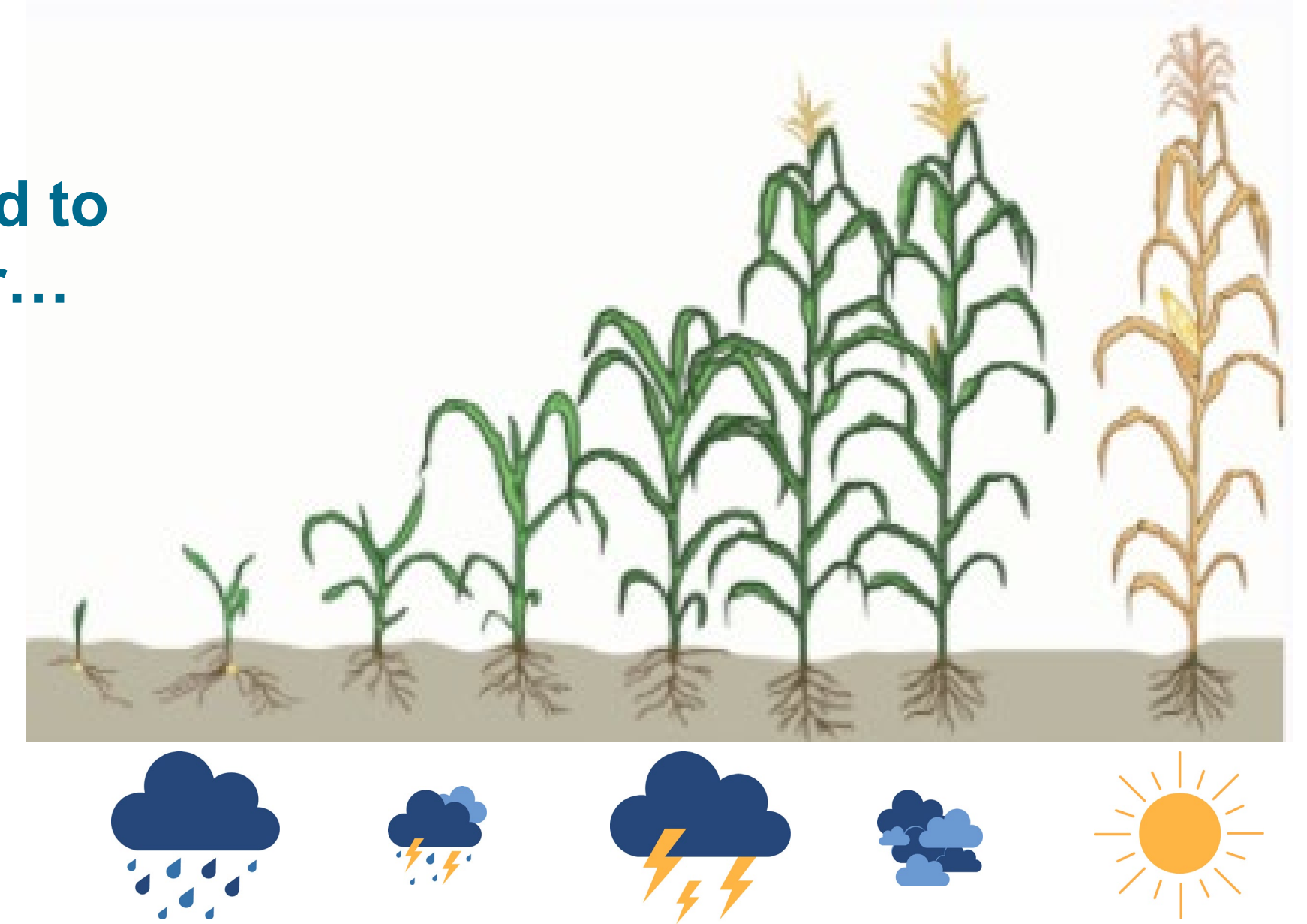
BEREXIA

GLOBAL ADVISOR FOR CHANGE



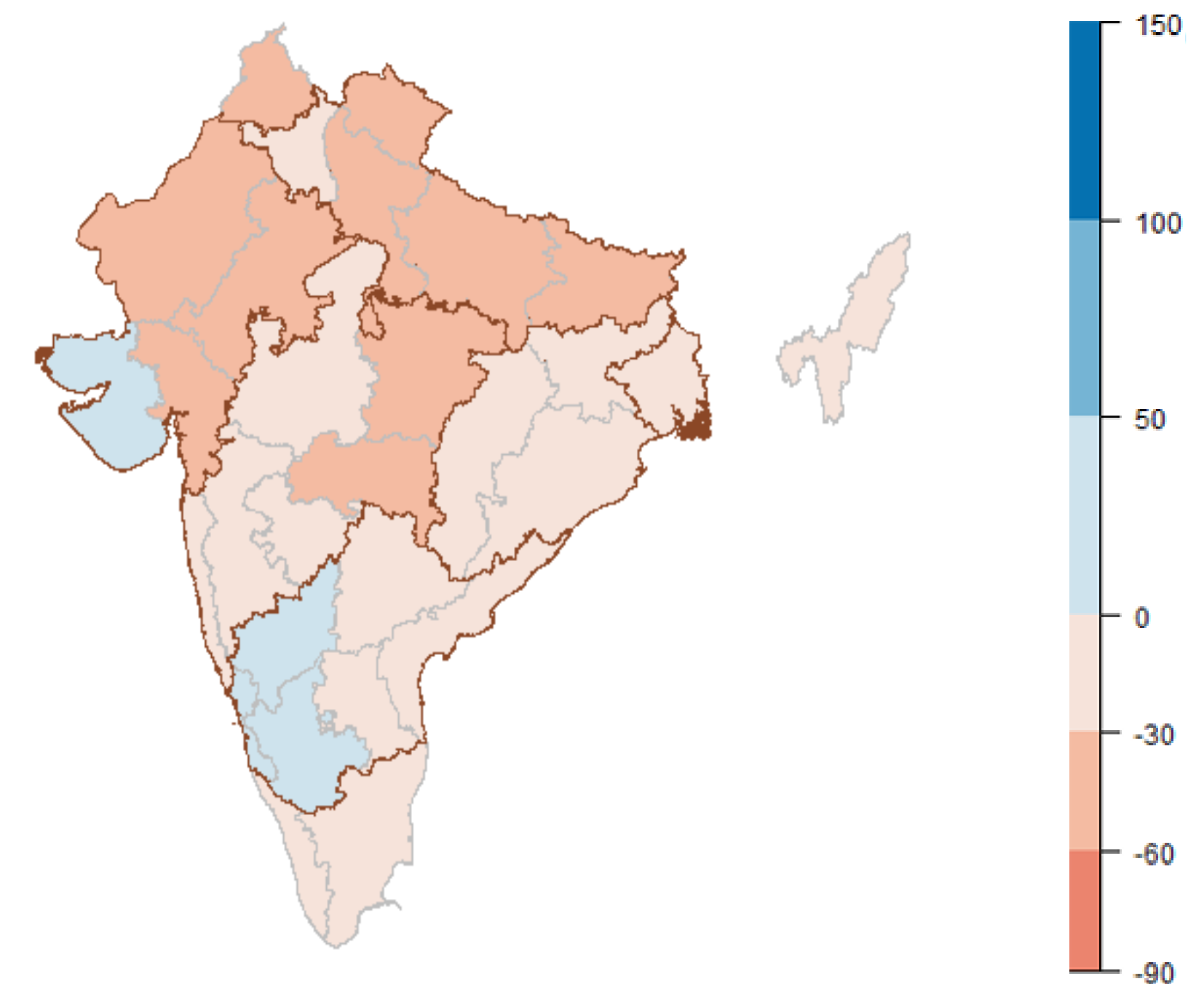
PRESENTATION OF AGRICULTURAL PRODUCTION RISKS

Exposed to
weather...

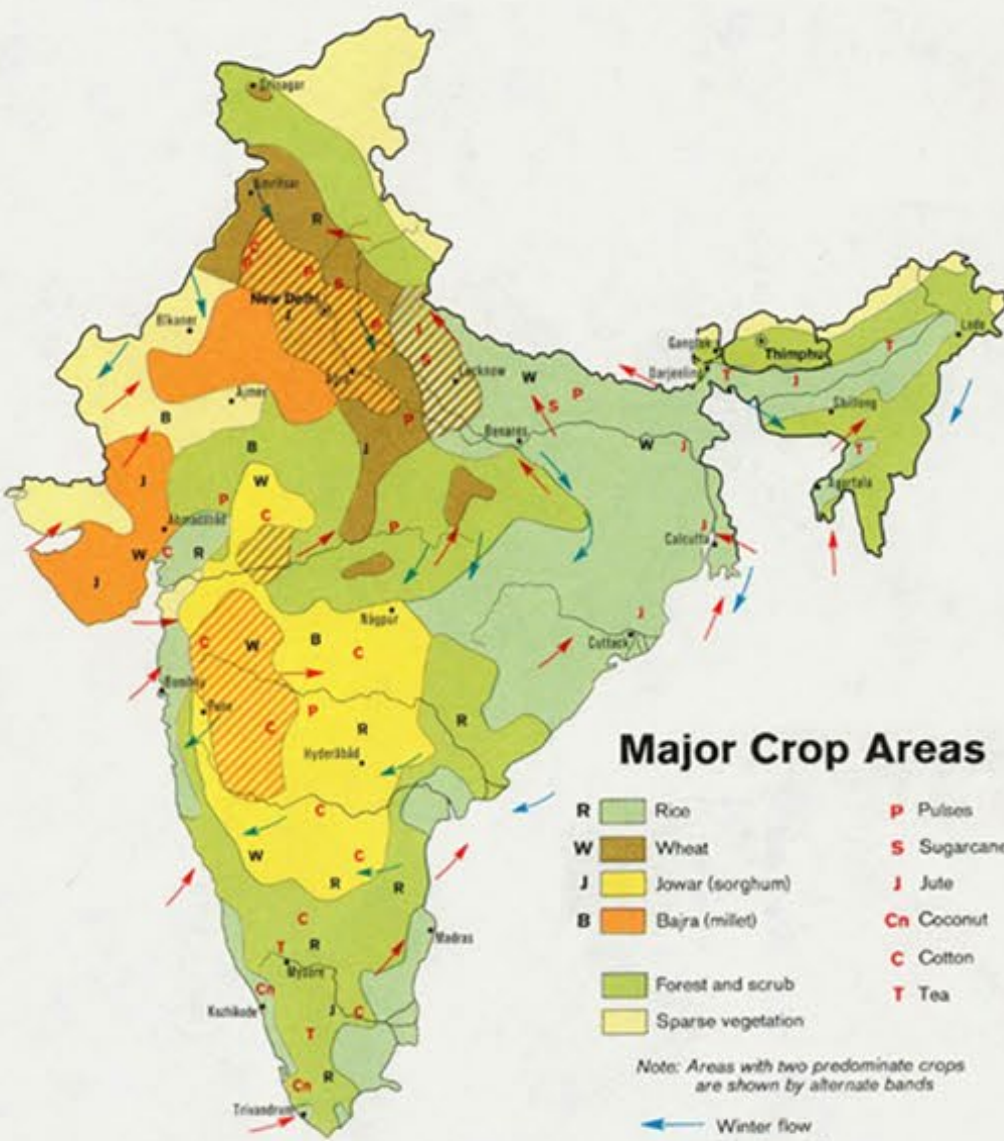


Example:
Drought
2009

Percentage
rainfall departure
per subdivision,
2009

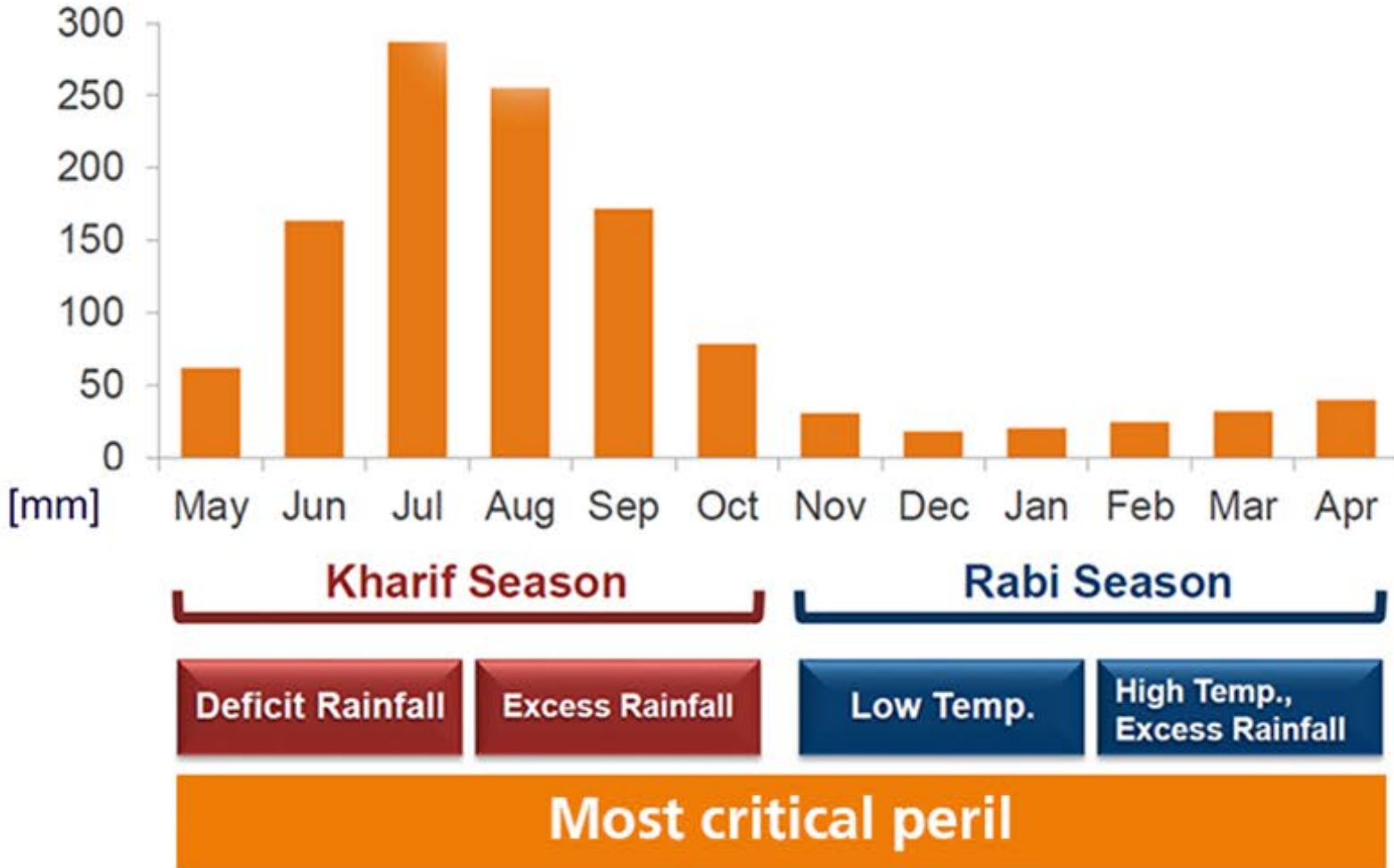


PRESENTATION OF AGRICULTURE IN INDIA



Cropping regions

Size of India nearly 6 times France



Average monthly rainfall



PRESENTATION OF THE DATA

2017_Andhra Pradesh_Kharif.xlsx
2017_Andhra Pradesh_Rabi.xlsx
2017_Bihar_Kharif.xlsx
2017_Bihar_Rabi.xlsx
2017_Chhattisgarh_Rabi.xlsx
2017_Gujarat_Kharif.xlsx
2017_Gujarat_Rabi.xlsx

Each file traces yields insured provided in a given **year**, in a given **state**, for a given **season**

A **glossary** will be provided with further definition of the dataset

| State | Cluster | District | Sub-District | Block | GP | Season | Crop | Area Sown (Ha) | Area Insured (Ha) | Per Ha (Inr/Ha) | Area Insured (Ha) | Itemnity Le | 2010 Yield | 2011 Yield | 2012 Yield | 2013 Yield | 2014 Yield | 2015 Yield |
|----------------|---------|----------|--------------|-------|-----------|--------|-------|----------------|-------------------|-----------------|-------------------|-------------|------------|------------|------------|------------|------------|------------|
| Andhra Pradesh | 1 | Anantapu | AGALI | | AGALI | Kharif | Arhar | 41 | 1 | 25000 | 32058 | 0.8 | 165 | 88 | 161 | 70 | 295 | 89 |
| Andhra Pradesh | 1 | Anantapu | AGALI | | AKKAGALA | Kharif | Arhar | 41 | 1 | 25000 | 32058 | 0.8 | 165 | 88 | 161 | 58 | 232 | 120 |
| Andhra Pradesh | 1 | Anantapu | AGALI | | HULIKERA | Kharif | Arhar | 41 | 1 | 25000 | 32058 | 0.8 | 165 | 88 | 161 | 67 | 295 | 89 |
| Andhra Pradesh | 1 | Anantapu | AGALI | | INAGALOF | Kharif | Arhar | 41 | 1 | 25000 | 32058 | 0.8 | 165 | 88 | 161 | 62 | 232 | 120 |
| Andhra Pradesh | 1 | Anantapu | AGALI | | KODIHALL | Kharif | Arhar | 41 | 1 | 25000 | 32058 | 0.8 | 165 | 88 | 161 | 70 | 295 | 89 |
| Andhra Pradesh | 1 | Anantapu | AGALI | | MADHUDI | Kharif | Arhar | 41 | 1 | 25000 | 32058 | 0.8 | 165 | 88 | 161 | 58 | 232 | 120 |
| Andhra Pradesh | 1 | Anantapu | AGALI | | NARASAMP | Kharif | Arhar | 41 | 1 | 25000 | 32058 | 0.8 | 165 | 88 | 161 | 58 | 232 | 120 |
| Andhra Pradesh | 1 | Anantapu | AGALI | | P.BYADIGI | Kharif | Arhar | 41 | 1 | 25000 | 32058 | 0.8 | 165 | 88 | 161 | 82 | 295 | 89 |
| Andhra Pradesh | 1 | Anantapu | AGALI | | RAVUDI | Kharif | Arhar | 41 | 1 | 25000 | 32058 | 0.8 | 165 | 88 | 161 | 62 | 232 | 120 |
| Andhra Pradesh | 1 | Anantapu | AMADAGUR | | AMADAGUR | Kharif | Arhar | 41 | 1 | 25000 | 32058 | 0.8 | 165 | 88 | 161 | 343 | 257 | 548 |
| Andhra Pradesh | 1 | Anantapu | AMADAGUR | | CHEEKIREV | Kharif | Arhar | 41 | 1 | 25000 | 32058 | 0.8 | 165 | 88 | 161 | 343 | 257 | 548 |
| Andhra Pradesh | 1 | Anantapu | AMADAGUR | | CHINAGAI | Kharif | Arhar | 41 | 1 | 25000 | 32058 | 0.8 | 165 | 88 | 161 | 98 | 273 | 394 |

Geographical & Seasonal data

Crop & Insurance data

Historical yield data

Sheets with 500'000 rows → 5 million yield values Have fun playing!



THE DATATHON OBJECTIVE IS TO...

BUILD A CLUSTERING MODEL SEGMENTING INDIAN AGRICULTURAL ZONES HIGHLIGHTING SIMILAR FACTORS AND BEHAVIORS FOR LOSSES

- Find clusters which **maximize** as much as possible **dependence within a cluster**
- Use the data we provided, but also search for **additional / external data**
- Clustering will be **evaluated** at **two levels** :
 1. Independent of administrative level (i.e., clustering can be performed at any level of granularity)
 2. At state level (i.e. a clustering of states will be expected)
- **Clustering** for **2 seasons** (Kharif / Rabi) independently
- **Remaining dependencies** between **clusters** and **season** should be estimated
- Last (but not least !), you should be able to **explain** the **factors** that are **driving your clusters**

DATAATHON

NOVEMBER 2021 - MARCH 2022

HICHAM DAHER

Head of Data Science
SCOR IT

SCOR
The Art & Science of Risk

BEREXIA

Global Advisor for Change



THE CHALLENGE IN A NUTSHELL

Prizes will be delivered to top 3 teams,
based on 2 major criteria



SCIENTIFIC



BUSINESS



To deliver those awards the jury will assess all along the challenge during the different touch points the following criteria



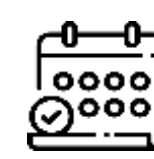
Scientific
Rigor



Mathematical
Validity



Code
Cleanliness



Risk
Taken

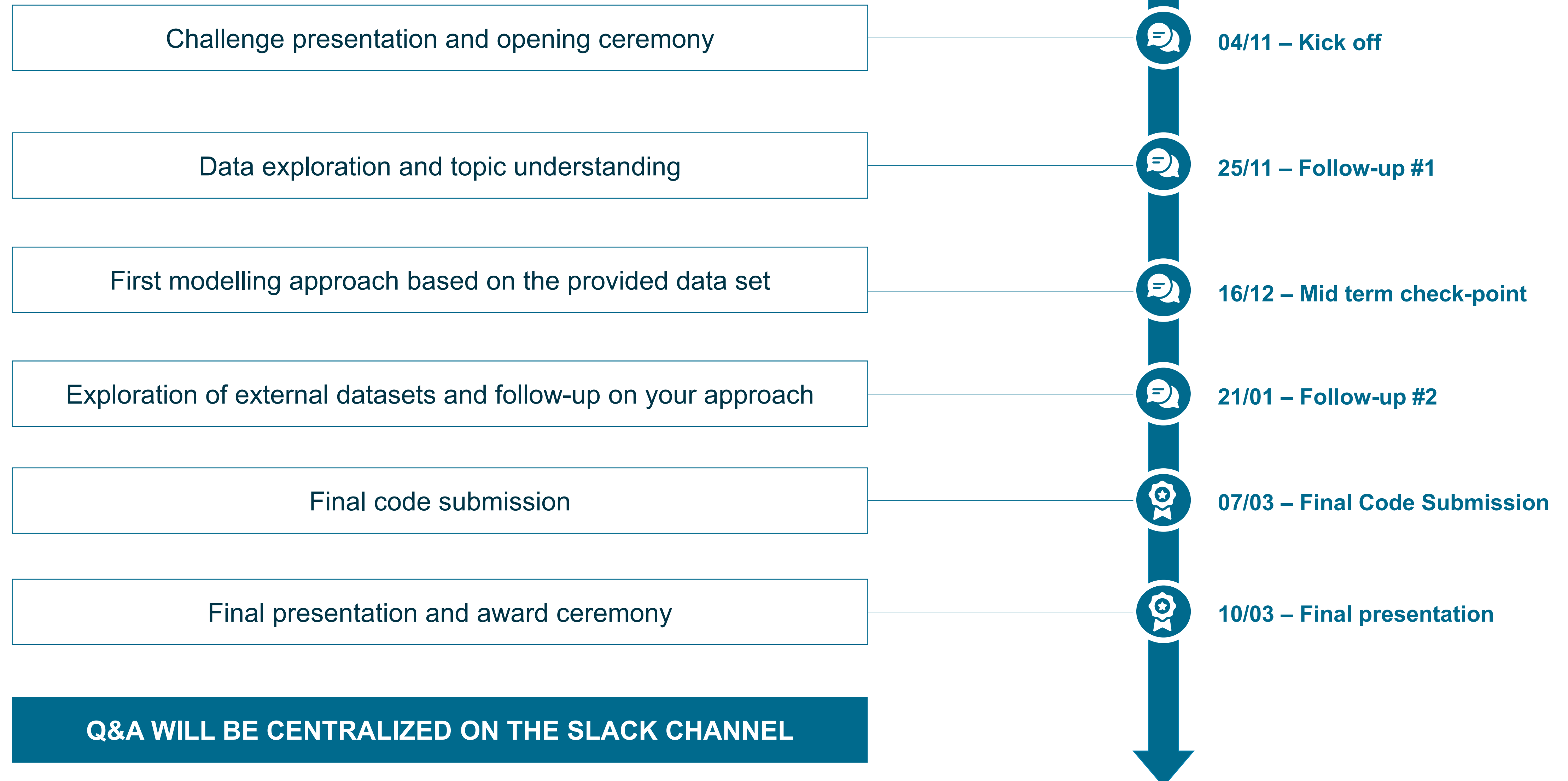


Analysis
Relevance



Visual &
Restitution

THE TIME FRAME



LOGISTICS

A slack channel to centralize all communications

You will receive an invite right after the event.

Teams should be formed by November 12th

Please send us your team members names and name a primary contact in each team.

The primary dataset will be published on November 04th

Submission in practice

We'll set up a git repository for you:

- One repository for each team.
- You will receive credentials to access to this repository
- You will be evaluated on the code and predictions pushed on this repository.

Credentials will be sent to each team as well as a tutorials.

MULTIDISCIPLINARY TEAMS REQUESTED!



DATAATHON

NOVEMBER 2021 - MARCH 2022

THANK YOU!

See you right now for Q&A and cocktail

SCOR
The Art & Science of Risk

BEREXIA

GLOBAL ADVISOR FOR CHANGE

