# SCOR DATATHON GLOSSARY

**Seasons**: There are 2 main seasons in India (**Kharif** and **Rabi**). In some southern states, there are even 3 seasons (Kharif, Rabi and “summer”)

* **Kharif**: monsoon season, bigger than Rabi, sowing with beginning of first rains (~June, may vary depending on state). Season lasts until ~ Sept/Oct, harvested at end of monsoon (Oct – Dec). Accounts for ~2/3 of insured crops
* **Rabi**: winter crop season, sowing after monsoon (~ mid November), harvested in April / May. Accounts for 1/3 of insured crops

**States**: Refers to Indian state or union territory. India has 28 states and 8 union territories. Not all territories are part of the insured program.

**Cluster**: As part of the insurance scheme, each Indian State defines its own “clusters”. Clusters are not in general official administrative divisions of India. Each insurance company participating in the insurance scheme can win in a tender process one or more clusters. Tendering used to be done on district level. District level tendering was abandoned because of the large number of districts and the additional administrative effort. Clusters are usually defined in such a way that the include Districts from different parts of the State (not neighboring Districts).

**District**: An administrative division of an Indian state (or territory). Each **cluster** (as defined above) contains several districts; one district belongs to only cluster (within a state).

**Sub-District**: A further (administrative) subdivision of districts. Some districts are divided into sub-districts, other in blocks (see below).

**Blocks**: A further (administrative) subdivision of districts, more “granular” than sub-divisions. Also called Tehsils or Talukas, they can be seen as a conglomerate of villages.

**GP**: Gram Panchayat, or “village”. This is the most granular administrative division in India.

**Crop**: The collected crop. Some crops are only grown in Kharif, other in Rabi.

**Area Sown**: Area sown by the designated crop. This value is most often provided at district level. If repeated, this level can be at a higher / lower level.

**Area Insured**: Area covered by insurance scheme. This value is most often provided at district level. If repeated, this level can be at a higher / lower level.

**SI per Ha**: Insured value per Ha in Indian Rupee.

**Sum Insured**: Area Insured \* SI per Ha

**Indemnity Level**: Threshold of indemnity level used in the **loss formula**.

**20xx Yield**: Yield harvested at a given year. The yields are often measured with a lag of 2-3 years, but the values that are obtained in the past 7 years will be used in the **loss formula** for a given year. In some cases, the yields will be provided at (average) district level, while in other cases, it can be given at a more granular division.

**Threshold Yield / Loss formula**: Within Insurance scheme, the threshold yield will be calculated by considering the 7 last years of yield data. Out of the 7 last years, remove the 2 lowest yield values and calculate the average between the 5 yield values. The **threshold** **yield** will be given by multiplying the indemnity level by this average. The losses are then evaluated as the normalized difference between threshold yield and actual yield for a given year, multiplied by sum insured.

Mathematically speaking, let be a given set of yield observations for the past 7 years, θ be the indemnity level, S the sum insured, L the calculated loss and the threshold yield corresponding to this indemnity level.

We exclude indices and designated by

We define the **threshold** **yield** as per following formula

Let x be the horizon at which we want to evaluate the loss.

The **loss** **formula** will be calculated as such