# Granularity

## What is granularity?

Granular data, as the name suggests, is data that is in pieces, as small as possible, in order to be more defined and detailed. The advantage of granular data is that it can be molded in any way that the data scientists or analyst requires, just like granules of sand that conform to their container. **Granular data can be aggregated and disaggregated to meet the needs of different situations.** \* If data is stored at the proper level it is very easy to summarize or aggregate into an overall message around the data itself

If data is not granulated, such as a name or address field being saved as a while, then it is very difficult for analysts to mine and analyze data because they are in large chunks. Granular data can be easily merged with data from external sources and can be effectively integrated and managed.

* Atomic level data is a way of making sure we have capability to summarize it.

## Chunky Data

* Too broad to discern meaning
* Lack of framework to provide context
* Interesting but meaningless

Example if you make an image chunky



### Increased grain

* Start to see meaningful patterns?
* Suggestion of framework
  + Data vs noise
* Increase tendency to add personal bias to data interpretation at this stage
  + I know what this is!

## Further Graininess

* Patterns emerge
* Data starts to become clearer
* May even suggest further enhancement required to fully interpret
  + I think I know but I want more detail to prove it
* Begin to intuit outcome
  + Resist this urge
  + Dont just think if you see a pattern that there is enough data there



## Still more Granularity

* Resolution of data may be sufficient to interpret now
* More intuitive results
* Still suggestive for further detail requirements
  + Ok. We’ve got something here. Do we need to go further?

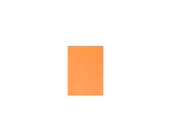


## Clear Granularity

* Framework of data now apparent
* Subject of data is clear within this framework
* Meaningful results can be inferred and validated at this stage

## Over Analysis

* Compulsion ot add as much data is available to the model, or to analyze down to minute details
* More is better right?
* Framework is lost
* Meaningful detail is lost
* No longer intuitive
* Patterns in data may be pure coincidence yet still present



* Too much detail does not give a clear framework anymore

## Summary, Granular data

* Smallest usable form (atomic)
* Can be aggregated/disaggregated as needed for analysis
* Easily merged with other external data to improve overall resolution of data model
* Easily managed in database
* Helps to minimize inferential biases and leaps of logic