CIIER
Centre for Innovative Industry Economic Research Inc

In Pursuit of Patterns

The process of identifying how things are happening, and how we can measure, predict, or simulate such processes.

All systems have patterns.
The currents in the seas; the wind and climate; voting intentions; movement of electrons; economic growth; virus mutations, all follow patterns, some of which we can discern, some of which we are still learning to identify, some, perhaps, that we will never completely understand.
Understanding how things evolve and change requires us to understand the patterns of process, and the factors applying to such patterns, direction, trend, acceleration, and countervailing forces.

There is, however, sometimes a confusion in the data science field between the process of pattern recognition, ie finding out what is happening, and that of pattern correlation ie confirming that the predicted pattern is discernible.
This is often reduced to a discussion on how to deal with "outliers" ie data elements that don't fit presumed ranges of acceptability. Outliers can, of course, be aberrant, through mismeasurement, misunderstanding, or extraneous impact. Outliers can also, however, be an indicator of a part of the pattern that has simply been overlooked by previous analysis, or a new and emerging part of the pattern showing change.

The other issue is that of identifying "missing factors", and here there is sometimes a tendency to restrict our analysis to only the specific data set presented, rather than to explore the availability of other data sets with correlating or indicative data that may improve our understandings.
An example of this aspect of analysis can be found in the various clinical Patient Outcome Registers, which are only recently being matched, in some cases, to demographic data freely available from the national census data held by the Australian Bureau of Statistics With such matching helping to identify location, gender and age based variances and trends that are less discernible purely from the Register data.

Data scientists with a primary background that does not include professional ICT analysis and programming, may tend to exclusively use statistical packages as if they were computer languages, without necessarily appreciating that statistical packages have significant presumed patterns embedded within their command structure.
Thus the underlying real pattern of a particular dataset may be less discernible, or may be interpreted as outliers that are sometimes simply discarded.

Thus, in practice, we often simply use patterns to match data to what has already been identified, whereas true data science is about finding new ways to explain the data we are presented with, and in exploring other data that may identify impacting factors.

Pursuing the pattern, rather than just presuming that we know what it is.

Centre for Innovative Industry Economic Research Inc. ABN 64 806 162 996
An ACNC registered not-for-profit Research organisation
PO Box 1132, Thornbury Vic 3071