



$$\text{Score} = \sum_{i=1}^n X_i * \text{Cos}(Y*i)$$

## Big Data assigned cohorts from variable memberships (Boolean vs Nolanean)

Published on March 6, 2016

[Edit article](#)[View stats](#)

**Tony Nolan OAM**  
Data Scientist

[48 articles](#)

One of the basic ways to work in large datasets, is to assign memberships to rows from their memberships in combinations of variables. Of course the problem is the number of variables that can be used. One of the quick ways to assign to a cohort is to use Boolean. Of course this limits the variable to be solely yes or no, 1 or 0, etc. While it is limited to about 8 to 10 variables, it is also limited that it treats every variable the same, so it is not really distance preserving. I.e the larger or smaller the numbers, don't carry extra meaning.

I sat down one morning, over 20 years ago, and thought how good it would be if there was a way to have numbers between 1 to 10 in each variable, and when they were added up, you got a unique number, a digital hash. But a digital hash that retained distance, etc. But when you apply normal maths to  $2 + 1 + 3 = 6$  or  $1 + 2 + 3 = 6$ , or  $3 + 1 + 2 = 6$ . I'm sure you get the idea.

The next step I tried, was to give a different weighting to each number in the sequence. The difference between each weighting, was a geometric progression, with an extension value. So it became impossible to have any combination of 1,2,3, etc come up to the same end number. I checked this out on graph paper, and ran some mathematical proofs, and did some testing on the ideal value to have as the cascading sequence value.

It soon became apparent, that the extension and secondary calculation for the x,y graph was not needed. In fact, I had invented a new variation of a Discrete Cosine Transformation (DCT), much to the shock of some mathematicians who said it should have taken me 20 years of PhD level work to come up with it.


Like Comment Share

[Messaging](#)

100 unique variables combinations. While there are some limitations, the value range of between 1 to 99 values in each variable. However by System of Systems methodology, I can compound and compress the entire membership behavior across a number of intersecting cohorts.

This methodology can be used on numeric and categorical variables, and can be reverse engineered, the calculations work in one direction only.

Published by



Tony Nolan OAM


Data Scientist

Published • 3y


Reactions



1 Comment



Add a comment...



Gene Sullivan • 2nd

Self-Learner/Self-Teacher at Autodidact U

3y ...

\*Yawn\*

<author's\_surname>ean Discrete-version-of-an-analog function contrasted with 2-value discrete-cum-binary-cum-Boolean function.

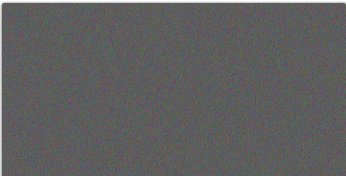
\*yawn\*

Like Reply




Tony Nolan OAM  
Data Scientist

More from Tony Nolan OAM



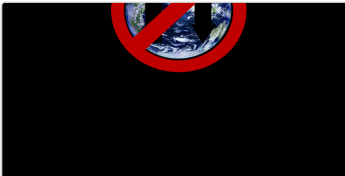
99.816% compression on Lorem Ipsum

Tony Nolan OAM on LinkedIn




One Solution to Help Detect Terrorists

Tony Nolan OAM on LinkedIn



No U Turn Ahead

Tony Nolan OAM on LinkedIn



Artificial Intelligence (AI): Join the dots to Empowerment.

Tony Nolan OAM on LinkedIn


See all 48 articles



- About
- Careers
- Advertising
- Small Business

- Talent Solutions
- Marketing Solutions
- Sales Solutions
- Safety Center

- Community Guidelines
- Privacy & Terms
- Mobile




Messaging

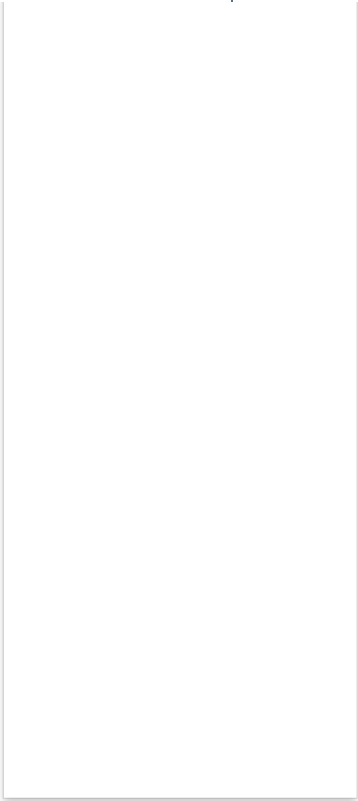


Search



 **Manage your account and privacy.**  
Go to your Settings.

LinkedIn Corporation © 2019



Messaging