Some Thoughts on Analytics in a Post COVID-19 Environment

Written by Richard Boire at Boire Analytics

In these most difficult times, the use of analytics is certainly not top of mind for most organizations unless it is being used to combat the virus. The challenging scenarios of meeting payroll and having access to cash are the obvious immediate priorities. But from a non analytical perspective, like most people, I am amazed by the many acts of giving and generosity that really speak to the better angels of our nature.

But we will overcome these challenges and being the constant optimist that I am, this will happen sooner rather than later. In this new post COVID-19 environment, it is not unrealistic to assume that the way consumers behave and think will be transformed significantly. Of course, this has ramifications when conducting analytics exercises. Virtually all data analytics exercises deal with historical and longitudinal data. The development of models, segmentation systems, and/ or reports all use historical data in their solutions. Given that much of the power of predictive analytics/machine learning solutions arises from longitudinal or historical data, this all begs the question of how we deal with data and specifically consumer behaviour data prior to, during, and after the COVID-19 crisis.

As I thought about this, I was reminded of the last crisis that seemed to galvanize our collective consciousness and made us reflect on what is truly important in our lives. That was the 9/11 crisis. During that time, this newfound awareness did impact consumer behaviour. My organization at that time was asked multiple times about the 9/11 impact on our analytics exercises especially the more advanced analytics exercises such as the many predictive models that we had built. In other words, would model performance be significantly compromised because of these changes?

Our perspective in evaluating model performance was to observe the increased targeting capability where top scored names are most likely to yield the desired behaviour and the bottom scored names are least likely to yield the desired behaviour. If we place these scored names into model deciles, then the top decile should have the strongest observed behaviour while the bottom decile should have the weakest observed behaviour. Essentially, the model can then be evaluated based on how well the model rank orders scored names based on the observed modeled behaviour.

Listed below is a graph example of a response model to target existing customers in the purchase of an insurance-related product.
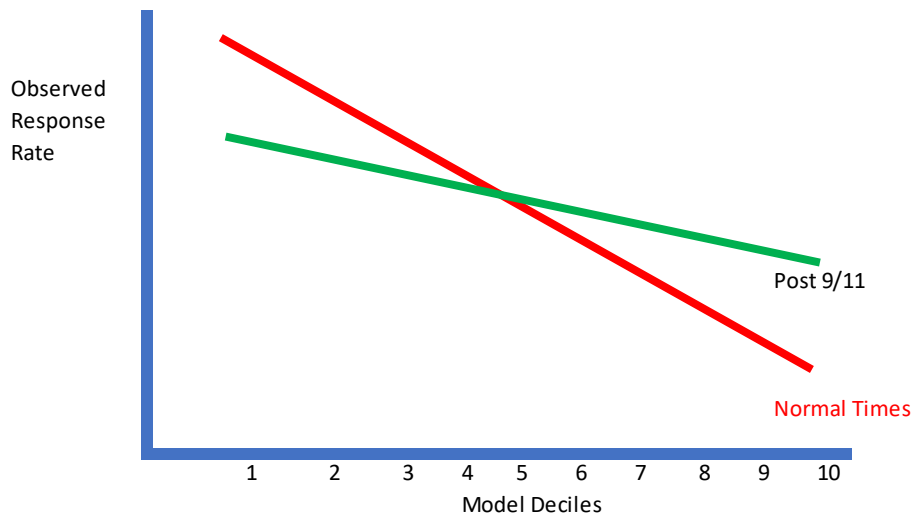
In this above example, we observe the same response model being applied during normal times and then being applied right after 9/11. In this example, the model has eroded as the slope of the line has decreased indicating the model's reduced rank-ordering capability.

Listed below is another insurance response model example of a situation where again there has been a change in performance before and after 9/11.
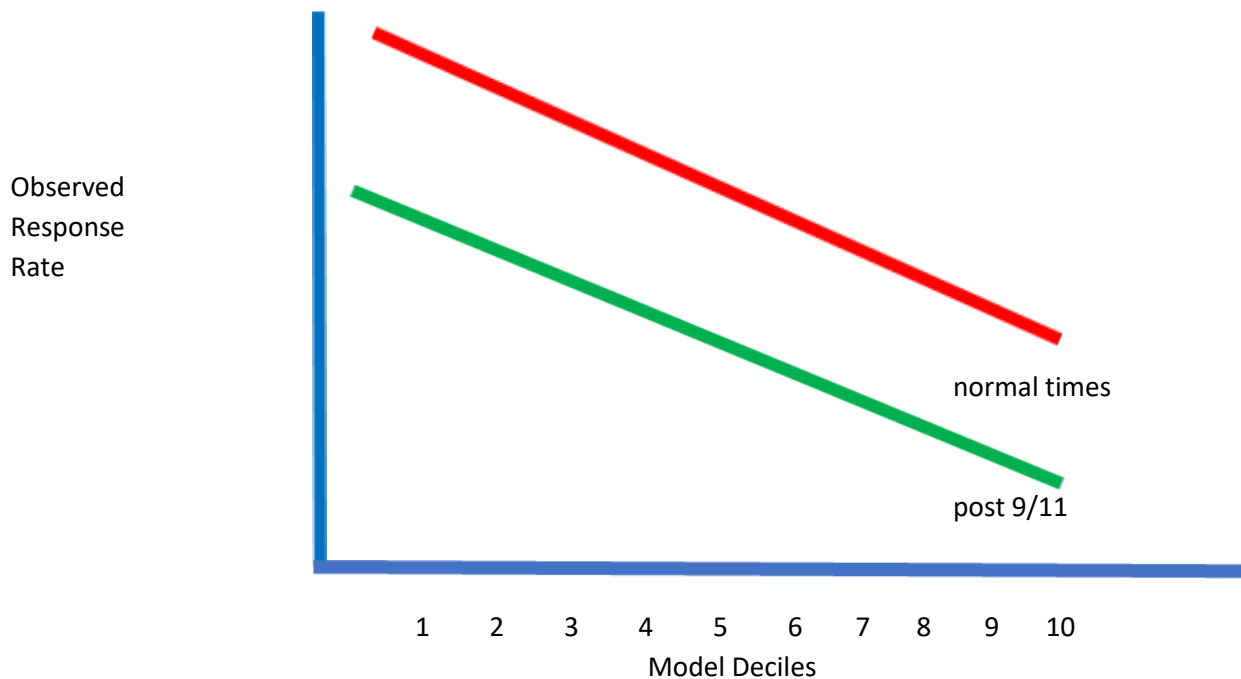
Yet in this scenario, we observe that overall response rate performance has deteriorated in post 9/11 but the model itself continues to perform as the slope of the line is still the same indicating no decrease in the model's rank ordering capability. But this scenario is more highly unlikely as model erosion really represents the most likely scenario after some major crisis.

These examples above serve to highlight the fact that analytics solutions can change quite dramatically and for the worse both from an overall perspective and from a modelling perspective. The erosion of model performance during and after a crisis is somewhat intuitive as consumer behaviour is not continuous throughout the population. Certain characteristics about consumer behaviour such as their age, where they live, income, etc. may all be impacted differently by an extraordinary situation. As data scientists who are always trying to mathematically assign weights to these variables based on patterns in the data, this becomes a flawed option as we know these data patterns have now undergone significant change.

So, what are the options for analytics practitioners? For many experienced practitioners, simplicity as a philosophy has always been core to what we do. Capitalizing on this notion of simplicity, one of the first concepts of basic analytics is the use of indexing where a specific consumer behaviour or demographic of an individual is compared to that same consumer behaviour or demographic of a group. A classic example of this is the use of RFM where each consumer is indexed on recency of last activity, frequency of activities, and the average amount purchased for a given activity. These indexes are then combined into one overall index. See example below.

	Recency of last activity(R)	Frequency of Activities(F)	Amount Purchased per Activity(M)
Customer A	2 months	30	40
Group	4 months	20	50
Index	2	1.5	0.8
RFM Index for Customer A	=4.4(2 + 1.5 +.8)		

In this simple example above, we are assuming that each behaviour (R, F, M) are equally weighted in calculating the overall index of 4.4. The use of an index is less sensitive to great changes in the data environment as we are simply trying to determine the relative difference between the individual's behaviour relative to the behaviour of the group. Given today's digital environment, one can translate these activities to R: recency since last visit to the website, F: # of visits to website, and M: average duration per visit

We have used this type of indexing or comparative approach multiple times and not always when there has been an extraordinary global event. For example, many organizations will undergo significant changes in their corporate strategy which will dramatically impact consumer behaviour and the accompanying data environment. In these situations, we have developed indexes based on customer value and change in customer value or behavior change.

Of course, in a more stable data environment, this indexed approach is sub-optimal as the advanced machine learning and deep learning technologies can really leverage the abundance of data that is now available at our fingertips. But as we enter this new post COVID phase, analytics practitioners need to be very cognizant of a more unstable data environment. As I have conveyed to fellow practitioners and students for many years, data science is about identifying patterns in the data. But if these patterns are undergoing significant changes, then a simple and pragmatic approach such as indexing can be a very useful targeting tool in a sub-optimal data environment.