HR Analytics – What Stopping Us And Where Do Go From Here?

Lyndon Sundmark

HR Analytics – What Stopping Us And Where Do Go From Here?

Introduction

Earlier this year I wrote a few blog articles on HR analytics. In one of those articles I indirectly expressed doubts, concerns, and readiness of the HR profession to capitalize on the promise of HR analytics and take rightful claim and ownership of it within the services that HR provides.

https://www.linkedin.com/pulse/why-hr-might-able-reinvent-itself-lyndon-sundmark-mba?trk=mp-reader-card

I mentioned a number of obstacles that might exist-historical and current- that could be standing in the way. Many of these are related to how HR views itself as a profession and a lack of readiness to see itself as a far more technical field and profession.

My sense is, though, that even if there was an immediate willingness by HR to embrace analytics and its 'technical' HR side, there would still be some immediate obstacles that would prevent uptake by HR.

What is the hold up?

There is still a problem of 'where we start'. Most people, to adopt new methodologies, require a 'show me'. That 'show me' has to be in the context that is relevant to them. In this case with HR, the context has to be HR examples. In any new developing field, there is often slow uptake on adoption of methodology because of a dearth of 'show me' examples.

There are at least three essential ingredients to a 'show me' for HR analytics;

- Collect data that illustrates what we are attempting to convey
- Show how to apply data science/ predictive modelling and HR analytics methodologies to it
- Share that with the HR community.

So what's the problem? We have oodles of HR data. Analytical methodologies are bursting on to the scene. The ingredients are there to do it, but there is little if any action.

Part of the problem is that there is interdependency of data needed and analytic procedures. What procedures we use will dictate the data we need and the form it needs to be in. We may have 'data' but it may not be the data we need. And how we know which analytical technique to use? We can't collect potentially relevant useful data for analysis, until we decide on the analysis. And we can't decide on the proper analysis until we understand what are analytical options are. And understanding those usually requires data to work with.

This presents us with a bit of a dilemma- analytical procedures require data, in order for us to learn. But the relevant data being able to be gathered being dependent on the procedure being chosen. And none of this being able to be shared with HR community until we have workable HR examples. And to encourage the forward movement of analytics in HR - we need those examples.

Resolving the Dilemma

In effect, the circular dependency of data and analytical examples has to be broken and a start has to occur somewhere. I think that a start can occur with an identification of a business problem that has HR implications, AND an inquiry. **'Is this an HR business problem that HR analytics can address and help solve'?** If we are courageous enough to ask the question, and if it truly requires an answer by the business we are likely to start the process of data gathering/collection and research to answer the question.

In fact if we truly accept some of the common definitions for workforce, HR, People Analytics

See https://www.linkedin.com/pulse/workforcehrpeople-analytics-hr-lyndon?trk=mp-reader-card we must ask the above question. Dr. John Sullivan's definition of People Analytics is particularly illustrative of this

From Dr. Sullivan's article:

http://www.eremedia.com/tlnt/how-google-is-using-people-analytics-to-completely-reinvent-hr/

"***People analytics*** *is a data-driven approach to managing people at work. Those working in people analytics* ***strive to bring data and sophisticated analysis to bear on people-related issues****, such as recruiting, performance evaluation, leadership, hiring and promotion, job and team design, and compensation …*

Data driven approaches, sophisticated analyses etc. require us to think outside the box and have the **spirit of inquiry regarding data and analytical methods** that can be brought to bear.

But even if we have the data, how do we begin to get our head around what are option for sophisticated analyses?

The Basics

I think there are some basic building blocks of understanding that can be tremendously useful. These building blocks cover:

- The Role and Understanding of Measurement
- The Importance of Measurement Scope
- The Purpose and Role of Statistics and Statistical Analysis to HR
- The Role of Statistical Software, Data Science /Machine Learning/Predictive Modelling to HR Analytics

Measurement

It begins with the understanding the definition and the role of measurement. It's important to understand that measurement is woven into the very fiber of HR DNA whether we acknowledge this consciously or not. Let's look at a definition of 'measurement'

From https://en.wikipedia.org/wiki/Measurement

Measurement is the assignment of a number to a characteristic of an object or event, which can be compared with other objects or events.

You could probably substitute 'description' for 'characteristic' in the above definition too.

When you think of above definition, the very act or recording HR information into information systems is 'measurement'. This is because to store the information into the system we assign numbers or descriptions/categories to it. We have employee 'numbers', gender, age, birthdates, names, titles, salaries etc. These are either numeric pieces of information or descriptions of things or categories.

That definition says nothing about the rules of measurement, or how good the measurement is. These are important concerns too, but at the most basic level, once we are storing information we are measuring. **So much for the argument that HR is a non-technical field**. And why do we store information and 'measure'? - To describe and understand end explain the world around us.

If we embrace the reality that HR too is about measuring-we have one of the first building blocks of heading in the direction of sophisticated analyses mentioned above.

HR Measurement Scope

Once we understand the centrality of measurement to what we do in human resources management, it's important to get our heads around the scope of what can be measured. Indeed, what can be measured is almost limitless. What can we, should we be paying attention to? And once we decide on what to measure, how or where does it stand compared to other things we measure? How do we see and keep track and keep sense of that we measure?

One possible framework, and there could be others, that could be useful is one I introduced earlier this year in the following article

https://www.linkedin.com/pulse/what-does-data-driven-hr-look-like-lyndon-sundmark-mba?trk=mp-reader-card

In it, I had suggested that 'metrics', the things we measure, typically fall into 1 of 3 categories for HR purposes:

- HR Activity
  - These metrics describe what is going on with the people in our organization. Examples are counts of current employees, turnover counts, hire counts, turnover and hiring rates, absenteeism, accident and injury rates, benefits participation rates, training enrollments, employee churn, grievances counts and rates. If it describes what is going on with our employees and what they are doing, or what is happening to them they are in this category
- HR Process Efficiency and Effectiveness
  - These metrics concern themselves with HR business processes, the demand for them, the length of time to provide HR services within our HR business processes, customer feedback on HR Services, waiting time to receive service. They aren't about people activity which makes it different than the category above. They are about measuring HR services or business activity.
- HR Methodologies
  - These are metrics or measures which are generated or used or could be used in the actual methodologies we use to carry out our HR functions or services. It's when we define the goal of our function or service and try to improve on reaching the goal, have a better or more accurate outcome in the goal and changing our methodologies or processes to improve the outcome. And its measuring what we are striving for and along the way to getting there. The purpose is to do what we are doing even better. It might include measures necessary for more accurate job classification, better sourcing of candidates in recruitment, better screening of candidates, and better selection of leaders within the organization.

If we have a framework to see our HR measurement in, its becomes a whole lot easier to see what we haven't done, what we have done, and what we could do with respect to bringing sophisticated analyses to bear in HR. Having a framework is a second good building block

Purpose of statistics

Even though statistics, statistical methodology and statistical software and tools have been around for decades and predate PCs, a lot of people misunderstand statistics and statistical analyses and their purpose.

Perhaps this is because in most people's minds and experiences, when you mention statistics or statistical analyses, the first thing than comes to mind is a boring summary table of numbers or confusing business chart, or summary data on a spreadsheet. I think data warehousing tools, unintentionally, also continue to foster this misunderstanding. We are simply given more and easier tools to slice and dice sometimes what appears to be boring business data. The predominance of introduction to charts and summary data and use of spreadsheet for this purpose in financial applications doesn't help this either.

If, in the previous section, at the heart of all data collection is measurement (without mentioning what we are measuring for) - then at the heart of statistical analyses and methodology is understanding the world around us. We want to answer the question of what are we measuring for: inevitably to provide an answer or solution to a business question or problem.

Yes, at its most basic level, statistics and statistical analyses is about descriptive statistics, such as counts percentages, and averages. But it is so much more than that. While it would take a statistics course to fully understand the typical repertoire of statistical procedures at our disposal and what they do, their purposes fall into probably just a few broad categories:

- **Describing what we have or are looking at.** This is the purpose of descriptive statistics and charts. Much HR data analysis in organizations doesn't go beyond this-either because the organization doesn't demand it, or the skills are present in HR to do more. Examples here are employee counts, hiring counts, termination counts etc.
- **Examining relationships in our data**. How different things we are measuring are related to each other. If we have information on employee age and absenteeism rates , do our measurements show these two things to be related? If so, does absenteeism go down with age, or up? Or is there no discernable relationship?
- **Determining whether what we are seeing in our data is universal right across the organization**. This could apply to both descriptive statistics and relationship. Is something universally true or does it vary across/between the categorical things we capture/measure? In the above example, does absenteeism vary by gender, or organizational level, or by job.
- **Determining whether differences we see are statistically significant**. Are the differences we are seeing attributable to the different categories or could they be due to random chance?
- **Predicting accurately something that we don't or have from something we do know or have-concurrently or in the future.** This is an extremely important consideration in business. To the extent to which we can predict the future based on what our past data and history tell us, allow us to pro-act and to a certain extent actively manage our future.

- **Understanding whether what we are measuring is related to time, and if so how?** There is a whole category of statistical methodologies, known as time series analysis, who purpose if to answer that question- teasing trend, seasonal, cyclical fluctuations out of our data. Stock market data is a good candidate example of this type of data and analysis.
- **Predicting the best fit category for something for a new item where categories preexist for existing items.** In HR, a perfect example of this is job classification. A new item (job) needs to be classified into a job level and/or job family. Other jobs are already classified to family or level. But we now have a new job that has been created, and it needs to be classified fairly as to level and family based on its similarities and differences to other jobs in the same of different classifications. By the way, statistics and statistical analyses are a natural fit for this HR function, but have been in very little organizational evidence over the decades.
- **Determining the optimal number of categories for some data where categories don't preexist**. Are their natural patterns in our HR data that help suggest where groups or categories naturally occur? In HR, sometimes salary broad banding is an example of this. We may have had previous categories, but they were no longer clearly differentiating well. So what does our data itself tell us about naturally forming groups or categories?
- **Determining the upcoming choices people will make, based on past choices**. In HR an example might be open enrollment for benefits. Looking at past choices people have made, a virtual grocery cart or checkout, predict what future choices they might make on complementing benefits etc. These are normally called 'recommender' systems

This isn't intended to exhaustive, but it does summarize a wide swath of purposes that statistical analyses are intended to serve. **If we are ever going to reach the promise of HR analytics and sophisticated analyses we have to understand that statistics, and statistical analyses are our servant and tools in this picture. And we have to envision how we can use them. This isn't optional.**

Role of Data Science /Machine Learning/Predictive Modelling/Data Mining

From the previous two sections, if you accept the importance that measurement and statistical analysis play in HR analytics and sophisticated analysis, then you will understand more clearly the role of data science. Machine learning, predictive modelling and data mining play. These methodologies are very much statistical analysis methodologies. But they are more than that because they attempt to more formalize a methodology around an 'inception to completion' analysis approach rather than just a specific type of statistical analyses or procedure.

These emerging methodologies all have a common denominator – **finding patterns in data**. The desire for looking for patterns in data come from wanting to solve or understand a problem. In the context of HR, it comes from wanting to solve a business problem or create new opportunities.

There are no guarantees up front, that there will be patterns in our data that are 'business meaningful' and relevant. But there no guarantee either that the data we collect is meaningless. We have to make the decision to look at it with these powerful methodologies and tools.

It would probably take several books to go into detail on these methodologies, and that isn't the purpose of this blog article. But I will try to illustrate further their relevance, first of all with web links for books, sites, training etc., and then with some examples of how these could be used in HR. While I will share some examples, my intent is to really spur HR into further creative thinking on application of these methodologies to HR business problems.

Some links

-There is no particular importance attached to these as compare to many others. They just happen to be ones I am familiar with.

Data Science and Machine Learning

I have been at the time of writing this blog article been taking a free online edX course

https://www.edx.org/course/data-science-machine-learning-essentials-microsoft-dat203x

This covers Microsoft's state of the art tool for machine learning –AzureML. They have dozens of example templates that give an idea of the possible uses on machine learning to find patterns. I will share some of these shortly – below.

Predictive Modelling

A book that I have come across in my travels for predictive modelling is Applied Predictive Modelling by Kuhn and Johnson

http://www.amazon.ca/Applied-Predictive-Modeling-Max-Kuhn/dp/1461468485/ref=sr_1_1?s=books&ie=U
1&keywords=applied+predictive+modeling

If uses the R statistical language and software and its 'caret' package

Data Mining

http://www.amazon.ca/Predictive-Analytics-Data-Mining-RapidMiner/dp/0128014601/ref=sr_1_2?s=books&
2&keywords=rapidminer

This book covers Rapidminer- which is software specifically designed for data mining.

I have used all 3 of these at one point or another. Again there are many other software packages and books out there. These are 3 I have come across.

Patterns in the Data

So if all of these methodologies have at their 'heart' finding patterns in the data, why should we care about this? The primary reason is prediction- whether concurrently or for the future. Sometimes being able to predict allows us to solve a business problem (including HR related) or help us to discover new opportunities.

To give an idea of the possibilities, a perusal of the sample experiment templates in AzureML gives us a glimpse of the possible application of machine learning

[http://gallery.cortanaanalytics.com/browse/?categories=$$\"Experiment\"$$&examples=true]{.underline}](h

xperiment

¨&examples=true]{.underline}](http://gallery.cortanaanalytics.com/browse/?categories=%5b%22Experiment%2

You may have click for 'new' experiment to see existing examples.

Looking at the keyword for samples, finding patterns in the data can take the form of:

- Classification of items into categories
- Anomaly Detection
- Clustering like with like
- Fraud detection
- Predicting/forecasting – using regression
- Predicting/forecasting –using time series analysis
- Predicting/forecasting - churn/turnover
- Text Classification

We need to envision how these might apply to various functions in HR

- Classification is a 'natural' for job classification- in terms on intent
- Anomaly detection- might be useful in measuring the performance/delivery of our HR services to our clients. Did any service take abnormally long to deliver that shouldn't have? How would we know this? Anomaly detection

- Clustering- might be useful for salary broad banding- creating new job families and levels based on characteristics of similar work. Clustering like with like
- Predicting/forecasting using regression – determining competitive salaries for positions that are not similar to key comparators.
- Predicting/forecasting –using time series analysis- being able to plan ahead of time for the cost of maternity leaves
- Predicting/Forecasting churn or turnover. In retail sales, business are concerned with customer churn- those that will leave them for competitors. Incentivizing the customer sometimes will prevent that. For HR, does predicting who will turnover (especially roles and people you don't want to lose) merit the organization attention? Is it important or not?
- Text Classification- suppose you had historical data on successful hires and not successful hires. Suppose you had their original resumes. Supposed there was a way to discover patterns in the text of the resumes? Suppose that those patterns were predictive of those who were successful hire and those that weren't? Is what worth something to organizations?
- Recommending 'like' or 'similar' services to employees based on current or past choices. These would be in category of recommender systems.

As was indicated previously, there are no guarantees of patterns in the data and its predictive value. It's either there in the data or it isn't. But if we aren't looking, how will we know? If we aren't looking and our competitors are (yes HR has competition) are we at a competitive disadvantage in the HR marketplace?

Unabashedly, my motive here is to stir HR into thinking about these technologies and their application in their 'world', pick up familiarity with the tools, and start examining their data. As HR experts what **other areas** in your HR business have application for this? Is training and development an area? What about health and safety/workers compensation?

The Importance of Sharing Examples

As indicated above, the technologies are there to do sophisticated analysis in HR/People/Workforce Analytics. And as indicated previously, lots of HR data exists. But it is the right kind of data? Is it tailored to the business questions we are asking? Are we sharing examples so that other can learn?

I think the correct terminology here might be 'open' data. I recently came across the following article on it, describing it in another context.

http://www.zdnet.com/article/open-data-open-mind-why-you-should-share-your-company-data-with-the-world/

Traditionally, and rightly so, much HR information is confidential. And it's also rightfully protected by privacy legislation. I guess the question is, is it feasible, legal, and possible to share out HR data on an anonymous basis that doesn't violate a person's privacy and contravene privacy legislation? Is it possible to create HR datasets for which:

- Others from the HR community can learn from and practice on to do proof of concepts for their own organizations?
- The additional sets of eyes beyond the organization, may be beneficial?

This already exists in other fields of endeavor and data. An example of this is Kaggle

https://en.wikipedia.org/wiki/Kaggle

Could this apply to HR as well? Can we move HR Analytics ahead by sharing data and real life data science/ predictive modelling, data mining examples? Should a website be created for people to upload examples of their HR data science projects to share knowledge with others? What say you?