

How Do I Start Preparing Myself for the Field of People/HR Analytics?

Lyndon Sundmark, MBA

2016-10-13

Table of contents

Introduction	1
Start with A Good 'Foundational' Definition	2
Become familiar with the basic building blocks of People/HR Analytics	3
Knowledge of HR and HR functions and business processes	4
Knowledge of Information technology with respect to data warehousing, data retrieval, and human resource information systems typical content	4
Knowledge of business statistical analysis – particularly with respect to HR data, HR questions, HR problems, and HR measurement and metrics.	5
Knowledge of the scope of what's measureable in the HR context.	6
Knowledge of the Data Science framework, and how to apply it.	8
Read, Network, and Get Hands On Practice	10
Closing	11

Introduction

A little over a year ago, Hendrik Feddersen (<https://www.linkedin.com/in/hendrikfeddersen>)

and I co-authored a short article on how to start a HR Analytics project within your organization.

<https://www.linkedin.com/pulse/helpful-tips-how-start-hr-analytics-project-hendrik-feddersen?trk=prof-post>

That article and a number of my previous blog articles on People Analytics in R, have been generating a fair amount of interest on LinkedIn, also when some have been published again on <http://AnalyticsInHR.com> . (And I'm honored by that).

Some of that interest has resulted in requests for informal mentoring as well along with questions of 'how do I get started' in the field of People/HR Analytics?- or how do I prepare myself?. I have shared my thoughts and my own experience in response to those requests. So I thought I would share some of that. I would like to follow up that article with another one- this one- on how do I start preparing myself for the field of People/HR Analytics?

This has a little different scope than the article that Hendrik and I co-authored. That one was intended to have the context of 'within your organization'. This one intends to have the context of the field in general.

I would preface all that share below with an acknowledgement up front that- I consider myself but one voice in this field. Others may have had different journeys/ experiences that have led to their work in this area. So I leave it up to your judgement as to whether you find my comments useful in your situation and experience.

I think there are several things that help in preparing oneself for this field. They are listed in summary below, and are expanded on in the rest of this blog article.

1. Start with a good foundational definition.
2. Become familiar with the basic building blocks of People/HR Analytics.
3. Read, network, and get hands on practice.

Start with A Good 'Foundational' Definition

This is important for at least one primary reason.

Depending on how you define People/HR Analytics will determine what you pay attention to in your learning and what you concentrate on. If you look askance at any Google search on People Analytics or HR Analytics- you will find an incredible number of links on this subject including many books on this- probably too many in a lifetime ;-). Many will cover HR metrics and scorecards. These ARE important. To me personally, a lot of where we are today in People/HR analytics comes from those historical roots and are still part of the picture. If 'you' see People/HR analytics as 'just that' then that's what your concentration for preparation will be based on, and you might be tempted to leave it at that.

The problem with metrics and scorecards is that, in and of themselves, they don't necessarily guarantee any action or decision making (if warranted). They can be fundamental building blocks in People/HR analytics, but by themselves are an incomplete picture.

A more complete picture is found at what seems to be at the heart of any ‘analytics’ endeavor- regardless of context- being ‘data driven’. And being ‘data driven’ for a purpose- that of decision making and taking action where warranted. This is very much consistent with Dr. John Sullivan’s comments on People Analytics.

http://www.ceo-cf.com/wp-content/uploads/2014/06/07_How-Google-Is-Using-People-Analytics-to-Completely-Reinvent-HR.pdf

The general theme seems to be ‘HR ‘data driven’ decision making and management’.

Wanting answers to the business questions and their HR implications, or HR issues affecting the business bottom line and basing these on what our data is telling us -so that our decision and actions are as informed as possible, is at the heart of what People/HR Analytics is about and provides a ‘more complete’ picture/definition.

If you accept that definition, your concentration on your preparation will take you into HR metrics and scorecards- yes- but it will require you to prepare yourself in many other areas as well.

So having good foundational definitions is important because it will affect your subsequent preparation choices. It’s your choice. Currently, my opinion as a result is, that ‘data driven’ HR decision making serves as a great foundational base. It recognizes the historical roots of HR analytics, but also demands more of us going forward.

I share this because when any field is relatively new, terminology can often be thrown around inconsistently leading to a lot of confusion. Many years ago I read a comment/thought that said: any discipline will rise or fall based on the reliability and validity on which its observations are based. A sobering thought. I want People/HR Analytics to transform the way Human Resources management is conducted and practiced-that all our decisions come as a result of being data driven. Our organization deserve no less.

Become familiar with the basic building blocks of People/HR Analytics

If you accept the above foundational definition, then there are at least 5 major building blocks of skills to support development and preparation in this area:

- Knowledge of HR and HR functions and business processes
- Knowledge of Information technology with respect to data warehousing, data retrieval, and human resource information systems typical content
- Knowledge of business statistical analysis – particularly with respect to HR data, HR questions, HR problems, and HR measurement and metrics.
- Knowledge of the scope of what’s measurable in the HR context.

- Knowledge of the Data Science framework, and how to apply it.

Knowledge of HR and HR functions and business processes

As was mentioned previously- ‘analytics’ itself, regardless of context, is about being ‘data driven’ – data analysis with a purpose. People/HR Analytics is really ‘that’ but within the HR context.

To understand the HR context means to understand HR. And to understand HR, it means understanding HR functions and business processes. And if I can be permitted to play around with the words -it means understanding and seeing HR ‘functions’ as the HR ‘business processes’ that they really are.

Too much ‘traditional’ thinking has seen HR as only a series of business ‘functions’. Every major function in HR is a business process- complete with inputs, steps in the process, and outputs. Why is this important?

To be ‘data driven’ in HR means that we are collecting data for the purpose of ‘measuring’. Seeing HR as inter-related business processes gets us focused on at least one major a category of HR measurement. (This will be elaborated upon later.) As well, you can’t really understand ‘data driven’ HR unless you live and breathe an understanding of HR itself. HR ‘IS’ the context.

Knowledge of Information technology with respect to data warehousing, data retrieval, and human resource information systems typical content

People/HR Analytics is absolutely ‘dependent’ on HR information as its retrieval and its processing. As the risk of stating the obvious, with today’s technology the existence of HR information isn’t the problem. The problem is that we are drowning in HR data, because there is so much. So a knowledge of getting at the data and getting at it efficiently and processing it efficiently is paramount. This is where a knowledge of data warehousing and retrieval is important.

If you are lucky, in your own organization, you may be able to depend on others in IT to do this and be the experts here. Even if that is the case, you are expected to be the HR expert and be able to talk with IT in their language to get the job done. This means that to be effective in People/HR analytics you need to know as much as you can of this area which is ‘relevant’ to your role in HR Analytics. Usually, initial efforts in organizations in HR Analytics are ‘proof of concept’ or ‘skunkworks’ projects that are ‘informal’ and ‘off the books’ because no resources have been allocated for this at an early stage. In that case you may have to be in this role at a rudimentary level yourself. The rudimentary level will often be at ‘at least’ a SQL commands level for retrieval of information.

As well, you have to be familiar enough with your Human Resource Information System and its contents to know what kinds of HR questions can be answered by your HR data.

In any case, being able to ‘talk the talk’ and ‘walk the walk’ with the information technology dimensions of People/HR analytics, requires the building block of knowledge in this area.

Knowledge of business statistical analysis – particularly with respect to HR data, HR questions, HR problems, and HR measurement and metrics.

Another ‘at the heart of analytics’ is analysis, measurement, and statistics. This isn’t optional. If you want rigorous, robust, and informative ‘data driven’ decision making this is an essential part of the picture. (If you hate statistical analysis- get over it. This skill is absolutely essential to data mining and analytics.)

Too much of ‘analytics’ has the propensity for organizations to be satisfied with ‘cool graphics’ especially interactive ones. The same goes for drilldown and slicing and dicing of information. Don’t get me wrong. These ARE part of the analytics picture, but by themselves are ‘superficial’.

To really understand what is going on in your data, you must delve into statistical analysis. And I tend to recommend learning it with business examples.

- You need to understand your data, your world, your HR in terms of Independent variables and dependent variables- what you are trying to predict, and what you are predicting with.
- You need to know your data in terms of measurement- what are categorical, ordinal, and continuous variables, and how that affects the particular statistical analyses and algorithms you use.
- You literally need to see all of HR and HR information as measurement- seeing HR as business processes is ‘part’ of this.
- You need to understand that different statistical analyses, algorithms, and procedures are designed to answer different HR questions. So you must understand the lay of the landscape from what to choose from. You must also have the HR questions to begin with to drive the effort. And you need to know what types of statistical procedures answer what types of questions
- You must understand that statistical analysis and HR analytics isn’t JUST about ‘Predictive’ analytics but ‘descriptive’ analytics as well. Too much of what is written on analytics these days would give you the impression that ‘analytics’ is about predicting. That’s because it’s ‘cool’ and gets people’s attention. And yes IT IS COOL and IT GETS PEOPLES ATTENTION. But again, if you limit your understanding to just this view- it’s superficial. Statistical analysis and the resulting being ‘data driven’ can be as a result of descriptive analytics as well- descriptive statistics. And with regard to prediction,

another misunderstanding is to assume that statistical prediction is all about predict the future only. Wrong. Statistical prediction for categorical ‘dependent variables’ is to predict a best fit category for some new observation in a population. That’s current time not future. So prediction can be present and future.

- Interactive graphics have critical role to play in communicating to others what is going in the data, and in an effective way. But it’s the statistical analysis that helps separate the wheat from the chaff, separate what is statistically significant in the data and what isn’t. Graphics by themselves don’t do this.

To live and breathe statistical analysis you have to be familiar with statistical packages. There are many out there. SPSS and SAS are commercial packages that have been around for decades. They are good- but they cost. R has also been around for decades and its free- although its more command based and has a higher learning curve for some. I tend these days to prefer R because it’s free. Free helps, when you want to be empowered to learn analytics.

Knowledge of the scope of what’s measureable in the HR context.

There is a tendency because of the historical roots of HR analytics being partly steeped in HR metrics and HR scorecards, to assume that ‘that’ is it with respect to HR analytics and what we measure. But again if we limit our picture to just that, then once again its superficial and only part of the picture.

If we step back and try to get a more extensive lay of the land, we find that there are at least 3 major areas of measurement in HR, that can form the basis of analytics and being data driven. These are:

1. What is going on with the employees in our organization? Some of our HR decision making will be based on the employee picture.
2. What is going on in our HR operations? Some of our HR decision making will be based on how efficiently and effectively our HR processes. ““
3. What is going on in our HR decision making and policies itself? How do we make HR decisions? How good or bad are our decisions and actions? Some of our HR decision making will be based on this – if we want to be extensive data driven.

What is going on with the employees in our organization?

The best way to describe this category is that it is the traditional HR Metrics for the most part. Another way to think of HR metrics, is that it concerns itself with measurement of various aspect of the employee life cycle in organizations including but not limited to:

- Applicant counts
- Interview counts

- Hires
- Employee counts
- Training activity
- Grievance activity
- Health and Safety activity/injury
- Employee engagement surveys
- Terminations
- Exit Interviews

I have shared this link before, but a good starting point for understanding the kinds of things that can be part of traditional HR metrics is the following:

<http://www.hrma.ca/wp-content/uploads/2012/05/HR-Metrics-Standards-and-Glossary-v7.3.pdf>

When the above types of metrics are calculated (often with data warehousing technology to make it efficient so that we don't spend inordinate amounts of time preparing the data), then both predictive and descriptive analytic techniques can be applied to these metrics to answer all sorts of business questions. We want to be data driven, by letting what our data is telling us through proper statistical analysis, help guide the decisions and actions we make.

This is what most organization think of first when they think HR analytics. It is a good starting point. But just don't end it there.

What's going on in our HR operations?

Being 'data driven' should mean that we are data driven in our HR operations as well. All that we do in People/HR Analytics should add value to the organizations. How efficient and effective we are in conducting our HR activities - is one part of this.

Traditionally, and alternatively for that matter, this could be thought of as quality improvement/continuous improvement of HR operations- the voice of the customer. This is a whole other area of study even on its own. But we can be 'data driven' in how we carry out our HR business operations and activities. So this could be perceived as a legitimate part of the People/HR analytics part of the picture.

The primary ingredient or enabler of gathering this kind of information – is HR Request Tracking systems. You need to monitor every single transaction from inception to completion, cradle to grave, that comes into HR. and you need to monitor where it came from, who worked on it, when it came in and when it was completed.

Think of your HR business as being like some the global delivery businesses, who can track any package in real time all the time, both for its needs and the customer needs. You get the idea.

Tracking every request – helps you to be ‘data driven’ in improving your HR services to your employee who are your internal customers. In fact, you cannot have ‘data driven’ operation without the ‘data’. That should be painfully obvious. Does your organization do this?

What is going on in our HR decision making and policies itself?

What this is getting at is the ‘infusing’ of People/HR analytics into the actual carrying out of our HR activities themselves in real time. Being data driven in real time. The previous two categories might tend to have the flavor of after the fact. And that’s ok. But being able to use analytics in real time is even better.

Some examples:

- Taking features and characteristics of our known job descriptions and job classifications to predict the best fit of a new job description to a best fit job classification using People/HR analytics
- Taking our existing HR information on existing employees and terminations, along with engagement survey data, and exit interview data to see if there are any patterns in the data with respect to who leaves or who stays, and if so, predict likely future terminations before they occur. Or better yet through identification of patterns in the data, make changes in HR policies to prevent the turnover in the first place.
- Taking our existing HR information on absenteeism, and see if there are any patterns in the data. If so, adjust our policies and decisions where possible to discourage absenteeism more effectively.

In each case, its recognizing that infusing analytics in our decision making in real time is another potential part of the ‘total’ landscape of People/HR analytics.

The important lesson here, is that the application of People/HR analytics in many ways is only limited by your imagination. Don’t limit it, by thinking that its only traditional HR metrics (as important as these are).

I guess a final message in this section, is to review why ‘data driven’ is so important. In its absence, organizations still take action. They always have. The issue is ‘informed’ action. Sometime the problem is taking action when no action is needed. Other times it’s not taking action when action is needed. If we aren’t ‘data driven’ how do we KNOW?

Knowledge of the Data Science framework, and how to apply it.

So far we have been hitting on the principle of being ‘data driven’ very hard in this blog article- for good reason. It goes right to the DNA of what analytics is about.

With that in mind, knowledge of one other area is helpful, and I mean REALLY helpful. That area is the Data Science Framework.

One of the challenges in People/HR analytics is the issue of how do I both structure my thinking and my efforts/endeavors for maximum likelihood of success? I think the answer to that is part is to have a good framework to help guide the process.

Data Science, as a process, is that kind of helpful framework- because its intent is enable being ‘data driven’ in a structured way.

There are a number of books out there on the subject. Many are not necessarily written for the context of HR. You have to apply the framework to an HR problem.

Some of my previous blog articles illustrate this:

<https://www.linkedin.com/pulse/people-analytics-example-using-r-lyndon-sundmark-mba?trk=mp-author-card>

<https://www.linkedin.com/pulse/people-analytics-using-r-employee-churn-example-sundmark-mba?trk=mp-author-card>

<https://www.linkedin.com/pulse/people-analytics-r-job-classification-revisited-sundmark-mba?trk=mp-author-card>

One book that I came across that I found useful regarding R and Data Science was:

Practical Data Science with R by Nina Zumel and John Mount.

<https://www.manning.com/books/practical-data-science-with-r>

The basic steps in the framework from that source are:

1. Define a goal.
2. Collect and manage data
3. Build the model
4. Evaluate and critique model
5. Present results and document
6. Deploy model

The following is an elaboration of that from one of my previous blog articles:

1. Define a goal

As mentioned above, this means identifying first what the HR management business problem is you are trying to solve. Without a problem/issue we don’t have a goal.

2. Collect and Manage data.

At its simplest, you want a ‘dataset’ of information perceived to be relevant to the problem. The collection and management of data could be a simple extract from the corporate Human Resource Information System, or an output from an elaborate Data Warehousing/Business

Intelligence tool used on HR information. For purpose of this blog article illustration we will use a simple CSV file. It also involves exploring the data both for data quality issues, and for an initial look at what the data may be telling you

3.Build The Model.

This step really means, after you have defined the HR business problem or goal you are trying to achieve; you pick a data mining approach/tool that is designed to address that type of problem. With absenteeism as an HR issue, are you trying to predict employee with propensity to high absenteeism from those who aren't? Are you trying to predict future absenteeism rates? Are you trying to define what is normal absenteeism from that which is atypical or an anomaly? The business problem/goal determine the appropriate data mining tools to consider. Not exhaustive as a list, but common data mining approaches used in modelling are classification, regression, anomaly detection, time series, clustering, association analyses to name a few. These approaches take information/data as inputs, run them through statistical algorithms, and produce output.

4.Evaluate and Critique Model.

Each data mining approach can have many different statistical algorithms to bring to bear on the data. The evaluation is both what algorithms provide the most consistent accurate predictions on new data, and do we have all the relevant data or do we need more types of data to increase predictive accuracy of model on new data. This can be necessarily repetitive and circular activity over time to improve the model

5.Present Results and Document.

When we have gotten out model to an acceptable, useful predictive level, we document our activity and present results. The definition of acceptable and useful is really relative to the organization, but in all cases would mean, results show improvement over what would have been otherwise. The principle behind data 'science' like any science, is that with the same data, people should be able to reproduce our findings/ results.

6.Deploy Model

The framework presented goes a long way in structuring your thinking and activities and I have found it quite useful personally. At the end of the day you want to increase the likelihood that your People/HR analytics activities provides value to your organization. This framework can help.

Read, Network, and Get Hands On Practice

I said at the beginning of this article that I think People/HR Analytics as a pervasive field is just in its infancy. If your sense of things on that front is similar, then I think the following will help you prepare for this field:

Read as widely on this subject as your time permits. I think much of the ‘thinking’ in this area is still very fluid because it is a relative new field or discipline within the HR context (just my personal opinion). As you are reading, how is reading from one source jiving with what another source is saying? As you are reading, are you getting a lay of the land? Do you know what part of the picture they are talking about? Is what you are reading, expanding your knowledge or the field?

Network with others who have the same or similar passion for this field. This could happen in a number of ways including attendance at HR Analytics conferences, linking up professionally on LinkedIn etc. The network spurs additional thinking, ideas, and discussion which can move the organization forward.

Get Hands On Practice This of course can provide a dilemma if you don’t have access to HR data

If you are in an HRIS role that has legitimate access to your HR information, and you have your organization’s permission- problem solved. But if you aren’t, you may have to track down HR datasets on the internet that have been shared publicly. There aren’t many. And it may be fake data to illustrate the ‘how to’. If you don’t have access to HR data, try to practice relevant data mining algorithms and analysis on data from another context.

This will become a wider issue in the near future. For people to gain the People Analytic skills requires access to meaningful HR data- fake or not. If there is a minimum of data that is made available, that fully complies with privacy laws around the globe, I think this will limit the field’s ability to move forward.

Having the necessary tools, is not a barrier in this field. If you choose to learn R as your statistical tool – it is free.

Closing

I said in the introduction, that I was going to preface what I wrote with an acknowledgement that I am but one voice among many in this field. Others probably have different experiences or journeys from mine. And that’s totally valid.

If the sharing of this information, helps guide others in their People Analytics journey, then I am grateful.