

Building and Scaling a People Analytics Practice with Limited Resources: 10 Guidelines for Success

Author: Craig Starbuck . November 2021

We can do anything with time and money, but the reality is that both are often in short supply for an emerging people analytics practice that has not yet proven its worth to the business. I routinely interact with data-driven folks in the people analytics space, both newcomers and seasoned practitioners alike, and one thing is apparent: most struggle to advance beyond descriptive analytics – if they have made the leap from reporting to analytics at all.

My purpose in publishing this article is to share with the community at large some guidelines to facilitate progress along the people analytics maturity curve and ensure an indelible impact that is realized by business leaders and sustained for the long-term. The material presented is neither a comprehensive set of guidelines nor a formula for success in every context, but my hope is that it will inspire and assist teams struggling to make some traction.

(1) Think 'Pro Employee'

'Pro employee' thinking is addressed first and for good reason. I love people analytics and find great meaning and purpose in what I do. I believe this stems from a conviction that employees' lives are in some way improved as a result of actions taken in response to the insights uncovered. Whether it is shedding light on an area of the business struggling with work-life balance or identifying developmental areas of which a group of leaders may be unaware, people analytics ideally improves employee well-being and effectively, the success of the business. It is important to embrace a 'pro employee' philosophy, as newfound knowledge could also have damaging repercussions if shared with the wrong people or if findings are disseminated without proper instruction on how to interpret and take action (e.g., disparate impact).

One way to an error on the side of caution when considering whether or not to disseminate insights is to ask the following: "With this knowledge, could the recipient act in a manner that is inconsistent with our 'pro employee' philosophy?" If the answer to this question is not a clear "no", discuss with your HRBPs, OD, and/or other HR leaders what was uncovered, and together, determine how best to proceed. The decision may be to not share the findings with the intended audience at all or to develop a proper communication and training plan to ensure there is consistency in how recipients receive the insights and take action in response.

(2) Coalesce Around a Shared Vision

While building and scaling a people analytics function requires deep technical expertise, it also requires strong leadership. We all want to experience meaning and purpose in our work – to feel that we are working for something bigger than ourselves. Therefore, we need a clear vision of what could be to propel us forward. Coalescing the team around this vision is critical, though not necessarily easy. It requires leadership.

Change is hard, and expecting every member of a reporting team to be ecstatic about venturing into terrain which they likely know little about may be unrealistic. The reality is that some may not be interested in analytics and where the team is headed, but be sure to treat everyone with dignity and respect. If you are serious about advancing on your roadmap, do what you can to help the individual(s) move on to a role for which their skills and interests are better suited and then move forward. Without a team of people who are committed to the vision, there will be a degree of continual and unnecessary resistance to change, and that will only hinder progress.

(3) Never Compromise Quality for Greater Velocity

Everyone is familiar with the old adage: garbage in, garbage out. All it takes is one instance of compromised quality to damage your reputation and cause consumers of your insights to view all findings as a suspect. Be sure quality is atop your list of core values, and guard your team's reputation at all costs. If users do not trust the insights provided, they will question what they receive which may, in turn, result in requests for additional reports to 'tick-and-tie' in order to gain confidence in the data. This is wasteful to both you and your user community.

To be clear, by 'quality I am referring to results, which is dependent on data integrity in the source systems, proper data preparation steps, and many other factors. The majority of the data scientist's time is spent on data preparation (data collection, cleaning, and organizing, building training sets, mining for patterns, refining algorithms, etc.). If tight controls do not exist within the source application to support data integrity, data preparation efforts can only go so far in delivering reliable and valid findings. It is often the analysts who identify data integrity issues due to the nature of their work; therefore, close relationships should be formed with source application owners to put into place validation rules to proactively preclude the entry of problematic data or at the very least, exception/audit reports to identify and address the issues soon after the fact. Close relationships with application owners can also facilitate application changes that will help reduce laborious data preparation steps. For example, if the source application collects information on employees' education via free-form text entries, it may make sense

to discuss populating a selection list of schools to free analysts from having to scrub “U”, “University”, “Univ.”, etc. to produce a clean, unique list. These enhancements can save you significant amounts of time down the road.

While the allure may emerge to curtail important data preparation steps or make incorrect assumptions about the quality of data in the source and jump into modeling prematurely, resist the urge to take shortcuts. Ensure experienced analytics professionals are involved in the initial development of a roadmap so that decision-makers who may not be as familiar with the technical minutia are better informed when creating timelines. If leaders broadcast deliverables that are not realistic, it will likely result in dangerous levels of pressure being applied to those doing the analysis which will increase the likelihood of shortcut exploitation to hit milestones. Be methodical in your approach and ensure you are progressing commensurate with a coherent and practical analytics roadmap (descriptive >> predictive >> prescriptive). If quality falls to the bottom of the priority list, all other efforts are exercises in futility and the people analytics function will face a significant risk of failure.

(4) Automate, Automate, Automate

The role of automation in scaling a people analytics practice cannot be overstated. So many organizations haven’t made the leap from operational reporting to a proactive analytics focus because they are myopic in their thinking. With the technology and tools available today, I believe everyone should be continually thinking about automation. If you are performing a manual task on a recurring basis, there is a very high probability that the task can be automated; it is just a matter of whether or not it makes sense, which can be determined via a simple cost-benefit analysis. An introductory course in VBA may be all you need to build some macros and free yourself from all of those manual steps taken each week to deliver reports per your senior leaders’ specifications.

With respect to reducing ad-hoc report request volume, there is hope. By spending some time categorizing the report requests received over the last x months, and then building and automating reports and interactive dashboards to satisfy the majority of the needs, reporting responsibilities can be quickly and significantly reduced. Get serious about viewing every report request through the following prism: “What can we implement to ensure we never again receive a request for this data?” To be clear, I am not an idealist who believes that all ad-hoc report requests will eventually disappear, because that likely will never happen. However, this helps to frame a proper perspective of the report request queue so that analysts are not simply living in the moment, reacting to the deluge of data requests. Proceeding from this angle not only requires the development of an appropriate reporting solution but an effective change management process as well (e.g., canned responses that inform requestors on how the data can be obtained without the assistance of reporting resources).

Change is hard but if you have 30 different versions of an employee changes report, it may be time to build a single report that returns the superset of data from all 30; then, educate users on the filter and hide features in Excel to remove unwanted rows and columns. Having support up the chain is instrumental in addressing the pushback when decommissioning the superfluous versions. Rest assured, over-customizing and maintaining many versions of the same report will only hinder progression along the people analytics maturity curve.

(5) Cultivate a Learning Organization

It would be an unfortunate miss to omit from this article the talent gap that needs to be addressed when transitioning from reporting to analytics. There often is not a team of experienced data scientists sitting idle and prepared to hit the ground running with a set of requirements, nor is there usually the budget to hire them. Most commonly, there exists a leadership opportunity (one that I find exciting) to augment the skill set of the existing reporting team. If the members have an interest and possess initiative, the skills can usually be learned. Consider starting up a weekly “tech talk” with an environment that is safe, informal, and highly interactive to develop the requisite skills your team needs to succeed. Not only is this a cost-effective alternative to formal training, but it is also a great team-building exercise.

Change is inevitable but in our field, the rate of change is increasing exponentially. People analytics leaders are pushing the envelope daily to advance the discipline. To this end, I believe a key to success is continual learning – regardless of the current skill set of the team. I would much prefer to possess an understanding of more tools and techniques than I have time to deploy than to be approached with a high-stakes project but lack the know-how to adequately undertake it. I also believe it is important to expand horizons and continually add to our knowledge base; the moment we stop, we are behind. There are many useful statistical learning techniques (e.g., support vector machines, survival models, clustering, NLP, etc.) that are not widely employed in the social sciences but are being introduced by newcomers with computer science, statistics, and other technical backgrounds. It is important to understand the utility of each.

To be sure, those infographics being circulated on the characteristics of the superhero data scientist are unrealistic if the implication is that they exist within a single individual. No one person can be exceptionally proficient at all things, and we need to team with others with specialized skills to offset our shortcomings. Consider scheduling an analytics collaborative on a recurring basis, wherein resources from the organization’s various analytics pockets come together (physically or virtually) to share current and upcoming projects. I have made many connections this way and thanks to the diversity of thought, I gained exposure to new ways of looking at problems. As an added benefit, there may be synergies that

can be realized due to breaking down the silos. So, make friends with the marketing analytics folks down the hall and consider the portability of their work on customer defection models to your work on employee attrition risk models.

(6) Build Relationships with Key Decision Makers

People analytics teams must position themselves as strategic business partners in order to have the intended effect on the organization's success and secure a seat at the table. When you begin having business leaders approach you to collaborate on how to obtain insights of a non-trivial nature vis-à-vis their people (rather than sending a report request so that they can work on it themselves), you have likely turned the corner and are succeeding in your plans to deliver real value to the business. This requires meeting regularly with business leaders to understand their challenges and identify viable initiatives to aid in overcoming them – a shift from a reactive to a proactive orientation. It also entails keeping abreast of industry trends and anticipating future needs to bring to these meetings relevant insights to inform future projects.

(7) Move Beyond Excel

You may be wondering why I am suggesting an upgrade from Excel, and effective skillsets, in an article intended to assist those with 'limited resources. I have included this because in most cases Excel is inadequate when moving beyond descriptive analysis, and the software I am recommending is free. While many commercial-grade analytics toolsets are very costly, R is an open-source statistical software package that can be downloaded free of charge. It is incredibly powerful, and there is a package for just about any statistical technique you wish to utilize. It is also widely used in highly regulated environments. What I like most is the actively engaged user community, from which I am continually learning.

Everyone learns differently, but I will offer some resources that assisted in my understanding of statistical learning techniques and their application in R in case readers wish to follow a similar path. To be clear, the path outlined here does assume a fairly solid foundation in statistics and probability as well as computing. If you lack a foundation in stats, I recommend taking a course that covers both descriptive and inferential statistics to become comfortable with measures of spread and central tendency, variable types, measurement scales, hypothesis testing, sampling methods, probability distributions, correlation, simple/multiple linear regression, and analysis of variance. If you lack a computing background, I suggest at least an entry-level programming course become familiar with basic computing concepts (variable declarations and assignments, arrays, loops, conditional and iterative statements, etc.). You do not need to be an expert coder to execute data analysis packages in R, but it will be helpful to understand the basics. A course in linear algebra will be helpful as well, but it is not required to learn R. With a

proper command of stats and computing, there are four books that I found instrumental in building a foundation in data science techniques and deploying them within the R environment:

1. **Multivariate Data Analysis** - *Authors:* Joseph Hair, William Black, Barry Babin, and Rolph Anderson
2. **Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking** - *Authors:* Foster Provost and Tom Fawcett
3. **An Introduction to Statistical Learning: with Applications in R** - *Authors:* Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani (A free download of the PDF is available at <http://www-bcf.usc.edu/~gareth/ISL/>)
4. **Applied Predictive Modeling** - *Authors:* Max Kuhn and Kjell Johnson

Multivariate Data Analysis was a required text for my Ph.D. and marked my first foray into more robust social science methods such as hierarchical multiple regression, logistic regression, factorial ANOVA, principal components analysis, and structural equation modeling. This is an excellent text, of which I have a highly favorable opinion. The second text, Data Science for Business, was one I read to gain exposure to key data science concepts beyond what is conventionally employed in social science research (e.g., overfitting problems, supervised vs. unsupervised learning, cross-validation, ROC curves, etc.). This text is more conceptual than technical and easy to comprehend. With respect to the latter two texts, I found these quite helpful both in expanding my repertoire of techniques as well as applying the methods covered in the first two texts within the R environment. Though these texts are math-light, depending upon the strength of your statistical foundation most will find that they are not easy reads.

In addition to these texts, I learn a lot from www.r-bloggers.com as well as the LinkedIn group, “The R Project for Statistical Computing”. The community is very willing to share and assist others with troubleshooting code. You might begin in more of a spectator role and as your knowledge increases, transition to an active contributor on the threads. There are many other useful online resources as well, which I hope readers will add.

If you learn best from education of a more formal variety, check out the panoply of options compiled by KDnuggets: <http://www.kdnuggets.com/education/index.html>.

(8) Compute Smarter Variables

Level one people analytics tends to utilize only the delivered fields from the HRIS (e.g., location, job profile, org tenure, etc.), but a good next step is to derive smarter variables from these fields. These can

then be used to slice and dice turnover and engagement data differently, use as inputs in attrition risk models, etc. Below are some ideas to get you started:

- Number of jobs per unit of tenure (larger proportions tend to see greater career pathing)
- Office/remote worker (binary variable dummy coded as 1/0)
- Local/remote manager (binary variable dummy coded as 1/0)
- Hire/Rehire (binary variable dummy coded as 1/0)
- Hired/acquired (a proxy for culture shock effects)
- Gender isolation (ratio of employee's gender to number of the same within immediate workgroup)
- Generation isolation (comparison of age bracket to most frequent generational bracket within immediate workgroup)
- Ethnic isolation (ratio of employee's ethnicity to number of the same within immediate workgroup)
- Difference between employee and manager age
- Percentage change between last two performance appraisal scores (per competency and/or overall)
- Team and department quit outbreak indicators (ratio of terms over x months relative to average headcount over x months)
- Industry experience (binary or length in years)

At this point, it would not hurt to reread the content related to the first guideline: Think 'Pro Employee'. While people analytics ideally improves the lives of employees and generates solid returns for the business, providing 'interesting' insights to managers without the proper training on how to interpret and act on them can do more harm than good. In fact, it may not be appropriate to share the finding with managers at all. If it is found that generationally isolated employees, for example, are less engaged or more likely to term than those who are not, this finding may be best used to tweak the curriculum for leadership development programs to raise awareness, with supplementary instruction on how to apply the research findings in leaders' particular contexts. If there is any question as to how findings may be acted upon by managers, always discuss with HRBPs, OD, and/or key HR leaders before disseminating.

Also, remember to compute variables consistent with a need (e.g., is there reason to believe generationally isolated employees are more likely to term?). There is certainly a time and place for undertaking data mining initiatives with no a priori theories about what may be uncovered; however, more often than not, our efforts should be tied to specific hypotheses the business needs to be tested, which have sound theoretical underpinnings.

(9) Transcend 'HR' Analytics

I personally prefer labels such as people analytics, workforce analytics, or talent analytics over HR analytics when describing the discipline which Bersin by Deloitte defines as “the use of measurement and analysis techniques to understand, improve, and optimize the people side of the business.” I see the distinction being more substantive than semantical. HR analytics seems to suggest a practice that seeks answers to HR’s questions and perceived challenges (which may or may not be aligned with what the business cares about) rather than proactively undertaking projects that support key people’s challenges limiting the success of the business. As well, HR analytics tends to leverage only HR data sources rather than integrating data from the myriad of often underutilized sources available within an organization. People analytics, on the other hand, ideally leverages many data sources, HR and non-HR, which house potentially valuable data points throughout the lifecycle of each employee. People analytics seeks to support non-trivial challenges the business needs to overcome by converting ambiguous problem statements into clearly defined research questions and testable hypotheses. Per these definitions, HR analytics is positioned as a relatively elementary stage on the people analytics continuum.

The failure to test HR’s deep-seated assumptions is precisely why employee engagement has received so much controversy over years past. It has long been accepted as a fact that engagement has a positive impact on a company’s bottom line but very little has been done to substantiate this claim. Testing this theory, of course, requires data beyond what is usually housed in HR systems. Engagement aside, people analytics initiatives often position ‘success’ or ‘performance’ as an outcome variable (e.g., identifying why top performers leave, predicting on-the-job performance pre-hire, etc.), but what makes an employee a top performer is a loaded question. While people analytics utilize the HRIS, onboarding survey data, engagement results, exit survey data, and other data that often lives in HR, it should also leverage sources that can provide a more comprehensive and valid measure of ‘success (e.g., customer leads by the employee, the success of customers served by the employee, sales by employee, etc.). The additional predictors also aid in separating the signal from the noise (e.g., economic factors explaining variance in sales, customer retention issues, etc.). While valuable insights can be captured from performance appraisals, I believe this type of subjective data is best married with hard numbers such as sales figures when modeling employee success for customer-facing roles. For other roles, performance is of course more difficult to quantify and the way in which success is defined is highly situational. In any case, more data sources can only increase the robustness of the measure and provide more broadly accepted conceptions.

The effective utilization of many data sources requires collaboration with other pockets of analytics across the organization as well as working closely with SMEs to understand the data and leverage it appropriately. With the utilization of additional data sources comes a significant risk if one believes he/she can fully understand the intricacies of data from all systems. So, be sure to make some friends across departmental lines and ask for help before attempting to incorporate data about which you are relatively unfamiliar.

In addition, I encourage thinking about everything as another potential data source – apart from the conventional data sources such as CRM, ERP, etc. The world is awash with data and likes it or not, data is collected on just about everything we do today. So, consider what is relevant and appropriate (always error on the side of caution when unsure and think ‘pro-employee’), and be creative. Sure, the data generated from the source may be unstructured and require munging; however, the value it may provide may very well warrant the effort of converting those free-form comments into dummy-coded quantitative variables for use in models. Also, that analytics collaboration may very well lead to knowledge of data sources you never knew existed.

One word of warning with respect to strategies for collecting additional data: caution should be exercised when incenting people for it. It is plausible that many won't feel comfortable providing an honest assessment of what is being requested (e.g., perceptions of one's direct superior) but since they really want the free month of casual days, there exists a risk that they will open a survey, provide a straight line or socially desirable responses, and invalidate the results. At least in the case of survey research, I am not an advocate of incenting participation to achieve a higher response rate unless the data requested is unequivocally non-sensitive in nature.

(10) Facilitate and Demonstrate the ROI

Last, but certainly not least, is the need to facilitate and demonstrate the impact of people analytics. Implicit in demonstrating a return is that there is a measurable impact to be shown; therefore, it is important to first highlight the importance of facilitating actionable results.

If we believe a project will not result in findings that are actionable, why should we conduct the project in the first place? That is obvious, yet the extent to which the findings are actionable is not always easily ascertained upfront because of the uncertainty around what will be found. However, the following questions may prove helpful when designing the project: “If this hypothesis is supported, what actions can be taken? If it is not supported, what actions can be taken? If we measure *X* on our next pulse survey, what can we do if the score is unfavorable?” If the answer to these questions is, “nothing, but it would be interesting to test/measure”, focus on something that is more likely to benefit the business.

One way to facilitate greater impact is ensuring decision-makers fully understand the findings being presented so that there is clear direction on how to proceed. If there is ambiguity around what was found due to using technical nomenclature that is confusing for the target audience, the potential for impact is limited. Replace “variable *X* explained significant variance in variable *Y*” with “*X* drives *Y*” or “*X* influences *Y*”. Effective storytelling with data is key, yet often difficult for highly technical folks.

Also, it may be tempting to employ the most flexible and sophisticated techniques in your toolbox but if the results produced by them are difficult to explain to a non-technical audience, you may want to entertain an alternative if the project does not warrant exceptional robustness and precision (e.g., for classification, most will find decision trees more intuitive than the log odds from a logistic regression model). Furthermore, the importance of effective visualizations in data storytelling cannot be overstated. While you may be comfortable with numbers, there is a significant risk that the meaning you wish to convey will be lost in translation without visuals. Always know your audience and tailor the presentation accordingly.

With respect to demonstrating the ROI, I know of very few organizations that are doing this. We should not assume that our efforts are delivering what we presumed they would at the outset; we should validate the outcomes empirically. If turnover was an issue and action plans were rolled out as a result of attrition risk models, follow up to see if the treatment had an effect on a turnover by comparing rates at time 1 and time 2. If it is leadership training outcomes that are of interest, track changes to scores on relevant engagement drivers (if surveys are administered frequently) to gauge the size of the effect and whether it is sustained over time.

The ability to quantify the impact of our efforts can go a long way when justifying the need to grow a people analytics function or when programs are on the chopping block and their impact can be substantiated. As well, insight into the efficacy of actions will inform future action plans when analytics reveal similar issues that warrant attention. Be sure not to do something purely because other organizations are doing it or because it is currently the ‘popular’ thing to do in people analytics circles. Take on initiatives because they have the potential to generate insights that are important to the business; after all, supporting the success of the business is why the people analytics function exists, and we should never lose sight of that.

Please join the conversation: I look forward to the dialogue around the material covered here. What are some of the most unconventional data sources you have leveraged for people analytics projects and did the resulting insights prove significant? What are some ways in which you have demonstrated the ROI of your initiatives, and what was the response? What are some additional guidelines that may position a people analytics function for success?

The post "Building and Scaling a People Analytics Practice with Limited Resources: 10 Guidelines for Success" was first published by Craig Starbuck, Ph.D. *here* <https://www.linkedin.com/pulse/building-scaling-people-analytics-practice-limited-10-craig/>

About Craig Starbuck, Ph.D.

I am passionate about transforming people's data into information and insights that issue clear imperatives for organizations. I am driven to support organizations in making data-driven decisions about their workforce and enhancing the employee experience. I find intrinsic meaning in leadership – coaching and mentoring people to facilitate growth and unleash hidden potential. My appreciation for workplace diversity and its role in driving innovation and gaining a sustainable competitive advantage is reflected in the diverse teams I have built within multiple organizations. My goal is to enable organizational transformation through a deeper understanding of human capital and the conditions under which performance is optimized.

Personality Assessments -DiSC: DMBTI: ENTJSDI: Red (Performance)Enneagram: 1) Achiever, 2) Reformer, 3) Challenger

<https://thehumancapitalhub.com/articles/building-and-scaling-a-people-analytics-practice-with-limited-resources-10-guidelines-for-success>