Deloitte.

Don't go on autopilot
Site selection in an age
of ubiquitous data

In site selection, data choices drive outcomes. Data without insights or experience can yield the wrong results.

We are in a time of easily accessible location data, with more sources, maps, and visualizations available to everyone, often on a free basis. For companies evaluating new locations for their operations, or considering a change in footprint, it seems to be the Golden Age of Data. Wages, demographics, real estate, tax conditions, and other location factors are available to all. With smart phones and tablets, it is increasingly easy, with just a few dabs and swipes, to pull information on every county and major city in the US. Data availability makes it possible for anyone to make location decisions.

Unfortunately, while the sources of location data are numerous and readily accessible, not all data is of equal value or accuracy. In addition, when it comes to screening locations for a particular investment, the experience to set the appropriate boundary conditions or screening thresholds is essential to narrow the field to locations that can support a thriving operation. Furthermore, understanding which location variables make a difference and which do not requires a deep understanding of the criteria that drive success for any given asset type. All of these complicating factors can often lead what might otherwise be a straightforward screening process to very different results depending on the data sources chosen, the data elements analyzed, and the boundary conditions used to perform the location analysis.

Prior to engaging in any site selection process, a hypothesis-based approach that defines the key success factors for the operation is essential. These success factors should in turn drive the data that should be

leveraged in the screening process. Often, the data inputs used to conduct a location analysis come from government agencies, such as the Bureau of Labor Statistics, or the US Department of Labor. Other common sources include the US Census, state and local governmental data, and even crowdsourced information and other free online resources. These sources can often have highly variable levels of accuracy. Even in the case of typically reliable data from a source like the BLS, there are important assumptions that need to be accounted for in a location evaluation.

First and foremost among these caveats is that it is important to understand the coverage and level of detail in any given piece of data that is being analysed, namely the geographical jurisdiction (CBSA/MSA, city, county, state, etc.) and for occupational data, whether the data is for major occupational groups, individual occupations, or groups of occupations. Second, the establishment of boundary conditions, or screening thresholds, is critical to realize results that are relevant to a particular site selection evaluation. Even slightly different boundary conditions (setting a minimum labor force size at 100,000 vs. 50,000 for example) can result in vastly different outcomes. Finally, the data itself needs to be sourced thoroughly to vet its accuracy and timeliness – there are numerous sources available in the public domain that rely on outdated or unreliable information. A test of "data quality" is essential to ensuring that the results of a location screening are based on accurate data.

"This mission is too important for me to allow you to jeopardize it." HAL 9000, from *2001: A Space Odyssey.*
Convenient data does not always equal useful data.

Questions to ask about location analysis data:

- Is the source reliable?
- Is the data consistently collected?
- How recent is the data?
- Is it the right data: Does the data predict our success factors?
- Have we set the right boundary conditions when using the data?

To illustrate these distinctions, we created a screening model for a potential hypothetical project scenario: A back office support center for a life sciences business. This center will have approximately 250 employees engaged in a mix of finance and IT activities along with a pharmacy support staff, primarily comprised of licensed pharmacists. The first model uses purely public data from a variety of sources, while the second model utilizes a mix of public and subscription-based data. The results of the screens are found below:

Table 1. Screening results using generally available, public data

Top metro areas (out of 395)	Labor force size	Number in business and financial occupations	Unemployment rate (%)	Number of bachelor's degrees awarded (annual)	Salary index for business and financial occupations
Orlando, FL	1,079,670	54,370	4.40	448,045	0.90
Albuquerque, NM	370,270	19,040	5.70	185,994	0.88
Columbia, SC	356,160	17,890	4.90	165,958	0.79
Tallahassee, FL	158,760	16,570	4.60	83,691	0.74
Salem, OR	148,920	8,510	5.60	61,556	0.85
Cedar Rapids, IA	139,720	7,180	3.60	54,519	0.86
Gainesville, FL	122,820	6,570	4.10	63,897	0.81

Sources: US BLS OES Survey, US Census, NCES Data
Parameters (>100,000 Labor force, 5% of MSA employment in business and financial occupations, >3.5% unemployment rate, >50,000 bachelor's degrees awarded, and <=0.90 salary index)

Table 2. Screening results using both public and subscription data

Top metro areas (out of 395)	Labor force size	Computer programmer LQ	Accountant LQ	Pharmacist LQ	Eri salary index	Relevant graduates
Tampa, FL	1,398,132	1.25	1.04	1.22	0.98	1,395
Pittsburgh, PA	1,234,266	1.32	1.02	1.06	0.98	1,837
Indianapolis, IN	1,047,448	1.20	1.03	1.06	0.99	1,321
Buffalo, NY	559,943	1.44	1.07	1.09	0.99	2,965
Omaha, NE	487,655	2.20	1.13	1.35	0.97	1,150
Sioux Falls, SD	140,300	1.09	1.67	1.66	0.94	1,196

LQ: Location Quotient is a ratio that allows an area's distribution of employment by industry to be compared to a reference area's distribution, in this case the U.S.

Sources: US BLS OES Survey, US Census, NCES Data, ERI, ESRI
Parameters (>100,000 Labor force, >1.0 Location Quotient for each occupation, <1.0 ERI Salary Index, >1,000 Relevant Graduates)

The choices of data elements and boundary conditions are critical. Many users of site selection data don't know which factors will drive operating success.

Clearly, the outcomes of each set of data screens are vastly different. In fact, even though the same basic location factors, talent availability, and labor costs were used in both screens, there is not one metro area that is common in the results. This has everything to do with the type of data and the boundary conditions set for each.

Knowing what drives a successful location for an investment is critical. It is important to assess whether the results emerging from a screening process align with the hypothesis established at the outset of the process, and whether specific results either confirm or refute the hypothesis. In this example, several of these communities could potentially be viable options for such a center, but determining which would merit field investigation and due diligence is where experience and judgement matter. It is imperative that the results be tied back to the hypothesis established at the beginning of the screening process to determine whether the results make logical sense.

Understanding what type of data to use, what boundary conditions to set, and then how to interpret the results is a critical part of the site selection process; entrusting location decisions to an uninformed mix of data is perilous.

We have found over time that when the wrong location data is used, or if the right data is used with the wrong boundary conditions, the resulting screen of locations is often suboptimal. This is particularly the case when large groups of skill sets are used in the screening process (i.e. Business and Financial Occupations) instead of a more nuanced look at individual skill sets (i.e. Financial Analyst and Accountant), or a broad measure of education (bachelor's degree output) is considered instead of a review of targeted degrees relevant to a project. These are but two examples of why experience matters in the site selection process, especially in setting up the screening criteria which helps guide the process.

Locating easy to find information online is not necessarily the answer when making critical decisions that could affect the company's future – careful vetting of that information is critical. That's why, in our opinion, it's not quite time to only rely on easily accessible data, but rather a combination of the right data with experience and insight to drive the right decisions.

There is no substitute for judgment and experience in validating the results from a site screening process.
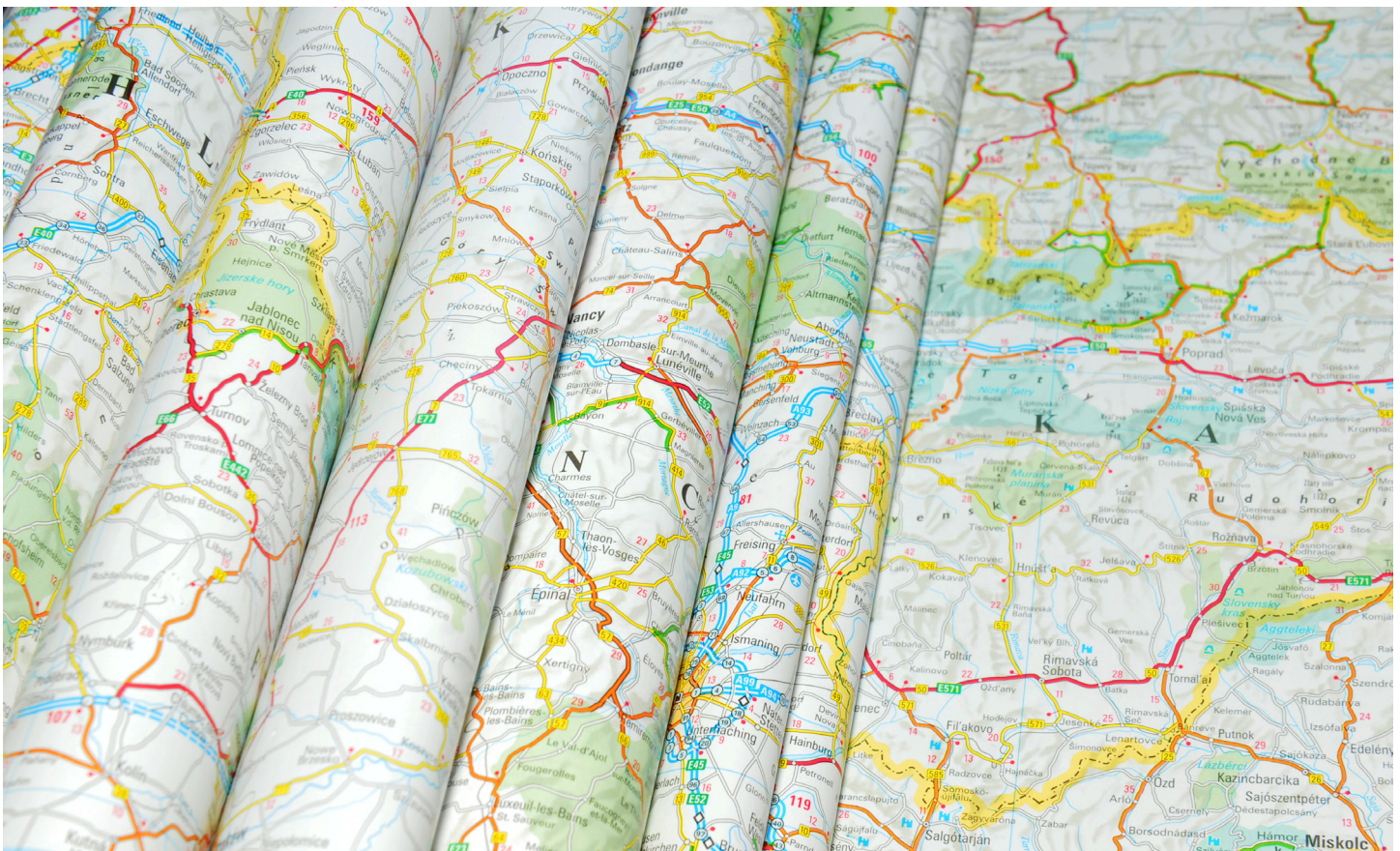
Contacts

Darin Buelow
Principal
Deloitte Consulting LLP
dbuelow@deloitte.com

Jayanth Iyengar
Manager
Deloitte Consulting LLP
jiyengar@deloitte.com

Seth Balkanyi
Manager
Deloitte Consulting LLP
sbalkanyi@deloitte.com

Deloitte.