

ALGORITHMS, ASSUMPTIONS AND IGNORANCE

JUST HOW RELIABLE ARE THE NUMBERS IN YOUR MARKET ANALYSIS?

Michael Stumpf, AICP, CECD – Principal, Place Dynamics LLC

Economic developers are increasingly interested in developing the retail sector in their community and seek to understand their prospects through a market analysis. Planners are also embracing market analysis as an integral preface to new plans, especially for corridors, neighborhoods, and business districts. Of course, developers will also seek out a market analysis in the early stage of a project. This has created a huge opportunity for data vendors who provide inexpensive “market reports” purporting to show the demand, existing sales, and leakage within a trade area. These reports make it easy to provide numbers, but unfortunately, those numbers are likely not very good.

The market reports provided by vendors have been a windfall for planning and economic development consultants, most of whom have little or no experience in market analysis. It is a simple matter, though, to download a report and repackage the results as a market analysis that can be given to a client.

Limitations of the methodology used to generate these reports and issues of data quality are little understood by most users, including many consultants. Error is introduced and compounded with each step in the calculations, and can lead to some wildly inaccurate results. This is especially true in small markets such as rural communities and downtown or neighborhood business districts.

While no estimates are perfect, the best outcome will be achieved with a custom model using local data and assumptions that can be modified based on first-hand observations and market knowledge. Aside from this, there are still considerations of how the numbers are interpreted, the impact of competition, and the potential to expand the market.

HOW DO RETAILERS SELECT SITES?

It may be informative to consider how retailers are conducting their site searches. Today’s chain retailers use a highly sophisticated process to select sites for expansion. If the kind of analysis provided by the vendors is used at all, it is only an initial screening tool. Instead, these retailers are using data collected internally from customer in-store and internet

transactions, loyalty programs, and other sources to build complex market models.

In its most basic form, the data can be used to pinpoint the locations of existing customers within a market. But the chains have also associated customer sales and local demographic characteristics, creating detailed matrices that can be used to quantitatively evaluate *their* specific market potential at any possible site.

This raises a question: If the chains already have better numbers, then why conduct a market analysis? Two obvious answers are to inform the user, whether that is a community, business district, or developer, and to provide information to non-chain businesses. Even more importantly, the market analysis should identify the information that is not known by the businesses targeted for recruitment. Examples might include visitor traffic, anticipated growth, or competitive factors that do not show up in the data.

UNDERSTANDING PREDICTIVE ANALYTICS

Predictive analytics uses information about consumers and behavior to predict future spending patterns. Essentially, if the average consumer purchases X dollars worth of a product, we would expect the aggregate purchase in the market to be X times the number of consumers in the market. This is further refined by adjusting the amount spent by characteristics of the consumer such as age, race, income, type of market, and other variables.

Consider a proposition: Two homes of a similar size and age, on equally-sized lots in the same neighborhood, should have more or less equal value. On the face of it, this seems like a reliable statement and is the basis for the “fair market value” generated by websites such as Zillow.com. These sites rely on basic assumptions in developing algorithms to calculate expected home values based only on data from sources like recent home sales in the vicinity.

Realtors rightly point out that every property, neighborhood, and market is different, and simply plugging numbers into a formula is not a reliable way to determine a price. What about other factors not apparent in the data? What if one home is beautifully restored while another has not been updated since the 1970’s? What if one is next to a conservancy while the other is on a four-lane road in an industrial area? While obvious to us, the computer cannot “see” these differences and will suggest similar values for the homes, significantly over- or under-estimating the true value of the home.

A recent article in the Wall Street Journal¹ tackled this issue, noting that the home values these sites determine could vary significantly from true values, and could also swing widely from month to month depending on factors such as new sales. Assumptions on which the models rely are only a part of the problem. Data may also be unreliable, particularly at the local level where there may be few recent sales, or where some variables may not be reported. Unfortunately, the typical person visiting one of these websites does not understand the methodology and its inherent flaws. The result is that pricing decisions made by both buyers and sellers relying on these estimates are often out of line with the market.

But what about retail market analysis? Several companies have been providing data-driven market reports for years. At a low cost, developers, planners, economic developers and others can purchase a report that claims to offer analysis of local demographics and growth, market potential, estimated sales within the market, and demand for retail by sector. Almost every consultant, from national firms that tout their automated report generation to local planning firms, are simply repackaging the numbers generated by these algorithms.

These market reports have been around for so long that they are seldom questioned. But like the home value websites, they should be. While each provider has their own methodology and may use differing raw data sources, the analysis ultimately rests on assumptions that are untested against the realities of the local marketplace.

WHY MARKET REPORTS GET IT WRONG

Errors can be introduced into the analysis in predictions of market demand and estimates of existing sales. Additionally, a



Two homes, practically identical and adjacent to each other, can have very different values. When analyses rely solely on numbers without any “ground-truthing” they are bound to generate unreliable results.

good deal of experience is needed to correctly define the market area and interpret the results.

Estimating Market Demand

Estimates of market demand are based on demographic information collected by the Census Bureau through the Decennial Census and the American Community Survey. This data is projected to the current year and then usually for one additional five year increment. Opportunities for errors occur in several places either due to problems with the data or reliance on past trends that no longer hold true.

- Missing populations. Population counts in the source data can be over- or underinflated for many reasons. One of the most common is an undercounting of the population in the Census. This can also happen in the methodology that allocates portions of the population to trade areas.
- Growth or gentrification. Many areas experience rapid changes that are not captured by the data sources except every ten years through the Census. Because the changes occur at the neighborhood level the error will be most pronounced in analyses of small areas. Examples include greenfield sites that become developed, and older built-up areas that may be redeveloped more densely or undergo gentrification changing the market demographics.

The oil fields of North Dakota offer a great illustration of this problem. The population of this area shrank for decades, and projections based on this trend showed a continued decline. In reality, oil exploration that ramped up in the mid 2000’s was drawing about 40,000 workers into the region. Projections that see this increase will now predict continued explosive growth. But over the next decade as labor-intensive oil exploration gives way to well servicing, there will be much less demand for workers and the population should start to decline.

- Under-reported income. Over the past decade or so there has been a great deal of research into under-reported income, especially in poor urban areas. Under-reporting

¹ The Fuzzy Math of Home Values, November 14, 2011

Case Study – Alamogordo, New Mexico

Alamogordo is a great example of how a vendor market report will generate inaccurate results. Situated along a lonely highway the Spanish Conquistadors named the *Jornada del Muerto* (or *Journey of Death*), isolated by mountains to the east and west, and hours from the nearest larger communities north and south, it might seem like an easy place for which to conduct a market analysis. But a number of factors complicate the issue:

- The region has an undercounted Hispanic and Native American population;
- Population growth is accelerating as the city is rapidly becoming a retirement destination with an identical climate, but lower costs than places like Tucson;
- Force realignments will result in 1,500 military personnel and their families being transferred to the air base five miles south of the city;
- The base is also home to the German Air Force's training operations, a unique demographic group 1,200 strong;
- There are over 7,000 vacation homes in the trade area, an astounding number that hints at the scale of tourism as a contributor to retail sales; and
- A handful of new stores have been announced, with the most important of these being a new Albertson's opening a year after the market analysis was completed.

Missing any of these factors will cause significant errors in the analysis, but the reports generated by data vendors take none of them into consideration. The result is a gross underestimation of the market potential and sales projections that fail to account for known new businesses.

also occurs frequently in rural areas and among certain populations, such as some Native American communities.

- Special populations. Several unique population groups have characteristics that can be very different from the general population, yet most of the data vendors lump these groups in with the rest of the trade area.
 - Military personnel. Military installations are very diverse. Factors influencing potential spending include the base mission, distribution of personnel living on the base, and base facilities competing with businesses off of the base. Additionally, deployments may result in fluctuations in the trade area population.
 - College students. Not all college students spend the same way. Differences can be observed between spending patterns of graduates or undergraduates, those

living on-campus or off-campus, and even based on the kind of school (private college, flagship state university, second tier college, etc.). The data vendors do not make these distinctions. College students are treated the same as any other household, even when the students may be living elsewhere four or five months out of the year. Additionally, college students may have low incomes, but spend more than their income suggests, as they receive money from other sources like loans, grants, and parents.

- Prison populations. Trade areas that include a prison are going to have the prisoner population counted as part of the trade area population. Hopefully they are not leaving the prison to shop local businesses. Seriously, though, including the prisoners inflates the trade area population and simultaneously skews income levels lower, painting a very inaccurate picture of the market.
- Assumptions used in projections. Population projections need to be based on many assumptions, some of the most important being birth and mortality rates, employment and wage trends, and inflation. These can vary widely across regions, but the data vendors need to have a single, easily identified source for the assumption. While it may be preferable to use local birth rates, for instance, the vendors will compromise and use national, or at best, state-level statistics.

Where these kinds of errors exist, they can only be corrected by adjusting the source data, which is something that none of the automated processes allow. So, even before starting to estimate market potential, existing sales, and leakage, there are likely to be problems with the population estimates and projections on which these numbers will be based.

Another problem is that the vendors only provide an estimate of the resident market potential, or in other words, the demand generated by people living in the trade area. Additional demand can be generated by visitors and by people who work in the district, but live outside of the trade area. The contribution of these groups can be very significant, and leaving them out of the analysis is a serious flaw.

Estimating Existing Sales

After estimating the market potential, the second phase of the analysis generates estimates of existing sales in the market. These sales estimates are compared against market potential to generate the leakage report (gap analysis).

Again, there is a methodological question to ponder. Every one of the data vendors compares the sales estimates of *all of the businesses in the trade area* to the market potential in the trade area. Is this what you want, or do you want to understand the market share held by businesses in your study area? The latter approach is going to be more informative in understanding how the district functions in the bigger marketplace. The data vendors and the consultants who rely on them can't provide this insight.

The next issue to explore is the reliability of the sales estimates. Developing them requires an inventory of the businesses in the district and/or trade area. For this, the data vendors turn to existing databases supplied by others. Again, there is no on-site verification of the list, and concerns about quality are widespread.

In 2014 this issue was tackled by the Initiative for a Competitive Inner City², which conducted a test using one of the more popular business lists.

We visually checked nearly 1,600 businesses and found startling inconsistencies between the commercial database and the walking inventory. Only 65% of businesses were visually confirmed through the walking inventory, 30% were not found, and the remaining 5% were duplicate entries or had incorrect addresses. In addition, we discovered 380 additional new businesses that were not included in the database.

These findings are not inconsistent with our own experience and that of others who have examined them in detail. Locally-owned businesses (the ones found most often in downtown and neighborhood business districts) are the most likely to be missing from the database.

It can take a couple or more years for a new business to be included in the list, and just as long for it to be taken out if the business closes. Even when a business is listed, the data associated with it is often wrong. This is important as the number of employees is often used as the variable on which sales estimates are based.

Average sales by store type and average sales per employee are the most common statistics used by most market data services in their estimation of sales. The basic idea is that if a store has a given number of employees, it is possible to multiply that by the average sales per employee in the respective retail sector, and the result will provide a close approximation of actual sales at the establishment.

Once again, the average can be misleading. Costco and Walmart's superstores are classified in the same industry, where the average annual sales per employee is about \$300,000. Although both stores will have roughly the same number of employees, in reality, Costco receives about \$577,000 in sales per employee while Walmart receives about \$225,000³. The result is that data services tend to overestimate Walmart sales while grossly underestimating Costco's. The same issue is applicable to stores across every retail segment.

Just as with market potential estimates, the only way to ensure reasonably accurate estimates of existing sales is through direct observation of local circumstances. This means verifying the businesses and generating sales estimates based not on

industry averages, but on data specific to the chain or business type. That level of detail requires an expert analyst and is not something that can be automated.

Local knowledge can also contribute to more accurate projections. Going into an analysis, the analyst will often know of planned store openings or closings that will have an impact on future sales. These can be factored into a customized market analysis model, but not into the estimates provided by vendors.

INTERPRETING THE RESULTS

A leakage analysis or gap analysis is the best known of the reports generated by a market analysis. The leakage analysis compares the estimated market potential in the trade area to the estimated sales of known businesses in that same trade area. As noted earlier, this is not the same as knowing your market share, or how much of the potential market is being captured by the businesses in your study area.

Where existing sales are less than the market potential, there is said to be "leakage" from the trade area. For many consultants the sectors with the most leakage are identified as the best opportunities and analysis essentially stops at this point. There is quite a bit more that is needed.

A Little Thing Called Competition

Leakage is a misleading term. All demand is being met somewhere in the marketplace. Draw your trade area large enough and you can show all demand being met by existing businesses. More than anything, the ability of a business to be successful comes down to competitive advantage. Only a portion of that advantage is attributable to capturing unmet demand. A prospective business can expand the market, tap a market outside of the area (such as through internet sales), or take sales from businesses already in the market.

Understanding competition is critical to the market analysis. What are the competitors in the market? These competitors may be located in the trade area or in communities surrounding the region. In this context, it is important to understand that how the trade area is drawn will influence the leakage estimates. Draw it to one side of the street and the businesses on the other side of the street will not be calculated into the sales estimates, even though they obviously draw into the trade area. This again points out the limitations of an automated analysis, and reinforces the idea that leakage reports are not the best tool for determining what the market can support.

Good retailers will not be scared by competition, and will even consider a market in which the leakage report indicates a surplus. The operative question is whether the competition is weak or vulnerable to the particular competitive qualities of the potential retailer. In other instances, though, the presence of a number of competitors can be seen as evidence of a

² A Roadmap for Inner City Business Data Collection, Initiative for a Competitive Inner City, 2014

³ Marketwatch estimates retrieved on March 13, 2015

market niche. Arts districts with several galleries are an example of this.

Losing Sales to New Businesses

Cannibalization is the term used to describe when a new business takes sales from an existing business in the market. Communities and business districts need to be sensitive to this possibility, but perhaps not be averse to it. Possible outcomes might include any of the following.

- In cases where the market is growing, existing businesses may see a temporary drop in sales, then recover as new consumers move into the trade area. This is the common experience in fast-growing suburban locations.
- New businesses may cannibalize sales mostly from elsewhere in the trade area or beyond, but not from within the study area. This might occur where a business that is not represented in the study area's business mix is added.
- Attracting a new business may have the impact of transferring sales from existing businesses in the study area. This could be minor or could result in the loss of one or more existing businesses.

Given the potential for a negative impact on existing businesses, cannibalization should be addressed in most analyses for public agencies.

About the Author



Michael Stumpf started his career as a geographer at the Census Bureau before taking his data and analytical skills to the retail site selection community. As an economic developer he continued to hone that practice, helping existing businesses to expand and attracting new retail and restaurants to the communities in which he worked. As a consultant he has served both the private commercial development sector and public clients, providing the market research and strategies for communities, business districts, and real estate development projects across the United States.

For more information contact:

Michael Stumpf, AICP, CEcD
Place Dynamics LLC

www.placedynamics.com

michael.stumpf@placedynamics.com

Minimum Sales Levels

One last issue that is usually missed is that the market share that can be captured by a potential new business must be sufficient to support that business. This is a financial analysis. How much revenue can be anticipated compared to typical costs to operate the business? If the business cannot attain a profit, it is not viable, regardless of the amount of leakage in the trade area.

A CALL FOR BETTER ANALYSIS

If the data from online market analyses is so poor, then why is it so widely used? Part of the answer is simply cost. Dedicating staff or hiring a qualified consultant can cost thousands. Data vendors provide numbers at a fraction of the cost.

Another consideration is that most people do not know how poorly these analyses reflect actual market conditions. Most users do not have the market or industry knowledge to question numbers they see in computer-generated reports, and do not have access to the finer-detailed estimates (such as sales estimates for a particular store) that might make them question the results. Unfortunately, this lack of understanding extends to many consultants as well.