Sabal Strategy

Data Accuracy

What It Is & Why it Matters

Sabal Strategy is committed to providing the highest level of accuracy possible in the data we provide to our customers. Our data partners range from major national compilers to specialty and managed list providers. Collectively, we believe when the data is more accurate, our customers see a higher return on their investment (ROI). And better ROI leads to greater customer satisfaction, higher customer retention, and more loyal customers.

Data Accuracy

We know there is an inherent problem with data: it is subject to error. We recognize data alone may not accurately reflect an individual. Why? Because the data comes from many sources each of which is an opportunity for errors. This is why you need to understand data accuracy.

Data can be defined as something given or admitted, indicating the transmission of information. Consumer data is given or compiled from a variety of sources, including:

- Memberships
- Online behavior
- Online and offline surveys
- Public records
- Product Registrations
- Purchase behaviors
- Subscriptions
- Other public and proprietary data contributors

Because of this massive data funnel, there is an inherent level of information variances and inaccuracies. People can misunderstand survey questions, fill in the wrong blank by accident, or check boxes to make a happy face on the form. This is why our compilers and data partners have robust data quality control and updates in place to reduce incorrect information assigned to individuals and households as much as possible. But there is no data source that can claim 100% coverage or accuracy!

Data is also defined as a collection of facts from which conclusions can be drawn. In other words, consumer data can be "modeled" by using other consumer data. That means we can make sound conclusions based on the demographics and geographic information of a model household. For

example, household income may be estimated, in the absence of self-reported data, based on the age, marital status, and neighborhood of a household. A data model may also need to consider economic or market changes, which can generate additional fluctuation in the data's accuracy. The bottom line is that there is no such thing as perfect data, only varying levels of data accuracy.

The national data partners that Sabal Strategy uses place the highest importance on data accuracy. This is why a continuous effort is made to acquire the highest quality and freshest (most recently updated) data for our customers. To do this, our compiler partners make use of more than 35 contributors, representing thousands of sources, for their supply of data.

Data Accuracy Variables

Did you know U.S. businesses waste more than $3 trillion each year because of undeliverable direct mail, incorrect personalization, excess postage costs, and fines for Do Not Call violations? Before you call that "the cost of doing business," we'd like to point out that it's a needless waste of your resources.

For example, by applying additional postal hygiene solutions, we can close the gap from the industry's 10-15% "Postal Undeliverable" acceptable rate to 1-2%. Mail that isn't delivered because of bad addresses can be due to faulty capitalization, incorrect ZIP code plus 4, missing or incorrectly formatted suite or apartment numbers and other addressing errors that cost you money.

Our experience has shown that it matters which hygiene software is used because each one has a unique set of business rules for determining vacant, undeliverable, and bad addresses. So it is vital you understand the level of address correction your mailing requires in order to achieve the best delivery rates (First Class mailing requires a higher level of address accuracy for delivery than Standard rates).

As for demographic accuracy such as age, income, net worth, IPA and others, the statistics on accuracy vary because of many factors, including data contributors and sourcing.

No matter which type of data we're talking about, the industry average on the expected amount of data variance is 15%. That means a marketer should expect a 15% variance on any given attribute being used. Taken one by one, the variance can be significantly higher (the variance for a home's market value can be 30%) or lower (an individual's exact age variance is only 5%).

Each attribute's variance is based on its compilation method and how frequently it is updated. Examples across all major compilers:

- **Exact Age** is collected from online & offline self-reported data by consumers, historical DMV information, voter registration and other public records. It has a fairly low variance depending on which Age attribute is used (head of household vs. specific member in the household). There is a very small error rate associated with the age attribute but when it does occur it can be significant. For example, a single record on file may confuse a minor's name with another adult who matches all the filters in the same household. To guard against this type of error, we use exact age / year-of-birth as an individual level element, and it's aged on file automatically.

- **Household Income** is mainly self-reported through online and offline surveys, registration data, etc. Census data and modeling is also used to assign income in cases where self-reported information is not available. This data can be derived from sophisticated models based on dozens of affluence measures that include important factors that impact wealth such as tax rates, the local cost of living, household composition, and life stage. Income models are updated once a year. Self-reported data is updated as new information is reported.

- **Net Worth** is based on Claritas' or proprietary models that estimates a household's assets including:
    - Value of a primary residence
    - Cash value of insurance policies
    - Collectibles
    - Real estate
    - Liabilities including:
        - Mortgages,
        - Automotive loans
        - Student loans
        - Lines of credit
        - Credit card balances

  Net worth is sourced from actual dollar measures that are reported through a survey of financial behaviors and also a variety of demographic data elements, including:
    - Age
    - Geography
    - Homeownership
    - Income
    - Presence of children
    - Other factors.

  Net worth data is updated annually.

- **Income Producing Assets** (IPA)is a proprietary model by Claritas that estimates the value of a household's liquid assets. This typically includes:
    - Cash
    - Checking accounts
    - Savings products
    - Investment products
    - Other assets that are considered easy to redeem and move

  IPA is sourced from actual dollar measures reported through a survey of financial behaviors and a variety of demographic data elements that may include age, income, presence of children, homeownership, and other factors. Typically, financial marketers benefit most from IPA as they can easily determine the amount of financial resources available to be moved to alternative financial products. Some marketers use IPA as a wealth measure similar to net worth. The IPA data is updated annually.

- **Home Market Value** is powered by a sophisticated proprietary model that identifies the estimated value of a home. Multiple sources of data feed this model, including:
    - Household-level data (accounting for 62% of sources) including:
        - Values as reported by the County Assessor's office
        - Comparable prices of homes sold in the last 3 to 15 months
        - Self-reported data
        - Income
        - Vehicle values
        - Other demographic values.
    - Geographic-level data (accounting for 38% of sources), including:
        - Aggregated home market values (at the Zip + 4 and Zip Code)
        - Census home market values to provide data when other sources are not available

  The home market values are updated annually and applied at the address level.

- **Ethnicity Group** includes language spoken, race and religion. These are defined and assigned through the use of a proprietary model from E-Tech that identifies given names that are unique to specific ethnic groups as well as surnames and geographic locations

- **Homeowner/Renter** identifies whether the occupant of the residence is the homeowner or a renter. In some files, the designation of homeowner or renter can be further clarified by confidence levels that include:
    - Verified Homeowner (has at least one contributing source that is County Assessor or County Recorder based)

- Highly Likely Homeowner (identified through multi-sourced compiled records)
- Probable Homeowner (typically single-sourced)

It is important to remember the national compilers we depend upon spend massive resources to ensure the accuracy of their database sourcing and updates. In addition, at Sabal, our value proposition is our deep knowledge of each database and their unique strengths and challenges across various data elements. This enables us to choose the right combination of data elements from each compiler to build a multi-compiler prospect audience offering comprehensive coverage and quality for our clients.