Business-Facts

Methodology

Table of Contents

Base File Creation and Sources

Business-Facts® is developed using the infoUSA™ (formerly American Business Information, ABI) data file as its base source. The infoUSA™/ABI database of over 12 million U.S. businesses and is well known in the business-to-business industry for accuracy, coverage, deliverability, and depth of information.

The infoUSA™ base data is mined from over 4,900 Yellow Page directories published each year by Regional Bell Operating Companies and independent phone companies. infoUSA™ also compiles records from over 500 Business White Pages, regular White Pages, Federal, State and Municipal government Blue Pages, annual reports (publicly-held companies) industrial and regional business journals. Specific input data sources include:

- Yellow Pages
- Business White Pages
- White Pages
- Edgar – Sec
- Major Company Web-Sites
- Annual Reports
- State Manufacturers Directories
- Industry Directories
- City Directories
- Big Business Directories
- *The Wall Street Journal*
- Major Metro Newspapers
- *IPO Reporter*
- *NASDAQ* Fact Book
- *American Stock Exchange Facts Book*
- *Standard & Poors Stock Guide*
- County Court House records for new business formations

After compiling the base file, infoUSA™ runs the business information through several routines to ensure that the final product will be dependable for use by salespeople and marketers interested in using its contents for mailing campaigns. For instance, infoUSA™ processes the business file through Group 1 CASS Certified Software for address and ZIP code verification. Furthermore, each month infoUSA™ runs the files against the National Change of Address (NCOA) and the Delivery Sequence files to ensure that business addresses are current and accurate.

Telephone Verification/Enhancement

infoUSA™ calls existing and new businesses at least once each year to verify information, collect additional data, and correct suspected errors. At least four attempts are made to contact each business. Large businesses (i.e., those with 100 or more employees) are phoned twice a year to verify their information. This activity results in 20 million out-bound telephone calls each year by infoUSA™. Telephone verification concentrates on the following data elements:
- Address
- Name of owner or manager

- Number of employees
- Primary business activity
- Fax number

Enhancement of the infoUSA™ Base File by Claritas

All of the data compilation, validation, and verification techniques outlined above result in a robust, thorough base file for use in creating the Business-Facts database. For most clients, the Business Facts database is released quarterly, however the Claritas process for creating Business Facts incorporates monthly changes from infoUSA™. This results in a database that is either concurrent with or, at the most, one month older than the content source data. Furthermore, each Business Facts record is flagged as to whether it is a record without change (ownership, contact name, geography codes, employment, sales and brand/fran codes) or new to the database.

When the infoUSA™ base file is delivered to Claritas, additional enhancements are made to improve matching capabilities and provide greater integration into a range of applications, like mapping. For instance, employment data, sales figures, and point/geo codes (latitude, longitude, census tract, etc.) remain incomplete when Claritas receives the data. The database development team at Claritas adds value to the business data by populating fields that are left empty by infoUSA™, providing the most precise point coding and creating an improved SIC roster.

Below are descriptions of the most significant enhancements made by Claritas to the infoUSA™ base file in the creation of Business-Facts. More complete information follows:

- Claritas ensures that 100 percent of the business establishment records are appended with the most accurate latitude and longitude coordinates available. In addition, tract and block group assignments are made using a point-in-polygon routine.

- The only actual sales volume available is for publicly held companies. Therefore, only a minority of businesses on the infoUSA™ base file has sales volume data when the file is delivered to Claritas. Claritas models sales volume based on actual data from a national sample of approximately six million businesses to provide most of the absent data. This enhancement results in virtually all records having an estimated sales volume, except for government records.

- infoUSA™ provides local employment figures for approximately 80 percent of US businesses. From this actual employment information, the Claritas employment model estimates a "most likely" number of employees for the remaining 20 percent of US businesses.

More on the Point Coding Enhancement

For geocoding purposes, Business-Facts uses spatial addresses. Mailing addresses are carried on the master version of Business-Facts as well. These addresses are confirmed and improved in the production cycle. infoUSA™ uses the USPS National Change of Address database (monthly) and the USPS Delivery Sequence file to verify the validity of addresses. Claritas then assigns the most accurate latitude and longitude, based on the qualified, verified address. The resulting Business-Facts database is 100 percent lat/long coded, with the most accurate level of point geocoding the input address allows.

The latitude and longitude coding process assigns lat/long codes to every new, geographically changed, or imprecisely coded record (from the previous release) on each Business-Facts release. If there is no change to the address record, then the lat/long is perpetuated.

For records with 500 or more employees, the Claritas staff investigates large employment locations without precision lat/long assignments. Claritas uses current software and phone calls if necessary to ascertain the actual location of these records. Business-Facts' lat/longs are assigned to five decimal places of precision, which means this calculation is accurate to within approximately 75 feet.

Tract and block groups are assigned using a point-in-polygon routine. The record is assigned its tract/block group based on its spatial dispensation within a tract/block group's boundary. The block group's boundaries are built by Claritas from the US Census Bureau's TIGER street files.

Employee Count Data

infoUSA™ provides local employment figures for approximately 80 percent of US businesses. With this actual employment information, Claritas models employee counts for the remaining 20 percent of US businesses without employment figures. The Claritas employment model estimates a "most likely" number of employees for records.

The primary determinant for achieving the objective of a "most likely" number of employees is the distribution of businesses by size range. The starting point is a cross-tab detailing the number of businesses and the number of employees for each Claritas Standard Industrial Classification (SIC) code by the following size ranges: 1-4, 5-9, 10-19, 20-99, 100-499, 500-999, 1000-4999, 5000+. Depending on the percentage distribution of employment within a SIC code group, the industry is included in the appropriate strata for estimation purposes.

Secondly, if the SIC has a propensity to have large businesses, little or no modeling will be done in the SIC, since the actual number of employees will be present in most large employers' records.

Although it is difficult to predict the absolute effect on the Business-Facts file of modeling employee counts on the number of total employees, preliminary estimates suggest that this model will increase the number of employees by five to seven percent. If so, the outputs from this model will increase the overall number of employees in the Business-Facts database by roughly one million or .008 percent. It is important to remember that this model is only applied to records where no actual employment figure exists. And its use, moreover, provides users of the Business-Facts database with better coverage and more information for modeling and other analytical applications.

Sales Volume Data

infoUSA™ provides actual sales information only for publicly held companies. Sales figures, therefore, for all other companies must be estimated. The infoUSA™ model uses aggregated sales divided by the aggregated number of employees by SIC to arrive at sales per employee. This aggregated method does not factor in the distribution of employees by SIC, and may allow large corporations to have a disproportionate effect on the sales-per-employee estimate.

In comparison, Claritas models sales volume based on actual data from a national sample of approximately six million businesses. The model is calculated at the four-digit Standard Industrial Classification (SIC) code level with the results applied to those records for which no actual sales volume figures exist. This enhancement results in virtually all records having an estimated sales volume, except for government records. However, since one number is being applied to remaining businesses in a four-digit SIC industry, analyzing a business as a member of a sales range is more appropriate than looking at the specific sales volume for that business.

As with the employment information, Claritas' clients will benefit from the combined actual/modeled sales data with better coverage and more information for modeling and other analytical applications.

SIC Rosters

Business-Facts uses a universe of 1,604 SIC codes compared with over 10,000 SIC codes in the original infoUSA™ file. The first four digits of the SIC Code are standard as defined by the U.S. Government Department of Labor. The last two digits of the SIC Code on infoUSA™ roster are infoUSA™ defined sub-categories providing further resolution to the standard governmental SIC Codes.

Claritas shortens the infoUSA™ roster by combining unpopulated, or little populated 6 digit infoUSA™ defined SIC codes into our more implementable Claritas SIC 6 digit categories. In the Claritas SIC consolidation process of the infoUSA™ SIC roster, the first four digits of the standard government SIC Code remain intact.

Claritas retains the original infoUSA™ SIC code hierarchy in the event there is a requirement to identify which infoUSA™ SIC Codes constitute Claritas Business-Facts SIC Codes.

Data Elements

The Business-Facts® database includes the following data elements:

Business Name #	Primary SIC Code #	County Code *
Address	Secondary SIC Codes #	Tract Code *
City	Employment Here #	Block Group Code *
State	Employment Total	Foreign Ownership Flag
ZIP Code	Sales Volume Here *	Deliverability Flag *
ZIP+4 #	Sales Volume Total	Recency Code
Telephone Number	Location Type	Franchise/Chain/Brand/Specialty #
Executive Name	Company Type	Year 1st in Yellow Pages
Executive Gender	Fortune Flag	Business at Home Flag
Latitude *	Industry Size Indicators	Stock Exchange Label
Longitude *	Stock Exchange Code	Stock Exchange System

* Claritas data element
# Claritas enhanced data element

Note: This is not a complete list of data elements available on the Business-Facts® file. In addition, not every data element listed is delivered through Claritas software platforms.

Latitude, longitude, block group code, tract code, county code, and ZIP+4 are elements added to the infoUSA™ base file in order to create Business-Facts. Moreover, sales volume here is created by Claritas with a proprietary model and is appended to the business records. Select data elements list above are enhanced by Claritas are as follows:

Data Element	Enhancement
Business Name	Business names, particularly chain names, are standardized for uniformity.
ZIP+4	ZIP+4s are corrected when necessary.
Primary/Secondary SIC Codes	infoUSA's 10,000 SIC categories are consolidated into Claritas' 1,604 SIC categories. This enhancement reduces the number of poorly populated SIC codes that are too fragmented to constitute a market.
Employment Here	Actual employment is available on 79% of all business records. The employment figures for the remaining records are estimated using Claritas' proprietary employment model.
Sales Volume Here	Actual sales volume is available on approximately 25% of all business records. The sales figures for the remaining records are estimated using Claritas' proprietary model.
Franchise/Chain/Brand/Specialty	Besides the codes received from infoUSA, Claritas builds additional codes based on client demand. An additional 300 codes for chain, franchise, and specialty codes result.
Linkage Consistency	Efforts are taken by Claritas to ensure that all family records get the correct total sales and total employee information.
"Best" Record Build	Claritas processes all records for each business location. The record build rules insure that primary SIC's are correctly identified, the "best" field information is retained, i.e., ad size, brand/franchise codes, etc.

Business-Facts® and NAICS

The Bureau of Labor Statistic of the U.S Department of Labor recently implemented the North American Industrial Classification System (NAICS). The 6-digit NAICS code is intended to replace the 4-digit SIC code whose origin dates to the 1930's and offers a better picture of the structure of the twenty-first century economy. The NAICS system provides detailed information available about many individual industries that were previously clustered together.

NAICS codes are currently available on the Business-Facts database. Both SIC and NAICS codes will be available allowing Business-Facts users ample time to adopt the new industry classification system and revise operational processes that currently depend on SIC codes.

TETRAD COMPUTER APPLICATIONS INC.

www.tetrad.com

1465 Slater Road, PO Box 5007
Ferndale, WA 98248-5007

Tel 360-734-3318
Fax 360-734-4005

Call toll-free 1-800-663-1334

Email: info@tetrad.com

nielsen

TETRAD Computer Applications Inc. is a Licensed Distributor of Nielsen Claritas Data:

- Pop-Facts
- Business-Facts
- Consumer Buying Power
- Retail Market Power
- PRIZM Segmentation
- Net Worth & Income Producing Assets
- Financial Clout
- P$YCLE Financial Segmentation

01-2009