

## FORECAST FOCUS



# Collecting, Cleaning and Trending Raw Data

By: Tiffany LaReau, Human Numbers

**Analyzing raw data often requires as much time and effort as producing the final forecast — but it's time well-spent.**

The most important element for developing an accurate contact center forecast is the data. You must take as much care in the initial steps of collecting and cleaning the raw data as you do to your analysis to ensure that you're accounting for all of the scenarios and behaviors that may influence your final forecast. There are times when analyzing the raw data set requires as much energy as producing the forecast. Since the root source of a good forecast lies with the integrity of the base data, the extra effort is well-spent.

In this article, I'll give an overview of how to collect, clean, trend and record the raw data so that you can begin your forecasting activities (my favorite part!). Let's take a look at the four steps involved.

### **Step 1: Data Collection**

The forecasting process begins with collecting a good data set. The first step is establishing a process to collect interval-level data. Accessing raw data is not always the easiest thing to do, but that's no excuse for ignoring its importance. If I'm working with a call center that is brand-new and doesn't yet have historical data, I usually will rely on data from a

similar industry's distribution percentage combined with the call center's own good judgment, at least until I have that first good week's worth of data from which to sample. I've also encountered situations where the call center does not use standard ACD reports that supply me with nicely summarized interval data. In those cases, I have to create a custom reporting solution to get the data I need to generate forecasts.

The type of data to collect begins with the answered calls. This is the mandatory data element. If you can also collect abandoned calls, that's even better, because you can use it to calculate two additional metrics: calls offered and percent abandoned. Additional counts that can prove helpful include any calls that were rerouted to a different group (also referred to as "flow-in/flow-out"), transferred calls, calls that were blocked, forced to a busy signal or redirected to voice mail. Although you may not necessarily use all of these volumes to calculate Total Calls Offered, these are ones that can indirectly impact the future forecasted answered calls.

Next, collect the handle-time data elements, which include talk time, after-call work (ACW)

*Continued on page 2*

Continued from page 1

time and hold time, usually from the total handle time. In some call centers, the ACW time is forced to stop at a certain threshold, so it's necessary to reevaluate the actual ACW time that is really occurring in the AUX or AVAIL state in order to truly measure total handle time. Another situational handle-time metric is ring time. If ring time is up to an agent's discretion, that can impact average handle time (AHT) differently depending on which agent is taking the most calls. I also like to collect raw data reports for positions staffed. It's not something that normally helps me with the forecasting process, but since I'm analyzing that raw data anyway, it's an easy way to see if any unusual activity is occurring outside of regular operating hours.

### Step 2: Cleaning

After the raw data is reporting and compiled in one place, it's time to clean the data. The purpose of cleaning the raw data is to categorize special activities separately from normal activity. For instance, one example of special activity is Independence Day — a U.S. federal holiday that is always celebrated on July 4th, but is observed on a weekday. Since the impact of July 4th can hit on a different date every year, a different day of the week and even a different week of the year, it's not safe to allow that data to remain in the normalized pattern set (unless your call center isn't affected by this special activity.) For a holiday like Independence Day, I recommend tagging it with the name "4th of July Saturday," which will make it easy for you to know when to reuse it in the future. And by incorporating the day of week into the name, you can also interchange it with other holidays that fall on a Saturday but which are observed on the preceding Friday. This is extra helpful if you're using a limited history set.

However, special activities are not limited to holidays. Include anything that has an abnormal influence on the volume, such as system outages, company outings and special promotions. Sometimes the cleaning will need to occur at the interval level; other times, at the daily level. Holidays are the easiest to clean, because the entire day is impacted and it happens on a regularly scheduled basis, but I have to dig deeper to find abnormal data

Figure 1. Weekly AHT Results

| Average o DOW |     |     |     |     |     |  |
|---------------|-----|-----|-----|-----|-----|--|
| Time          | Mon | Tue | Wed | Thu | Fri |  |
| 10:00         | 345 | 292 | 388 | 61  | 313 |  |
| 10:30         | 226 | 281 | 361 | 387 | 454 |  |
| 11:00         | 136 | 459 | 264 | 231 | 210 |  |
| 11:30         | 295 | 211 | 173 | 434 | 396 |  |
| 12:00         | 139 | 194 | 264 | 239 | 332 |  |
| 12:30         | 259 | 258 | 335 | 129 | 131 |  |
| 13:00         | 292 | 237 | 463 | 43  | 391 |  |
| 13:30         | 431 | 257 | 267 | 159 | 216 |  |
| 14:00         | 352 | 283 | 295 | 248 | 252 |  |
| 14:30         | 221 | 183 | 210 | 199 | 413 |  |
| 15:00         | 274 | 121 | 227 | 241 | 330 |  |
| 15:30         | 126 | 483 | 105 | 199 | 188 |  |
| 16:00         | 241 | 350 | 430 | 208 | 365 |  |
| 16:30         | 257 | 305 | 254 | 294 | 183 |  |

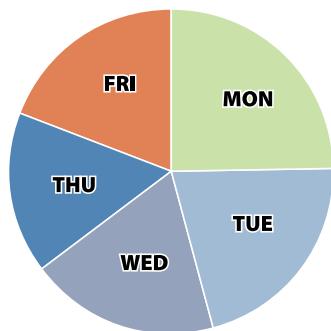
on my own. I have found success using standard deviation with conditional formatting rules to highlight extremes at the interval level. In addition, Upper and Lower control limits add additional value to this process because they naturally absorb recent movement into the formulas. The call behaviors can change periodically, so it's necessary that any methodology that you use is flexible and responds organically to those changes without requiring a major overhaul each time.

Figure 1, above, offers an example of weekly AHT results. I use a three-color data scale so that I can instantly see where the hot (red) and cold (green) spots are. Using a resource tool that highlights the extremes allows me to spend less time hunting for information and more time analyzing it.

Once the raw data set has been cleaned, it's referred to as normalized data. In my work, I like to keep the normalized data right alongside the raw data. I use the normalized data to calculate interval distribution percents, day-of-week percents, week-of-year percents and month-of-year percents (for future forecasts), and I reserve the actual data for calculating forecast accuracy (at the interval, daily, weekly and monthly levels). You may be tempted to measure your forecast accuracy using the normalized data

Continued on page 3

Figure 2. Day-of-Week Distribution



| Sum of Off |     | DOW    |        |        |        |        |
|------------|-----|--------|--------|--------|--------|--------|
| Year       | Wk# | Mon    | Tue    | Wed    | Thu    | Fri    |
| 2009       | 16  | 18.63% | 24.51% | 26.47% | 17.65% | 12.75% |
|            | 14  | 29.01% | 20.61% | 16.79% | 12.98% | 20.61% |
|            | 12  | 25.81% | 15.32% | 15.32% | 18.55% | 25.00% |
|            | 11  | 28.30% | 23.90% | 20.13% | 8.81%  | 18.87% |
|            | 10  | 21.97% | 18.94% | 18.94% | 19.70% | 20.45% |
|            | 9   | 25.74% | 22.06% | 16.18% | 18.38% | 17.65% |

Continued from page 2

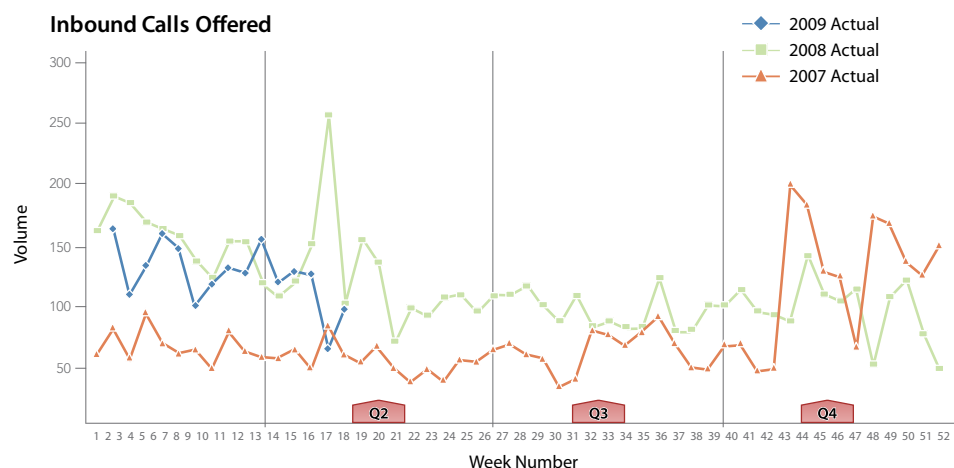
(which would produce better results), but don't give in — forecast accuracy is a result of what is happening and changing in the call center, and there is more to be gained by properly reporting on call behavior.

### Step 3: Trending

Before building out your forecast, I recommend using different methods to map out the raw data, which will give you a visual example of what is happening.

For the day-of-week distribution, I like to use a pie chart and a two-color shaded table, like the example shown above in Figure 2. For a weekly view, I prefer the full year-over-year chart, which is shown below in Figure 3.

Figure 3. Year-over-Year Chart



Seeing the results in this type of format is helpful because it will give you the opportunity to carefully consider which forecasting method to use. At this point in the process, most software solutions would move forward with their built-in Time Series method; however, the trending exercise proves to me that an alternative

forecasting method (in this case, a combination of Event Driven with an Altered Growth Rate would be the better choice). When the information is laid out so nicely, it allows you to realize special seasonality, bizarre call spikes and overall call behavior.

### Step 4: Journalizing

The final step is to document and archive everything that is special, out of the ordinary, and anything that you come across that you think might be helpful in the future. Journalize these activities using vivid descriptors with keywords that will make it easy to refer back to those special events or strange calculation methods.

I sometimes use the Comment sections that are available in Excel, Word and PowerPoint, but I find that comments are good if the descriptor is one sentence, fewer than 10 words. Anything longer really deserves its own page in a WFM Diary. A typical page would include the date, the group affected by any abnormal or special circumstances, a description of the problem and full details on the solution used, including calculations, analysis results and any necessary follow-up (use flags to follow up on topics and solutions that need future testing for validity).

There are so many different drivers that can affect volumes and handle times, and each of them has their own rate of impact.

Taking the steps to collect, clean, trend and document the data will enable you to understand why volumes behave the way they do, and it will better prepare you for the fun part: FORECASTING! 📊

## ABOUT US

**Contact Center Pipeline** is a monthly instructional journal focused on driving business success through effective contact center direction and decisions. Each issue contains informative articles, case studies, best practices, research and coverage of trends that impact the customer experience. Our writers and contributors are well-known industry experts with a unique understanding of how to optimize resources and maximize the value the organization provides to its customers.

### Pipeline Publishing Group, Inc.

PO Box 3467, Annapolis, MD 21403  
443-909-6951 | info@contactcenterpipeline.com | [www.contactcenterpipeline.com](http://www.contactcenterpipeline.com)

Tiffany LaReau is the Workforce Manager and Owner of *Human Numbers*, a forecasting and scheduling service provider.

✉ Tiffany@HumanNumbers.com  
☎ 678-494-1506

