PIPKINS℠

Enterprise Workforce Management ...Since 1983

Pipkins Forecasting Case Study:

Data and Forecasting Analysis

Part 1 Overview: Data Analysis

All call centers are unique. This is certainly true when considering historical data. When it comes to using this data to produce forecasts, the ‘one size fits all’ approach is unlikely to work in all circumstances.

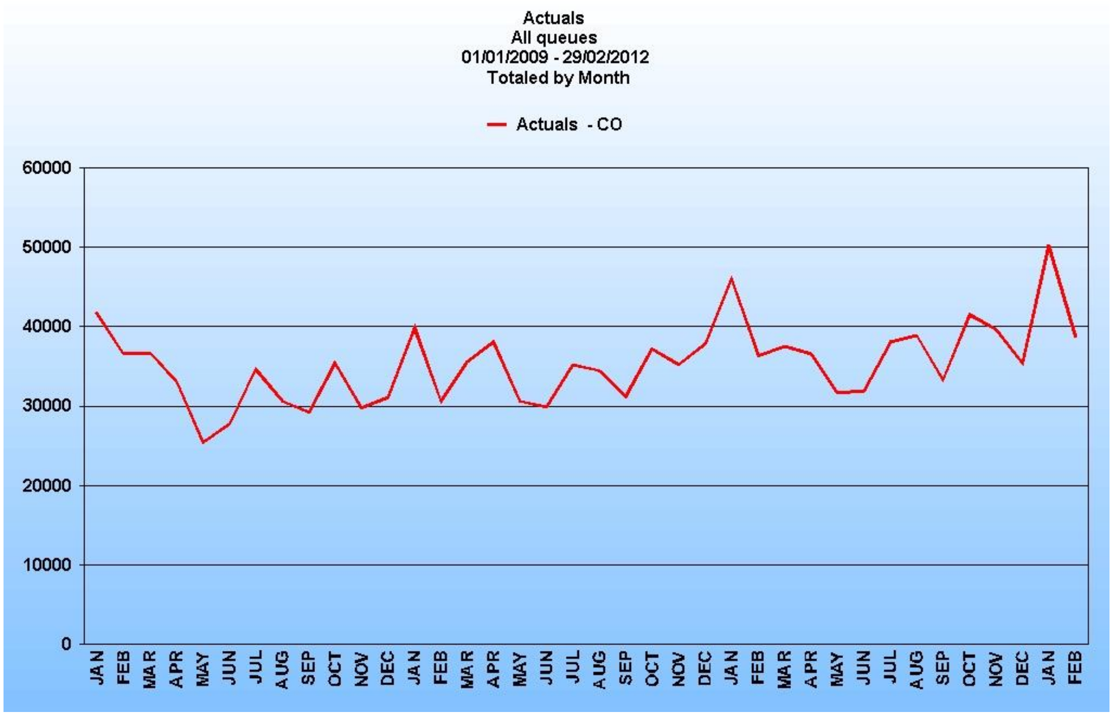
The Vantage Point forecaster is a versatile tool which can be tailored, using the appropriate directives, to produce the most accurate forecasts for any given call center. However, if the forecasting tool is to make the best predictions for an individual center, it is recommended that an analysis of the historical data is undertaken. Such an analysis was done for a Pipkins customer recently and the following case study is a summary of the key steps in the process with this first part focusing on the data analysis aspect.

Analyzing the Data

The first step is always to examine the historical data to look for certain factors, e.g. seasonality, growth, step changes, etc. The amount of available data is important when looking for patterns - ideally at least 2 year’s worth of data is necessary for such an analysis. In this case the customer had 3 years of data and all of this was used at some stage.

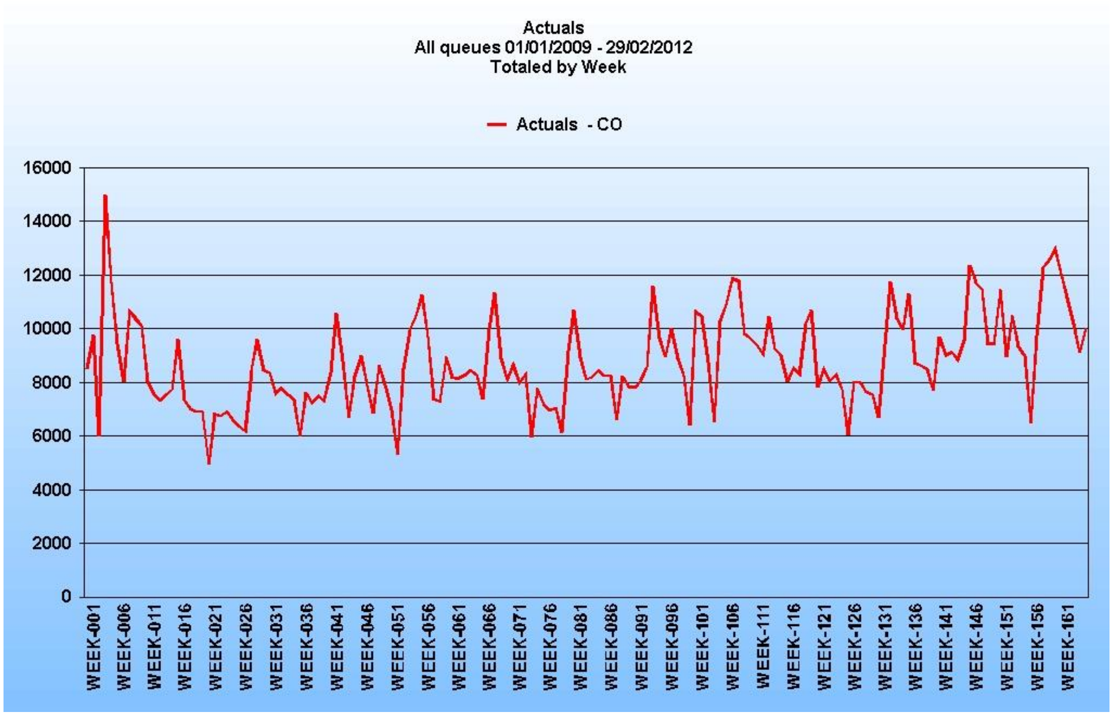
The initial analysis looked at all queues. This can often give a good view of the overall nature of the call center although this view will be drilled down as the analysis progresses.

Looking at the overall data from January 2009 on a monthly basis, this is the result:



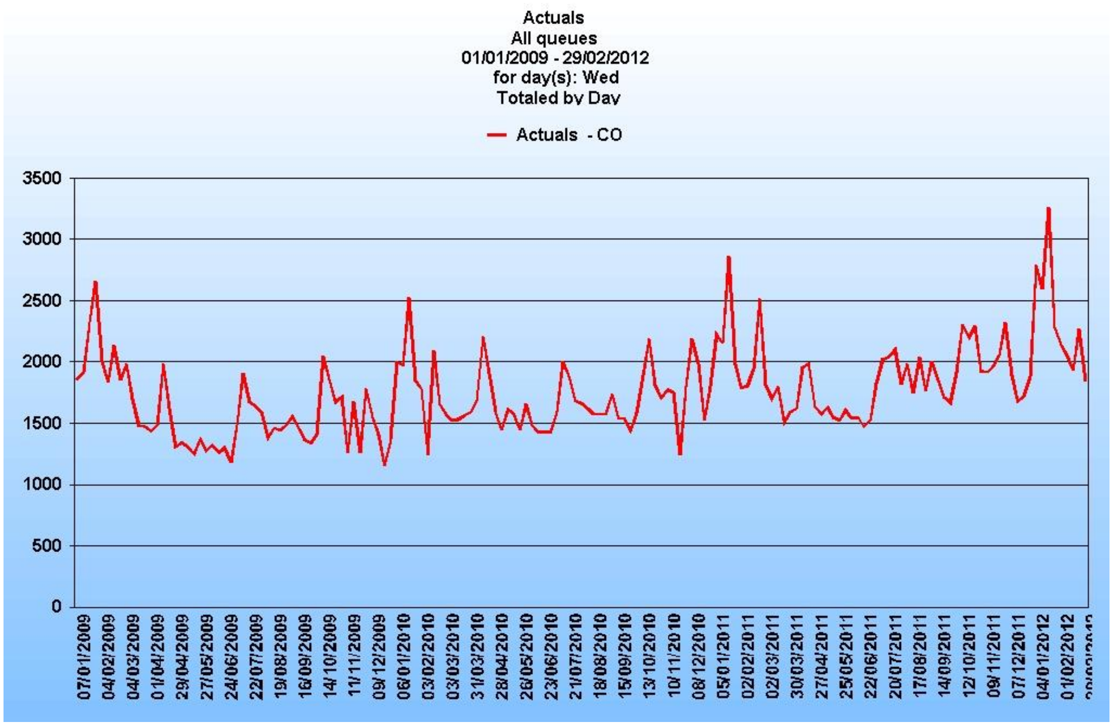
There appears to be some underlying growth, which seems quite steady. There are peaks in January of each year and also peaks in other months which are not as pronounced.

Looking at the weekly picture over the same period:

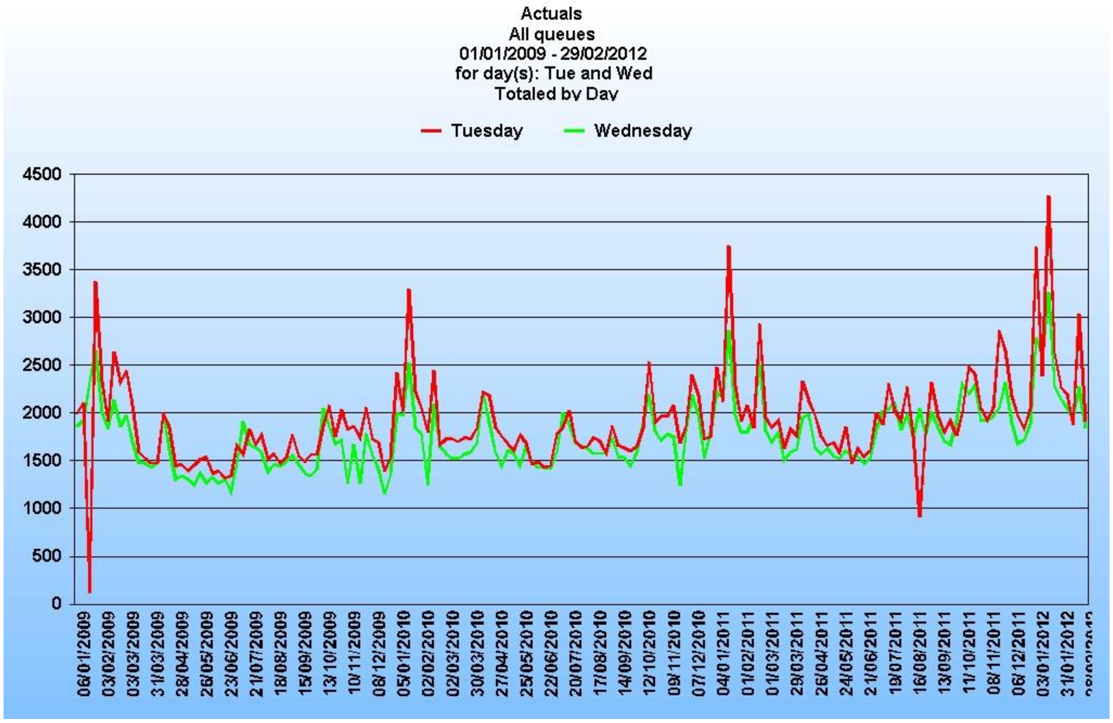


There are some significant weekly variations with some volumes changing by 20% or more week-on-week. Some of these will be due to obvious factors (Christmas and other public holidays) but it would be worth investigating further to try and discover other events, which result in large changes in call volume.

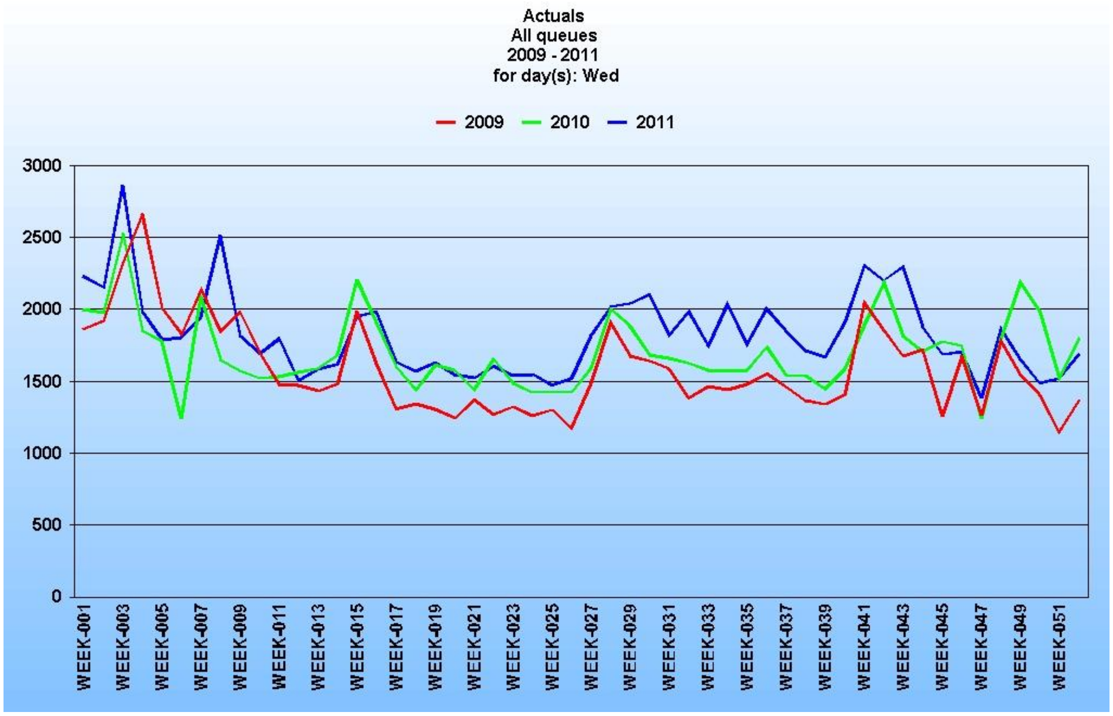
To this end, we can look at an individual day, in this case Wednesday:



This shows a similar pattern to the weekly graph, which indicates that the fluctuations may not be limited to particular days, but more likely affect the whole week – this can be confirmed by comparing different days of the week – here is an example showing Tuesday and Wednesday. In the main they follow the same pattern.

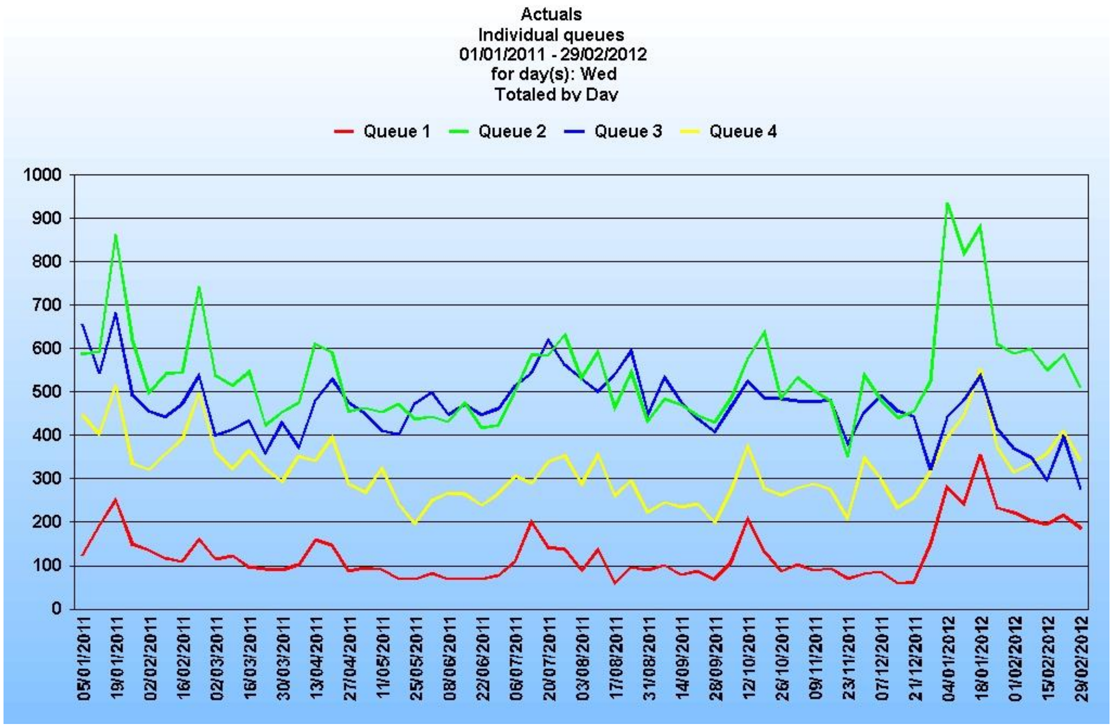


Looking at the monthly, weekly, and daily patterns show some obvious repeated patterns as mentioned above (e.g. Christmas) but it is not easy to determine recurring (seasonal) patterns from these graphs. To help determine seasonal trends we can also look for a year on year correlation by comparing them:



It appears from this that there is quite a strong seasonal component to call volumes. Some events do recur at the same time every year, some may be offset by a week (this can depend on where the year starts), some show little or no pattern year on year.

Further analysis can be done by looking at the higher volume queues to see if the seasonal peaks are recorded for all queues or limited to certain ones:

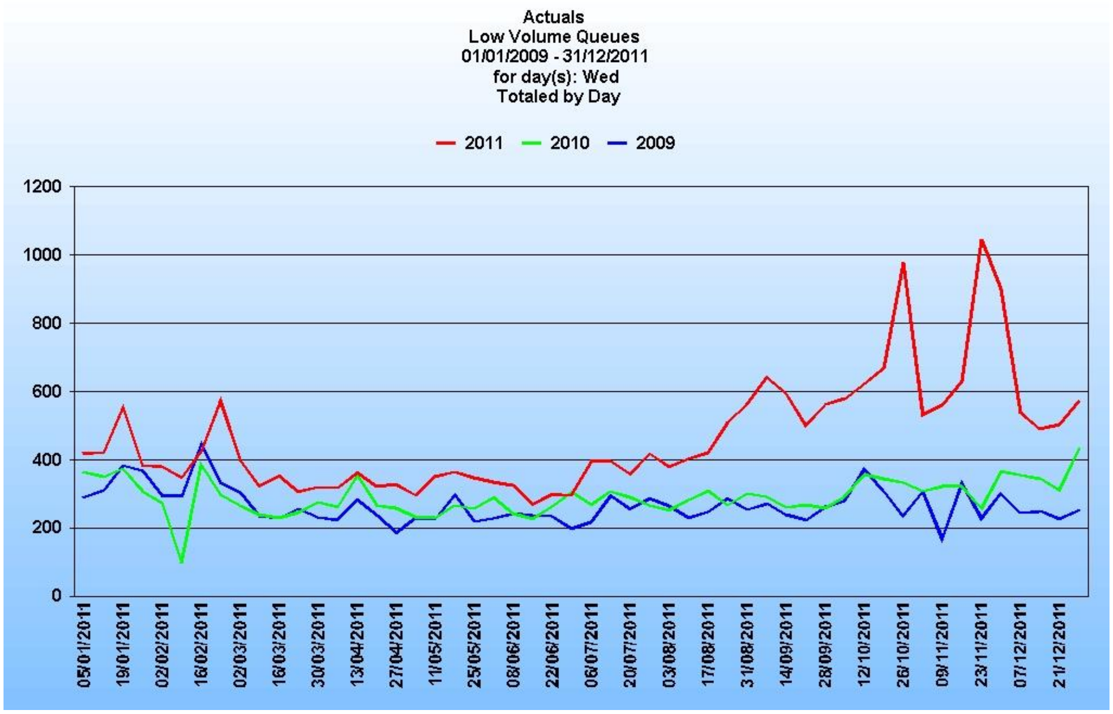


This shows that at least some of the peaks are much more pronounced for certain queues although all show some seasonality around the same times.

The above would indicate that, when forecasting, the appropriate directives should be used to enable this to be taken into account and any special events identified which can be predicted.

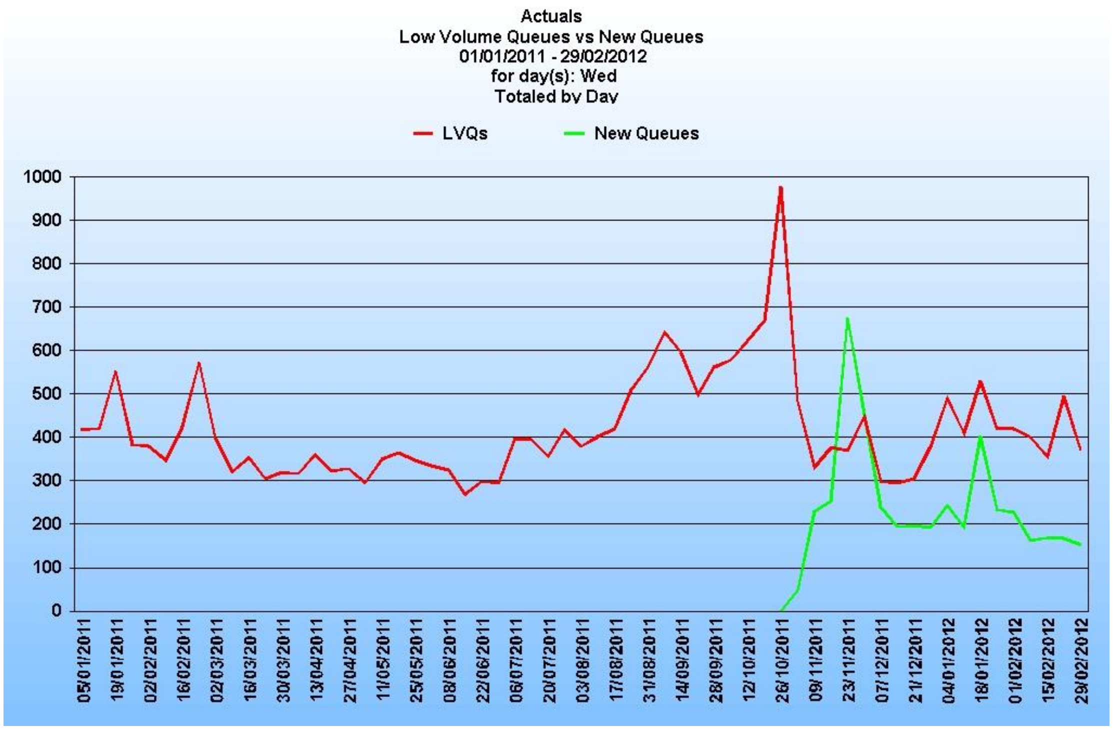
This customer has several low volume queues (LVQs) – in this case these were queues which accounted for less than 100 calls per day. These were grouped by removing the higher volume queues identified earlier from the list of all queues.

Looking at the low volume queues (as a group) shows a different picture from that already described above:

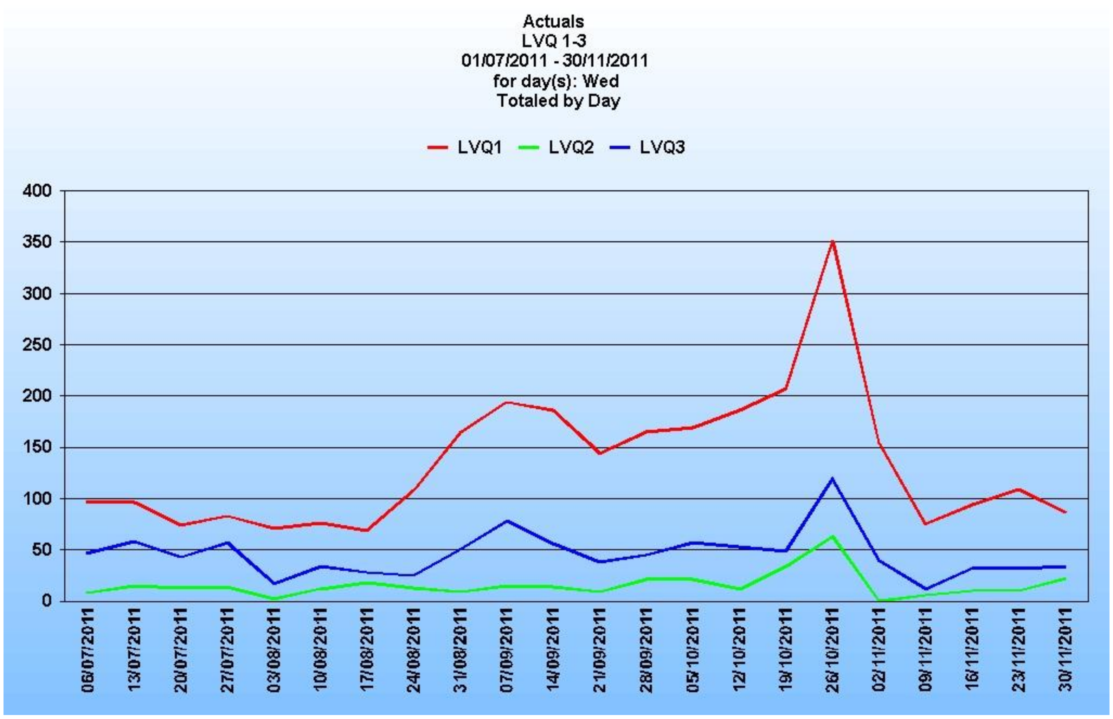


The seasonality indentified in the higher volume queues is not as recognizable here but there appears to have been a significant change around September 2011. Further analysis of this shows an increase of calls between the end of August and late October. Volumes returned to normal until two new queues were opened in mid-November, which resulted in the overall increase.

This is illustrated below by showing the two new queues separately from the other LVQs:



Further analysis of the busy period (September – October) showed that the increase was limited to 4 out of the 25 LVQs with the final peak in the last week of October:



As this period is much more pronounced than in previous years the customer would need to determine if this was likely to recur and account for it in the forecast.

It is worth pointing out that although it is normally recommended that low volume queues be aggregated with higher ones, some events which may only affect some smaller queues for short periods may be ‘lost’ if the queues are aggregated. It is therefore worth doing the analysis initially to identify such cases.

It is important to treat the new queues separately when forecasting, as the amount of historical data available is limited.

Conclusions

1.1 For the established higher volume queues there is a strong seasonal element and some underlying growth.

1.2 There are recently added queues which need to be forecasted separately.

1.3 There are some LVQs which may need to be forecasted separately if there is a evidence of periods of marked change in volumes.

Next we will consider how to use this information to tailor our forecasts.

Part 2 Overview: Forecasting Analysis

In order to determine whether forecasting directives are required and if so, which ones, it is necessary to consider the nature of the historical data. This was investigated in Part 1 above.

The default method for forecasting is to look for patterns in the historical data. If the forecaster identifies one or more patterns it will utilize this to calculate the forecast for the future dates. If no pattern is found then it will use a weighted average of recent historical data.

This method is often suitable for short-term forecasts (say 1 – 3 weeks into the future). However, there are other methods available which may be more appropriate. This needs to be determined. Once more the key to this will lie in the historical data.

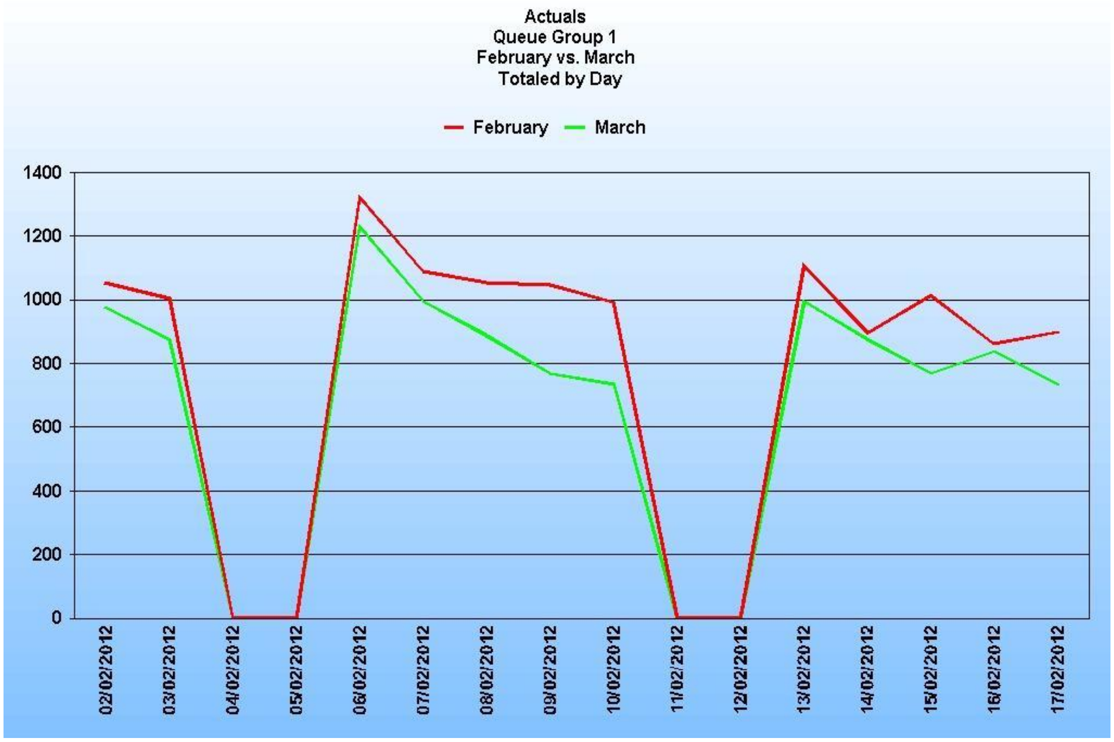
Methodology

For this exercise we wished to produce forecasts for the first three weeks of March 2012 based on the historical data up to the 29th February using different forecasting methods. In order to determine the accuracy of the forecasts, these were then compared to the actual data for the same period.

Initial comparisons were also made to compare the actual calls received in March 2012 with previous call actuals. The purpose of this was to give some indication about the nature of the call volumes month on month and year on year.

In this case, two comparisons were made – an annual comparison and a monthly comparison. Using actual data from March 2012 (up to the 20th) comparisons were made between Feb 2012 and Mar 2012, and also between Mar 2012 and Mar 2011 for a selection of queues, which are shown below:

First, comparing the last two months for the main queue group (minus the recent additions identified in Part 1).

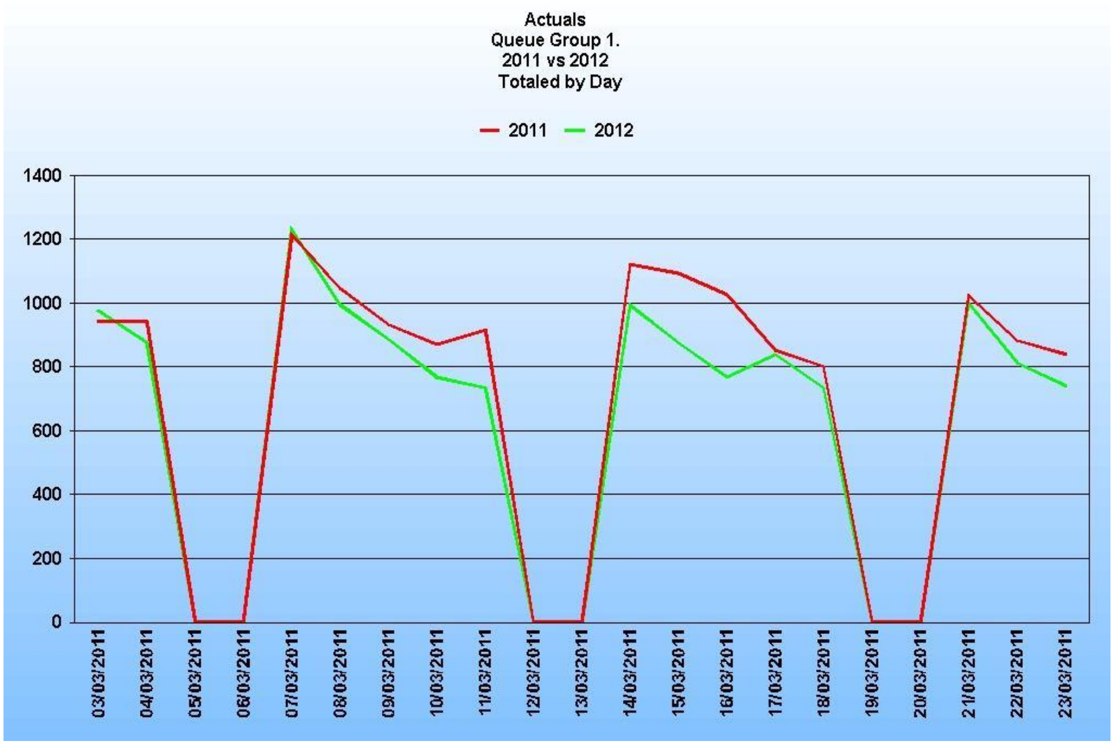


**See the section on recently added queues below*

Second, the same comparison for the largest queue in the group:

Conclusion: on most days March was less busy than February – on some days this difference was significant (>20%).

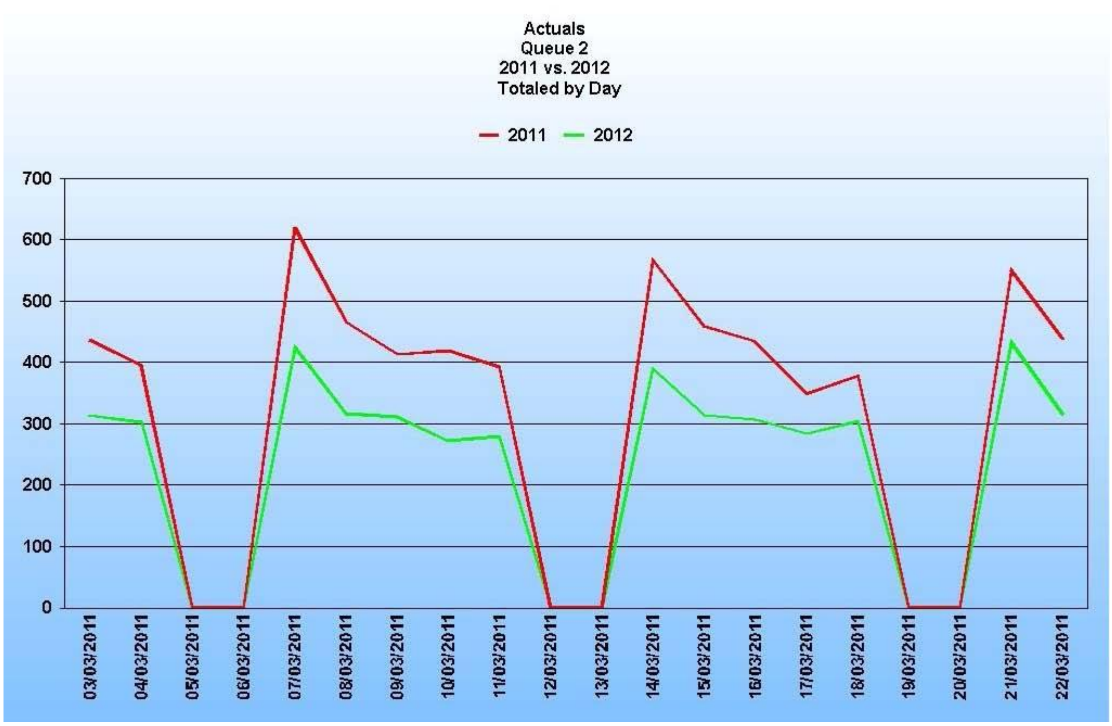
Next, comparing March 2012 with the previous year, March 2011:



(Note: this group excludes queues with no data in March 2011)

Conclusion: this again shows a significant variance (>20%) on some days but not others.

Looking at individual queues shows a larger year on year variation in volume:



PIPKINS SM
Enterprise Workforce Management ...Since 1983

Conclusion: in most cases March 2011 has higher volumes of calls received than March 2012.

Alternative options: using forecasting directives

Using a database with data up to 29th February, forecasts were created for the first three weeks of March 2012 using different directives. It was then possible to compare the forecasts with the actual data to determine which were more accurate. The results for the main queue group are summarized below:

	Default forecast - all data	Default forecast - omitting president's day week	*form forecasting data set by day of week in corresponding months of the year;*	*form forecasting data set by day of week in corresponding months of the year; renormalize forecasted data using centered current data points relative to the previous years;*	*form forecasting data set by day of week in corresponding months of the year; renormalize forecasted data using the actual data in the current year relative to the previous years;*	*form forecasting data set by day of week in corresponding months of the year; use the 4 most recent actuals data elements to renormalize the forecast;*
	Comparison of Actuals with					
Start Date	**Option 1**	**Option 2**	**Option 3**	**Option 4**	**Option 5**	**Option 6**
Thu 03/01/12	3.60%	3.80%	-2.90%	-7.40%	-6.50%	-11.80%
Fri 03/02/12	7.50%	20.70%	7.00%	1.60%	2.70%	-3.20%
Sat 03/03/12	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Sun 03/04/12	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Mon 03/05/12	-10.80%	7.80%	-0.70%	-6.80%	-5.20%	-10.90%
Tue 03/06/12	26.70%	4.40%	5.40%	0.50%	0.90%	-5.00%
Wed 03/07/12	32.80%	34.20%	5.70%	-0.60%	0.70%	-5.00%
Thu 03/08/12	30.30%	40.00%	13.90%	7.40%	8.60%	2.90%
Fri 03/09/12	57.10%	38.00%	25.20%	18.20%	32.00%	12.20%
Sat 03/10/12	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Sun 03/11/12	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Mon 03/12/12	29.70%	15.80%	13.10%	6.90%	8.50%	1.60%
Tue 03/13/12	46.00%	23.60%	25.30%	17.80%	19.10%	11.40%
Wed 03/14/12	41.30%	43.00%	34.50%	26.60%	28.00%	20.80%
Thu 03/15/12	20.30%	23.00%	1.90%	-3.80%	-2.60%	-7.70%
Fri 03/16/12	40.00%	54.40%	10.70%	4.50%	5.40%	-0.70%
Sat 03/17/12	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Sun 03/18/12	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Mon 03/19/12	10.90%	30.80%	2.90%	-2.70%	-1.30%	-8.20%
Tue 03/20/12	60.00%	23.50%	9.40%	3.40%	3.90%	-1.70%
Wed 03/21/12	51.30%	53.80%	14.20%	7.00%	8.40%	2.60%

(Note: the highlighted cells show where the variance from the actuals was less than 10%)

Similar results were obtained for individual queues.

These results indicate that the seasonal trends identified in part 1 of this case study were being repeated in March 2012 but that some account should be taken of year on year changes in volumes. Therefore, when generating forecasts, the best results would be obtained by using directives designed to recognize this trend

Understanding Forecasting Directives

form forecasting data set by day of week in corresponding months of the year;
[Option 3] This tells the forecaster to look back at previous years' data and base the forecast around those patterns.

The next three directives below are used to adjust the forecast based on recent history – to take account of changes in call volumes compared to the previous year:

renormalize forecasted data using centered current data points relative to the previous years;
[Option 4] This uses a variable amount of data for the comparison, dependent on how far ahead the forecast.

renormalize forecasted data using the actual data in the current year relative to the previous years;
[Option 5] This looks at current year's data and compares to the same period last year.

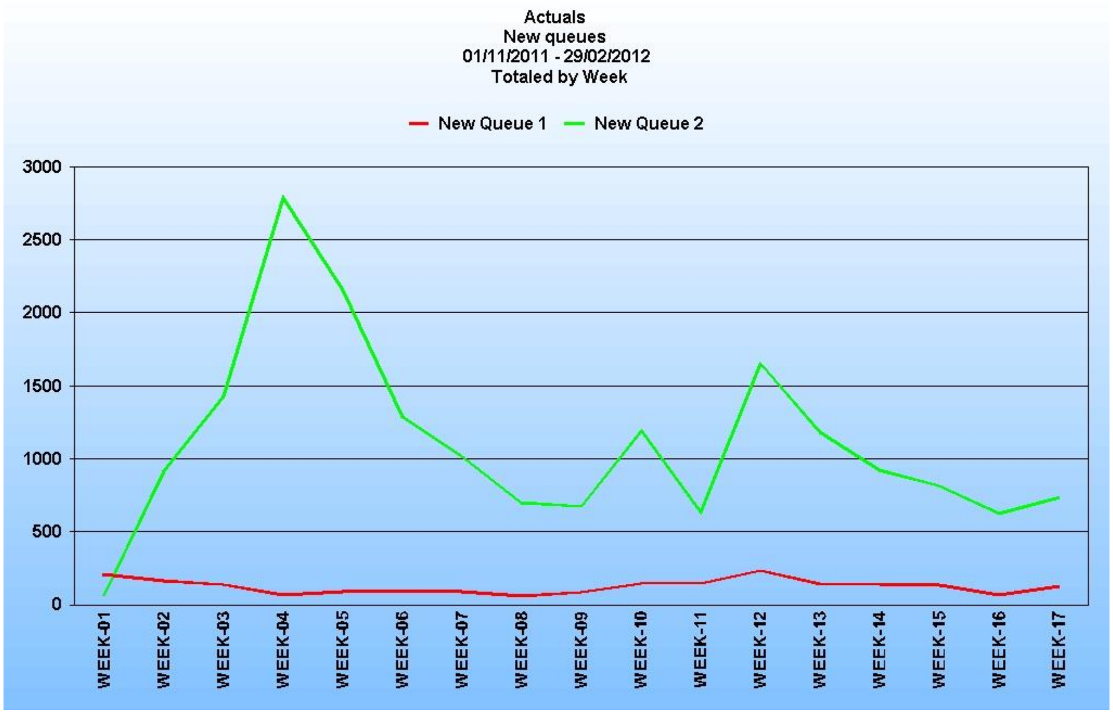
use the x most recent actuals data elements to renormalize the forecast;
[Option 6] This uses the last x weeks of data to carry out the comparison.

Options 4, 5 and 6 produce similar results and the most appropriate should be chosen depending mainly on the distance into the future the forecast was for.

Recently added queues

There are two queues for which there is only actual data from November 2011, which were identified in Part 1 of this document. The above forecasting method is not suitable for these queues so they should be excluded and forecast separately.

The data for these queues looks like this, firstly totaled for the week:



and secondly for a selected day (Friday):

Queue 1 has quite steady volumes, which are also fairly low. Queue 2 has, on the face of it, an unpredictable call volume although there are signs of more settled behavior since the end of January. The default forecasting method is recommended, with the 'auto detect growth trend' disabled, until more actual data is available. This will produce a weighted average of recent data.

Low Volume Queues

The actual data shows that, for this customer, all queues have data for each time step recorded, even if this is a zero. This makes things more straightforward for low volume queues (LVQs) as the forecaster will not attempt to reconstruct 'missing' data. The normal recommendation for this type of queue is to aggregate them with one or more larger volume queues – they will then effectively take on the properties of the larger queue(s) for forecasting purposes (see "A guide to LVQ forecasting" for details on this issue).

However, further analysis of these queues should be undertaken initially to see if there is any seasonality or other behavior which might require them to be considered separately. As we saw in Part 1, a couple of the LVQs did show a short-term rise in call volumes which would be lost if these were aggregated. It is up to the user to determine whether this pattern is likely to be repeated. If so then those queues should be forecasted separately – at least for the peak period.

Special Events

From the analysis of historical data done previously it was suggested that there may be identifiable special events which were affecting the call volumes at certain times of the year.

Some of these are easily identifiable (e.g. public holidays and Christmas/New Year) but others, if they exist, would require further analysis of the historical data and some local knowledge of predictable events which may be affecting volumes.

However, was also noted that there is a pretty strong year-on-year correlation of call volumes which will be considered using the aforementioned directives. It is possible that this yearly correlation may be sufficient to predict the volume changes without the need to identify each as a special event. The user may only need to identify those events which are not recurring annually.

Summary of Recommendations

1. Consider aggregating low volume queues with higher volume queues. Whilst not strictly necessary it may increase the accuracy of the forecast for these low volume queues.

2. Remove recent queue additions from the current queue groups and forecast separately as described above.

3. For all other queues use the directives outlined above. The directive '*form forecasting data set by day of week in corresponding months of the year'* should be used in conjunction with one of the normalizing directives as appropriate.