

Instant Customer Analytics using only Excel

By Ken Thompson, ServQ Alliance

There are some excellent specialized data analytics, data mining and data visualization tools available however we should not forget about Microsoft Excel, particularly the new versions, which offers powerful features to quickly slice and dice customer data from live systems to identify important trends and behaviors which can inform strategy. In this article and supporting screencast I share a 7-step plan, illustrated with a worked example, for *Instant Customer Analytics using only Excel*.



FIGURE 1 – Example Customer Data Analysis Summary

The example I will use is a retail banking customer dataset however the same principles and techniques apply equally well to any live customer dataset for any business sector.

To master this approach you simply need to understand the power of 3 excel functions - IF, COUNTIF (single criteria) and COUNTIFS (multiple criteria) along with their cousins SUMIF and AVERAGEIF.

If you don't want to learn how to use these 3 excel functions don't worry - you can also achieve steps 3-5 just using **excel pivot tables** provided you are prepared to invest a little time in understanding this powerful excel facility. Pivot tables are better shown than described so if you are interested in pivot tables I demonstrate them here in this 5-min screencast - <http://vimeo.com/34738136>

If you would like a copy of the worked example excel spreadsheet with all the formulae please contact me at www.servQ.co.uk.

My example spreadsheet, which would be extracted from your live customer system, has one row per customer containing just the following fields:

- Customer Reference
- Balance
- Region
- Interest Rate
- Date A/C Opened

The objective of the analysis is to understand this data to see if we can spot any useful business insights which could inform our strategy on customer acquisition and retention.

First Identify some questions you want answers to

Before you start any data analysis it's always a good idea to identify a number of questions you want answers to. At a minimum you need to answer these. However during the analysis the results will also suggest other questions you can answer which you won't have anticipated!

Step 1 - Get a good clean dataset to work with

There are a few things you need to be careful about concerning the copy of the live data typically provided for you by an IT person. You need to be clear in what you ask for particularly around the following:

- **Most recent month** - Make sure it is a full month's data - partial months may create misleading trends. So you might have to wait for the next end of month.
- **Oldest month** - Make sure the oldest month's data it is not also a "catch-all" for older data too - if you want older data make sure it is recorded as such rather than aggregated in with the oldest month.
- **Approximations** - Sometimes the exact data is not available but the programmer can make an approximation. For example, an approximate start date might be inferred from the date of the first transaction but might only be accurate to the month rather than the day. This is fine as long as you know what approximations are built in!
- **Multiple Record Types** - Sometimes you may need different record types e.g. Header Records for Customers and Detail Records for Accounts. Just make sure the different record types are clearly distinguished and can be processed independently of each other. Life can get tricky in excel if you have to keep jumping between record types within a single calculation!

Step 2 - Add the required Categorisation Columns

The problem with numeric fields such as balances or interest rates is that all we can really do with them in their raw state is to total them or average them. To go further we need to group them into bands and categories.

In the example I have created a new column (F) in the spreadsheet called "Balance Band" to group customers by balance by simply using the excel **IF** function as follows:

=IF(D2>1000000, "1M+", (IF(D2>500000, "500K-1M", (IF(D2>250000, "250-500K", (IF(D2>100000, "100-250K", (IF(D2>50000, "50-100K", (IF(D2>10000, "10-50K", "1-10K"))))))))))))

A similar problem arises with dates so I have created a new column (G) in the spreadsheet "Account Age" (in months) using the excel **DATEDIF** function.

	A	B	C	D	E	F	G
	Cust. Ref.	Region	Interest Rat.	A/C Balance	A/C Opened	A/C Balance Band	Account Age (M)
1	1	North	5.3	9542	03/06/2011	1-10K	7
2	2	North	13.8	70250	17/05/2011	50-100K	7
3	3	North	13.4	85000	17/05/2011	50-100K	7
4	4	East	2.6	42265	23/10/2011	10-50K	2
5	5	East	14.0	30000	21/02/2011	10-50K	10
6	6	East	13.4	507	26/11/2011	1-10K	1
7	7	East	13.4	40000	26/08/2011	10-50K	4
8	8	East	7.0	2089	21/04/2011	1-10K	8
9	9	East	7.0	2089	21/04/2011	1-10K	8
10	10	East	13.4	71258	26/09/2011	50-100K	3
11	11	East	13.4	6974	26/10/2011	1-10K	2
12	12	East	13.4	10494	26/10/2011	10-50K	2
13	13	East	13.4	36629	27/10/2011	10-50K	2
14	14	East	13.4	58476	27/10/2011	50-100K	2
15	15	North	13.4	5915	17/06/2011	1-10K	6
16	16	East	13.8	100000	20/10/2011	50-100K	2
17	17	East	7.0	2330	20/10/2011	1-10K	2
18	18	East	13.4	99100	31/08/2011	50-100K	4

FIGURE 2 – The Example excel spreadsheet

Step 3 - Perform the Basic Analysis

Now we are ready to begin the basic analysis of the data - first by simply *segmenting the data by a single field* such as Region or Balance Band. I do this using the excel COUNTIF, SUMIF and AVERAGEIF functions. It's important to do this first and to check that the totals (by rows and values) are correct before we move on to the more advanced analysis. It's a simple job then to graph the data using the excellent excel chart facilities.

For example, we can count the number of customers in North Region in Cell B3 with the formula '**COUNTIF(LIVE!\$B\$2:\$B\$101, \$A3)**' which uses the COUNTIF function in excel.

In our example the basic analysis (FIGURE 3) indicates that South Region accounts for 17% of customers by number but 30% by value and has an average customer balance 2-3 times greater than the other regions. It looks like South is unique in the fact that it is dealing with a small number of very valuable customers?

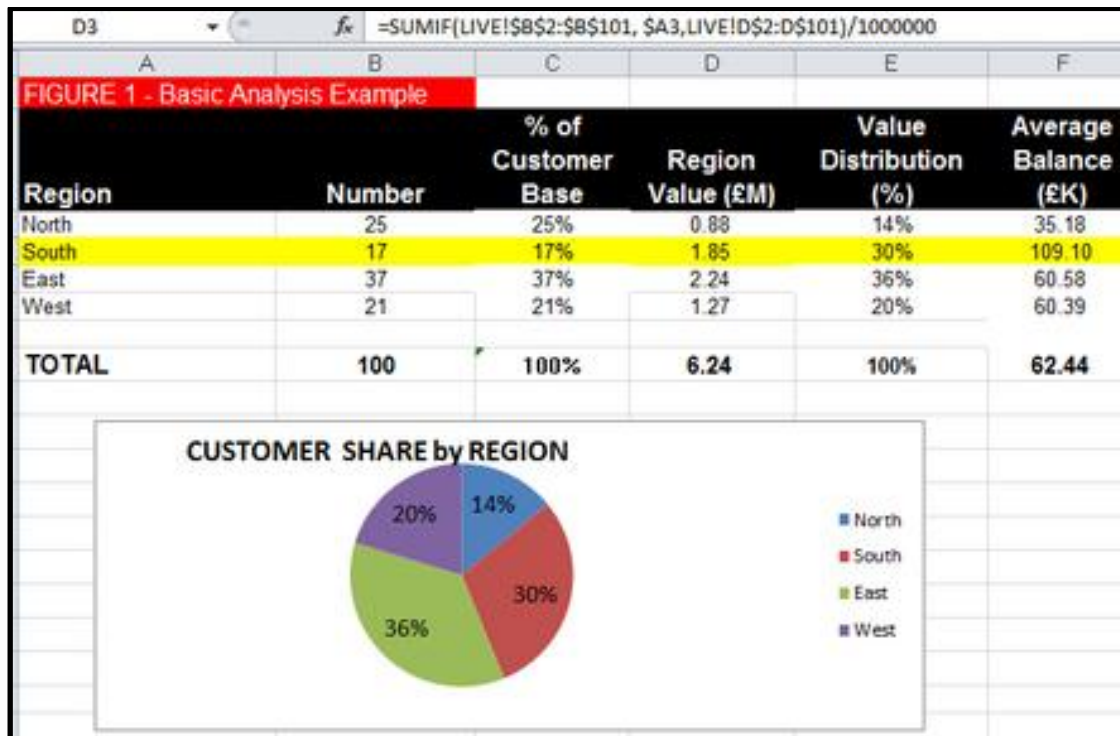


FIGURE 3 - Basic Analysis

Step 4 - Perform the Cross-Reference Analysis

It's unlikely the basic analysis will tell us very much we do not already know but we might be surprised. However when we start to do the cross reference analysis new insights will emerge.

This is where we segment the data by creating matrices of 2 fields with a **primary field** such as *account balance band* on the vertical axis which is then used to segment the totals or averages of the values of one or more **secondary fields** such as *interest rate* on the horizontal axis.

To achieve this we need to use an extended version of the COUNTIF, SUMIF and AVERAGEIF functions imaginatively named COUNTIFS, SUMIFS and AVERAGEIFS which allow us to supply multiple criteria. The simpler versions of the function are single criteria.

In our example the cross-reference analysis (FIGURE 4) shows that 56% of the total balances sit with just 7% the customers whose balances are greater above £250K but we are paying them less interest than the lower balance customers. Why is this? Does it make sense? Are we creating a problem? Again we repeat this kind of cross-reference analysis for all the fields in the data.

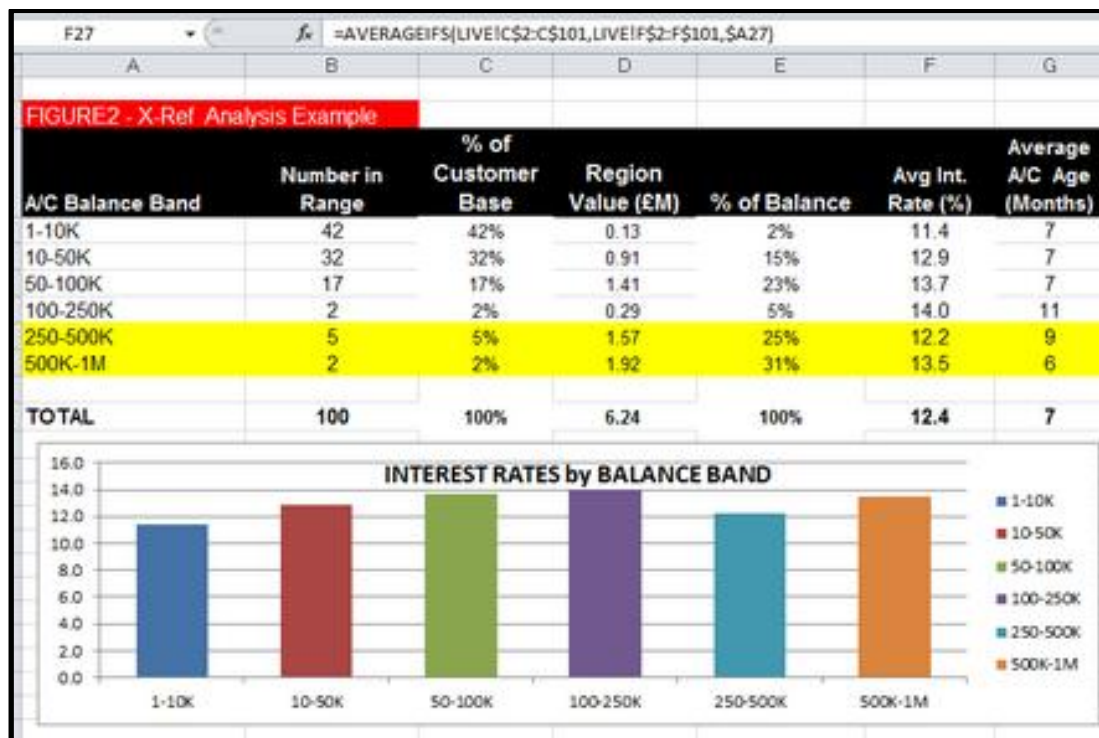


FIGURE 4 - Cross-Reference Analysis

Step 5 - Perform the Trend Analysis over Time

We now extend our analysis to include the time dimension to see how things change over time. For example, are the numbers and values of accounts opened growing over time or decreasing or static. How are the Regions doing against each other over time. Are a disproportionate amount of high-balance customers closing their accounts in the last 3 months. You get the idea.

For example we can calculate the average balance for each month for each region in Cell F473 with the formula '**SUMIFS(LIVE!\$D\$2:\$D\$101, LIVE!\$G\$2:\$G\$101,\$A47, LIVE!\$B\$2:\$B\$101,F\$46)/1000**' which uses the SUMIFS function in excel to sum across multiple criteria, in this case using "age of account" and "customer region".

In our example the Trend Analysis (FIGURE 5) shows that at an overall level the business is recovering well from a major problem. However when you look at this regionally it is clear that South Region's business has in fact totally collapsed but this has been compensated for by strong growth in East Region.

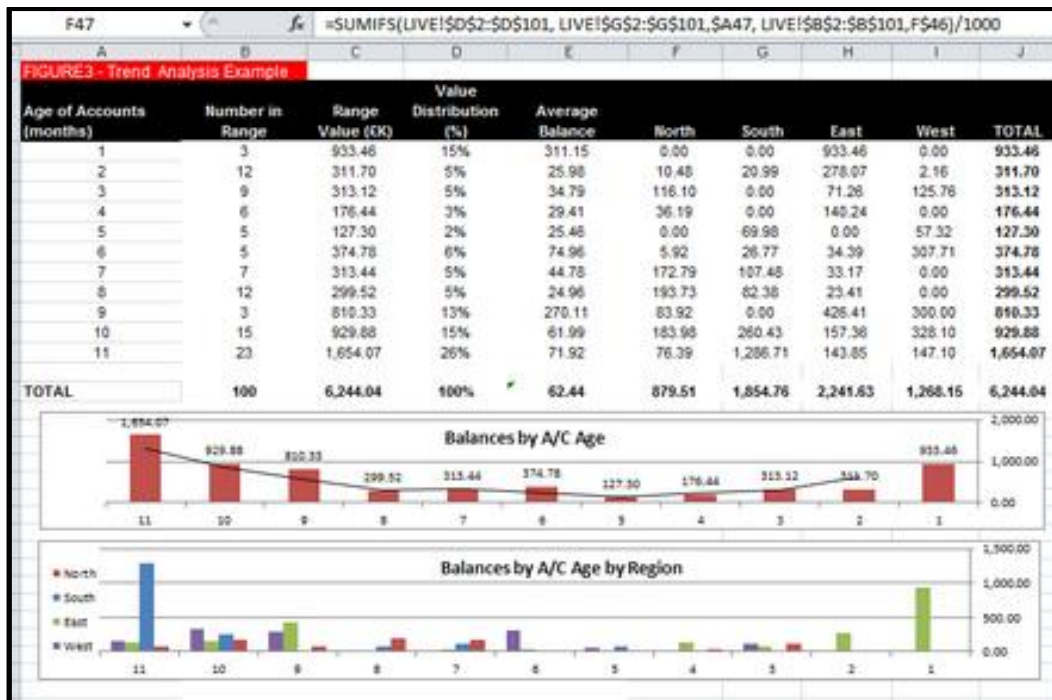


FIGURE 5 - Trend Analysis

Step 6 - Check your assumptions about the data

It's important never to do data analysis in a vacuum. First you need to check what assumptions you have made about the underlying data and confirm these with a person (e.g. in IT) who understands in detail the behaviour of the data in the live system which has been provided to you. For example, do closed accounts stay on the system forever or are they archived after 12 months? The answer to this will affect your results on account closure trends.

Step 7 - Review the Data Insights with Operational Staff

All you really have so far is a set of *potential insights* some of which may have real value to the business and others may be trivial, wrong or can be explained in other ways. You need to share these insights and the data you based them on with staff who are intimate with the operational details of the processes and you need to be open to have your conclusions challenged, destroyed, revised, improved or confirmed.

Finally ...Some Performance Tips

In my experience you can easily handle large data sets (up to 100,000 records) using these techniques. Once you go above 10,000 rows you should set formula calculation to "manual" rather than automatic to maximise spreadsheet performance. Above 50,000 rows you can make a copy of your spreadsheet with all your categorisation calculations converted to values (COPY, PASTE SPECIAL (VALUES)).

Above 100,000 records, if you need to, you can either create multiple spreadsheets with a consolidation sheet or else you can use the excel **RAND** function to extract a random subset (say 25%) of the data to analyse.

About Ken Thompson

Ken Thompson can be found at www.servQ.co.uk and delivers keynote conference speeches, workshop facilitation and in-house consultancy in key business areas including:

1. Creating High Performing Teams in enterprises including Virtual and Mobile Teams (based on the Bioteams Book)
2. Establishing effective Collaborative Business Networks enabling companies to co-operate effectively in areas such as sales and product development (based on the book - The Networked Enterprise)
3. How to use the latest social media technologies including blogging and online communities to promote enterprises, brand, organisation or event
4. Development of graphical on-line interactive Business Dashboards and What-if Simulators for organisations to support Performance Improvement, Strategy Development and Executive Team Development.