Analysis of the DWP working age customer base

Customer segmentation

17th November 2010

DWP Department for Work and Pensions

Executive summary

In response to the 2006 Capability Review, DWP made a commitment to understand its customers better and put them at the heart of everything it does. This analysis supports that commitment by creating a consolidated view of DWP working age customers from the wealth and richness of the data held in Departmental systems.

The analysis has drawn upon well-established principles to create ten segments based on shared characteristics such as length of time on benefits, demographics and situation using data drawn chiefly from DWP administrative data and supplemented with select external socio-demographic profiling information.

It provides DWP with an alternative to using type of benefit and life stage as the primary means by which to define customers.

1 Introduction

1.1 Context

The Department for Work and Pensions (DWP) is responsible for promoting work as the best form of welfare, helping people into work and supporting those who cannot work. In addition, one of the current Departmental priorities is to simplify the welfare system so that work always pays and the momentum of the system is always into work for those who can.

In response to the 2006 Capability Review, DWP made a commitment to understand its customers better and put them at the heart of everything it does. This analysis supports that commitment by creating a consolidated view of DWP working age customers from the wealth and richness of the data held in Departmental systems.

It provides DWP with an alternative to using type of benefit and life stage as the primary means by which to define customers. The ability to define and understand customers by other characteristics has become increasingly important with the decision to introduce Universal Credit (a single income-related payment).

1.2 Analysis objective

The objective was to identify distinct and appropriate working-age customer segments driven by Departmental objectives and based chiefly on DWP administrative data. In order to examine the capacity of this approach to inform a range of business needs such as strategy and operations, the ability to perform robust drill-down analysis and present the findings geographically were also requirements.

2   Methodology

2.1 Data

DWP Information Directorate (IFD) statisticians undertook the analysis working in partnership with the Departmental Customer Insight team. IFD is the central source of DWP information and plays a key role in helping to make the most effective use of this data, providing a shared information service to the whole of DWP, as well as meeting wider needs for DWP statistics. IFD develops and maintains an integrated source of administrative data that covers DWP customers and the outcomes which the Department helps to achieve for them; products and business processes; staff and many other aspects of its business.

The primary data source for this analysis drew upon elements of the Work and Pensions Longitudinal Study (WPLS). This data is based on 100% of claimants and covers information such as age and gender of claimant, duration of their spell on benefit and geographical locations of claimants.

This data was linked with a number of internal data sources. While chiefly based on Departmental data, the segmentation was enhanced by external data procured from Experian's Consumer Dynamics Database. This external data was used in the main to profile and describe segments. It supplemented Departmental data where no comparable information was available in order to give a more complete picture of our customers. Data sources are outlined in more detail in the annex.

The final database contains a cohort of around one million customers randomly selected from the September 2009 WPLS who were in receipt of a working age benefit.

2.2 Segmentation method

Segmentation is sometimes described as a combination of art and science. The technical analysis drew upon well established principles to create a number of robust segmentation options for consideration. Non-technical DWP staff with a wide range of responsibilities and experience were also involved in order to agree the most useful and business-relevant solution.

As with the majority of statistical analytical intelligence conducted by DWP this project was delivered using SAS Business Analytics. For each of the individuals in the cohort, information was appended that would drive apart and identify segments on the basis of getting people into work, paying benefits and supporting our customers. This selection struck a balance between including too much information in the build resulting in an unfocused solution which would be difficult to action, and too little which would not to cover off all the key aspects.

Individuals were then assigned to similar subsets based on the chosen information by performing (K-means) cluster analysis using the data mining application SAS Enterprise Miner. The chosen approach, endorsed both by internal and external peers, tries to form groups of observations such that the observations within a group are close to each other. By "close to each other", we mean that the distances between the members **within** a group should be small compared to the distances **between** the groups. It does this through iterative estimation and assignment to groups. The main advantage of this approach is the simplicity and speed when it is run on large datasets. This approach can also produce tighter clusters than other methods.

Determining the number of clusters in a data set is a frequent problem and is a distinct issue from the process of actually solving the clustering problem. The number of solutions must be small enough to be managerially relevant but large enough to warrant strategic attention.

Clustering via this approach is an unsupervised classification. There are no predefined classes and any choices are typically highly subjective including analytical consideration such as giving more weight to certain information rather than others. A number of cluster solutions were created for consideration. The solution eventually selected was deemed to be the most distinct and business-relevant.

The resulting analysis segments the DWP customer base based on shared characteristics such as length of time on benefits, demographics and employment history.

3   Outputs

The chosen solution consists of ten customer groups. There follows descriptions of the typical characteristics of each segment based on the data contained in the compiled dataset. The segment names and descriptions are intended to help segmentation users conceptualise the segment based on its tendencies rather than pigeon-hole customers by absolutes.

3.1   The ten segments

1. Young jobseekers

These tend to be predominantly males who are mostly still living at home with their parents in modestly-valued urban housing. Most have had some form of employment in the past and their current spell on benefits is typically the shortest of all the groups.

2. Recession-hit middle class

These tend to be affluent middle-aged workers, primarily men, who have recently left employment, possibly due to recession-driven redundancy. They typically live with their family in larger, mortgaged properties in the outer suburbs or rural locations. Half are claiming Jobseeker's Allowance (with smaller numbers in receipt of Incapacity Benefit), but very few are eligible for further means-tested benefits.

3. New to sickness benefits

These tend to be middle-aged claimants of sickness benefits and/or Carer's Allowance. Women are over-represented in this group. People in this segment mostly live in fairly low-value properties with their husband/wife, older children and possibly extended family. They typically have a moderate level of income and savings and a small amount of outstanding mortgage and do not tend to claim means-tested benefits.

4. In-and-out of work; better off

These tend to be working-age urban males with an intermittent work/benefits lifestyle. In terms of affluence, they are typically relatively comfortable but may have a poor credit rating, reflecting their lack of permanent employment. They are mostly single and tend to live either alone or with a friend/home-sharer in inner urban areas. Few rely on social housing however.

5. In-and-out of work; poor

These tend to be poorer working-age urban males with a low skills level and an intermittent work/benefits lifestyle. They are mostly single and tend to live either alone or with a friend/home sharer in inner urban areas. They typically have high levels of financial stress and nearly all claim supplementary means-tested benefits such as Housing Benefit.

6. Lone parents

These are almost exclusively single mothers aged between 18 and 44, living either alone or with friends/relatives. This is one of the poorest segments and as such almost all are in receipt of Income Support and supplementary benefits. Most have worked in the past.

7. Working class, 'retired' onto sickness benefits

This tends to be an older group mainly claiming sickness and/or disability benefits. They are typically of modest means, living in 'empty-nester' households, though a third lives on his/her own. These are generally people who have been on sickness-related benefits for one long spell and though they have typically had a strong work history their advancing age, declining health and a belief in a lack of appropriate work tend to diminish their ability and motivation to return to the job market.

8. Empty nesters supported by spouse

This is a predominantly middle-aged female segment either in receipt of or providing care. This segment is one of the least affluent types; however, few claim supplementary means-tested benefits, possibly due to a partner being in employment. This segment tends to be far from the labour market as many have had no recent employment history.

9. Middle aged, middle class, long term sick

This tends to be an older group claiming disability or sickness-related benefits. However, this is one of the most affluent segments, typically occupying high value housing and living a relatively comfortable existence financially. These tend to be people who have not worked for a long time due to the onset of chronic health problems. With a third of this type approaching retirement many may have mentally 'retired' already.

10. Long term benefit dependent

This tends to be an older and vulnerable group characterised by a very long spell on sickness and/or disability benefits. They are typically men and women who tend to live alone in urban social housing, with a large proportion living in inner cities and high rises. This is one of the poorest and most deprived segments, tending to be far from the labour market and highly benefit-reliant, with almost all receiving means-tested benefits.

3.2 Geographical presentation

The final database which sits behind the segmentation also permits the findings to be examined geographically. The following provides a regional comparison.

3.2.1 Proportions in each segment broken down by region.

Figure 1 below illustrates the similarities and differences in the customer base between regions.

The width of the bubble represents the proportion of the sample database within that particular segment in that area. For example the 'long term benefit dependent' is the largest segment in the North East.

The depth of colour of the bubble represents how that segment proportion compares against other areas. The darker the colour the more overrepresented that segment is when compared to other areas. This is based on an index score. For example in the North East the 'in-and-out of work; poor' are over-represented compared to other regions. This shows that while the largest segment in the North East is the 'long term benefit dependent', within this region you are move likely to find the 'in-and-out of work poor' than in any other region.

Figure 1. Proportions and indices of segments by region.
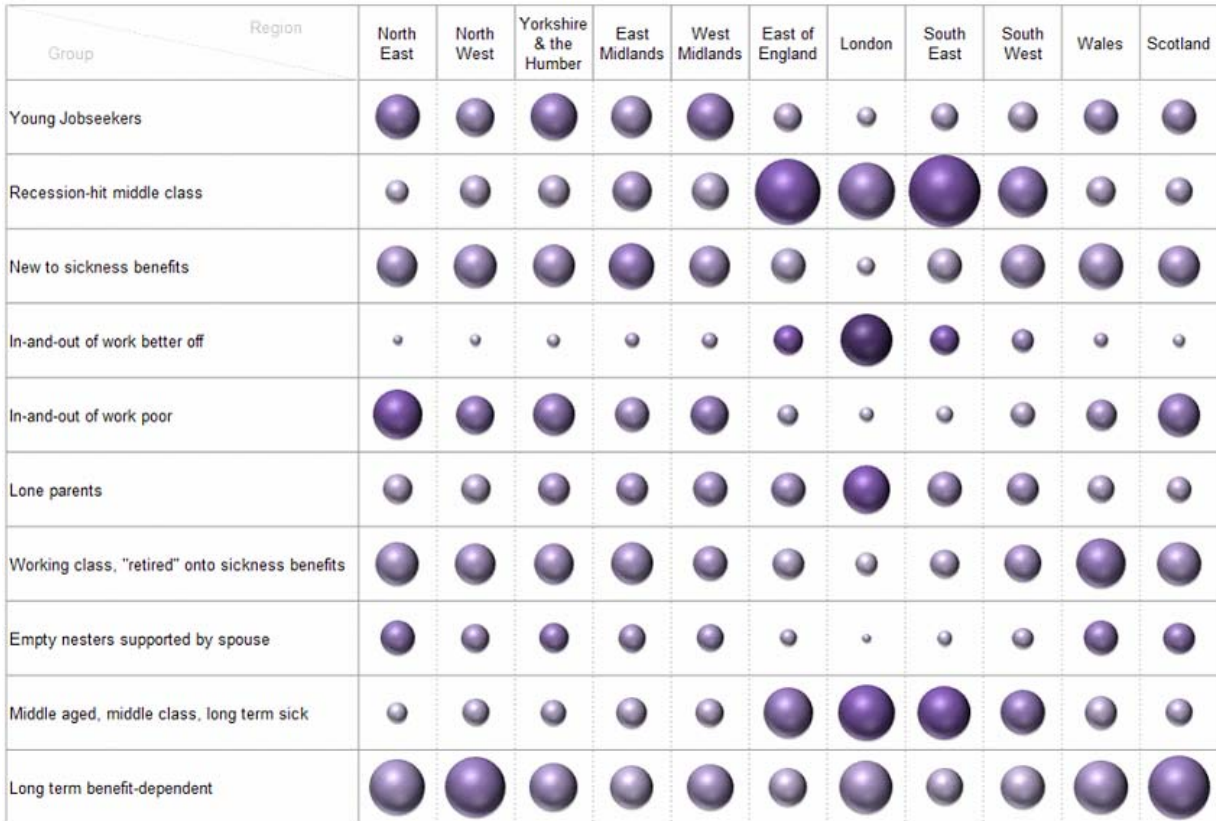
Indices are calculated by comparing proportions of segments in a given region against the proportions based on **the rest** of the sample combined (**excluding** the region in question).

For example:

If we have a score of 60% in region x and 30% on the remaining sample combined, the indexed score will be:

(60/30)*100 = 200. So region x has a score twice as high or 100% higher than the rest of the regions combined.

Similarly if we have a score of 10% in region x and 40% on the remaining sample combined, the indexed score will be:

(10/40)*100 = 25. So region x has a score four times lower or 75% lower than the rest of the regions combined.

Figure 1 highlights similarities between the northern regions of the country and the differences with those regions further south. It can also be seen that although London shares features with the neighbouring regions of East of England and the South East, it does possess certain distinctive characteristics of the customer base.

4 Summary

This analysis proves that it is possible to successfully create distinct and meaningful segments of DWP customers from DWP administrative data and which are not solely based on primary benefit type. The resulting segmentation is an important asset as DWP develops and implements plans for the Work Programme and Universal Credit. In addition, the experience gained and lessons learnt will be invaluable for future segmentation development within DWP.

The work continues to be developed and refined and the additional options it faciliates are being examined.

5   Annexes

5.1   Data sources

5.1.1 National Benefits Database (NBD)

The 100% NBD holds information about clients' claims and spells on the main DWP benefits from June 1999. The data contained comes from the different benefits systems.

5.1.2 Single Housing Benefit Extract (SHBE) Database

The 100% SHBE data is collected on a monthly basis and contains information about Housing Benefit and Council Tax Benefit claimants returned to the department by each Local Authority. The SHBE in its current format came into effect from April 2008 with collection commencing in August 2008 following the department wide data embargo.

5.1.3 Basic Skills

This dataset includes individuals who engage (via Jobcentre Plus intervention) in any part of the basic skills process. From April 2001, clients needing help with basic skills can access suitable training. The dataset is used in evaluating client literacy and numeracy levels and the numbers screened, attending independent assessment or starting a course.

5.1.4 Labour Market System Client Extract (LMS CE)

The client extract was introduced in 1998 and holds a record for every individual who has had interview details recorded by either front line or call centre JobCentre Plus staff. This data contains an indicator to show if a claimant has a low qualification level.

5.1.5 P45/46 Databse 2 (P45DB2)

The databases contain information taken from an HMRC administrative system. Data is collected via P45 and P46 forms returned by an employer when an individual joins or leaves an employment scheme (this may cover PAYE employment, JSA, IB or Occupational Pensions).

5.1.6 New Deal Evaluation Databases

The databases are derived primarily from information stored within the Labour Market System, but also use information from other administrative systems, such JSAPS, to provide statistics on starters, participants, leavers and destinations of New Deal.

5.1.7 Experian Consumer Dynamics Database

The database comes from Electoral roll data sourced directly from local authorities augmented with the data from responses to Experian lifestyle surveys, third party consumer data shared with Experian and public registers, such as company director lists, shareholder lists and County Court Judgement information.