Time to fly: An approach to segmentation and modelling customer dynamics in the travel sector

Received (in revised form): 4th March, 2004

Nigel Magson

co-founded Talking Numbers Data based solutions Ltd in 1996, to provide database consulting and data analysis services to commercial and not-for-profit organisations. With David Dipple, he has many years' experience in the application of data for marketing. He is a Fellow of the Institute of Direct Marketing.

David Dipple

is co-founder of Talking Numbers Data based solutions Ltd. Like Nigel Magson, David is a regular contributor to industry journals and speaks on database-related topics at conferences. He is a Fellow of the Royal Statistical Society.

Abstract The aim of this paper is to illustrate the modelling approach used by Talking Numbers, specifically focusing on customer behaviour and dynamics, within the travel sector. Although the approach is marketing-led, it follows clear stages, which are described with real life examples. The approach is not unique and can probably be described as a 'classical analysis approach'. The authors feel it is worth reiterating the benefits of this approach, and the amount of human intervention and the understanding which it entails by comparison with 'black-box' software, which claims to cut the model build process to a matter of minutes! This approach does not use necessarily complex 'black-box' analysis engines but emphasises the requirements for analysis definition and data preparation. There is scope to stage projects to be budget-conscious. The approach follows a number of stages, which are illustrated below.

TRAVEL SECTOR INTRODUCTION

The job of the travel sector marketer might be described as 'to create and convert demand according to the available supply at the best possible margins'. Travel companies operating within the UK consumer market, either directly or via the 'trade' ie tied or independent chains of travel agents, are seeking to attract existing and new customers (bookers) through a range of media.

The travel sector promotional cycle has a standard if not formulaic approach to securing bookings year on year.

Typically, the traditional showcase for companies is the holiday brochure, which follows a promotional cycle. This involves the rounds of brochure previews, enquiries, early booking offers, follow-ups and post-Christmas send outs of material, filling the available capacity. While for many operators within the industry the 'brochure marketing cycle' drives the targeting of activity in large pre-planned mailings in August/September, with a consequent consumer education to expect this cycle, the use of ongoing tactical activity plays an important part in filling specific

Nigel Magson
Talking Numbers Data
based solutions Ltd.,
Parabola House,
Parabola Road,
Cheltenham,
Gloucestershire GL50 3AH,
UK
Tel: +44 (0)1285 644220;
Fax: +44 (0)1285 644226;
e-mail: nmagson@
talkingnumbers.com;
ddipple@
talkingnumbers.com;
web:
www.talkingnumbers.com

occupancy. As the promotional season advances, tactical activity and often discounting is pursued to fill 'late availability'.

Since this brochure process is less well 'honed' for targeting pre- and post-peak periods, or niche segments and destinations, status triggered activity and the modelling of late availability is important to supplement bookings. Within the industry however, this is slow to emerge due to the practicalities and constraints on systems. Talking Numbers has initiated this approach with a number of clients. This will be covered in more detail later.

Systems

Most travel/leisure organisations are reliant on their reservation systems and the handling of associated booking details. The primary entity is the booking product and not necessarily the customer(s) that relates to the booking. As enquiry and brochure activity can arise from multiple media channels, and quite frequently can be outsourced, the integration and tracking of this information back to the booking can be limited. More organisations have now realised that a consolidated customer view is needed and have developed marketing databases to support direct marketing activity and analysis. There are few organisations within the travel sector performing customer relationship management (CRM), where front office access to data from a consolidated marketing database system is held and modelled to re-inform front office staff to influence at point of booking or to drive key communications. This will be realised by one of Talking Numbers' clients in 2004 using links created from the hosted database system to the front-end reservation system. There is a whole string of advantages to this,

starting with the empowerment of the salesperson through to the delivery of dynamic segmentation and updating based on behavioural response. (Nb, this goes way beyond holding a database of known booking history and communications, as held within many reservation systems, to coming up with dynamic matches of the types and times of holiday available which suit this customer from modelled algorithms). Needless to say there are considerable systems investment and inevitable integration issues to cope with, as well as ongoing tuning of algorithms and monitoring.¹

The travel sector has a number of key factors that affect the overall level of demand for bookings and the profitability of different organisations. These demand and supply factors can be summarised as:

- Time: seasonality, stay duration
- Price: booking value (price per person per night), discounting, period pricing
- Property/accommodation: mix and availability (occupancy)
- Geography: distance to point of holiday
- Customer booking behaviour and dynamics
- Customer and enquirer demographics
- The holiday experience — often described from attitudinal information from post-holiday questionnaires
- Competitor and market climate
- The climate! — external factors — economy, world events (eg 9/11) and weather.

Their description, importance and weight may vary across different travel sector clients, but these remain the core factors influencing demand and supply.

The inter-relationships between these factors can be understood and manipulated through analysis to provide a profitable (re)distribution of demand

Figure 1: Demand and occupancy curves

according to supply availability. This is undertaken through analysis and modelling within the customer database. The management of supply variables using yield management techniques, however — although part of the marketing mix — is rarely considered in relation to the main engine for generating consumer demand: the marketing database.

Modelled occupancy

The 'demand' led approach followed by most organisations often results in over- and under-demand for certain products/tours at points in time. As there is normally a constraint in capacity — linked to accommodation, flights etc — the way of regulating demand to this point is price. A 'supply' led approach of reconciling the supply or bought capacity with modelled demand and customer segmentation offers an alternative approach to generalised targeting techniques. Yield management² concepts can be applied to matching the remaining known or bought available supply with key segments, maximising the amount of return on these slots requires the modelling of customer and enquiry base data to fit a tactical

position. The modelling of discounting strategies in conjunction with the time point is what large package 'haul' operators have traditionally done but, these models have generally not linked cohesively back to the known customer or prospect base and engendered a consequent impact on communications and promotions.

Figure 1 illustrates imaginary demand curves and where promotions need to be targeted to raise occupancy at specific segment levels and with a specific property mix.

For operators, as the season develops, there are certain entities (destinations or products or properties) which have lower occupancy. The reverse profiling of these entities and the matching of them to prospects based on their segmentation and scoring, allows a greater focusing of availability to the customer base. Reverse profiling alters the normal question — what is the profile of customers and prospect? to: What profile is needed to fill specific marketing conditions?³

TALKING NUMBERS' ANALYSIS APPROACH

The following stages represent the approach Talking Numbers (TN) follows

in delivering analysis. Each of these stages are expanded in the paper, within the context of the travel sector.

- 1 Formulation of the analysis brief
- 2 Data preparation
- 3 The analysis
- 4 Interpretation
- 5 Assessment and review

These stages involve close collaboration between 'client' and analyst. The strength of the analysis lies in the close definition of the marketing requirement and a marketing understanding of the data to be gathered and defined by both sides. Whilst this can be more time intensive than loading all the data into a 'black box' engine and allowing a technical solution to emerge, the application of the results and, therefore, the longevity of the solution are stronger. This is arguably a 'traditionalist' approach. TN finds that the first two stages of the process account for between 60 and 80 per cent of time input into most projects.

FORMULATION OF THE ANALYSIS BRIEF

Albert Einstein commented that: *'The formulation of a problem is often more essential than its solution which may be merely a matter of mathematical or mental skill.'*³ The initial business and marketing objectives need close definition, as well as the development of the brief. The underlying business driver for the analysis, as well as its likely returns and importance, are key for the definition stage. Importantly, this is developed in close collaboration with the client. TN often cites the well-known first golden rule to applied mathematics:

'An approximate answer to the right question is worth a great deal more than the precise answer to the wrong question'

Within the travel sector, specific briefs or requirements can vary in breadth and the 'tightness' required of the outcome. For instance, a specific model aimed at raising the responsiveness from a brochure mailing has a specific requirement measurable by gains modelling. Alternatively, general investigations into a customer base and developing an ongoing communication programme through generic segmentation may have less clear bounds than a tactical application, although multiple application points.

Outcomes from activity can ultimately be measured by a series of relationship points and the cost and return metrics it takes to get there.

Figure 2 gives a diagrammatic representation of the enquiry/booking process. The position can be traced at each point in the enquiry process and the possible outcomes investigated. For example at the 'Follow-up brochure send out' stage there are two possible outcomes⁵ the first is that no action will be taken and the second (2d) is that there will be a follow-up phone call. Each outcome may attract different analysis profiles.

At the formulation stage the data requirements needed to support the analysis, the sources of data and an assessment of their quality and likely ability to deliver the required solution are examined. This will involve auditing data sources and doing preliminary descriptive analysis. The end solution and the deliverables from the analysis work should be defined.

At this stage, however, it is important for both 'client' and 'analyst' to be realistic in terms of expectation of what is achievable. It may be necessary to modify or even abandon the project after initial investigations within the data. This might occur where there are low levels of data or if little discrimination between

Figure 2: Diagrammatic representation of a model enquiry/booking process

the target population and a base population is uncovered. Even in these instances, however, preliminary investigations may provide insight into future requirements for data and help clients modify data strategies.

DATA PREPARATION

Data preparation is generally a significant process of any analysis project. In the holiday sector the resolution of multiple data sources and the creation of an agreed entity level can be complex.

The data may come not only from multiple sources but also from relationally-structured systems, where there are one to many relationships defined between tables. For analysis purposes, these structures generally require flattening to allow for ease of

creation of summarised data. Clearly these structures and the linking keys between tables need to be understood, but the information may also need 'lifting up' in a particular sequence so that fields are correctly aggregated.

Figure 3 illustrates table information built from a reservation system where sub-booking table information requires aggregation up to booking level before bookings are aggregated to the customer level. This level may, in fact, be the 'household', which is often different to individuals who inhabit the household. Although marketing database systems can cope with this distinction, often the resolution of this from a reservation system to a marketing database is poorly handled due to insufficient rule definition and data management techniques. It is worth noting that the same situation can

Figure 3: Simple entities from travel systems

Note: 'Booker' is an entity that includes enquirer in this instance.

occur between enquiries and enquirers, and whether they are attributable to a household or individual.

At this point in the data consolidation process it is also necessary to form a business rule set where certain information, if conflicting, is retained, or the information is held and treated differently. Again this a crucial difference between the way reservation systems are structured to 'view' data and that which might be taken by a marketer. For instance, the marketer might market just to one household. In the reservation system, however, on one occasion a household may produce a party, which is a couple; on another occasion the same household may produce a family party or, in fact, the household might be part of a group booking. Two bookings (back to back or with a very small time break) might be made for different products, but a decision might be taken that this represents a single holiday and, therefore, to process these two bookings into one holiday for analysis purposes. All of these instances may be valid for exploration and need recognising through rule definitions that differ from the way a booking would be treated in a reservation system.

There are other processes, that occur within the data preparation process. Handling missing data is an area in its

own right.⁴ Other processes may include cleansing of data to identify duplicates and consolidate information between them. The creation of the analysis dataset will normally also involve the addition of new variables: eg geodemographic data or proprietary profiling systems (eg the PRIZM system from Claritas mentioned in Figure 3); various recoding of 'old' data items; turning continuous data into categorical data; the ability to generate binary information for certain fields; integrating coding descriptions etc.

The data preparation process is unique for each set of client data and, for reasons already stated, will not necessarily produce a set of counts reconcilable with the pre-prepared information. The end dataset is usually replicable from similar supplies of data. The checking of the syntax used is very important to ensure, and have confidence in, the accuracy of this process. Field and record auditing and verification back to supplied information provide further means of testing the veracity of the preparation.

This process also provides the analysts involved with an understanding of the dataset which will be carried forward into the main 'analysis' phase. It will help start the formulation of the appropriate techniques to apply, as well

as identifying where there are 'suspect' areas within the information.⁵

THE ANALYSIS

The analysis phase of a segmentation project is where the analyst combines statistical and analytical expertise with marketing knowledge. It is important that all of these elements are used so that the resulting segmentation satisfies the marketing requirements. It is important that any analyst working on the project has a good knowledge of the marketing objectives of the client involved; the days of back room analysts 'beaver away' on unknown data whilst a consultant oversees the project should be consigned to the dustbin of history.

Historically, segmentation solutions have been produced from behavioural information that concentrates on recency, frequency and value (RFV), or from attitudinal information gleaned from focus groups or small scale telephone surveys. Both of these methods can produce useful descriptors but also have limitations.

Segmentation solutions produced using behavioural information are relatively easy to produce on a marketing database but may not give an agency sufficient differentiation for marketing communications. Similarly, segmentation solutions produced using attitudinal information give group descriptions much loved by marketing agencies (for example — culture vultures or happy campers), but are virtually impossible to apply to the marketing database for selection and targeting. To satisfy everybody is often a difficult task but is not unachievable.

Segment assignment

There are four main types of data encountered when producing a segmentation solution:

Interval

Interval variables are ones that can be added together, divided or multiplied. Some examples of interval variables are:

- Age
- Income
- Life time value
- Propensity score

For instance, a person aged 40 is twice as old as somebody who is £20; if a person has taken one holiday for £500 and another for £1,000 their total spending is £1,500.

Ordinal

Any data that can be ranked or ordered is known as ordinal data. Some examples of this data type are:

- Age group
- Satisfaction scale: not satisfied, neutral, satisfied
- Value band

There is, however, a distinction between ordinal and interval data. Ordinal data cannot be aggregated in the same fashion as interval data; that is, a certain amount of precision can be lost when information is grouped.

For example, if a person has had two holidays costing between £500 and £1,000; if the two values are combined it will not be known if the total holiday value will exist in the £500 to £1,000 or over £1,000 groups.

Categorical

Categorical data can be thought of as labels applied to data to make them more understandable. Categorical data items cannot be added together in the form of a sum.

Figure 4: A hypothetical value segmentation

Some examples of categorical variables are:

- Gender
- Marital status
- Name/address
- Key customer group

Spatial

Spatial data resemble categorical data in that labels are given. Spatial data differ from categorical data in that there is a positional element which is difficult to see unless a map is used. A simple table cannot usually explain the relationships between data points. Some examples of spatial data include:

- Postcode
- County

Spatial information is especially important for holiday companies as the information can often be hub-based data. Most holiday companies are trying to get customers to go to specific places or hubs; these can be holiday parks, areas or even airports. The distances that customers are willing to travel to a hub is affected by numerous factors such as type of hub, party make-up or even seasonality.

Combining information for segment formulation

Segmentation solutions are usually created using one of two methods. The first is to create a discrete solution where customers are assigned to a group depending on individual elements or a hierarchy. Figure 4 shows a hypothetical value segmentation using gender and income. A basic RFV segmentation is often created using this regime. When creating this form of segmentation it is often necessary to compress data into a small number of groups or the solution can become unwieldy.

Behavioural segmentation — An example

After initial analysis, it was agreed that a pre-determined construct of variables eg recency and frequency should be applied to the historic transactional file.

These provided a recognisable marketing construct of named segments for client and agency and also 'marketable sizing'. A snapshot method was used to trend movement and sizing of these groups, eg showing increase in the first timers (debutants) segment compared with devotees.

A wide range of descriptive and

Table 1: Model RF segments

Segment	Definition
Debutants	Customers who have only ever taken one holiday. This was also split into new and old debutants to be able to differentiate between newly-recruited customers and those who had not returned for some time
Devotees	Loyal customers, who have taken a minimum of three holidays with one of these taken in the last two years
Deserters	Any customer who has taken more than one holiday but none in the last three years
Drifters	Anyone that is none of the above

Table 2: Key metrics and their variance by segment type

Factors	Devotee	Deserter	Drifters	Total
Length of holiday (days)	13.30	12.80	12.47	12.76
Holiday value (ratio)	1.04	0.95	1.02	1.00
Party size	3.98	4.08	4.08	4.06
Time between booking and holiday (months)	7.51	6.51	6.61	6.76
Average LTV (ratio)	1.64	0.76	0.90	1.00
Number of holidays	4.97	2.48	2.70	3.09

Note: LTV-Lifetime value

profiling techniques was employed to understand these segments. See Table 1.

value to the total value. For example, the LTV of the devotee group is 1.64 times that of all multi-holiday customers.

Booking behaviour

There are a number of transactional data elements that can be used to derive key metrics and information relating to an individual's booking behaviour.

Table 2 shows a number of these derived metrics and their variance by segment type, eg booking lag to holiday. Debutants are not included within this Table as they have only had one holiday. Certain characteristics, when identified, can have interesting application back to the marketing programme. For instance, pre- or post-Christmas booking behaviour is a strong factor which was observed as quite 'staid', ie the customers are likely to remain with their opening behaviour, and hence this is useful in the timing of communications. Some of these variances can become even more marked at a sub-segment level.

The lifetime and holiday values are expressed as the ratio of the segment

Segmentation Levels

The example given above assigns an individual or household to a single segmentation group. The definition of appropriate entity levels is important within the travel sector and can make modelled solutions quite complex. Consider, for instance, that individuals can book with organisations offering a wide variety of different types of holidays — short break/long break, short haul/long haul, family/couple/group — sometimes:

- Just for themselves;
- For themselves and partner;
- For their family; or
- For extended family/group.

The enquiry may have come from a different point within the household. Should behaviour be modelled for individuals or households? This is a

Figure 5: Seasonality segments

business definition, but it is possible to create segmentation solutions where a 'customer' can exist in more than one group. This is often referred to as a fuzzy segmentation solution. This is one of the key differences between the 'black box' and the analyst, who can decide that in this instance a mathematical cluster solution is not appropriate. The analyst can choose to employ a fuzzy marketing segmentation model, which better suits the marketing requirement. This leads to 'understandable' marketing descriptions rather than a series of complex mathematical interactions which lack marketing focus. It uses an intuitive set of marketing rules that can be used to describe groups within the data. Although segmentation created in this manner may not be the optimal solution in statistical terms, it offers a better balance between marketing and statistical requirements.

For instance, consider a segmentation system that includes party size or life stage as one of the key factors, which is highly likely in a holiday scenario. It is known that a company's visitors might, in fact, come as a family for a particular

time (eg school holiday period), but that the same visitors might come again as a couple, or with an extended adult group party and the children might also come with grandparents. In this instance a fuzzy indicator or sub-segment approach would be needed to identify this marketing opportunity. (See Figure 5.)

Similarly, seasonality scoring is quite often kept as fuzzy: ie a propensity for a customer to come in a number of periods may be high or low.

Periodic assessment and evaluation of the segments, and overlay accuracy and performance, can be undertaken using standard statistical techniques such as discriminant and gains charts.

Creating fuzzy and multi-level segment solutions

Moving to this more sophisticated view of customers and enquirers requires the use of more advanced statistical analysis and modelling. A case study example is given below.

Previously the organisation had used a type of recency frequency (RF) segmentation whilst it had also used

Table 3: Lifestage by holiday frequency

Lifestage	Number of holidays		
	One	Two	Three +
Young	Young — one holiday		
Middle aged			
Older	Older — three holidays		

more sophisticated models for specific campaigns and tactical promotions.

Whilst an RF + lifestage system could be adopted along similar lines going forward, it has already been noted that there are a number of known factors which differentiate this business model to other holiday models. These include:

- As the product is point-based within the UK, catchments and proximity to site can be a strong factor for certain groups and at certain times of the year.
- Spend and profitability. Given the high occupancy rates achieved, one of the objectives of the segmentation is to ensure that holidaymakers with a higher on-site spend are attracted over those with a low spend.
- Break type, length and number. Mid week vs weekend, short vs long break, single vs multi.
- Accommodation type and mix for each destination point. This is potentially highly important as over-demand for one type can lead to inefficiencies, eg booking lags for larger groups are longer rather than smaller.
- Booking behaviour. There are a number of dynamics and metrics that can be understood, eg late booking propensity, repeat booking at site etc.
- Attitudinal information from the customer satisfaction questionnaires and other sources.

In terms of having a manageable segmentation system which provides ease

of reporting and understanding with sensible marketing economies of scale, there was a discussion about whether there should be a limit on the number of main segments, although these could be ‘cut’ as necessary by sub-segments or scores.

Multi-level segmentation

It was agreed with the client and their agency to understand the available data and profiles and to produce a segmentation model that satisfies both the criteria that it can be used on the marketing database and for marketing communications.

It was decided that an nxm dimensional segmentation model would be unwieldy and difficult to use. This is because the necessary number of potential segments would be too large to be meaningful or be multi-dimensional.

Table 3 shows the nine possible segments that can be created by combining life stage and numbers of holidays.

A multi-level segmentation uses a set of defined descriptors, sub-segments can be linked to these in the form of a dendrogram or tree-diagram. A multi-level segmentation system requires a top level solution that is used to anchor it.

From the profiling analysis that was carried out, it could be seen that a viable top-level descriptor would be customer lifestage: The PRIZM Household classification system from Claritas was used. These are factors that can be both accessed from the database and used to

Figure 6: Lifestage groups

differentiate marketing communications and target specific groups of customers. If the segmentation were restricted to this level it would be too blunt a tool and not detailed enough for fine-tuning marketing communications. For this reason sub- or pseudo-segments were also used.

The top-level segments used are:

- Young couples
- Young families
- Established families
- 'Empty nesters'
- Retired (see Figure 6).

These are then further broken down into sub-segments for descriptive and marketing purposes (see Figure 7).

One of the advantages of this method of segmentation is that retrospective analysis can also be carried out in order to understand the behaviour of both large and niche segments within the database. This analysis can take the form of size, value, responsiveness as well as

other summary statistics such as party make-up. The results can be further broken down by site if required.

The database can be set up to produce a multitude of informative reports that can be used to understand the dynamics within the population. It also enables the marketers to understand potential within groups that would otherwise not be part of the key target audience. For example, if the marketing objective is to maximise on-site spend, who are the best prospects for contact? (See Figure 7.)

Calculating expected payback

The segmentation can be incorporated to supplement conventional travel marketing metrics such as cost per booking. Party demographics, value and responsiveness can be used to understand the expected value of contacting a segment (see equations 1–3).

$$\text{Spend revenue} = \text{on-site spend} \times \text{party size} \times \text{responsiveness} \quad (1)$$

Figure 7: Life stage segmentation tree

$$\begin{aligned} \text{Holiday revenue} = & \\ & \text{average holiday costs} \\ & \times \text{segment size} \times \text{responsiveness} \quad (2) \end{aligned}$$

$$\begin{aligned} \text{Total value} = & \text{spend revenue} \\ & + \text{holiday revenue} \quad (3) \end{aligned}$$

This will give an idea of the payback in terms of on-site spend in terms compared with the cost of contact. It can also be linked with the expectation from the overall value calculations to produce a true picture of the value of a campaign.

Incremental revenue can also be calculated that shows the additional or reduced revenue received by contacting a segment.

$$\begin{aligned} \text{Incremental spend} = & \\ & \text{segment average spend} \\ & - \text{overall average spend} \quad (4) \end{aligned}$$

$$\begin{aligned} \text{Incremental holiday revenue} = & \\ & \text{segment holiday revenue} \\ & - \text{overall holiday revenue} \quad (5) \end{aligned}$$

Calculating campaign statistics

The actual segments that are used will depend on a number of factors, these include:

- Size of mailing
- Message
- Value and payback

For a campaign, it may be decided that 20,000 (size of mailing) young families are to be contacted (the message) and these people should either have a high-expected on-site spend or holiday value (value and payback). If the summary statistics are examined, the best prospects can be seen or calculated.

The segment value calculations can then be combined to produce global campaign statistics.

A customer can exist in more than one group, within this segmentation solution. This means that if the agency wants to contact different groups over time, a customer or prospect can receive

Figure 8: Modelled gains response curve for brochure mailing campaign

different messages depending on the creative mix. They are not pigeon-holed into a single restrictive group.

Moving to predictive techniques

All of the examples given so far are based on historic information. The next step is to try and understand the segment that a customer is likely to inhabit in the future. If it can be determined that a customer or prospect is likely to be a high-value prospect in the future, then more efficient marketing decisions can be made at the time of recruitment or contact. In a similar vein, poor prospects can be identified and communicated with accordingly.

When modelling, the overall impact of the model in terms of profitability can, and should, be calculated, and gains and net income curves created. These curves can be used to highlight the optimum number of prospects to be contacted and to produce break-even statistics. Through the use of methods such as these, it was possible to reduce the cost of recruitment by over one half for one client in the

travel sector. This was made possible by reducing the number of brochures produced and only sending brochures out to the best prospects.

Figure 8 shows the modelled response curve for the brochure mailing campaign comparing the actual responses (sorted by model score) compared with prospects contacted. It can be seen that the majority of responses came from high scoring customers. Over 55 per cent of all responses came from the top 25 per cent contacted. This gives cumulative gains in excess of two for this range.

The cost per booking can also be calculated for each section. In Figure 8 the difference in value between the top 10 per cent and bottom 10 per cent can be seen clearly. Through the use of expectation calculations, it is possible to produce similar figures prior to the actual campaign.

To produce these predicted segmentation solutions, it is often necessary to resort to more complex statistical methods such as regression. Logistic and standard regression models

were used to create a number of segment scores for another client. These included:

- Debutant to devotee upgrade score
- Seasonality score
- Late booking propensity
- Enquiry grading model

Each of these scores had specific application to the forthcoming programme and were to be used in conjunction with other selection variables within the database environment.

Certain of these scores might be discrete at customer level whilst others were 'fuzzy', eg seasonality. The holiday year was split into five periods and each customer's propensity modelled. Factors such as age of eldest child were used to create a regression score and thus predict likelihood to travel in a particular period. It is possible for customers to have high or low scores in a combination of different periods. This can be used for determining offers for particular periods and selecting those more likely to travel at these times.

Repeater upgrade model

This model aimed to predict those 'first timers' more likely to become repeat bookers. The model examined many factors but in the end a stepwise regression technique used the following elements:

- Month holiday booked
- PRIZM lifestage and affluence
- Value of holiday
- Completed post holiday questionnaire (PHQ)
- Length of stay

The model produced was very strong, with over 72 per cent accurate classification of an unknown test dataset.

As these variables are known very early in a relationship, the model can also be implemented on new first timers.

As the organisation had experienced an increase in the volume of enquiries produced from marketing activity without a corresponding increase in bookings, a way of grading enquirers was sought. The data collection strategy for this client had included the development and systematic collection of information at the point of enquiry for a holiday.

Initial investigative analysis showed that there were a number of answers to these current questions that were indicative of a booking. These included:

- Higher affluence (achieved from tagging an external demographic system);
- Age of eldest child (older the more predictive);
- Whether had visited Europe before for this type of holiday; and
- Accommodation preference.

These were combined into a grading model using stepwise regression.

The models were delivered back to the database, as well as through TN's profiling tool. Regular reviews were held and post-campaign analysis — using techniques such as CHAID (Chi-Squared Automatic Indicator Detection) — were used to refine segments.⁶

INTERPRETATION

The area of interpretation of analysis back into a marketing environment is key for the correct understanding and application of findings from the analysis. This is often the knowledge transfer process between analysts and marketers, and is reliant on the respective communication and skill levels.

Transparency in the analysis development process makes this stage

Figure 9: Talking Numbers' Profiler tool

more straightforward, especially in comparison to 'black box' techniques, where the relationships between data are often complex to uncover. This transparency is valuable in revealing key importance of information to be collected at different points in a customer relationship.

There are a number of typically used methods from reports, presentations, workshops etc. The physical implementation of scores back onto a database system and integration of scores with selections may involve a combination of different people.

These tools can be used to 'present back' scores and information from an analysis project, and allow marketers to explore their own data.

TN's Profiler tool can be used for planning and 'train of thought' analysis on the dataset (see Figure 9).

Changes over time

A segmentation system can be dynamic or static. The decision on this may partially be driven by mechanics, but it is also a marketing planning decision.

A dynamic system allows periodic or even 'real-time' change (ie at each data refresh). This is easier to implement with a rule-based system, eg where a new booking is recorded and the customer moves segment — for instance from being a first-timer or lapsed.

A static system will still need periodic refreshes, but effectively does not change until the data is returned.

There are actually marketing disadvantages and advantages to each type of system. Often the most effective solution is to create both static and dynamic solutions on the database. The static solutions can be examined to understand changes over time, whilst the

Predictor	Nodes	Split Type	F	D.F.	Adj.Prob.
Segment	3	Default	81.0995	2, 26...	0.000000000
year	4	Default	72.9917	3, 26...	0.000000000
hols to date	3	Default	25.8390	2, 26...	0.000000000
Length of holiday	3	Default	18.8385	2, 26...	0.000000023
Devotee score	2	Default	31.7807	1, 26...	0.000000057
master ref	*	Arbitrary	*	*	*
Mobile	*	Arbitrary	*	*	*
Tent	*	Arbitrary	*	*	*

Figure 10: CHAID variables

dynamic score reveals the state of the current segment position. The advantages and disadvantages of delivering analysis back into the appropriate systems environment in the form of rules-based or statistical optimisation is explored by Doyle.⁷

The static scores can be used to help understand the historic position at specific points in time and will give a good idea of how the marketing effort is working.

ASSESSMENT AND REVIEW

Analysis and models are not static, they require periodic assessment and modification. This is normally conducted through post-campaign analysis and assessment of models. These can be aided and supported by appropriate tools such as campaign libraries, tree-based analysis on response and selection variables (CHAID), trend analysis and even econometric forecasting models.

Example: Post-campaign analysis using CHAID

Figure 10 shows variables that are important in describing responsiveness

and the level of predictive significance. The Figure also ranks each variable in order of predictive power. The most important factor in determining response is the RFV segment — debutants to drifters.

Factors such type of accommodation used were not predictive of response in this case.

The numbers in Table 1 refer to the segment on the CHAID tree in Figure 11. The adjusted response level on the Node 0 (overall position) is 0.5015. The response has been adjusted to balance the responders and non-responders.

The response gain values can be calculated from these figures. For example, in Node 50, the responders account for 77.9 per cent of this segment compared with the overall 50.15 per cent. This gives a gain of 1.55 ($77.9/50.15$), which is an indication of the power of this segment in terms of responsiveness.

Segment change

Inevitably, the effect of marketing activity is to drive segment change; tracking and monitoring this change with various marketing metrics will allow the

Figure 11: CHAID tree analysis output

effectiveness of activity to be judged. The use of change data also becomes a further possibility for marketing models and to aid understanding, however. For example, understanding those people who did change segment due to activity will be useful for future modelling and communications.

Updating the scores to include new customers, new enquiries and bookings can be scheduled. Change positions in scores can be used to forecast trends and shifts within the base.

Updating segment scores

Where regression scores have been used to try and predict behaviour, it is important to remember that these models

and equations should be investigated over time to make sure that they have not become stale. All models should be periodically revisited to make sure that they are still performing well and to see if they can be improved.

When applying segmentation to contact strategy, universe sizings can be approximated and rule definitions developed to allow set-up on the database. This then allows separate segments to be identified and different contact strategies to be adopted.

The amount of splitting entertained by an analyst is one of practical resource issue and that of producing sensibly sized contact cells for the marketing channel. The use of alternative media channels, however — eg e-mail rather than direct

Figure 12: Summary of segmentation and modelling approach
For further reading, the text by Berthold and Hands⁹ and Berry and Linoff¹⁰ are to be recommended.

mail — may economically support smaller sized groups.

SUMMARY

Figure 12 describes the process steps discussed within this paper.

References

- 1 Doyle, S. (2001) 'Communication optimisation — The new mantra of database marketing. Fad or fact?', *Journal of Database Marketing*, Vol. 9, No. 2, pp. 185–191.
- 2 McMahon-Beattie, U. and Palmer, A. (2000) 'One for all or all for one: A comparison of everyday low pricing and yield management strategies in the hotel industry', *Journal of Targeting, Measurement and Analysis for Marketing*, Vol. 8, No. 3, pp. 249–258.
- 3 Kimes, S. (1997), 'Yield management: An overview', in Yeoman, I. and Ingold, A. (eds.) 'Yield management: Strategies for the service industries', Cassell, London, UK, pp. 3–11.
- 4 Koslowsky, S. (2002) 'The case of the missing data', *Journal of Database Marketing*, Vol. 9, No. 4, pp. 312–318.
- 5 Foss, B., Henderson, I., Johnson, P. *et al.* (2002) 'Managing the quality and completeness of customer data', *Journal of Database Marketing*, Vol. 10, No. 2, pp. 139–158.
- 6 Madgison, J. (1990) '*CHAID.LOGIT and Log-Linear Modeling*', Datapro Reports on Marketing Information Systems, McGraw-Hill.
- 7 Doyle, S. 'Rules-based engines or statistical optimisation: The intelligent way forward', *Journal of Database Marketing*, Vol. 10, No. 1, pp. 85–89.
- 9 Berthold, M. and Hands, D. J. (2002) 'Intelligent Data Analysis', 2nd Edition, Springer.
- 10 Barry, M. J. A. and Linoff, G. (1997) 'Data Mining Techniques for Marketing, Sales, and Customer Support', Wiley Computer Publishing, Chichester, UK.