Predicting Ski Lodging Selection Using Customer Databases

by

Tanja Prlincevic

B.Sc., University of Belgrade, 1993

PROJECT SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF BUSINESS ADMINISTRATION

in the
Faculty of Business Administration

SIMON FRASER UNIVERSITY

Summer 2006

APPROVAL

Name: **Tanja Prlincevic**

Degree: **Master of Business Administration**

Title of Project: **Predicting Ski Lodging Selection Using Customer Databases**

Supervisory Committee:

Dr. Robert E. Krider
Senior Supervisor
Associate Professor
Faculty of Business Administration
Simon Fraser University

Dr. Colleen M. Collins-Dodd
Second Reader
Associate Professor
Faculty of Business Administration
Simon Fraser University

Date Approved: August 8, 2006

STATEMENT OF ETHICS APPROVAL

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

(a) Human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

(b) Advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

(c) as a co-investigator, in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, BC, Canada

ABSTRACT

This research illustrates the advantages that the application of data-mining techniques can bring into designing a direct marketing campaign. Specifically, it presents results of two multinomial logit models. The models identified patterns in the data and established the relationship between characteristics of repeat visitors and their lodging preferences. Based on this relationship it is possible to predict what type of accommodation customers would prefer at a level better than chance and to customize marketing messages to meet customers' needs. Two of the strongest predictors of choice are income and length of stay for the previous visit.

The project also offers lift chart analysis of the multinomial models' results that led to better understanding of the two models' predictive powers, a visualization tool rarely used for interpreting multinomial models. This analysis revealed that the model identified 94% of all visitors that prefer Silver accommodation within 50% of the top scoring customers.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Objective

The project was sponsored by a large corporation specialized in destination resorts and adventure travel. The Company develops real estate and operates village-centred ski, golf, and seashore resorts throughout North America and Europe. The focus of this project is on the Company's ski resort division in North America.

The ski resort industry is unique in that an estimated 60% of resort guests are one-time visitors, as these skiers prefer to try a different mountain each season. Consequently, at the core of the Company's past marketing efforts are predictive models created to profile target customers in an attempt to increase the effectiveness of its marketing campaigns. Based on the models, the Company was able to target customers who were most likely to be interested in purchasing lodging. All potential customers received messages that included an offer for the same type of lodging.

However, as Slavens (2006) noted tailored messages tend to result in a higher response rate and to generate new and recurring revenue. In light of this fact, the Company management is considering moving away from a generic offering in its marketing campaigns and testing targeted messages that would feature accommodations with characteristics suitable for given customers. The results of this project will aid management in deciding whether or not to proceed in this direction.

The goal of this project was to determine, through appropriate predictive analytics techniques, the relationship between customers' attributes and preferred lodging type. Once this link is established, the Company would be able to make appropriate lodging offers to potential and existing customers.

The analysis was conducted on transaction data collected by the Company at one of its ski resorts in Colorado, and the Company intends to extrapolate the results of the project to other resorts in North America. Hence, this project would act as a stepping-stone for the Company in understanding the needs and preferences of its customers' for a certain category of lodging and for the design of more effective marketing campaigns with personalized messages.

1.2 Solution

The models built by the Company to date addressed the question of whether or not customers would buy any lodging. These were *binary discrete choice models*, commonly found in data mining software and applications, which are used to predict the choice a consumer would make between two distinct alternatives. These types of models are employed when trying to predict if individuals would purchase or not purchase an item, or donate or not donate, etc.

The binary discrete choice model cannot be applied in this project since resort visitors face more than two accommodation choices. In this case, it was necessary to build a *multinomial discrete choice model.* Unlike binary models, multinomial discrete choice models are rarely found in standard data mining software, though they are applicable to many practical situations. Consumers often select from several choices, differentiated by attributes such as brands and sizes.

Both the binary and multinomial discrete choice models are discussed in more detail in Chapter 3.

The original attempt to build the multinomial discrete choice model was focused on one-time customers. This model only included customers' attributes that could be obtained based on their addresses and neighborhood data. Since no strong patterns were found in the data of one-time customers, however, the focus of the project shifted to the repeat customers. To build this model, the data from past customer' transactions was included in the analysis.

For the purpose of this project the lodging types available at the resort were divided along two dimensions: ***unit type*** and ***medal classification***.

The unit type refers to the size of the accommodation with the following options: hotel, studio, 1-bedroom, 2-bedroom, 3-bedroom and 4-bedroom.

The medal classification captured difference in the amenities in various lodging units with the following options: bronze, silver, gold and platinum, with platinum being the most luxurious accommodation.

The resort visitors can choose among twenty different combinations of unit types and medal classifications. Given the small data set for the repeat customers, however, several choice combinations were under-represented in the set making modelling of these choices impossible. Instead, two separate models were built: the Medal Choices Model and the Size Choices Model.

2 HOTEL INDUSTRY PROFILE

According to the report issued by the American Hotel & Lodging Association in 2005, tourism represents the third largest retail industry in the United States. This sector is the largest service export industry and one of the largest employers in the U S. The tourism industry encompasses several interrelated businesses such as lodging establishments, airlines, restaurants, car rental firms, travel agents and tour operators.

According to the same report, the entire tourism industry generated $600 billion in sales in 2004. The lodging industry generated revenue of $113.7 billion in 2004, an increase from $105.3 billion in 2003. The total lodging industry gross pretax profit for 2004 was $16.7 billion.

Finally, the report offers a profile of a typical leisure traveler. In 2004, an equal number of people traveled for business and for pleasure. Fifty-one percent of leisure room nights were generated by two adults that were between 35 and 53-years-old with an average yearly household income of $72,600. Leisure travelers tend to make short bookings as 45% spend one night, 28% spend two and 27% spend three or more nights at a given location. Seventy four percent of leisure travelers travel by auto.[2]

2.1.1 Hotel Industry Trends

Until recently, the lodging industry was concentrated on developing physical products as well as services that companies' management teams believed customers needed and wanted. However, changes in the business environment, as well as the global economy ignited a profound, multidimensional business transformation that is affecting every aspect of the industry. As the result, the lodging industry is moving away from a product-focused, physical-asset-intensive industry towards a more customer-focused, brand-intensive one, with the goal to build customer loyalty that is durable and tangible.

Some of the major ***drivers*** of this transformation are (Dickinson , 2006):

- ***Physical constraints:*** It is becoming increasingly difficult and more expensive to differentiate a lodging establishment based on its distinctive architectural design and physical properties;
- ***Demo/psycho graphic:*** The lodging industry will soon simultaneously service two distinctly different demographic groups: Baby Boomers and Echo Boomers. The industry will have to transform itself to successfully meet the needs of the two largest groups in history;
- ***Costs:*** Both fixed and variable costs are on the rise. The most prominent changes are in labour costs and the customer acquisition costs, making the need to develop lasting customer loyalty greater than ever.

The transformation that is taking place will be reflected in the way lodging companies approach their daily operations. For the purpose of this project, the most important shift is occurring in the way lodging companies use customer information gathered at the time of a transaction.

Dickinson (2006) found that the lodging industry has made progress in gathering data about its customers but has done little to analyze and translate the data into knowledge that would drive the actions of the company. Such knowledge would enable a company to personalize its offerings to more closely match the needs and wants of its target market, and ultimately gain and sustain a competitive advantage over its competitors.

The relatively new process known as *data mining* makes identifying important variables and relationships located in these large consumer-information systems easier to manage.

Data mining is a largely automated process that uses statistical analysis to sift through massive data sets to detect useful, non-obvious, and previously unknown patterns or data trends (Magnini, 2003).

Unlike the traditional statistical modelling that emphasizes theory-driven hypothesis testing, data mining is much more machine-driven model building. Since a researcher, with preconceived ideas about the constructs that should be examined, supervises statistical modeling, relevant associations may be easily overlooked with this method. In applications of data mining, on the other hand, some mix of analyst supervision and automatic searches is usually used. As a result, data mining builds dependency hypotheses and, by doing so, can reveal important links (Magnini, 2003).

Data mining tools are used for various purposes within the hotel industry. Magnini et al (2003) identified the following applications of data-mining information in hotel marketing:

- Create direct-mail campaigns
- Plan seasonal promotions
- Plan the timing and placement of ad campaigns
- Create personalized advertisements
- Define which market segments are growing most rapidly

- Determine the number of rooms to reserve for wholesale customers and business travellers

In spite of the many advantages associated with data mining, the same researchers found that this technique is in its initial stage in the hotel industry. Only the early adopters such as Marriott Vacation Club International are gaining a competitive advantage over companies that are lagging to adopt the technique.

One of the most advanced applications of the data mining tools within the hotel industry can be found at the Harrah's Entertainment Casino Hotel. By applying the data mining tools on the consolidated data from all of its properties the chain was able to weed out thrifty visitors and develop a profile of valuable guests who will spend thousands of dollars on food, shopping, and gambling while staying at one of its resorts. Harrah's is able to predict how much each visitor is worth to the hotel and price their visitors' hotel rooms accordingly. Harrah's also uses data-mining to determine what type of services to offer to each valuable customer, to gauge its marketing efficiency, and to track detailed transactions.

One of the latest outcomes of Harrah's customer-focused data mining was its recent merger with the Caesars Casino in Las Vegas. Given its marketing expertise, Harrah's will be in charge of creating and implementing marketing and service-delivery strategies for the new merger.

3 DISCRETE CHOICE THEORY

Predictive analytics, a branch of statistical analysis, focuses on predicting future behaviour patterns based on the information extracted from data. In essence, predictive analytics identifies relationships between specific target behaviour and predictive variables, usually other persons' behaviours and characteristics recorded from past occurrences.

Discrete choice models are predictive analytics techniques for predicting the choice that consumers will make between distinct, discrete alternatives. The most common discrete choice model found in data mining is the *Logistic (Logit) Regression*, that is applied when consumers face two discrete choices such as donate/do not donate, purchase product/doesn't purchase, etc. This model can be used only if the variable of interest, *the target variable*, is either:

- a categorical variable with exactly two choices; or
- a continuous variable that has values in the range 0.0 to 1.0 representing probability values or proportions.

The binary discrete choice cannot be applied in this project as resort visitors choose one type of accommodation from several possibilities. Several other statistical models have been developed to address these types of situations; however, they are not commonly found in data mining.

The remaining part of this chapter offers an overview of the theory behind discrete choice modeling as well as a short description of the discrete choice model applied in the project: *the Multinomial Logit.*

3.1 Behavioral Theory

At the core of discrete choice modelling are theories that attempt to explain how an individual makes choices. The main difference among alternative theories is in the degree they idealize the decision-making process that results in the observed behaviour. Still, all theories share the general framework where a choice is viewed as an outcome of the following sequential decision-making process (Ben-Akiva & Lerman, 1985):

- Definition of the choice problem
- Generation of alternatives
- Evaluation of attributes of the alternatives
- Choice
- Implementation

A decision maker chooses among distinct alternatives (choices) by applying a *decision rule*. One of many proposed decision rules assumes that a decision maker attempts to maximize the utility of an alternative through his or her choice. A vector of attributes values expresses the utility or the attractiveness of an alternative. Attributes are assumed to be compensatory, meaning that the decision maker will put enough effort into the decision to consider tradeoffs among the attributes in choosing the preferred alternative. This can be modelled through a weighted sum of the alternatives, where the analyst estimates the weights by using the results of the decisions made.

Expressed in general terms, a decision maker *i* maximizes his utility by choosing u_{ij} that is the highest of all $u_{ir,\ r=1...J}$, where *j* denotes an alternative that belongs to the choice set of all alternatives $\{1,2\ldots J\}$.

A probabilistic choice mechanism was introduced in an attempt to explain inconsistencies and non-transitive preferences observed in experiments that presented participants with the same choice set several times. The probabilistic approach circumvents the need for more precise knowledge about individuals' decision processes. At the same time, it captures the effects of the unobserved variations among decision makers and the unobserved attributes of the alternatives.

The *random utility* approach suggests that the observed inconsistencies are the result of the analyst's observational deficiencies and that a decision maker always selects the alternative with the highest utility. Since the analyst does not know the utilities, they are treated as random variables. Choice probabilities of an alternative are equal to the probability that the utility of that alternative is greater, or equal, to the utilities of all other alternatives in the choice set. In deriving choice probabilities in the random utility approach, analysts assume a joint probability distribution of the set of random utilities. The distributional assumption stems logically from the observationally random elements that result in randomness of the utilities. Manski (1973) identified the following sources of utility randomness:

- Unobserved attributes
- Unobserved taste variations
- Measurement errors and imperfect information
- Instrumental variables

3.2 Multinomial Choice

In most practical situations, a customer chooses among K mutually exclusive alternatives of a product or of a service. Models built for these types of situations are generally known as *Multinomial Discrete Choice Models.*

Assuming that the probability that the two alternatives can have equal utility values is zero and that therefore are no ties, we can write (Ben-Akiva & Lerman, 1985):

$$u_{ik} = V_{ik} + \varepsilon_{ik} = \beta_j X_{ik} + \varepsilon_{ik}$$

where:

X_{ik} is the $(M\text{-}1) \times J$ vector of the attributes of the k^{th} alternative $(k=1,\ldots,M)$

β is a $J \times 1$ vector of unknown parameters

V_{ik} is a function of the observed utilities, and

ε_{ik} is a $J \times 1$ vector of residuals that captures all four sources of utility randomness.

The main goal in building a multinomial model is to estimate vector β, of the unobserved utilities u_{ik} and of the probability that the individual i will choose the alternative s. By estimating vector β, an analyst can estimate how frequently alternative s would be chosen in a given population of individuals.

The probability that the i^{th} individual will choose the s^{th} alternative is:

$$P_{is} = P\ \{V_{is} + \varepsilon_{is} > \max_{k=1,\ldots,M-1}(V_{ik} + \varepsilon_{ik})\}\ i=1,2,\ldots N$$

By introducing $F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M)$, the joint cumulative distribution function over the values ε_{ik} $(k = 1, 2, \ldots, M)$, and F_{ik}, the partial derivative of F for the k^{th} argument, the probability equation can be written as:

$$P_{ik} = \int_{\varepsilon=-\infty}^{\infty} F_k\,(\varepsilon_{ik} + V_{ik} - V_{i1},\ \varepsilon_{ik} + V_{ik} - V_{i2},\ \ldots,\ \varepsilon_{ik} + V_{ik} - V_{ip})\ d\varepsilon$$

Joint cumulative distribution function F can be specified in different ways resulting in several families of discrete choice models. Some of the most known ones are the *Multinomial Logit (MNL)*, the *Nested Multinomial Logit (NMNL)* and the *Multinomial Probit (MNP)*.

3.2.1 Multinomial Logit

Choice probabilities in the Multinomial Logit or *conditional logit* model are defined as:

$$P_{ik} = \frac{\exp\,(u_{ik})}{\sum_{j=1}^{M} \exp\,(u_{ik})}$$

where:

$u_{ik} = V_{ik} + \varepsilon_{ik}$ is the utility for alternative *k* and for individual *i*

V_{ik} often called, *representative utility*, is the systematic observable utility component of alternative *k* and for individual *i*.

ε_{ik} is the random error component associated with alternative *k* and individual *i*. It represents the difference between the true utility u_{ik} and the part of the utility that researcher captures in V_{ik}.

Multinomial logit models may use respondent characteristics as predictor variables, as well as alternative characteristics. Based on the individual's characteristics, the probability can be predicted that the person will choose alternative *k* from a choice set.

It is important to emphasise that all econometric models require identification restriction of some kind, or normalization, in order to be estimated. There is no actual scale for measuring utility and at the same time there is no agreed base value. The simplest way to address this issue is to nominate one of the choices as a baseline and then calculate log-odds for all other choices, relative to the baseline and let the log-odds be a linear function of the predictors.

4 METHODOLOGY

4.1 Data Collection

The original data for this study, supplied by the sponsoring company, consisted of three files: a lodging transactions file with 24,869 observations and 25 variables, a non-lodging transactions file with 570,515 observations and 6 variables and a demographic information file with 325,853 observations and 155 variables. The demographic data was gathered from three sources: census data, **P**otential **R**ating **I**ndex by **Z**ip **M**arkets (PRIZM) cluster survey information, and the individual customer data collected by the Company at the time of a transaction.

PRIZM is the system that classifies residential neighbourhoods based on over 500 demographic variables. The variables were obtained through surveys concerning products, media, and opinions. PRIZM breaks the 250,000 neighbourhood areas in the US into 40 types based on consumer behaviour and lifestyles (Churchill & Iacobucci, 2001).

The lodging transactions data file covered the lodging transactions from November 1, 2003 to April 30, 2005. Since this file contained information on the customers' accommodation choices, the variable we are trying to predict in this project (*the target variable),* it was used as a base for further data preparation.

The target variable describes the phenomenon of interest for which we want to make predictions about using the independent variables or attributes.

4.2 Data Analysis

Initially, the project was divided into two stages:

- The initial data exploration
- Model building or pattern identification including validation of the built model

During the course of the project these two stages were repeated several times, while trying to build models with different sets of choices. In the first attempt to model the one-time customer data, the choice set contained 16 different options. Since no strong patterns were found with such a large number of choices and with several choice combinations observed rarely, in the subsequent attempts the number of choices was collapsed. However, even with a smaller number of choices, no strong patterns were found.

Instead, the focus of the project was shifted to the repeat customer data. Since this data set was small, it was necessary to limit the number of choices used during the modelling stage. To address this issue, it was decided to build two different models: 1) with the unit types as choices, and 2) with the medal classifications as choices.

The following is a short description of the data analysis steps performed while building the two models on the repeat customer data: The Medal Choices and The Size Choices Models.

4.2.1 Data Exploration

Typically, ***exploration*** involves data preparation, which may consist of cleaning the data, transforming the data, and selecting subsets of records. Depending on the problem at hand, the next step might involve either simple a selection of the independent, predictor variables to be included in the modelling process. In more complex cases, the analyst might use a wide variety of statistical and graphical methods to identify variables for inclusion in the next stage.

4.2.1.1 Data Cleaning

The focus of this study was on the independent travellers. These visitors do not book their stays via tourist agents; but instead, they make reservations directly with the resort. To ensure that only the independent travellers are included in the analysis, all customers with unusually large number of transactions were removed from the data set at this stage. A large number of transactions indicated non-independent travellers (business, and wholesale customers) that should not have been included in the data.

The next step was to create a flat file with one record per customer with all related lodging and non-lodging transactions for each of the three files. To achieve this, the variables from lodging and non-lodging transaction files were transposed.

In the following step, all three flat files were merged into one file, removing all customer records with no personal data. As a result, the file contained data on 14,507 customers with lodging transactions and the complete personal data as well as the non-lodging transactions.

Once all of the one-time customer records were removed, the file contained only 2077 observations.

The following step in the data preparation was to run a series of descriptive statistical tests to identify any anomalous variables with possible out-of-range values or impossible data combinations.

The patterns of the missing data and their extent were also examined. The AGE variable was found to be missing in 50% of the cases. In such situations, an analyst would usually either remove the entire variable with the high percentage of missing values, or remove the cases with the missing values from further analysis. In this case, neither option was acceptable since the data set would have been too small if all observations with the missing age values were removed. At the same time, based on the literature review, AGE was expected to be a strong predicting variable and for this reason had to remain in the analysis.

To address this issue, all missing values were replaced with zero and a new dummy variable (DMAGE) was created with 1 representing visitors with age data and 0 representing visitors with missing age data. This method allowed us to retain records with missing age data, and use the age values when available, without forcing an artificial "zero age" effect.

4.2.1.2 Variable Selection

Given the total number of independent variables, the next step was to reduce their number in the modelling stage to a more manageable level. The existing hotel industry research was reviewed for relevant accommodation choice drivers. However, there appears to be no previous research focusing on this particular issue.

Still, the existing hospitality research literature on tourist segmentation offered some guidance as to possible significant predictive variables. As Inbakaran and Jackson (2005) noted, segmentation of tourists is useful when marketing to particular groups, and for lowering the costs and increasing the effective penetration of appropriate promotional material. Kotler (1991) classified some 14 segmentation variables into four major categories: demographic (age, gender, education, life cycle), geographic (trip origin, trip destination, distance), psychographic (personality, life-style, values, motives), and behavioural (status, usage rates, tourist activities/experiences).

Similarly, Morrison et al. (1996) found demographics (age, gender, marital status, education, income, employment), behavioural (pre-trip planning, tourist activities during stay), and psychographics (emotions, preferences and benefits sought) to be significant variables in segmenting four different types of resort visitors.

Riddington et al. (2000) found that the utility of a ski centre for an individual skier is determined by alternative specific variables and individual specific characteristics such as distance between the site and the individual's place of residence and expenditure per person per day.

In addition to the predictors identified through the literature review, binary choice models were developed for each medal and size choice in an attempt to determine other potentially significant variables that should be included in the modelling stage.

Furthermore, the large number of variables from the original data sets captured the same attribute for different intervals. For example, the Housing value was specified in 13 intervals ranging from under $20,000 to over $1,000,000. The variables were expressed as percentages of people in the area with the Housing value in a certain price range. Clearly, poor neighbourhoods will have no housing over a certain amount, and expensive neighbourhoods will have no housing under a certain amount. If we used these variables in their original form, we would not be able to detect their effect on visitors' accommodation choice. To address this issue, the Housing value and other similar variables were agglomerated for larger (pricing) intervals to ensure that enough contrast was present among the different levels for these attributes.

In order to incorporate unit size and medal choice of the last visit in the modelling process two approaches were considered:

1) Two 'Same As Last Visit' dummy variables were created that indicated when a customer made the same unit size or the same medal choice for the last two visits;
2) Medal and unit sizes were assigned arbitrary values in order to capture the ordinal information contained in these choices. Medal choices were assigned the following values: 1-Bronze, 2-Silver, 3-Gold, and 4-Platinum. Size choices were assigned the following values: 1-Hotel, 2-Studio, 3-One-bedroom, 4-Two-bedroom, and 5-Three-bedroom.

Since the second approach resulted in only slightly better overall results it will be presented in this report.

Also, since NLOGIT, the software used for modelling cannot manipulate nominal data several variables were transformed into dummy variables.

The list of transformed and created variables during this stage is found in Table 6-2 (see Appendix C).

The final stage of data preparation was to partition the data into training and validation sets. The files were then reformatted into a layout suitable for NLOGIT.

4.2.2 Modelling

The dependent, or target variable, for the Medal Choices model was the medal choice made by visitors at their last visit. The choice options for this model were:

- Bronze
- Gold
- Silver
- Platinum

The baseline for this model was the Platinum accommodation and the probability of making any other choice was compared against this reference cell.

The target variable for the Size Choices model was the size choice that visitors made at their last visit. The choice options for this model were:

- Hotel
- Studio
- 1-bedroom
- 2-bedroom
- 3-bedroom

The baseline for this model was the 3-bedroom accommodation and the probability of making any other choice was compared against this reference cell.

The independent variables consisted of selected variables from the original data sets as well as the variables constructed during the data preparation stage.

As previously mentioned in the brief theoretical overview, the choice probabilities for this model are defined as:

$$P_{ik} = \frac{\exp\,(u_{ik})}{\sum_{j=1}^{M} \exp\,(u_{ik})}$$

where:

$u_{ik} = V_{ik} + \varepsilon_{ik}$ is the utility for alternative *k* and for individual *i*

V_{ik} - the deterministic part of utility can be expressed as $V_{ik} = \beta_j X_{ik}$.

The final model was estimated by ***Maximum Likelihood*** using NLOGIT. The goal at this stage was to estimate unknown model parameters that will maximize the sample likelihood, or the probability of optimally reproducing the distribution of choices as observed in the data.

The null hypothesis for each variable was that the estimates of parameters equal to zero. The null hypothesis was rejected for variables with p-value of 0.05. Goodness-of-fit was assessed with the McFadden R2 (adjusted for degrees of freedom), and the Akaike Information Criteria (AIC) was used to compare different models and select the best fitting model.

The final list of predictive variables for the Medal Choice and Size Choice models can be found in Appendix A and Appendix B respectively.

4.2.3 Deployment

Usually, the final stage of the modelling process is the ***deployment,*** or ***scoring,*** of the new data. Deployment refers to the application of the model for predicting or classifying new data. However, this stage was not within the scope of this project.

5 RESULTS

5.1 Repeat Customer Profile

The project's primary objective was to build a predictive model and its secondary goal was to gain a better understanding of the repeat customers' past accommodation choices and examine the potential differences in characteristics between the repeat and the one-time customers.

A series of contingency tables were constructed for the last, and second last visits, to examine the repeat customers' movements across different accommodation types and the time of visit. The second last visit was used as a base to determine the percentage of customers who visited the resort during the same sub-season, and stayed in the same size and the same quality of accommodation during their last two visits.

The analysis showed that the majority of customers who visited the resort during the Regular and March seasons tended to return to the resort either in the same period or for one of the sub-seasons after Christmas. Contingency tables for repeat customers in the sub-seasons are found in Appendix D.

Furthermore, the analysis of repeat customers' movements across quality types indicated that three different categories of repeat customers should be considered:

- Centre group that stays in either Gold or Silver accommodation;
- Lower range group that stays in either Bronze, Silver or Gold accommodation;

- Upper range group that stays in either Silver, Gold or Platinum accommodation.

The results for the customer movement across the quality types need to be considered in the context that the majority of rooms offered at the resort are either Silver or Gold quality. Contingency tables with movements of repeat customers across the quality types are found in Appendix E.

Finally, the analysis of repeat customers' movements across unit types showed that the majority of customers that stayed in Hotel, 1-Bedroom and 2-Bedroom units during their second last visit tended to stay in an accommodation of the same size during their last visit. Also, it seems that customers that stayed in Studio units tended to upgrade to 1-Bedroom accommodation during their next visit.

The results for customer movements across unit types should be considered in the context that the majority of rooms offered at the resort are either Hotel, 1-Bedroom, or 2-Bedroom units. Contingency tables for the movement of repeat customers across the unit types are found in Appendix F.

A series of t-tests and cross tabulations were run to examine potential differences between the one-time and the repeat customers. The data used for this part of the analysis covered the period from November 01, 2004 to April 30, 2005.

The results of the t-tests (Appendix G) show that:

- One-time customers tend to stay in slightly more expensive accommodation than the repeat customers;
- One-time customers tend to stay longer than the repeat customers;
- Repeat customers tend to live closer to the resort then do the one-time customers;

- Repeat customers tend to reside in areas with higher median housing value;
- Median household income of the repeat customers is higher than that of the one-time customers;
- Repeat customers tend to spend more on recreation than do the one-time customers;
- Repeat customers tend to spend more on travel then do the one-time customers.

The results of the cross tabulations (Appendix G) reveal that:

- One-time customers tend to purchase package arrangement while repeat customers tend to travel independently;
- Repeat customers tend to sign up for a snow school more often than do the one-time customers;
- Close to 54% of the repeat customers belongs to the core enthusiasts (CE) target group, compared to 41% of the one-time customers.

5.2 Models

As previously mentioned, the two models were built on the repeat customer data, using NLOGIT software. No customers with only one visit were included in this analysis.

The Medal Choices and the Size Choices models were validated based on *the hit rate matrices,* created by NLOGIT software. The *hit rate* reflects the number of customers for whom the model predicted the choice alternative that is the same as the known actual alternative chosen by the customer. To compute the hit rate, customers were each assigned the alternative for which the customer's choice probability is the highest. The hit rate is computed on validation set; the data set not used to estimate the model parameters.

Since the ultimate purpose of this project was to improve the effectiveness of a direct marketing campaign, it was necessary to find a way of reaching as many as possible target customers without having to contact the entire database. This was crucial for maximizing the return on investment for the campaign.

For this reason, the next step in the result analysis was to examine the model's predictive ability in a greater detail by computing lift charts for each medal category similar to the charts computed in binary predictive models.

A *lift chart* provides a visual presentation of the *lift* that measures the effectiveness of a predictive model. Lift is calculated as the ratio between the results obtained, with, and without the predictive model.

Lift is usually computed by dividing the validation dataset into deciles (ten even groups), into which visitors are placed, based on their predicted probability of responding positively or making the desired choice. The highest scoring visitors are put into decile 1, etc.

These results are then plotted so that the horizontal axis represents the percentage of customers to be contacted. The vertical axis captures both incremental and cumulative number of expected positive results or *hits*. Usually, these two curves are compared to *the baselines*, the results of a random approach where, for example, we would expect to get 50 % of the possible positive results by contacting 50 % of the customers in the database. The greater the area between the lift curve and baseline, the more the model was able to concentrate hits in the top deciles.

Although we know of no previous applications of lift charts to multinomial choice models, computing lift charts for each medal and size choice proved to be useful by permitting a better understanding of the model's predictive power.

5.2.1 Medal Choices Model

The detailed, final list of variables for this model, generated by NLOGIT software, can be found in Appendix A.

The most significant variables predicting a difference in the likelihood of choosing Bronze, Gold, or Silver over a Platinum room are presented in the three tables that follow.

Bronze Medal Choice

Table 5-1: The Most Significant Variables for the Bronze Choice

Variable	Description	Coefficient	P[Z]>z
BxQPR	Quality for previous visit	-.9075038848	.0004
BxSPR	Size for previous visit	-.5223972454	.0155
BxITA	AC Target Segment	1.697237797	.0458
BxITC	CE Target Segment	1.749765820	.0082
BxITL	LL Target Segment	2.899162960	.0048
BxDIS	Distance	-.1178523118E-02	.0270
BxHHO	Percent of housing value over $750,000	-4.069577282	.0196
BxHI1	Percent of household income under $150,000	-10.68956747	.0357
BxHIN	Percent of household income over $150,000	-16.39293733	.0468
BxSIN1	Percent of 1 and 2 Person households	-24.38201579	.0172
BxTRA	Average travel expenditures per HH	-.5417223723E-02	.0506
BxLDG	Average lodging expenditures per HH	.3747439299E-01	.0058

Interpretation:

- Customers who stayed in higher medal and larger size accommodations during their last visit are less likely to stay in Bronze accommodations than in Platinum accommodations.
- Visitors from larger distances are less likely to stay in Bronze accommodations than in Platinum accommodations.

- Customers with higher income are less likely to stay in Bronze accommodations than in Platinum accommodations.
- Single people or couples without children are less likely to stay in Bronze accommodation than in Platinum accommodations.
- Customers who travel often are less likely to stay in Bronze accommodations than in Platinum accommodations.

All of these variables make sense, however, the model also found that the luxury lifestyles (LL) segment, one of the Company's target segments, is more likely to stay in Bronze accommodations than in Platinum accommodations. **Luxury Lifestyles** segment was created by grouping three PRIZM clusters: 1-Upper Crust, 2-Blue Blood Estates or Elite Super-Rich Families, and 3-Movers & Shakers or Executive Families.

Since Luxury Lifestyle encompasses the richest part of the US population, it is counter-intuitive that these customers would prefer Bronze to Platinum accommodation. Also, it is counter-intuitive that customers that spend more on lodging are more likely to stay in Bronze accommodation than Platinum.

Gold Medal Choice

Table 5-2: The Most Significant Variables for the Gold Choice

Variable	Description	Coefficient	P[Z]>z
GxQPR	Quality for previous visit	-.6820360985	.0010
GxSPR	Size for previous visit	-.3667198935	.0430
GxAVV	Average vehicles	-1.423814541	.0455
GxHIN	Percent of household income over $150,000	-15.50073870	.0241

Interpretation:

- Customers who stayed in higher medal and larger size accommodations during their last visit are less likely to stay in Gold accommodations than in Platinum accommodations.
- Customers who own larger number of vehicles are less likely to stay in Gold accommodations than in Platinum accommodations.
- Customers with higher income are less likely to stay in Gold accommodations than in Platinum accommodations.

Silver Medal Choice

Table 5-3: The Most significant Variables for the Silver Choice

Variable	Description	Coefficient	P[Z]>z
SxSPR	Size for previous visit	-.6297957248	.0005
SxITL	LL Brand Segment	1.850957514	.0273
SxSIN1	Percent of 1 and 2 Person households	-17.12506840	.0472

Interpretation:

- Customers who stayed in larger size accommodations during their last visit are less likely to stay in Silver accommodations than in Platinum accommodations.
- Single people or couples are less likely to stay in Silver accommodations than in Platinum accommodations.
- Surprisingly, customers from the LL brand segment are more likely to stay in Silver accommodations than in Platinum accommodations.

5.2.1.1 Validation of Medal Choices Model

Given the size of the data set and the small number of observations with Bronze and Platinum choice; validation set for Medal Choices model consisted of observations with only Gold and Silver choices. As well, the small counts made the results somewhat noisy.

Gold Choice Validation

As illustrated in Table 5-4 the medal model hit rate for the Gold choice is 61%. The NLOGIT software correctly assigned Gold choice to 64% of customers that actually selected Gold accommodation during their last visit.

Table 5-4: Gold Validation Hit Rate Matrix

Gold Validation Hit Rate Matrix			
		Predicted	
		Yes	**No**
Observed	**Yes**	275	153
Observed	**No**	84	95

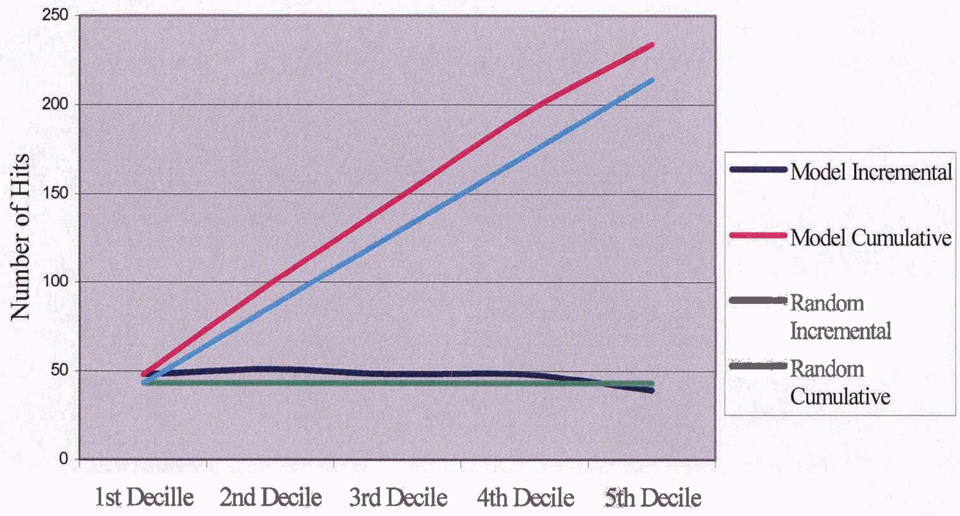
Hit Rates 275/359 95/248
Total Hit Rate: 370/607=61%

Once the customers were sorted in descending order of their probabilities for the Gold choice it was found that the model performed slightly better in the first 5 deciles. As can be seen from Table 5-5, the model correctly identified 55% (234 out of 428) of all visitors that would prefer Gold accommodation within 50% of the top scoring customers. The lift chart for these results can be seen in Figure 1.

Table 5-5: Incremental and Cumulative Gold Medal Hit Rates per Decile

Gold Medal Hit Rates				
Decile	Hits Incremental	Hits Cumulative	Hits Random Incremental	Hits Random Cumulative
0.1	48	48	42.8	42.8
0.2	51	99	42.8	85.6
0.3	48	147	42.8	128.4
0.4	48	195	42.8	171.2
0.5	39	234	42.8	214

Figure 1: Gold Medal Choice Lift Chart



Silver Choice Validation

As illustrated in Table 5-6 the medal model hit rate for the Silver choice is 62%. The software correctly assigned Silver choice to 53% of customers that actually selected Silver accommodation during their last visit.

Table 5-6: Silver Validation Hit Rate Matrix

Silver Validation Hit Rate Matrix			
	Predicted		
		Yes	No
Observed	Yes	86	76
	No	153	292
Hit Rates		86/239	292/368

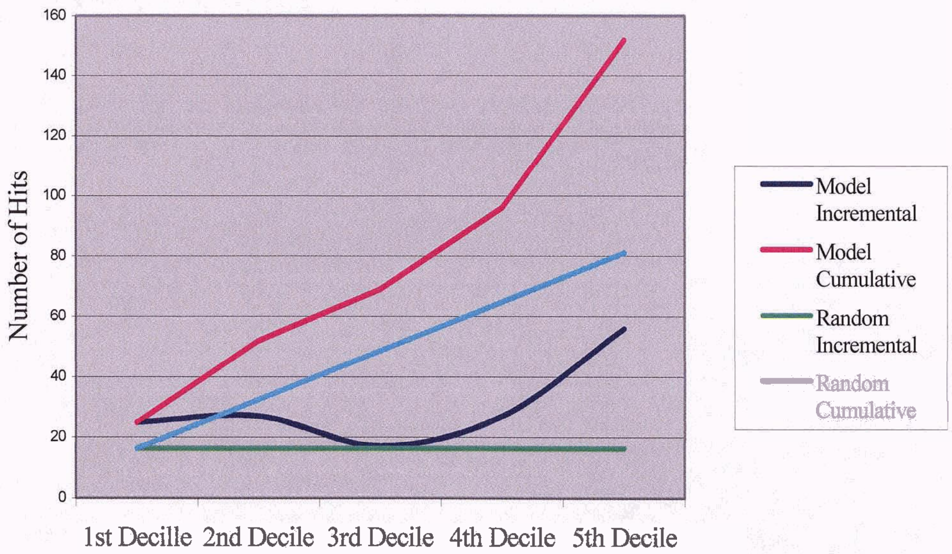
Total Hit Rate: 378/607 = 62%

By sorting customers in descending order of their probabilities for the Silver choice it was found that the model performed better in the first 5 deciles. As Table 5-7 illustrates, the model correctly identified 94% (152 out of 162) of all visitors that would prefer a Silver accommodation within 50% of the top scoring customers. The lift chart for these results can be found in Figure 2.

Table 5-7: Incremental and Cumulative Silver Medal Hit Rates per Decile

Silver Hit Rates				
Decile	Hits Incremental	Hits Cumulative	Hits Random Incremental	Hits Random Cumulative
0.1	25	25	16.2	16.2
0.2	27	52	16.2	32.4
0.3	17	69	16.2	48.6
0.4	27	96	16.2	64.8
0.5	56	152	16.2	81

Figure 2: Silver Medal Choice Lift Chart



5.2.2 Size Choices Model

The detailed final list of variables for this model, generated by NLOGIT software, can be found in Appendix B.

The most significant variables predicting a difference in the likelihood of choosing Hotel, Silver, One-Bedroom or Two-Bedroom over a Three-bedroom accommodation are presented in the tables that follow.

Hotel Size Choice

Table 5-8: The Most Significant Variables for the Hotel Choice

Variable	Description	Coefficient	P[Z]>z
HxSPR	Size for previous visit	-.9769094404	.0000
HxGUE	Number of guests for previous visit	-.2979772957	.0021
HxHI	Percent of household income under $150,000	-8.437542703	.0205
HxHIN	Percent of household income over $150,000	-15.04019161	.0045
HxSPT	Average spending on recreation per HH	.3413614483E-02	.0095
HxNLO	Non-lodging expenditure during previous visit	-.4413332266E-03	.0149

Interpretation:

- Customers who stayed in larger accommodations during their previous visit are less likely to stay in Hotel accommodations than in 3-Bedroom accommodations.
- Customers who stayed with larger number of guests during their previous visit are less likely to stay in Hotel accommodations than in 3-Bedroom accommodations.
- Customers with higher incomes are less likely to stay in Hotel accommodations than in 3-Bedroom accommodations.
- Customers who spend more on recreation are more likely to stay in Hotel accommodations than in 3-Bedroom accommodations.

- Customers with higher non-lodging spending during their previous visit are less likely to stay in Hotel accommodations than in 3-Bedroom accommodations.

Studio Size Choice

Table 5-9: The Most Significant Variables for the Studio Choice

Variable	Description	Coefficient	P[Z]>z
SxSPR	Size for previous visit	-.7284347548	.0001
SxLEN	Length of stay for previous visit	.8283037289	.0002
SxAMO	Net amount for previous visit	-.2739396126E-02	.0092
SxHI	Percent of household income under $150,000	-11.41051586	.0056
SxHIN	Percent of household income over $150,000	-16.26554709	.0061
SxSPT	Average spending on recreation per HH	.3996677120E-02	.0071
SxVIS	Visited other types of resorts #or users per HH	.7328920820E-01	.0032

Interpretation:

- Customers who stayed in larger accommodations during their previous visit are less likely to stay in Studio accommodations than in 3-Bedroom accommodations.
- Customers who stayed longer during their previous visit are more likely to stay in Studio accommodations than in 3-Bedroom accommodations.
- Customers who paid more for their last visit are less likely to stay in Studio accommodations than in 3-Bedroom accommodations.
- Customers with higher income are less likely to say in Studio accommodations than in 3-Bedroom accommodations.
- Customers with higher spending on recreation are more likely to stay in Studio accommodations than in 3-Bedroom accommodations.
- Customers who visited other types of resorts during the last year are more likely to stay in Studio accommodations than in 3-Bedroom accommodations.

One-bedroom Size Choice

Table 5-10: The Most Significant Variables for the 1-bedroom Choice

Variable	Description	Coefficient	P[Z]>z
BxLEN	Length of stay for previous visit	.4237701756	.0167
BxRAT	Room rate for previous visit	-.7001895756E-02	.0071
BxGUE	Number of guests for previous visit	-.1933496500	.0220
BxHI	Percent of household income under $150,000	-7.518343637	.0288
BxHIN	Percent of household income over $150,000	-12.72297702	.0113
BxSPT	Average spending on recreation per HH	.3186884386E-02	.0108

Interpretation:

- Customers who stayed longer during their previous visit are more likely to stay in 1-Bedroom accommodation than in 3-Bedroom accommodations.
- Customers who paid higher room rate during their previous visit are less likely to stay in 1-Bedroom accommodations than in 3-Bedroom accommodations.
- Customers who stayed with larger number of guests during their previous visit are less likely to stay in 1-Bedroom accommodations than in 3-Bedroom accommodations.
- Customers with higher income are less likely to say in 1-Bedroom accommodations than in 3-Bedroom accommodations.
- Customers with higher spending on recreation are more likely to stay in 1-Bedroom accommodations than in 3-Bedroom accommodations.

Two-bedroom Size Choice

Table 5-11: The Most Significant Variables for the 2-Bedroom Choice

Variable	Description	Coefficient	P[Z]>z
BBxLEN	Length of stay for previous visit	.2991779096	.0588
BBxVIS	Visited other types of resorts #or users per HH	.5497460785E-01	.0093

Interpretation:

- Customers who stayed longer during their previous visit are more likely to stay in 2-Bedroom accommodations than in 3-Bedroom accommodations.
- Customers who visited other types of resorts during the last year are more likely to stay in 2-Bedroom accommodations than in 3-Bedroom accommodations.

5.2.2.1 Validation of Size Choice Model

Studio choices were very sparse in the data, making it difficult for the model to detect variables affecting them. With much larger data sets, the studio choices could be oversampled, allowing effects of any predictor variable a better chance of being detected.

Hotel Choice Validation

Table 5-12 shows that the size model hit rate for Hotel accommodation is 80%. Furthermore, the software correctly assigned Hotel choice to 63% of customers that actually selected Hotel accommodation during their last visit.

Table 5-12: Hotel Validation Hit Rate Matrix

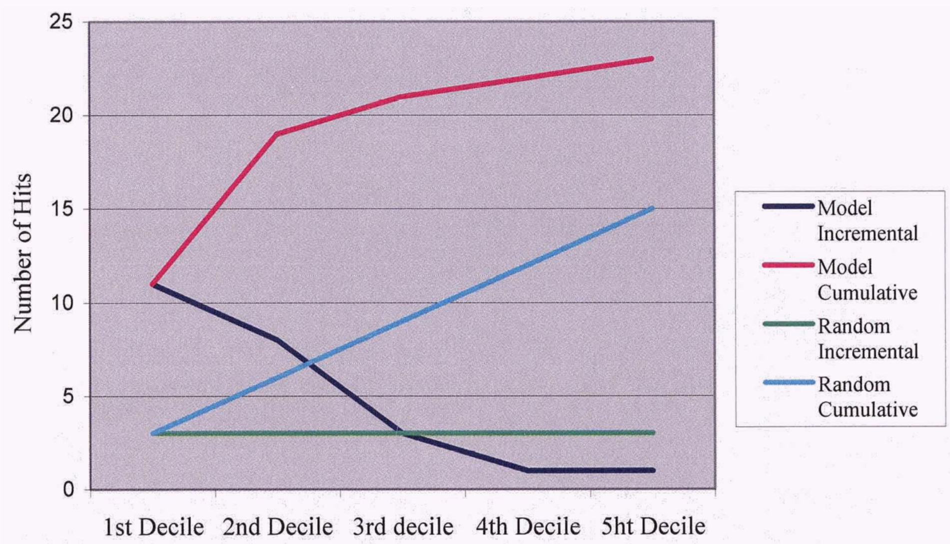
Hotel Validation Hit Rate Matrix			
	Predicted		
		Yes	No
Observed	Yes	19	11
	No	88	387
Hit Rates		19/107	387/398
Total Hit Rate:		406/505 = 80%	

Once the customers were sorted in descending order of their probabilities for the Hotel choice it was found that the model performed better in the first 5 deciles. As Table 5-13 shows, 80% (23 out of 30) of all visitors that selected the Hotel accommodation during their last visit were found within 50% of the top scoring customers. The lift chart for these results can be seen in Figure 3.

Table 5-13: Incremental and Cumulative Hotel Size Hit Rate per Decile

Hotel Hit Rates				
Decile	Hits Incremental	Hits Cumulative	Hits Random Incremental	Hits Random Cumulative
0.1	11	11	3	3
0.2	8	19	3	6
0.3	3	22	3	9
0.4	1	23	3	12
0.5	1	24	3	15

Figure 3: Hotel Size Choice Lift



Studio Choice Validation

Table 5-14 shows that the size model hit rate for the Studio accommodation is 96%. However, the software correctly assigned Studio accommodation to 1 customer that actually selected Studio accommodation during their last visit. As noted above, this reflects model difficulty in picking up the very sparse studio data.

Table 5-14: Studio Validation Hit Rate Matrix

Studio Validation Hit Rate Matrix			
		Predicted	
		Yes	No
Observed	Yes	1	18
	No	1	485
Hit Rates		1/2	485/503

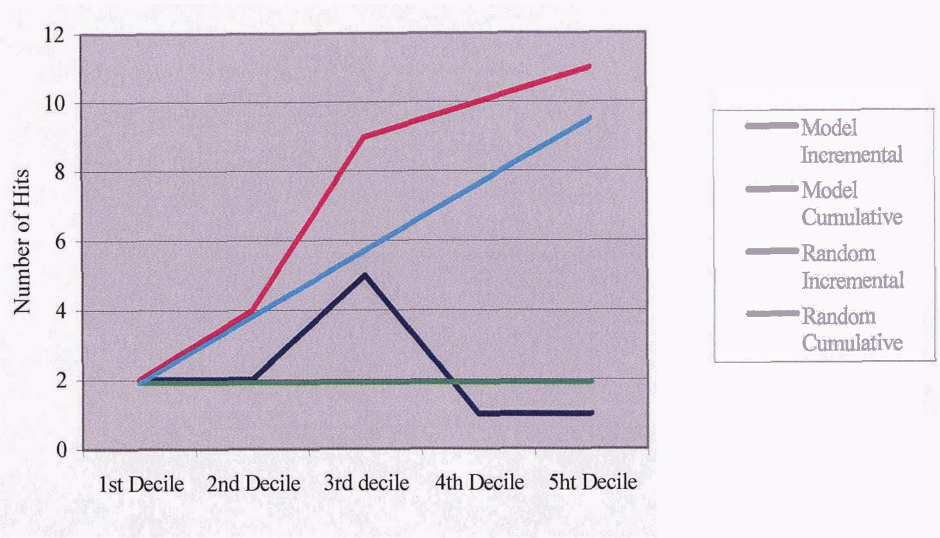
Total Hit Rate: 486/505 = 96%

Once the customers were sorted in descending order of their probabilities for the Studio choice it was found that the model performed better in the first 5 deciles. As Table 5-15 shows, 58% (11 out of 19) of all visitors that selected Studio accommodation during their last visit were found within 50% of the top scoring customers. The lift chart for these results can be found in Figure 4.

Table 5-15: Incremental and Cumulative Studio Size Hit Rate per Decile

Studio Hit Rates				
Decile	**Hits Incremental**	**Hits Cumulative**	**Hits Random Incremental**	**Hits Random Cumulative**
0.1	2	2	1.9	1.9
0.2	2	4	1.9	3.8
0.3	5	9	1.9	5.7
0.4	1	10	1.9	7.6
0.5	1	11	1.9	9.5

Figure 4: Studio Size Choice Lift



One-bedroom Choice Validation

Table 5-16 shows that the Size Model hit rate for 1-bedroom accommodation is 53%. The software correctly assigned 1-bedroom choice to 47% of customers that actually selected 1-bedroom accommodation during their last visit.

Table 5-16: 1-bedroom Validation Hit Rate Matrix

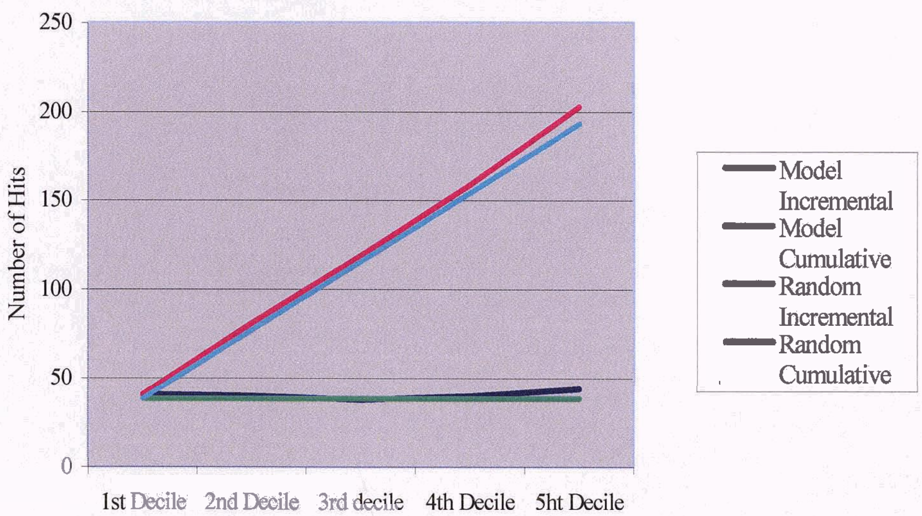
1-Bedroom Validation Hit Rate Matrix			
		Predicted	
		Yes	**No**
Observed	**Yes**	184	202
	No	39	80
Hit Rates		184/223	80/282
Total Hit Rate:		264/505 = 53%	

By sorting customers in descending order based on their probabilities for 1-bedroom size accommodation it was found that the model did not perform much better in the first five deciles. As Table 5-17 shows, 53% of all visitors that selected 1-bedroom accommodation during their last visit were found within 50% of the top scoring customers. The lift chart for these results can be found in Figure 5.

Table 5-17: Incremental and Cumulative 1-bedroom Size Hit Rate per Decile

One - bedroom Hit Rates				
Decile	**Hits Incremental**	**Hits Cumulative**	**Hits Random Incremental**	**Hits Random Cumulative**
0.1	41	41	38.6	38.6
0.2	40	81	38.6	77.2
0.3	38	119	38.6	115.80
0.4	40	159	38.6	154.4
0.5	44	203	38.6	193

Figure 5: 1-bedroom Size Choice Lift



Two-bedroom Choice Validation

Table 5-18 shows that the Size Model hit rate for 2-bedroom accommodation is 67%. The software correctly assigned 2-bedroom choice to 54% of customers that actually selected 2-bedroom accommodation during their last visit.

Table 5-18: 2-bedroom Validation Hit Rate Matrix

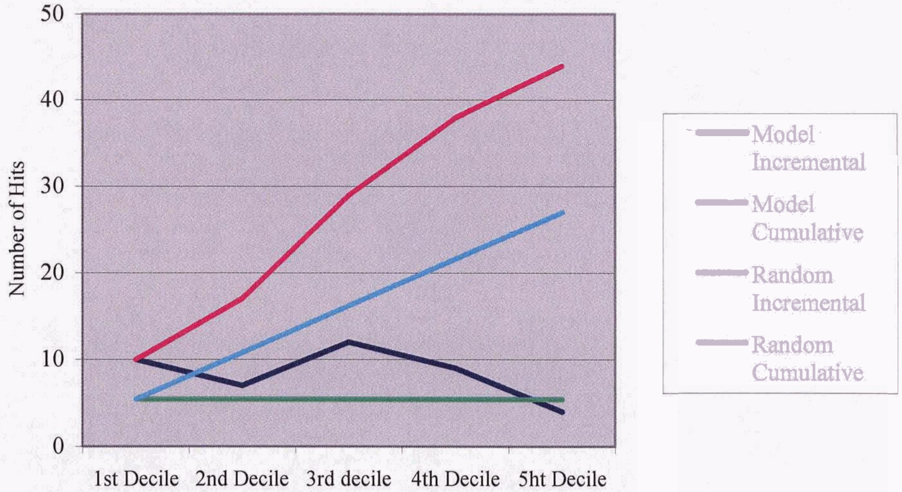
2 Bedroom Validation Hit Rate Matrix			
	Predicted		
		Yes	No
Observed	Yes	29	25
	No	140	311
Hit Rates		29/169	311/336
Total Hit Rate:	340/505 = 67%		

Once again, by sorting customers in descending order based on their probabilities for 2-bedroom accommodation it was found that the model performed better in the first 5 deciles. As Table 5-19 shows, 78% (42 out of 54) of all visitors that selected 2-bedroom accommodation during their last visit were found within 50% of the top scoring customers. This is an improvement of 28% over random approach. The lift chart for these results can be found in Figure 6.

Table 5-19: Incremental and Cumulative 2-bedroom Size Hit Rate per Decile

Two - bedroom Hit Rates				
Decile	**Hits Incremental**	**Hits Cumulative**	**Hits Random Incremental**	**Hits Random Cumulative**
0.1	10	10	5.4	5.4.
0.2	7	17	5.4	10.8
0.3	12	29	5.4	16.2
0.4	9	38	5.4	21.6
0.5	4	42	5.4	27

Figure 6: 2-bedroom Size Choice Lift



Three-bedroom Choice Validation

Table 5-20 illustrates that the size model hit rate for 3-bedroom accommodation is 97%. The software assigned 3-bedroom choice to 2 visitors that actually selected 3-bedroom or larger accommodation during their last visit.

Table 5-20: 3-bedroom Validation Hit Rate Matrix

3-Bedroom Validation Hit Rate Matrix			
		Predicted	
		Yes	**No**
Observed	**Yes**	2	14
	No	2	487
Hit Rates		2/4	487/501

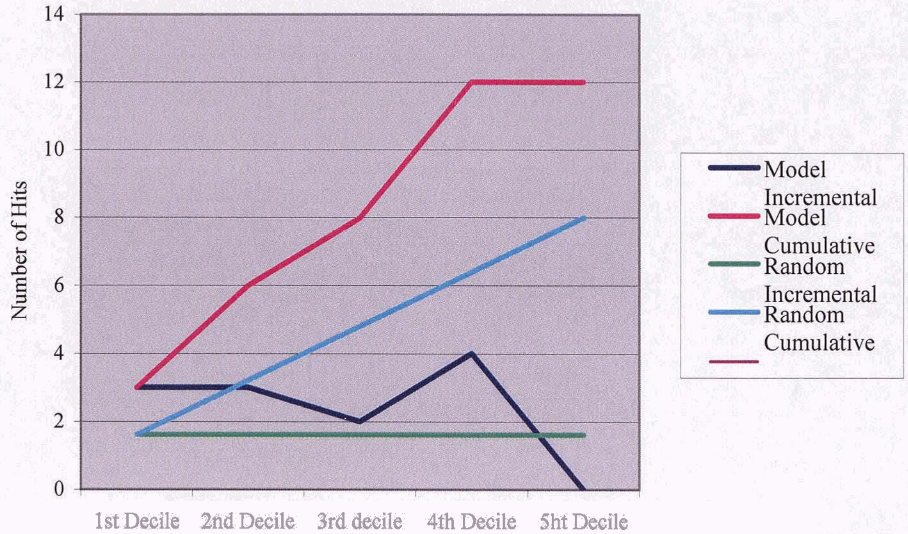
Total Hit Rate: 489/505 = 97%

Similarly to other size choices, the model performed better in the first 5 deciles. As Table 5-21 shows, 75% of all visitors that selected 3-bedroom or larger accommodation during their last visit were found within 50% of the top scoring customers. The lift chart for these results can be found in Figure 7.

Table 5-21: Incremental and Cumulative 3-bedroom Size Hit Rate per Decile

Three - bedroom Hit Rates				
Decile	**Hits Incremental**	**Hits Cumulative**	**Hits Random Incremental**	**Hits Random Cumulative**
0.1	3	3	1.6	1.6
0.2	3	6	1.6	3.2
0.3	2	8	1.6	4.8
0.4	4	12	1.6	6.4
0.5	0	12	1.6	8

Figure 7: 3-bedroom Size Choice Lift



6 LIMITATIONS AND FURTHER RESEACH

As the results indicate, if applied in practice, the Size Choices and the Medal Choices models would bring a certain level of improvement to marketing campaign effectiveness. However, it is important to note that several issues might have had an effect on the predictive power of these models. This chapter discusses these limitations and offers suggestions how to overcome them in the future research.

The majority of the available customer attributes were derived from relating customers' addresses to neighbourhood variables such as PRIZM clusters or neighbourhood census variables. These variables indicated percentages of the population in a given area with the certain characteristics. This fact most certainly introduced a certain level of error as these variables only indirectly reflected individual characteristics of a resort visitor.

It is quite likely that the predictive power of either model would be much stronger if the personal characteristics on an individual level were used in the analysis. One way of obtaining this information is by surveying repeat customers or by purchasing third party data.

In addition, given that the majority of accommodations offered at the resort are Gold, or Silver medal, and Hotel, 1-bedroom, or 2-bedroom size units, it is possible that the data set does not reflect customer's true choice but the availability at the time of the reservation. One way to address this is by setting up an experiment aimed to determine customer's true choice and then modelling the data obtained via such experiment.

In addition, both multinomial models were built mostly on the individual specific variables. The only available alternative specific variable was RATE or price customers paid for their accommodation at the second last visit. This means that the medal choices were used more as signals that were supposed to indicate a certain level of accommodation quality. It is quite likely that customers didn't fully understand the difference among medal levels and used price as the only quality signal.

One way of addressing this issue is through conjoint analysis that would reveal customers' true accommodation needs while on ski vocation. Once the alternative specific attributes are determined it would be easier to link different customers with specific accommodation characteristics they might seek.

APPENDICES

Appendix A and B present NLOGIT outputs for the two models detailing all significant variables as well as the variables that even though not significant improved the predictive power of either model. As previously mentioned, the null hypothesis for each variable was that the estimates of parameters equal zero. Significant variables are those for which the null hypothesis was rejected as their p-value was equal or less than 0.05.

Coefficients for significant variables indicate an increase or decrease in probability of selecting a certain alternative relative to the chosen base choice. For example, for the Bronze medal choice:

- AVHSIZ – Average Household Size has a negative coefficient of -6.690175282 indicating a decrease in probability of choosing Bronze over Platinum accommodation with an increase in a household size.
- ITAE – Active Entrants Segment has a positive coefficient of 1.697237797 indicating an increase of probability of choosing Bronze over Platinum lodging for the customers belonging to this target group.

Finally, the list of predictive variables and their descriptions for both models can be found in Appendix C.

Appendix A

```
DISCRETECHOICE;Lhs=TARGET;Choices=b,g,s,p;Rh2=ONE,QPRE,SPRE,SUBMAR,LENP
RE,AMOPRE,RATEPRE,GUEPRE,DVMARSEG,AGE,DVNAGE,AE,CE,LL,DIS,AVHSIZ,AVVHCL
S,MEDHINC,LHOUSE,HHOUSE,HI1015,HINC,SINGLE,LGFAM,HHLDSU18,SPT_REC,TRAVE
L,LDG,EDU,AFFIND,VISRESOR,NLODAMT;Crosstab$
Normal exit from iterations. Exit status=0.
CuMedal.xls
+---------------------------------------------+
| Discrete choice (multinomial logit) model   |
| Maximum Likelihood Estimates                |
| Model estimated: Apr 14, 2006 at 06:36:16PM.|
| Dependent variable              Choice      |
| Weighting variable                None      |
| Number of observations            1405      |
| Iterations completed                 6      |
| Log likelihood function      -1376.789      |
| Log-L for Choice   model =  -1376.78871     |
| R2=1-LogL/LogL*  Log-L fncn  R-sqrd  RsqAdj |
| Constants only   -1512.3041  .08961  .06839 |
| Chi-squared[93]          =    271.03069     |
| Prob [ chi squared > value ] =    .00000    |
| Response data are given as ind. choice.     |
| Number of obs.=  1405, skipped   0 bad obs. |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] |
+---------+--------------+----------------+--------+---------+
 A_B         36.08508706       13.688615     2.636   .0084
 BxQPR1     -.9075038848       .25472517    -3.563   .0004
 BxSPR1     -.5223972454       .21579840    -2.421   .0155
 BxSUB1      .3399259102       .43704694      .778   .4367
 BxLEN1      .2120365526       .24644275      .860   .3896
 BxAMO1   -.1171171329E-02  .12029114E-02    -.974   .3302
 BxRAT1   -.2737201984E-02  .41856999E-02    -.654   .5131
 BxGUE1      .2240273923       .12929264     1.733   .0831
 BxDVM1      .3858222229       .46031252      .838   .4019
 BxAGE1    .1248951680E-01  .19221550E-01     .650   .5158
 BxDVN1      .2708152776       .95164635      .285   .7760
 BxAE       1.697237797       .84969955    1.997   .0458
 BxCE       1.749765820       .66151352    2.645   .0082
 BxLL       2.899162960       1.0280706    2.820   .0048
 BxDIS1   -.1178523118E-02  .53293073E-03   -2.211   .0270
 BxAVH1     -6.690175282       3.4965883    -1.913   .0557
 BxAVV1     -.4734838600       .85090282     -.556   .5779
 BxMED1    .1503651805E-04  .14125047E-04    1.065   .2871
 BxLHO1     -1.387589374       1.0937761    -1.269   .2046
 BxHHO1     -4.069577282       1.7430688    -2.335   .0196
 BxHI11     -10.68956747       5.0889472    -2.101   .0357
 BxHIN1     -16.39293733       8.2461117    -1.988   .0468
 BxSIN1     -24.38201579       10.233184    -2.383   .0172
 BxLGF1      20.98474049       10.907163     1.924   .0544
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
BxHHL1	-6.624024240	6.7653908	-.979	.3275
BxSPT1	-.5583400419E-03	.24731947E-02	-.226	.8214
BxTRA1	-.5417223723E-02	.27711080E-02	-1.955	.0506
BxLDG1	.3747439299E-01	.13576515E-01	2.760	.0058
BxEDU1	1.749203269	1.6322865	1.072	.2839
BxAFF1	-.1205865299E-01	.84291880E-02	-1.431	.1525
BxVIS1	-.3127875508E-01	.34631836E-01	-.903	.3664
BxNLO1	-.3164056553E-03	.22526265E-03	-1.405	.1601
A_G	26.91285106	11.280045	2.386	.0170
GxQPR2	-.6820360985	.20704707	-3.294	.0010
GxSPR2	-.3667198935	.18119233	-2.024	.0430
GxSUB2	-.2904312183	.37662989	-.771	.4406
GxLEN2	-.2148390132	.18333211	-1.172	.2413
GxAMO2	.1054576562E-02	.71619163E-03	1.472	.1409
GxRAT2	-.3986248855E-02	.28052539E-02	-1.421	.1553
GxGUE2	.5697757595E-01	.10298255	.553	.5801
GxDVM2	.2560452086	.37657357	.680	.4965
GxAGE2	.1648893559E-01	.16001528E-01	1.030	.3028
GxDVN2	.9361517284	.78705762	1.189	.2343
GxAE	.5730872596	.72659536	.789	.4303
GxCE	.3740851987	.52159857	.717	.4733
GxLL	1.514296811	.82903552	1.827	.0678
GxDIS2	-.4760940255E-03	.41546388E-03	-1.146	.2518
GxAVH2	-4.345858395	2.8110030	-1.546	.1221
GxAVV2	-1.423814541	.71193568	-2.000	.0455
GxMED2	.1730143789E-04	.12235219E-04	1.414	.1573
GxLHO2	-.1136317550	.89372530	-.127	.8988
GxHHO2	-1.775018743	1.2852783	-1.381	.1673
GxHI12	-8.007362082	4.2558363	-1.882	.0599
GxHIN2	-15.50073870	6.8730854	-2.255	.0241
GxSIN2	-16.29200822	8.6059284	-1.893	.0583
GxLGF2	11.75591281	8.8203829	1.333	.1826
GxHHL2	-5.172669650	5.8480639	-.885	.3764
GxSPT2	.1343717894E-02	.19472546E-02	.690	.4902
GxTRA2	-.3791479588E-02	.26266686E-02	-1.443	.1489
GxLDG2	.2207680000E-01	.11978785E-01	1.843	.0653
GxEDU2	.3589251694E-02	1.3261751	.003	.9978
GxAFF2	-.1659982938E-02	.69062653E-02	-.240	.8101
GxVIS2	-.2649799154E-02	.29618605E-01	-.089	.9287
GxNLO2	-.1539299597E-03	.13702129E-03	-1.123	.2613
A_S	28.55689107	11.309814	2.525	.0116
SxQPR3	-.8694660119E-01	.20654190	-.421	.6738
SxSPR3	-.6297957248	.18190598	-3.462	.0005
SxSUB3	-.1768235618	.37872935	-.467	.6406
SxLEN3	-.2764927589E-01	.18816382	-.147	.8832
SxAMO3	.6627494548E-03	.75925088E-03	.873	.3827
SxRAT3	-.4693792754E-02	.29408803E-02	-1.596	.1105
SxGUE3	.6319604532E-01	.10546599	.599	.5490
SxDVM3	-.6499640122E-02	.38262857	-.017	.9864
SxAGE3	.7759539152E-02	.16060480E-01	.483	.6290
SxDVN3	.5680614847	.78870831	.720	.4714
SxAE	.4910299078	.73495799	.668	.5041
SxCE3	.5382578627	.52586370	1.024	.3060

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
SxLL	1.850957514	.83877718	2.207	.0273
SxDIS3	-.5330702255E-03	.41907592E-03	-1.272	.2034
SxAVH3	-5.096292109	2.8263475	-1.803	.0714
SxAVV3	-.7131842398	.71283286	-1.000	.3171
SxMED3	.1798548845E-04	.12441328E-04	1.446	.1483
SxLHO3	.4480583238	.89805419	.499	.6178
SxHHO3	-2.095759087	1.3156706	-1.593	.1112
SxHI13	-5.885348329	4.2770999	-1.376	.1688
SxHIN3	-13.13676623	6.9013191	-1.904	.0570
SxSIN3	-17.12506840	8.6303824	-1.984	.0472
SxLGF3	10.11940399	8.8483573	1.144	.2528
SxHHL3	-4.743531482	5.8628109	-.809	.4185
SxSPT3	.1660846192E-02	.19616716E-02	.847	.3972
SxTRA3	-.3330638626E-02	.26360821E-02	-1.263	.2064
SxLDG3	.1601989176E-01	.12051503E-01	1.329	.1838
SxEDU3	-.5818775579	1.3306295	-.437	.6619
SxAFF3	-.5350029281E-02	.69523276E-02	-.770	.4416
SxVIS3	.1374774763E-02	.29733041E-01	.046	.9631
SxNLO3	-.2033650848E-03	.14678565E-03	-1.385	.1659

(Note: E+nn or E-nn means multiply by 10 to + or -nn power.)

Medal.xls

Appendix B

```
DISCRETECHOICE;Lhs=TARGET;Choices=h,s,b,bb,bbb;Rh2=ONE,SPRE,QPRE,LENPRE
AMOPRE,RATEPRE,GUEPRE,AGE,DVNAGE,AE,DIS,FAMILIE,AVVHCLS,POPU20,LHOUSE
HHOUSE,HI1015,HINC,SINGLE,SPT_REC,EDU,EXEWEE,VISRESOR,NLODAMT;Crosstab$
Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| Discrete choice (multinomial logit) model   |
| Maximum Likelihood Estimates                |
| Model estimated: Apr 15, 2006 at 07:39:25PM.|
| Dependent variable               Choice     |
| Weighting variable                 None     |
| Number of observations             1509     |
| Iterations completed                  6     |
| Log likelihood function       -1919.129     |
| Log-L for Choice   model =  -1919.12883     |
| R2=1-LogL/LogL*  Log-L fncn  R-sqrd  RsqAdj |
| Constants only   -2215.0944  .13361  .11961 |
| Chi-squared[92]          =    591.93108     |
| Prob [ chi squared > value ] =    .00000    |
| Response data are given as ind. choice.     |
| Number of obs.=  1509, skipped   0 bad obs. |
+---------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
A_H	3.056604026	3.3918646	.901	.3675
HxSPR1	-.9769094404	.17285489	-5.652	.0000
HxQPR1	.5122636013E-01	.20638493	.248	.8040
HxLEN1	.1230886684	.19216299	.641	.5218
HxAMO1	.2582492745E-03	.74245903E-03	.348	.7280
HxRAT1	-.3974412008E-02	.28334927E-02	-1.403	.1607
HxGUE1	-.2979772957	.97044874E-01	-3.071	.0021
HxAGE1	-.1828488311E-01	.15740392E-01	-1.162	.2454
HxDVN1	-1.423714845	.77731606	-1.832	.0670
HxAE	.8828183799	.66426768	1.329	.1838
HxDIS1	-.5055366564E-03	.32928194E-03	-1.535	.1247
HxFAM1	.2542428097E-03	.27912233E-03	.911	.3624
HxAVV1	-.8009955988E-01	.50959869	-.157	.8751
HxPOP1	.8117866526	4.3786646	.185	.8529
HxLHO1	-.7218265296E-01	.72371582	-.100	.9206
HxHHO1	1.431992579	1.2395592	1.155	.2480
HxHI11	-8.437542703	3.6419146	-2.317	.0205
HxHIN1	-15.04019161	5.2958340	-2.840	.0045
HxSIN1	2.954596605	2.6232769	1.126	.2600
HxSPT1	.3413614483E-02	.13155536E-02	2.595	.0095
HxEDU1	-1.525325115	1.1483173	-1.328	.1841
HxEXE1	-.3276912048E-02	.29346530E-02	-1.117	.2642
HxVIS1	.4179699471E-01	.22842250E-01	1.830	.0673
HxNLO1	-.4413332266E-03	.18126280E-03	-2.435	.0149
A_S	2.246579453	3.8551465	.583	.5601
SxSPR2	-.7284347548	.19094271	-3.815	.0001
SxQPR2	.1750348263	.23325514	.750	.4530
SxLEN2	.8283037289	.22204698	3.730	.0002

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
SxAMO2	-.2739396126E-02	.10519315E-02	-2.604	.0092
SxRAT2	-.6090558196E-03	.35768280E-02	-.170	.8648
SxGUE2	-.7478680542E-01	.11044986	-.677	.4983
SxAGE2	-.1605265973E-01	.17436784E-01	-.921	.3572
SxDVN2	-1.273024308	.86139961	-1.478	.1394
SxAE	1.152906850	.71252046	1.618	.1056
SxDIS2	-.4493737533E-03	.36650883E-03	-1.226	.2202
SxFAM2	.4863442991E-03	.28698693E-03	1.695	.0901
SxAVV2	.1392624692E-01	.57919168	.024	.9808
SxPOP2	-5.164880260	5.0334125	-1.026	.3048
SxLHO2	.7494739233E-01	.82547496	.091	.9277
SxHHO2	.5412441382	1.4587741	.371	.7106
SxHI12	-11.41051586	4.1195034	-2.770	.0056
SxHIN2	-16.26554709	5.9345807	-2.741	.0061
SxSIN2	.9668856678	2.9872614	.324	.7462
SxSPT2	.3996677120E-02	.14857005E-02	2.690	.0071
SxEDU2	-2.216880633	1.3162636	-1.684	.0921
SxEXE2	-.6391335402E-02	.32784636E-02	-1.949	.0512
SxVIS2	.7328920820E-01	.24873202E-01	2.947	.0032
SxNLO2	-.2170843745E-03	.18787606E-03	-1.155	.2479
A_B	1.783368824	3.2469848	.549	.5828
BxSPR3	-.2744058262	.16781594	-1.635	.1020
BxQPR3	.6047569336E-01	.19332013	.313	.7544
BxLEN3	.4237701756	.17703709	2.394	.0167
BxAMO3	-.6809489879E-03	.66734420E-03	-1.020	.3075
BxRAT3	-.7001895756E-02	.26006873E-02	-2.692	.0071
BxGUE3	-.1933496500	.84388164E-01	-2.291	.0220
BxAGE3	-.1966385825E-02	.14980457E-01	-.131	.8956
BxDVN3	-.3286659217	.74539836	-.441	.6593
BxAE	.8206634606	.64121317	1.280	.2006
BxDIS3	-.2446132338E-03	.30056455E-03	-.814	.4157
BxFAM3	.1740650851E-03	.26216860E-03	.664	.5067
BxAVV3	.4877141992	.48688315	1.002	.3165
BxPOP3	-4.566980969	4.2561593	-1.073	.2833
BxLHO3	-.3151284215	.66935078	-.471	.6378
BxHHO3	.1211721683	1.1764489	.103	.9180
BxHI13	-7.518343637	3.4387450	-2.186	.0288
BxHIN3	-12.72297702	5.0211695	-2.534	.0113
BxSIN3	1.250747758	2.5020764	.500	.6172
BxSPT3	.3186884386E-02	.12499928E-02	2.550	.0108
BxEDU3	-.8125202193	1.0915321	-.744	.4566
BxEXE3	-.3961115518E-02	.27303586E-02	-1.451	.1468
BxVIS3	.4041747775E-01	.21723030E-01	1.861	.0628
BxNLO3	-.2170141197E-03	.13366599E-03	-1.624	.1045
A_BB	-2.124818738	3.1699488	-.670	.5027
BBxSPR4	-.1896749514	.16473786	-1.151	.2496
BBxQPR4	-.2516675123E-01	.18762168	-.134	.8933
BBxLEN4	.2991779096	.15834066	1.889	.0588
BBxAMO4	-.8930087099E-04	.50382312E-03	-.177	.8593
BBxRAT4	-.3727818247E-02	.20931281E-02	-1.781	.0749
BBxGUE4	-.4600495544E-01	.75591776E-01	-.609	.5428
BBxAGE4	-.2079653266E-01	.14631072E-01	-1.421	.1552
BBxDVN4	-1.135255641	.72697419	-1.562	.1184
BBxAE	.2092748320	.63532029	.329	.7419

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
BBxDIS4	-.3717699935E-03	.29317119E-03	-1.268	.2048
BBxFAM4	.2958148729E-03	.25276804E-03	1.170	.2419
BBxAVV4	.7250710375	.47515676	1.526	.1270
BBxPOP4	1.656877034	4.1269023	.401	.6881
BBxLHO4	-.5036856664	.65011819	-.775	.4385
BBxHHO4	-.4301025814E-01	1.1183645	-.038	.9693
BBxHI14	-5.551780979	3.3379282	-1.663	.0963
BBxHIN4	-6.715390516	4.8767383	-1.377	.1685
BBxSIN4	4.426888864	2.4472664	1.809	.0705
BBxSPT4	.1876752052E-02	.12162565E-02	1.543	.1228
BBxEDU4	-1.718616830	1.0645201	-1.614	.1064
BBxEXE4	-.1701251964E-02	.26113121E-02	-.651	.5147
BBxVIS4	.5497460785E-01	.21144643E-01	2.600	.0093
BBxNLO4	-.4640840963E-04	.11144580E-03	-.416	.6771

(Note: E+nn or E-nn means multiply by 10 to + or -nn power.)

Appendix C

Table 6-1: Alphabetic List of Variables

Variable	*Description*
AFFLUENT	Affluence Index
AE, CE, LL	Target Segments: AE – active entrants, CE – core enthusiasts, LL – luxury lifestyles
AGE	Customer Age
AMOPRE	Lodging transaction amount for second last visit
AVHSIZ	Average household size
AVVHCLS	Average Vehicles
DIS	Average distance of customer's zip code to the resort
DVMARSEG	Dummy variable that identifies independent travellers
DVNAGE	Age dummy variable: 1 representing visitors with age data and 0 for visitors with age data missing
EDU	Education Index
EXEWEE	Weekly exercises

Variable	***Description***
FAMILIE	Average number of families
GUEPRE	Number of guest sharing the same lodging for the second last visit
HHLDSU18	Percent of households with Person 18 or under
HHOUSE	Percent of houses valued over $750,000
HI1015	Percent of households with income between $100,000 and $150,000
HINC	Percent of households with income over $150,000
LDG	Average lodging expenditures per household
LENPRE	Length of stay for the second last visit
LGFAM	Percent of households with more than 5 persons
LHOUSE	Percent of houses valued between $300,000 and $750,000
MEDHINC	Median household Income
NLODAMT	Amount spent on non-lodging transactions during second last visit
POPU20	Percent of population under 20
QPRE	Medal classification of lodging for second last visit

Variable	*Description*
RATEPRE	Average daily rate per unit during second last visit
SINGLE	Percent of household with 1 or 2 person
SPRE	Size of the second last visit lodging
SPT_REC	Average sports and recreation expenditure per household
SUBMAR	Dummy variable that indicates if customer visited the resort during Regular season for the second last visit
TRAVEL	Average travel expenditure per household
VISRESOR	Percent that visited other resorts in the last year

Table 6-2: List of transformed and new variables

Variable	*Transformation*
AFFLUENT	Was created by adding the following variables: Any_Domestic_Travel_by_Airplane, Belong_to-a_Business_Club, Belong_to_a_Country_Club, Belong_to_a_Health_Club, Play_Golf_1_Yr and Stay_Hilton_on_Vacation
DMAGE	Dummy Variable with 1 representing visitors with age data and 0 for visitors with age data missing
EDU	Was created by adding up the following variables: pct_edu_ba, pct_edu_doct_c, pct_edu_mast_c and pct_edu_prof_c
HIHOU	Was created by adding up pct_val7501mlc and pct_val_1milpc
HIINC	Was created by adding up pct_hi_150_250c, pct_hi_250_500c and pct_hi_500p_c
Brand_Segment	Was split into three dummy variables for AE (active entrants), CE (core enthusiasts) and LL (luxury lifestyle)

Variable	*Transformation*
	brand segments
LGFAM	Was created by adding up pct_h_5_pers_c, pct_h_6_pers_c and pct_h_7_pers_c
Line_of_Business_Code	Was transformed into a dummy variable to indicate whether customer purchased any show school services
LHOUSE	Was created by adding up pct_val_300400c, pct_val_400500c and pct_val_500750c
NLOGAM	A new variable that indicated a total amount of visitors spending on non-lodging transactions
NUMEXE	Was created according to the following formula: Exercise_1_Time_Wk + 2*Exercise_2_Time_Wk + 3.5* Exercise_3_4_Time_Wk + 5* Exercise_5_Time_Wk
SINGLE	Was created by adding up pct_h_1_pers_c and pct_h_2_pers_c
SMHOU	Was created by adding up pct_h_3_pers_c and pct_h_4_pers_c
VEHICLE	Was created by adding up pct_vehicle_4pc and pct_vehicle_5pc

Appendix D

Table 6-3: Number of Repeat Customer per Sub-season

		Last Visit							
		Early	**Pre-Xmas**	**X-mas**	**Regular**	**March**	**April**	**Out of Season**	***Row Totals***
Second Last Visit	**Early**	72	30	18	91	50	16	4	**281**
	Pre-Xmass	6	57	13	41	21	7	3	**148**
	X-mass	5	11	35	62	37	10	0	**160**
	Regular	35	24	19	486	252	140	28	**984**
	March	20	13	13	60	177	58	6	**347**
	April	9	1	6	24	43	48	4	**135**
	Out of Season	19	10	10	48	12	12	96	**207**

Table 6-4: Percentage of Repeat Customers per Sub-season

		Last Visit							
		Early	**Pre-Xmas**	**X-mas**	**Regular**	**March**	**April**	**Out of Season**	***Row Percentage***
Second Last Visit	**Early**	**25.62%**	10.68%	6.41%	32.38%	17.79%	5.69%	1.42%	**100%**
	Pre-Xmas	4.05%	**38.51%**	8.78%	27.70%	14.19%	4.73%	2.03%	**100%**
	X-mas	3.13%	6.88%	**21.88%**	38.75%	23.13%	6.25%	0.00%	**100%**
	Regular	3.56%	2.44%	1.93%	**49.39%**	25.61%	14.23%	2.85%	**100%**
	March	5.76%	3.75%	3.75%	17.29%	**51.01%**	16.71%	1.73%	**100%**
	April	6.67%	0.74%	4.44%	17.78%	31.85%	**35.56%**	2.96%	**100%**
	Out of Season	9.18%	4.83%	4.83%	23.19%	5.80%	5.80%	**46.38%**	**100%**

Appendix E

Table 6-5: Number of Repeat Customers per Quality Type

		Last Visit				
		Bronze	**Silver**	**Gold**	**Platinum**	***Row Total***
Second Last Visit	**Bronze**	40	38	38	2	118
	Gold	39	343	809	55	1246
	Platinum	3	11	18	17	49
	Silver	50	460	321	17	848

Table 6-6: Percentage of Repeat Customers per Quality Type

		Last Visit				
		Bronze	**Silver**	**Gold**	**Platinum**	***Row Percentages***
Second Last Visit	**Bronze**	**33.90%**	32.20%	32.20%	1.69%	100%
	Gold	3.13%	27.53%	**64.93**	4.41%	100%
	Platinum	6.12%	22.45%	36.73%	**34.69%**	100%
	Silver	5.90%	**54.25%**	37.85%	2.00%	100%

Appendix F

Table 6-7: Number of Repeat Customers per Unit Type

		Last Visit						
		Hotel	Studio	1-Bedroom	2-Bedroom	3-Bedroom	4/5-Bedroom	*Row Totals*
Second Last Visit	Hotel	215	33	121	55	11	0	**435**
	Studio	38	61	69	25	2	1	**196**
	1-Bedroom	121	52	579	186	29	2	**969**
	2-Bedroom	39	23	156	284	49	5	**556**
	3-Bedroom	7	2	17	35	27	6	**94**
	4/5-Bedroom	1	0	4	3	0	4	**12**

Table 6-8: Percentage of Repeat Customers per Unit Type

		Last Visit						
		Hotel	Studio	1-Bedroom	2-Bedroom	3-Bedroom	4/5-Bedroom	*Row Percentages*
Second Last Visit	Hotel	49.43%	7.59%	27.82%	12.64%	2.53%	0.00%	100%
	Studio	19.39%	31.12%	35.20%	12.76%	1.02%	0.51%	100%
	1-Bedroom	12.49%	5.37%	59.75%	19.20%	2.99%	0.21%	100%
	2-Bedroom	7.01%	4.14%	28.06%	51.08%	8.81%	0.90%	100%
	3-Bedroom	7.45%	2.13%	18.09%	37.23%	28.72%	6.38%	100%
	4/5-Bedroom	8.33%	0.00%	33.33%	25.00%	0.00%	33.33%	100%

Appendix G

T-test: Difference in RATE between One –Time and Repeat Customers

Group Statistics

	DVRepeatVisit	N	Mean	Std. Deviation	Std. Error Mean
RATE	One Time Customers	4627	207.67	111.655	1.641
	Repeat Customers	418	189.96	108.299	5.297

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
RATE	Equal variances assumed	.303	.582	3.113	5043	.002	17.709	5.689	6.557	28.861
	Equal variances not assumed			3.193	500.514	.001	17.709	5.546	6.814	28.605

T-test: Difference in Length of Stay between One-Time and Repeat Customers

Group Statistics

	DVRepeatVisit	N	Mean	Std. Deviation	Std. Error Mean
LENSTAY	One Time Customers	4627	3.17	1.801	.026
	Repeat Customers	418	2.00	1.425	.070

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
LENSTAY	Equal variances assumed	113.818	.000	12.850	5043	.000	1.164	.091	.986	1.341
	Equal variances not assumed			15.604	544.987	.000	1.164	.075	1.017	1.310

T-test: Difference in Distance between One-Time and Repeat Customers

Group Statistics

	DVRepeatVisit	N	Mean	Std. Deviation	Std. Error Mean
DISTANCE	One Time Customers	4627	753.36079	604.919356	8.892989
	Repeat Customers	418	254.39430	415.870712	20.340901

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
DISTANCE	Equal variances assumed	186.387	.000	16.514	5043	.000	498.966493	30.214001	439.7339	558.1990
	Equal variances not assumed			22.476	589.705	.000	498.966493	22.199944	455.3659	542.5670

T-test: Difference in Median House Value between One-Time and Repeat Customers

Group Statistics

	DVRepeatVisit	N	Mean	Std. Deviation	Std. Error Mean
MED_VALUEC	One Time Customers	4627	266560.22	165045.214	2426.349
	Repeat Customers	418	312036.03	173706.913	8496.283

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
MED_VALUEC	Equal variances assumed	1.602	.206	-5.371	5043	.000	-45475.814	8466.837	-62074.49	-28877.13
	Equal variances not assumed			-5.147	487.49	.000	-45475.814	8835.949	-62837.05	-28114.57

T-test: Difference in Median Household Income between One-Time and Repeat Customers

Group Statistics

	DVRepeatVisit	N	Mean	Std. Deviation	Std. Error Mean
MED_HINC_C	One Time Customers	4627	80782.66	41000.492	602.753
	Repeat Customers	418	86060.26	40900.299	2000.499

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
MED_HINC_C	Equal variances assumed	.014	.905	-2.521	5043	.012	-5277.600	2093.602	-9381.969	-1173.230
	Equal variances not assumed			-2.526	495.781	.012	-5277.600	2089.332	-9382.636	-1172.563

T-test: Difference in Recreation Spending between One-Time and Repeat Customers

Group Statistics

	DVRepeatVisit	N	Mean	Std. Deviation	Std. Error Mean
spt_rec	One Time Customers	4627	1950.642002257	818.6763195063	12.0354540382
	Repeat Customers	418	2088.969216869	810.4200956750	39.6389425827

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
spt_rec	Equal variances assumed	.049	.825	-3.311	5043	.001	-138.327214	41.7776771	-220.2296144	-56.4248147
	Equal variances not assumed			-3.339	497.049	.001	-138.327214	41.4258122	-219.7185022	-56.9359270

T-test: Difference in Travel Spending between One-Time and Repeat Customers

Group Statistics

	DVRepeatVisit	N	Mean	Std. Deviation	Std. Error Mean
travel	One Time Customers	4627	2058.872450140	898.5693018310	13.2099698924
	Repeat Customers	418	2227.847645270	866.9718387823	42.4049787533

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
travel	Equal variances assumed	.194	.659	-3.693	5043	.000	-168.975195129	45.76148663	-258.68759248	-79.26279777
	Equal variances not assumed			-3.804	501.436	.000	-168.975195129	44.41492460	-256.23747206	-81.71291819

*Crosstabulation: DVRepeatVisit * MARSEG*

DVRepeatVisit * MARSEG Crosstabulation

Count

		MARSEG				Total
		P	S	T	X	
DVRepeatVisit	One-Time Customers	2375	229	1213	810	4627
	Repeat Customers	75	46	225	76	422
Total		2450	275	1438	886	5049

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	214.683[a]	3	.000
Likelihood Ratio	217.378	3	.000
N of Valid Cases	5049		

There is a weak relationship between these two variables.

*Crosstabulation: DVRepeatVisit * SSCHOOL*

DVRepeatVisit * SSCHOOL Crosstabulation

Count

		SSCHOOL		
		0	1	Total
DVRepeatVisit	One-Time Customers	4020	606	4626
	Repeat Customers	351	71	422
Total		4371	677	5048

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.620[b]	1	.032		
Continuity Correction[a]	4.305	1	.038		
Likelihood Ratio	4.354	1	.037		
Fisher's Exact Test				.036	.021
N of Valid Cases	5048				

There is a weak relationship between these two variables.

*Crosstabulation: DVRepeatVisit * ITWBRANDSEGMENT*

DVRepeatVisit * ITWBRANDSEGMENT Crosstabulation

Count

		ITWBRANDSEGMENT				Total
		ITAE	ITCE	ITLL	ILOTH	
DVRepeatVisit	One-Time Customers	329	1897	620	1781	4627
	Repeat Customers	22	227	37	136	422
Total		351	2124	657	1917	5049

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	27.420[a]	3	.000
Likelihood Ratio	27.436	3	.000
N of Valid Cases	5049		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 29.34.

REFERENCE LIST

American Hotel Lodging Association. (2006). The 2005 lodging industry profile. Washington, DC: American Hotel Lodging Association

Ben-Akiva, M.E. & Lerman, S.R. (1985). Discrete choice analysis: Theory and application to travel demand. Cambridge, MA: MIT Press

Churchill, G.A. & Iacobucci, D. (2001). Marketing research: Methodological foundations (9th ed). Chicago, IL: Dryden Press

Dickinson, C. B. (2006). *Lodging Hospitality.* Journal of Hospitality Industry, 60-63

Inbakaran, R. & Jackson, M. (2005). *Understanding resort visitors through segmentation.* Tourism and Hospitality Research, 6(1), 53-71

Loveman, G. (2003). *Diamonds in the Data Mine*. Harvard Business Review, 109-113

Magnini, P. V. (2003). *Data mining for hotel firms: Use and limitations.* Cornell Hotel and Restaurant Administration Quarterly, 94-105

'*Off-the-Charts Results*'. (2004, Jun). Baseline, p26-32

Riddington, G. (2000). *Modelling choice and switching behaviour between Scottish ski centres*. Applied Economics, 32, 1011-1018

Rodriguez, G. Lecture Notes. Retrieved Jun 28, 2006 from http://data.princeton.edu/wws509/notes/c6s2.html

Slavens, R. (2006). *Relationship building essential*, Vertical Insight, 91(8), 36-38

Wawro, G. (Jun 2006). *Advanced topics in maximum likelihood categorical choice models and models for qualitative dependent variables with repeated observations*. Retrieved Jun 29, 2006 from http://www.columbia.edu/~gjw10/icpsr_notes.pdf