**University of Kwazulu-Natal**

# Develop a Predictive Model that drives Business Strategy to determine Property Sales within the Real Estate Industry based in KwaZulu Natal.


**By**

**Mischa Naidoo**

**961045758**


A dissertation submitted in partial fulfillment of the requirements for the degree of Master of Business Administration


College of Law and Management Studies

Graduate School of Business & Leadership


Supervisor:

Dr. ME Hoque

2017

UNIVERSITY OF KWAZULU-NATAL

INYUVESI YAKWAZULU-NATALI

**College of Law and Management Studies**

# SUPERVISORS PERMISSION TO SUBMIT THESIS/ DISSERTATION FOR EXAMINATION

<table>
<tr><td colspan="2">Name: Mischa Naidoo</td><td colspan="2">Student No: 961045758</td></tr>
<tr><td colspan="4">Title: Develop a Predictive Model that drives Business Strategy to determine Property Sales within the Real Estate Industry based in KwaZulu Natal.</td></tr>
<tr><td colspan="2">Qualification: Master of Business Administration (MBA)</td><td colspan="2">School: Graduate School of Business and Leadership: UKZN</td></tr>
<tr><td colspan="2"></td><td>Yes</td><td>No</td></tr>
<tr><td colspan="2">To the best of my knowledge, the thesis/dissertation is primarily the student’s own work and the student has acknowledged all reference sources</td><td>X</td><td></td></tr>
<tr><td colspan="2">The English language is of a suitable standard for examination without going for professional editing.</td><td>X</td><td></td></tr>
<tr><td colspan="2">Turnitin Report %</td><td colspan="2">5%</td></tr>
<tr><td colspan="4">Comment if % is over 10%:</td></tr>
<tr><td colspan="2">I agree to the submission of this thesis/dissertation for examination</td><td>X</td><td></td></tr>
<tr><td colspan="4">Supervisors Name: Dr. Muhammad Hoque</td></tr>
<tr><td colspan="4">Supervisors Signature:</td></tr>
<tr><td colspan="2">Date: 30 November 2017 (Original Submission)<br>Resubmission: 12 March 2018 ( Minor corrections required)</td><td colspan="2"></td></tr>
<tr><td colspan="4">Co- Supervisors Name:</td></tr>
<tr><td colspan="2">Co- Supervisors Signature:</td><td colspan="2"></td></tr>
<tr><td colspan="4">Date:</td></tr>
</table>

# DECLARATION

I Mischa Naidoo declare that:

- The research reported in this thesis, except where otherwise indicated, is my original work.

- This thesis has not been submitted for any degree or examination at any other university.

- This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

- This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

    a) their words have been re-written but the general information attributed to them has been referenced;

    b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.

    c) Where I have reproduced a publication of which I am author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.

    d) This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed:________________________________

# ACKNOWLEDGEMENTS

I am thankful to my family for their inspiration during the past 2-and-a-half years, for tolerating my absence and calming my stress. I would like to thank my children for understanding the sacrifice that I had to make during my MBA.

I would like to express my sincere thanks to my supervisor, Dr. ME Hoque, for his excellent guidance, encouragement and advice from inception to the conclusion of this dissertation. His support, assistance, collaboration and critical feedback provided me with the drive to pursue the writing of this thesis. All of his advice helped me to re-organize and polish the thesis.

To my friends who encouraged me during my studies, I am grateful for your support.

I would also like to thank my work colleagues for taking time off from their schedules to assist my study. Without you, this would not have been possible.

Most of all, I am deeply grateful to my parents for their support, encouragement, and spirit while I studied for the Master of Business Administration.

Finally, I give deep thanks to the Lord for walking with me during this journey and guiding me on the straight and narrow path.

# ABSTRACT

The study focuses on property owner's attitudes, behavior and personal drivers when deciding to sell their real estate property. There have been limited studies performed around types of behavior that drives property owners to sell their property. Very little information around this topic exists in South Africa and this poses a risk for property buyers and estate agents of a residential property. Determining the drivers that influence a property owner to sell their property will generate property stock for estate agents in terms of identifying when an individual in their life cycle will sell their property. Apart from estate agents obtaining property stock from property sellers, they will also have the opportunity to sell the property seller another property. This quantitative study seeks to examine three suburbs within KwaZulu Natal residential property market and formulate a regression model to best predict what motivates an individual to sell their real estate property. The research included a seven-year sample period of residential property sales from 2010 to 2017; along with South African citizen data sourced from Home Affairs and public domain information and adopts a regression analysis to interpret the data at the relevant significance level. The 80:20 rule based on the Pareto principal was used to split the data in a test and train dataset, with the train subset being used to the build the predictive model and the test dataset to evaluate the model accuracy.

The results from the analysis applied on the three suburbs within KwaZulu Natal indicates a good fit with an accuracy of 73.4% prediction of properties that are highly likely to go on sale. Variables applied to the study that are found to be statistically significant include: 1. The price of the property; 2. Age of the property; 3. Property owner's age, gender, lifestyle indicator (LSM) and loan finance credit risk score; 4. Historical property sales data and 5. Population suburb density. The relevant results were then interpreted and recommendations provided to property estate agents that indicated why an individual would sell their property.

Key words: Consumer behavior, lead generation, predictive analytics, selling decision.

# TABLE OF CONTENTS

# LIST OF FIGURES

**Number** **Description** **Page**

# LIST OF TABLES

**Number** **Description** **Page**

# LIST OF ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| LSM: | Living Standard Measurement |
| KZN: | KwaZulu Natal |
| CIPC: | Companies and Intellectual Property Commission |
| SA: | South Africa |
| SQL: | Structured Query Language |
| CON: | Consumer |
| COM: | Commercial |
| DE: | Deeds |
| HA: | Home Affairs |
| API: | Application Interface |

# CHAPTER ONE: Research Overview

## 1.1 Introduction

Residential property is an important segment of the property market in South Africa and is the driver for economic growth. Consumer spending has formed a significant element of the gross domestic product and is directly affected by changes in a consumer's needs and wants (Hawkins et al., 2013).

According to Mooya (2016), real estate studies are generally based on neoclassical economic theory that is adopted when it comes to buying or selling a property i.e. people make rational decisions in order to maximize their utility. Other researchers have tried to examine the influence that real estate buyers and sellers have on the demand for real estate by understanding the internal and external forces that drives individuals to make their decisions.

By integrating the human element of decision making like 'tastes and preferences' with the financial decision factors, will lead to a better understanding and prediction of decision makers actions in the real estate market (Ajzen, 1991). The purpose of this study is to integrate consumer behaviour theories to help improve real estate studies, specifically around what drives a property owner to sell their property. Although there are several techniques to estimate a consumer decision to purchase a property, further research is required to provide insight into what drives a property owner to sell their property.

This chapter presents an overview of the research project, background to the study, contextualizes the research and defines the problem statement. Research objectives and hypothesis are expressed which provides insight into how the research problem was explored.

## 1.2 Motivation for the Study

Consumer behaviour is the study of individuals or organizations in selecting, purchasing and selling of goods and services to satisfy their needs and desires. Loudon and Della Bitta (1993) stated that on a macro level, marketers are interested in the drivers for those behaviors, society's values and practices that affect a consumer's interaction with the marketplace.

Within the real estate environment, there is a need for insight on consumer selling behavior. The interest was to find out what drives a property owner to sell their property and with this understanding, how will this improve an estate agent's marketing approach. This will allow agencies to segment their customer database against properties that will improve product uptake with reduced sales and marketing effort. Consumer behavior research is the primary driver of any good strategy that provides actionable insights and ensures business success. According to Sorescu et al. (2011), marketing is understanding your consumers really well which then allows you to create valuable products, services and information to solve their problems.

Real estate data originates from multiple data sources that reside on different platforms and within different formats. This makes it challenging to consolidate data into one central location so that it can be analyzed. Another challenge faced related to data being too complex due to its format being structured, unstructured or semi-structured, which makes data aggregation complicated. This was the problem that underpinned this research.

This study will unpack the various sources of data that is related to the Real Estate industry in order to create predictive models, which will contribute to an improved understanding that will drive business strategy to determine property sales. In this data-intensive world, predictive models are becoming more valuable than ever in order to make sense of what is around us in terms of estimating, planning or assessing future events. With this understanding, real estate agencies can improve and optimize their approach when generating leads by targeting specific consumers that have a high probability of selling their property.

### 1.3 Focus of the Study

The focus of the study is based on consumer behavior that drives decision-making relating to real estate buying and selling. Most studies focus on the consumer buying behavior relating to property purchase. This study focuses on seller behavior and the psychological and financial factors that influence that behavior. Sellers set reservation prices of their property which reflect that strategies, but in most cases, reservation prices are almost always unobservable. A large focus area relates to the seller's lifecycle in terms of 'what moves' an individual to sell e.g. getting married. Real estate companies need to leverage these marketing tactics to elicit a seller's response by motivating them to sell their property. This will help them drive their business strategy.

This study will be conducted on sellers behavior within the real estate industry geographically located in KwaZulu-Natal. The population data will be a snapshot of three suburbs including Umhlanga, Mount Edgecombe and Phoenix.

### 1.4 Significance of the Study

Estate agencies are faced with a plethora of consumer and property data and if leveraged correctly, places them in an ideal position to generate 'Hot' leads from a marketing perspective. The intelligence that estate agencies can derive from this variety of data include property evaluations, investment analysis, determining property asking prices and identifying which properties are available for sale. Data that estate agents are exposed to include housing attributes, location attributes, consumer preferences, economic climate changes and wealth profiles. This study provides a model to property agents to help determine opportunities around when a consumer is most likely to sell their property.

The empirical evidence produced by this study was created using a regression analysis cluster technique. This evidence will be used to create a predictive scorecard model that will plug into business strategies in order to determine when a property owner will sell their property. A model of this nature, to the best of my research knowledge, has never been attempted in South Africa. Therefore, finding global research with this methodology has proven to be challenging with no pre-existing studies being identified.

An important feature of this model is that it combines consumer behavior attributes with attributes relating to when the property was purchased by utilizing large amounts of Home Affairs and Deeds data specific to three suburbs within KwaZulu-Natal. The application of this study has the potential to significantly benefit the real estate industry including investor and buyers of residential properties for evaluating whether a property owner would be interested in selling their property.

This study contributes to the existing body of knowledge by enriching current literature on what drives property owners to behave in a certain manner when selling their property. In addition, the study aimed to contribute to the limited research done on residential property within KwaZulu-Natal by utilizing the plethora of consumer and property data.

## 1.5 Problem Statement

The challenge that faces real estate agencies is identifying the probability that an individual, who owns a property, is likely to sell their property. According to Mooya (2016), real estate agents whom are generally commission earners, accumulate property stock by cold calling customers, referrals or by simply canvassing areas by driving within suburbs and viewing property 'For Sale' boards. Without having a full appreciation of the consumer's lifetime stage of ownership, it often results in wasted effort both in time consumed and expenses incurred in terms of phone calls made. Property lifetime stage ownership refers to the length of time in months for a property owner to no longer own the property i.e. from the time the consumer obtains first employment, getting married, starting a family or retiring.

An individual's behavior relating to sale of property is the business intelligence that real estate companies lack, which leads to smarter strategic-level thinking at the very top of the organization (Peter, 2016). which relates to the problem statement of this study. This is largely due to the vast streams of data from multiple sources that creates complexity in compiling the data into a central repository that is capable of interrogation. This leads to a lack of business reports that will drive strategies to improve decision-making.

The main research question, which relates to the aim of the study, was to determine: Can a statistical model predict an property owner's behavior to sell a property?

## 1.6 Research Sub-Questions:

- Does the introduction of the model result in financial benefits and/or cost saving in the long/short term?
    - Is the data representative of the target population of all formal properties in the South African market and will the data add value in terms of the correlation of Property Sales and information on hand?
    - What is the predictive accuracy of the model and can the underlying drivers of the prediction be understood?
    - Does the model prediction offer additional accurate information to assist in operational or strategic decision-making?

## 1.7 Objectives

The research, relating to the objective of the study intends to develop an integrated model that allows estate agents to identify market segments of consumers that are willing to sell their property. This will open opportunities for agents to not only obtain real estate stock to sell, but also offer the consumer a property as a replacement. The aim is to reduce sales effort; resource cost and creates efficiencies in the sales process.

The research relating to the objective of the study is intended to develop an integrated model that allows estate agents to segment their property listings to a property owner base that has a higher propensity of selling the property.

According to Yin (2014), performing a case study of this nature is preferable in situations whereby:

1. The primary research question pertains to ‘how’ or ‘why’ scenarios and
2. There is little influence over behavioral events.

This study aimed to answer the question of how to use ‘the propensity to sell a property’ model to predict a property owner’s behavior based on historical customer sales data, population suburb density, active companies in the area, number of houses sold in the last 5 years, average sales growth in property sales and the stage of a consumer's life cycle. The business aim is to initiate reduced sales effort, resource cost and to create efficiencies in the sales process.

1.7.1 **The overall objective is to determine if models can be created to predict property owner's behavior when selling real estate located in KwaZulu-Natal submarkets.**

1.7.2 Examine if the variety of real estate data sourced, based on KwaZulu-Natal submarkets, can be consolidated into a usable format that adds value to business strategy.

1.7.3 To determine if suburbs within KwaZulu-Natal can be measured in terms of propensity to sell.

1.7.4 To make recommendations on which business models will improve decision making within the real estate industry

## 1.8 Research Hypotheses

In order to achieve the overall research objective and sub-research objectives, the following research hypotheses were formulated:

*H0a*: Models cannot be created to predict property owner's behavior when selling real estate located in KwaZulu-Natal submarkets.
*H1a*: Models can be created to predict property owner's behavior when selling real estate located in KwaZulu-Natal submarkets.

*H0b*: The variety of real estate data sourced, based on KwaZulu-Natal submarkets, cannot be consolidated into a usable format that adds value to business strategy.
*H1b*: The variety of real estate data sourced, based on KwaZulu-Natal submarkets, can be consolidated into a usable format that adds value to business strategy.

*H0c*: The consumer population identified within KwaZulu-Natal submarkets cannot be measured in terms of determining the propensity for a property owner to sell their property.

*H1c*: The consumer population identified within KwaZulu-Natal submarkets can be measured in terms of determining the propensity for a property owner to sell their property.

## 1.9 Methodology

Quantitative methods were chosen for this study because consumer behaviour decisions related to real estate transactions have been studied by previous researchers for many years and the results are readily available. The object of this research is to test the hypothesis by using variables, which are quantifiable and measurable in order to form generalizations from the population data.

## 1.10 Outline of Thesis

This study will comprise of five chapters, which are introduced below in Table 1.1.

Table 1.1 Chapter Outline

| Chapter | Heading | Description |
|---|---|---|
| Chapter 1 | Introduction of the research | Chapter introduces the background, research problem, research questions, research hypothesis and a brief explanation of methodology adopted. |
| Chapter 2 | Literature Review | This chapter presents a comprehensive literature review of international and local studies based on consumer behaviour relevant to the research objectives and research hypotheses. It also presents an overview of the South African housing market. |
| Chapter 3 | Research Methodology | This chapter presents and defines the data analysis approach in order to answer the research objectives and research hypothesis. This chapter justifies the research paradigm, the research methodology, data collection methods on the population data and finally explains the data analysis |

| | | |
|---|---|---|
| | | procedure. |
| Chapter 4 | Analysis of Data | This chapter presents the results of the study and unpacks the statistical models employed using the statistical software R. It explains the results in relation to the context of the literature review presented in chapter two. |
| Chapter 5 | Conclusion and Recommendations | This chapter discusses and presents the conclusions of the study as well as the research limitation and recommendations of the study |

## 1.11 Delimitation of the study

The first delimitation is the South African residential housing market is highly complex and there are many factors that influence property owner's behaviour relating to the sale of their properties. For example, interest rate fluctuations directly affect the housing market. However, this study does not examine all factors and focus specifically on the property owner's life cycle.

The second delimitation is that the scope of the study relates solely to the domain of residential property, which is a subset of the entire property portfolio which does not factor commercial, industrial or retail properties.

Finally, the population data that was used in this study was for three coastal submarkets of KwaZulu-Natal. The primary reason for using these submarkets is that it represents a diverse split of consumer groups in terms of lifestyle, social and cultural backgrounds. In addition, there was a lack of research performed against these areas and finally, there was ample data available.

## 1.12 Limitations of the study

The real estate data was obtained in both a structured and unstructured format. According to Blumberg and Atre (2003), unstructured data denotes text heavy data that does not have a data defined layout or organized in a format that can be easily processed. Due to this, data had to be manipulated in order to render it into a readable format. As such, there were high efforts in data quality maintenance, however the data format may not always be consistent inflicting distorted results.

The second limitation in the study was missing data; especially address information from the Deeds data source. Measures were implemented to populate the missing data using other sources of data e.g. Google Maps.

## 1.13 Assumptions

The following assumptions where made in the study:

- The data used within this study sourced from the data supplier's database is accurate.
- Segmentation based on consumer behaviour for marketing purposes is not enough to form a comprehensive picture because consumers can fit into a certain segment but still behave totally different to their assumed peers within the same segment.
- The study indicates that consumer-buying behavior to purchase a product or service follows the five-stage model. The study assumes that with some modification, the five-stage model can also be applied to a consumer in terms of consumer behaviour who wishes to sell a product or service.

## 1.14 Summary

This chapter has presented the perspective of the study that aims to find out if a predictive model can be developed to determine property sales within real estate with the aim of driving business strategy. The background of the study, research context, detailing the problem statement, the research hypotheses and justification for the research were explained. Subsequently, the research methodology was defined, followed by delimitations of scope and assumptions made in the study. The next chapter reviews various sections of literature and unpacks the concepts and complexities that drive consumer behaviour.

# CHAPTER TWO: The Review of Literature

## 2.1 Introduction

Homeowners are required to make purchasing decisions that maximize their wealth and utility whilst attempting to balance the price of the house and their income constraints. With tastes and preferences being equal, the resultant consumer actions are used to imply these preferences. Understanding the psychology behind consumer behavior will benefit the real estate industry from a marketing perspective in terms of real estate analysts better predicting the behavior of purchaser decisions.

This study presents a review of the consumer behavior literature relevant to real estate and suggests how these concepts could expand real estate study. With a broader understanding of the behavior of consumers within the real estate market would help uncover some of the most common consumer behavior concepts and theories that specifically relate to what drives a consumer choice process in the case of strategic house selling.

In view of the existing literature exploring consumer decision making, the purpose of this research is twofold: (a) to develop a model of strategic decision making regarding prefabricated house selling; and (b) to explore this model and how it will provide business intelligence to the real estate industry. According to Soloman (2013), most studies today focus on consumer purchasing decisions with little attention being given to what drives a customer to sell a product. Soloman (2013) went on to state that consumers share similar characteristics when buying or selling products or services that drives their decisions. From a real estate marketing perspective, understanding consumer behaviour is good business that will drive intelligence into the business strategy. This understanding would help improve this real estate study.

## 2.2 Consumer behaviour

### 2.2.1 Definition of consumer behaviour

From a cognitive perspective, consumer behaviour can be defined as the activities that people engage in when selecting, purchasing and using products and services to satisfy needs and desires. Such activities involve mental and emotional processes, in addition to physical actions (Kardes et al., 2013).

According to Krishna (2015), research has indicated that there is diversity within a consumer’s knowledge in terms of prices of goods and services. A consumer’s purchasing behaviour or buying habits is subjected to not only the current price of the product, but also the expected price of the goods or services in the future. It is a physiological process related to the emotions of the consumer starting with recognizing a need. This need can change with the slightest change in the market, atmosphere and trends.

The neoclassical economic model of consumer behaviour hypothesizes that a consumer's choice and purchase behaviour can be termed as utility maximization, subject to a budget constraint. This refers to: 1. Consumers’ trying to get value for money; 2. Consumer face budget constraints; 3. Consumers have clear preferences for goods and services; 4. Consumers must choose amongst alternatives with their limited income (Keita, 2012).

However, according to Sent (2004), consumers in general adopt a ‘satisficing strategy’, as opposed to maximizing of utility function. Due to the challenge of constructing models based on ‘satisficing strategies’, the aim will be to model economic decision-making in the form of a mathematical logic. For simplicity, consumer purchasing behaviour are activities undertaken when obtaining, consuming and disposing of goods and services (Kotler and Armstrong, 2014).

### 2.2.2 Consumer choice and decision making

Consumer behavior has much to do with similarities between the utility theory in economics and attitude theory in social psychological behavior (Betsch and Haberstroh, 2012). According to Soloman (2013), the field of consumer behaviour is broad: it is the study of the method involved when consumers select, purchase, use or dispose of products, services or experiences to satisfy needs or desires. Betsch and Haberstroh (2012) stated that consumers purchase goods and services to derive benefits for their use. While the study of economics focuses on outcomes, consumer behavior emphasizes the purchasing process.

According to Nwanko et al. (2014), a consumer's involvement in the purchasing decision depends on the type of product and its relationship to the consumer, which dictates the type of information that the consumer needs to process. This process can be broken into low involvement purchases, which tend to be made by habitual decisions and high involvement purchases, which tends to be lengthy or more involved decisions (Nwanko et al., 2014).

Real estate purchases decisions are considered to be high involvement goods that would require complex decision-making. The decision process outlines the mental information process of the consumer's perceived need for the product in which they search for information; evaluate alternatives, purchase and finally the post purchase evaluation.

Other studies have indicated that a consumer's home buying decision can be referred to as several consumption activities such as choosing, buying and using housing products (Opoku and Abdul-Muhmin, 2010). According to Hoyer and MacInnis (2010), marketers today recognize consumer behaviour and psychology traits as being their primary focus in terms of consumption of goods, services, time and idea. However, Kotler et al (2012) argued that it is not easy to figure out the consumer's mind and there are reasons that cannot explain why consumers purchase some goods or services.

Kotler et al (2012) indicated a variety of factors that influence consumer behavior which include: culture; social class, reference groups, family, roles and status, age and life-cycle stage, occupation, economic situation, lifestyle, personality and self-concept, motivation, perception, learning, beliefs and attitudes.

An important factor to be considered that is missing from Kotler's list is the price of the product. A shopper appreciates that regardless of what they look to purchase, there is a price associated to it therefore price is not something they will obtain value from (Jones et al., 2012).

The value that a consumer will aim for is frugality in spending, which applies different meanings to different shoppers. For example, wealthier shoppers select a higher quality cut of meat; enthusiastic shoppers buy in bulk to save more; low income shoppers choose smaller package sizes and their options include shopping on sales, couponing as well as choosing non-purchase options like borrowing or create homemade solutions (Jones et al., 2012).

*Brand Loyalty* is the continuous purchasing manner of a consumer that helps a company determine customer preference, purchase intention and maintain profitability (Pappu and Quester, 2016). Attributes that a loyal consumer possesses include positive attitudes towards the brand, inattentiveness towards the price of the preferred brand, unpleasant feelings and recommendations to others about the brand. These elements lead to an improvement in the equity of the brand.

Linked to brand loyalty is *Customer Satisfaction*, which is a person's feelings of pleasure and disappointment that results from comparing the products, perceived performance to expectations. This highlights that if a customer is satisfied with the brand functions, they display loyalty to the brand and become insensitive to the price as they will buy the item at any cost. Monitoring and measuring customer satisfaction is the key to customer retention (Pappu and Quester, 2016).

Finally, *Perceived Quality* is another mechanism to establish consumer behavior. Quality can be defined as the consumer judgment of superiority or excellence of a product or service and is a determinant in the price that is paid for the product or service (Malik, 2012).

### 2.2.3 Ethical analysis of targeting consumers' needs and wants

According to Kotler and Keller (2012), behavioral targeting is providing companies with the ability to target consumers and to find the best match between advertisements and prospects. However, with the power that marketers hold in terms of available information, is the ethical dilemma in terms of misrepresentation, violation and general 'creepiness'. Hence ethics is receiving great attention and government considers if industry self-regulation will be sufficient or will stricter measures be required (Kotler and Keller, 2012). According to Lantos (2015), ethics is the study of morality, which means right, good and proper decisions. In addition, ethical issues arise when there is potential harm or benefits caused to a group of individuals.

### 2.2.4 Five stage decision making process

The consumer decision-making process is a series of events that leads to choices between alternatives. In terms of understanding consumer behaviour, business today needs to understand the decision-making process, which encompass consumer experience in studying and selecting products or services (Kotler and Keller, 2012).

According to Hawkins et al (2013) in Figure 2.1, consumers undergo a five-stage decision process when purchasing a product that includes: problem recognition, searching for information, evaluation of alternatives, purchase decisions and post purchase behaviour. Consumers may avoid or repeat certain stages (Quester et al, 2011).

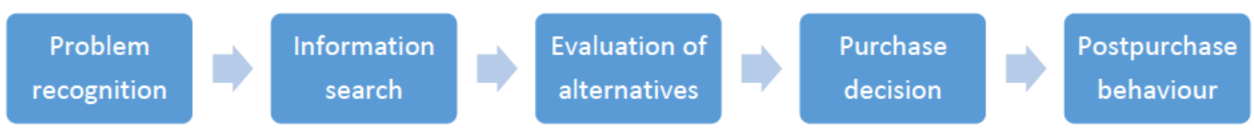



Figure 2.1 Five-stage purchase decision-making process

Source: Adapted from Hawkins, D.I., Mothersbaugh, D.L. and Best, R.J., 2013. *Consumer Behavior: Building Marketing Strategy*. 12$^{th}$ ed. New York: Irwin McGraw-Hill.

Although the 5-stage model applies to a decision-making process, the same steps, indicated in Figure 2.2, could apply to a property owner's decision-making process to sell a product, which in this case relates to their real estate property.

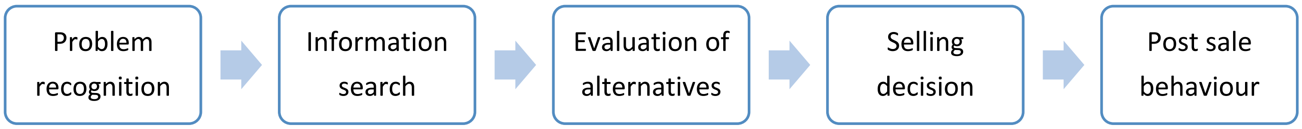



Figure 2.2 Five-stage selling decision-making process

Source: Adapted from Hawkins, D.I., Mothersbaugh, D.L. and Best, R.J., 2013. *Consumer Behavior: Building Marketing Strategy*. 12th ed. New York: Irwin McGraw-Hill. – **Re-applied by author**

### 2.2.4.1 Problem recognition process

According to Kotler and Keller (2012), a consumer's buying process is initiated by either external (e.g. colleague's car) or internal stimuli (e.g. thirst, hunger etc.). It is important for business to understand what triggers these circumstances by seeking information about consumers or initiating consumer interest into products. Bruner (1988) indicated that the problem recognition involves the collaboration between an individual desire and actual state of mind. The former refers to how an individual would like their needs to be met versus the degree in which the perceived need is being met.

According to Kassel (2016), the psychology behind selling a property may include

- Current home is too small – First time homebuyers may outgrow their 'starter' home due to increased family size.
- Up-scaling the home – Based on the financial improvements, people would often want a more expensive and grander upscale home.
- Job transfer – Relocation makes it necessary for people to up roots and move e.g. prefer not to spend too long in traffic.

- Personal relationships – moving in with a partner or change of marital status (married or divorce) can mean parties may need to sell due to affordability or bad memories.
- Empty nest – The children may have grown up and moved out and owners may want to downsize. The older you get, people may want to down size as they move closer to retirement.
- Lifestyle change – People may get tired of owning a home and prefer to travel. For example, as they move past a certain age they find a calling that is meaningful to them therefore owning a home becomes less of a priority and lean more to realizing a dream.
- Death in the family – According to Weintraub (2017), a death in the family often results in next of kin selling the property or transfers the title of the property into a trust to avoid property proceedings.

According to Quester et al (2011), consumer's may not necessarily move to the next stage of the purchasing decision as factors such as time, budget and a qualified importance becomes critical to the consumer's desired lifestyle. For example, property purchasing which is dependent on a property being sold is reliant on personal assets and budgetary requirements. If the consumer cannot afford the deposit or monthly repayments on the new property then they will not proceed to the information stage.

#### 2.2.4.2 Information search process

Post the problem recognition process, consumers attempt to collect pertinent information from their long-term memory to determine if a solution can be established by comparing the characteristics of each solution (Schiffman et al., 2012). Past experience, time since last purchase and satisfaction will contribute to a consumer's reliance on internal information. However, due to consumers rarely purchasing real estate, they in most cases have limited knowledge and turn to external information when selecting a new property in terms of location, price, etc. Marketing literature establishes that consumer information search behaviors differ markedly with product involvement (Schiffman et al., 2012).

According to Duncan and O' Sullivan (2012), research into the functioning of housing markets generally focuses on market outcomes which include house price changes and household movement patterns. This de-emphasizes the importance of imperfect information and means that partly informed households typically engage in search activity prior purchasing or selling a property.

According to Verma (2012), an individual would be exposed to many strategies when making a decision, which depends on the amount of information search and extensiveness of alternative evaluations. For example, a first-time homebuyer may require extensive information search due to the lack of previous information, brand knowledge and the high risk associated with the wrong choice. Information sources could be the Internet, real estate agents, friends and local newspapers.

#### 2.2.4.3 Evaluation of alternatives

Information search provides consumers with an information hub for making assessments. The following discussion describes mechanisms by which consumers use information in evaluating alternatives. According to Daugherty and Hoffman (2014), a consumer's evaluation criterion of products and services includes characteristics or attributes of either what they can benefit from or the price they are willing to sell it for. Consumers adopt evaluation criteria in the selling decision such as price and after sales service.

Buying behavior in terms of attitude and beliefs is strongly influenced by a consumer's experience, which determines how to handle the product. This perception and opinions gained from the alternative provides input in the most favorable option that satisfies the need for solving the problem (Kotler, 2012).

Certain consumer choices will hold more weight than other factors and will have a bigger influence on the choice made (Farzana, 2012). To this point, for a consumer who is concerned with the property price in relation to their affordability will use this criteria as the primary decision factor. In other words, if the consumer finds a property with a single garage that's within their affordability bracket, will choose this over a property with a double garage but at a higher price.

#### 2.2.4.4 Selling decision

The evaluation stage has indicated that consumers will form preferences from a set of product choices. As well as forming an intention to buy the preferred product (Kotler and Keller, 2012). Choosing to buy a product involves adapting different criteria that best fits their needs during the purchasing decision (Blackwell et al, 2012). The same thought process applies to a selling decision.

To further add, consumers from different market segments might make buying decisions based on their perception of attributes they consider important, rate the alternative on each attribute and make the choice based on the highest positive score (Kotler and Armstrong, 2012). Again, a consumer selling a product will exhibit the same characteristics. Relating this mindset to real estate, a consumer may choose an apartment that has a convenient location over an apartment that has more features like a pool and double garage, even though both apartments may be the same price.

#### 2.2.4.5 Post sale behavior

According to Paul and Hogan (2015), post sale behavior is the final stage in the consumer decision process when the customer established if there is satisfaction or dissatisfaction with the decision. The feelings and emotions that the consumer experiences at this stage will significantly influence whether they will sell their product again (the product in discussion relates to the sale of a property). Consumer experiences will also influence if they will use the same service provider, being the real estate agent.

Oliver (2015) states that consumers more often than not establish a loyalty to service providers than to physical products due to the challenge of evaluating alternatives before actually experiencing the service. The advantage of brand loyalty is that it can bring additional benefits discounts, customized service and insights into customer preferences.

Cognitive dissonance is another form of seller’s remorse at this stage of the process relating to psychological tension or anxiety. Cognitive dissonance refers to conflicts in attitudes, beliefs or behaviors that produces a feeling of discomfort. The feeling of discomfort may lead to an alteration of one's attitude or beliefs to reduce the discomfort and restore balance. In other words, Festinger (1957) suggest in the cognitive dissonance theory that humans have an inner drive to keep attitudes and beliefs in harmony and avoid disharmony.

A consumer may be compelled to question whether they have made the right decision at this stage. If the consumer is dissatisfied at this stage of the decision-making process, then they will be back at the problem recognition phase. Alternatively, a satisfied consumer will skip the next decision step and move straight to the decision based on their loyalty.

### 2.2.5 Influences on consumer behavior

According to Kotler and Keller (2012), consumer's response is influenced by motivation, perception, learning and memory. Consumer characteristics are influenced by cultural, social and personal factors that influence consumer-selling behavior. Quester (2011) indicated that a consumer's decision to purchase or sell could be split into individual differences and environmental influences.

#### 2.2.5.1 Individual Influence

According to (Schiffman et al, 2012), individual influences include consumer personality, demographics, lifestyle, motivation, knowledge and beliefs and occupation and income.

a. Personality from a consumer perspective can be defined as an individual's unique consistent response to environmental provocations. It is traits and behavior of individuals that differentiates an individual from another individual (Kotler and Keller, 2012). A consumer's personality may to some extent affect their choice in real estate investments. A risk taker will buy a property that is off-plan that will take a few years to develop with an intention to sell it. This is due to the anticipation of reaping the rewards of a good return on the investment.
b. Demographics may be defined as the subdivision of markets into unique homogeneous subgroups of customers, where any subgroup can conceivably be selected as a target

market to be met with a distinct marketing mix (Nandamuri, 2012). Different customer attitudes can be grouped into segments with a different approach applied to each segment. Demographics include age, gender, education, income, occupation and family structure. Demographics are critical for real estate marketing as age; income and family structure are important parameters when deciding on selling a property.

c. Lifestyle refers to the distinctive way that a consumer lives, including how the individual expends time and money in social activities as well as expressing interest and opinions (Schiffman et al, 2012). Consumers will purchase properties according to their lifestyle habits. An elderly couple will settle for a home that is single story to save them from being exposed to a staircase. A single person who values social interaction will prefer an apartment that has a swimming pool, gym and entertainment center.

d. Motivation is the force that drives a person to buy a good or service and begins when a need is aroused to satisfy their physiological and psychological needs. The area between the desired and actual state of mind is a condition called a drive (Hawkins et al, 2011). When the drive within the consumer is activated then the consumer will take action to satisfy the need. A real estate buyer may be trying to satisfy an actual need for space and a desired need for status.

e. Knowledge is the information that a consumer depends on to evaluate and select products and services (Schiffman et al., 2012). Consumers that have a recognized need will more than likely look out for information that can satisfy that need. To this point, real estate agents will be challenged to market to consumers unless they have a desired need to purchase a property. Property managers will then resort to using family brand names for all properties, as this will stimulate the consumer's memory network.

f. Economic circumstances have a considerable effect on consumer behavior. Consumers with more disposable income can afford luxurious goods and 'want' based purchases (Kotler and Keller, 2012). It is important to note that purchasing and selling habits are influenced by various factors therefore income alone does not determine the level of consumption. Possessions can be interpreted as a representation of status, with wealth reflected in the items purchased. Therefore, status affects the choice of where to live, the type of home and neighborhood. Consumers that want to move into a higher social status might choose a house or neighborhood that higher-class individuals would choose.

### 2.2.6 Environmental influences

According to Hoyer and MacInnis (2010), environmental factors include **culture, reference groups, social class and family.** Environmental information is primarily focused on personal experiences of buyers and then reflected in the form of marketing and non-marketing sources of information. External factors mean consumers are motivated to buy products by elements outside of individual which including socio-cultural influences and effective marketing strategies (Moslehpour et al, 2014).

a. Culture includes social beliefs, values, attitudes, behavior patterns shared by members of society and transmitted from one generation to the next through socialization (Lantos, 2015). Culture affects all five stages of the selling decision process i.e. need recognition, information search, search for alternatives, selling decision and post-sale decision. Research indicates that cultures would influence people to make a decision in addition to motivating them to think in a certain way. Most importantly, the culture differences influence consumer behavior in terms of cultural self-definition (Moslehpour et al, 2014).

   Cultural norms that affect the real estate demand include having a formal dining area so that everyone partakes in a meal together. Families that have their aging parents to care for may want a granny cottage or an extension relative to the existing building to house the elderly. This need for individualism warrants the need for customised properties.

b. Reference groups can be a person or a group of people who are real or imaginary that influences attitudes, beliefs and values of other people. Any influencer can be referred as a reference group – for example family, colleagues or social media (Ling et al. 2015).
   Informational influences involves the use of influential people that assess the merits and legitimize the purchase or selling of a given product or service. Word of mouth is also important when it applies to reaffirming a consumer's decision, especially with young adults seeking referrals from older people. Real estate consumers are influenced by a unique set of reference groups. A consumer will invest in property based on their preference to be near family that will assist in taking care of the children. Couples that enjoy social entertaining will choose a property that has been designed for entertainment.

c. Social class are divisions created in a society comprising of individuals or families that share similar economic positions such as wealth, status, values, lifestyle and interests. An individual's social class can be determined by their occupation, personal performance, possessions and value orientation. The distinction between social classes is representative in the consumption patterns with a noticeable impact on the evaluation of alternatives phase (Quester et al., 2011).

   The most important decision reflecting a family's social class is the decision on where to live in terms of type of home and neighborhood location. Home design is important so that it reflects the living style based on social class i.e. furniture choice.

d. Family is regarded as one of the most influential reference group. Family tends to participate in the consumer behavior related to purchasing or selling because of their expertise, cultural values or role structure within the family. Demographic dimension of a population relates to the family and household structure such as marital status, employment situation, family size and age bracket (Ling et al., 2015). With regards to real estate, the decision maker is usually the person with the financial authority and power to choose how the money will be spent.

   Changes in household structure have led to housing decision making being made jointly by the husband and wife (Anderson et al, 2017). Thus the real estate agent needs to factor both spouses during the sales approach. Families progress through distinct phases called family life cycle which include young singles, young married without children, married with children, middle aged married with children, older married and older not married (Pride and Ferrell, 2011). This creates a wider scope for housing options during the life cycle. For example, young marrieds without children that work would look for apartment style living close to fast food takeout and supermarket amenities. Middle ages couples with children would seek housing that suites more luxurious living habits and services.

### 2.2.7 The relationship between Consumer Behavior and the housing market.

The housing market, which forms part of the real estate market, is an engagement between buyers and sellers who exchange real property rights for other assets (Burnside et al, 2015). Burnside et al (2015) pointed out that in the past, real estate studies have been based on supply and demand economics principals. This suggests that constrained by a given income, a consumer is expected to make real estate decisions that will maximize their utility and wealth with tastes and preferences taken as given. Burnside et al (2015) also suggests that the study of real estate would profit from a marketing and lead generation perspective, if elements of consumer behaviour concepts that include sociology and psychology were considered. These concepts would enable real estate analysts to predict consumer behavioral habits within the real estate realm.

Buying behaviour habits exhibit the following characteristics:

#### 2.2.7.1 Purchasing intention

Wu and Teng (2011) stated that purchase intention can be defined as a customer's potential in the future purchase of a product or service. According to Sidi and Sharipah (2011), purchase intention follows several meanings, which include:

- A customer's willingness to buy products or services;
- Customer's purchase intention in the future; and
- Customer's decision of repurchase.

If an individual has a high purchase intention, this will increase their performance behaviour in order to achieve their purchasing goal. In the same token, it could be said that the customer exhibits the same habits that Sidi and Sharipah (2011) stated, when it comes to selling a product. For example: 1. The willingness to sell a product or service; 2. Customers selling intention in the future and; 3. Customer's decision to repurchase. If the customer has a high intention to sell a product, this will increase their performance behavior. Kotler and Armstrong (2014) stated that purchasing a property can be characterized as a high involvement purchase which is potentially risky and costly and will therefore require thoughtful cognitive deliberations.

#### 2.2.7.2 Property attributes

Property attributes consist of features including house design, building quality, house type (double story), house finishing (air conditioning, finishes), age of the house, interior and exterior designs, which are expected to influence an individual's house purchase decision (Haddad et al, 2011).

Housing features are important factors in establishing consumer's selling choice as it affects the price, which the house is valued at. Consumers look for internal attributes when purchasing a house e.g. quality of building, interior and exterior design, which is important to a consumer when purchasing a property. As such, these factors must be taken into consideration when selling a property.

#### 2.2.7.3 Living space

Gosling and McCunn (2013) defined living space with attributes that include living room size, kitchen size and its contents, floor area, number of bathrooms as well as bedrooms. Stevenson and Prout (2013) asserted that there is a strong relationship between living space features and a consumer's house decision-making and pricing. All other factors remaining equal e.g. property location; people generally sell their properties at a higher value which contains more living space. A property with large gardens will be attractive to families with young kids or pets and will therefore command higher prices.

#### 2.2.7.4 Monetary

Housing affordability has a direct correlation to household income, housing expenditure and household size. Findings from Bramley (2012) noted that household income, wealth, price of the house, demographic aspects and interest rates are determinants in explaining the affordability of households. Buying or selling a house is one of the biggest financial investments and makes up a large portion of a consumer's wealth.

#### 2.2.7.5 Location

One of the most important factors affecting an individual's choice of a house is the location. Location is defined as the distance to the central business district, schools, work, public services and retailers (Ajayi, 2012).

According to Gelfand et al (2004), housing location is also found to affect the house price and help a buyer develop a preferred way of life by identifying the opportunities available for study, work and access to services and facilities. Developed infrastructure in the housing area contributes to the rise or fall of the house price. Specifically, if a house is close to a school, shopping mall, bank, transportation facility, hospital, restaurant, church, temple, airport or any other place that can provide convenience to the people staying in that area, the house will undoubtedly possess a high housing value which is an important factor to consider when purchasing a house. This will directly affect a seller's decision when pricing the house or the time taken to sell the property.

#### 2.2.7.6 General economic performance

The real estate industry is sensitive to economic cycles. There is a significant negative relationship between real interest rate and housing return. Interest rate hikes increases the cost of borrowing as people will have less disposable income as more will be spent on interest payments (Sunil, 2015). In addition, economic fundamentals like real income affect a consumer's purchasing power, borrowing capacity of households, interest rate, house prices and supply of properties. Consumers generally tend to save as personal discretionary income is affected, hence they will tend to not want to commit to long-term debt in the form of home loans.

In order to operate effectively in today's dynamic global real estate market, it is important to understand the complex processes and variables that underpin them. Across different areas, there is a difficulty in identifying a uniform set of attributes that can be used to determine a better understanding of the value of real estate markets as well as the salient characteristics to consumers in these markets (French, 2015). Knowing these attributes as a real estate agent is vital to create that link between property for sale and the consumer's salient characteristics. This will be the basis of creating models to help predict consumer behaviour.

#### 2.2.7.7 Deceased estates

According to Adam (2011), deceased estate generally satisfies a lot of the requirements that buyers are looking for which includes a motivated seller and location in a well-established suburb. Deceased estates have a general perception of an old property and not modern townhouses. They therefore represent good buying opportunity for a renovator or a low-cost property in a good area. Adam (2011) stated that whoever is executing the will, is generally motivated to sell the property.

#### 2.2.7.8 Living Standard Measurements (LSM)

Living Standard measurement (LSM) is derived from a group of variables that segments consumers or people according to the living standards into homogenous groups by combining demographic characteristics. Some criteria used in the calculation include as possession of major assets, access to services, access to basic necessities e.g. fridge (Haupt, 2001). According to SAARF (2004), LSM is defined as the scale that indicates the socio-economic status of the individual. LSM is an important criterion for estate agents in terms of targeting and understanding their consumer base with regards to segmentation modeling.

#### 2.2.7.9 Credit rating

Credit information contained within credit reports issued by Credit Bureaus contributes towards the calculation of a consumer's credit score. Credit reports contain information relating to financial history, account information, payment history, age of account, judgments and defaults which generate credit scores serves as a guide when a consumer wants to apply for credit e.g. buying a home or car (Transaction Capital Credit Health, 2013). Generally a poor or unfavorable score dissuades lending institutions from affording debt or if credit is provided, it is usually issued at a much higher interest rate (Capitec, 2017).

## 2.3 Integration of consumer behavior theories and the effects on real estate

Real estate studies in general makes assumptions that a consumer's real estate transaction is an attempt to maximize their utility (Koklic and Vida, 2009). Real estate business today needs to include the human element of decision making i.e. tastes and preferences by integrating the elements of consumer behavior with the financial economic approach. This will lead to a better understanding of decision maker actions within the real estate realm.

### 2.3.1 Decision making process

Harris (1988) defined a decision as an ongoing process, which evaluates alternatives, related to a goal or objective, which impels the decision makers to choose a particular course of action. Decision making which is the heart of the management process of thought and deliberation that leads to a decision (Mintzberg, 1979). Decision-making is an integral part of the management and organization to achieve predetermined goals, which highlights competency and differentiates a manager from a non-manager (Dean and Sharfman, 1996).

Management understanding of emotions has a powerful effect on decisions. Dean and Sharfman (1996) argue that emotions increase the creative problem solving that facilitates the integration of information. Organizations that place a focus attention on consumer behavior will help better position their products and services, for example:

- Marketing strategy: manage market segmentation, new product identification, consumerism, marketing mix, etc.
- Product development: Consumer behavior helps organizations decide what products or services to offer by identifying a need that has not been satisfied.
- Customer service: By understanding how customers behave in relation to products being sold will lead to a better service being offered which will increase the chance of repeat customers and referrals.

### 2.3.2 Predictive models

According to Wyman et al (2011), economics is a description of the economic behavioral science of people under various conditions. According to Kalechosky (2016), predictive models use data to make decisions and take actions using models that are empirically derived and statistically valid. Predictive models use statistical procedures to predict future behaviour. There is a reliance on predictive models on how both buyers and sellers will behave in respect to a property that is available for sale.

The model that will form part of this study will be the ‘Scorecard to sell a property’. This is defined as the likelihood that a property will sell. Attributes that will feed into this model include the investment potential of the consumer, average lifespan of the property over time, demographics and past property acquisition of the current owner of the property.

The scorecard within this study is based on a current learning module such as logistical regression, which factors in a wide range of variables. The system then generates an outcome value, which is then translated into a scored result set. According to Fletcher (2011), data scoring will use analysis and data modeling outputs to drive business and create scenarios that can be tested, hence creating new products and new ways of thinking. Trends, dips and highs in data, migration patterns and hotspots displayed in heat maps are vital areas of distinction that will set predefined milestones which will instantly increase efficiency.

According to Kalechosky (2016), scorecards can be used to identify market segments for targeted marketing. They can also be used to create market segments specific to business needs and requirements. The scorecards provide a base for improved business intelligence that will drive business strategy. An integral part of building scorecards is the data pre-processing, which is used to identify incomplete, noise and inconsistent data.

#### 2.3.2.1 Implications of scorecard models within real estate include:

- Real estate intelligence that is derived from valuation models include: property evaluation, property financing limits, investment analysis, determining the property asking price and listing real estate for sale.
- According to Sunil (2015), scorecards help improve product design and marketing messages tailored for targeted marketing by adopting the right message to the right individual. This can be achieved by identifying patterns that leads to purchase/ selling behavior that will improve personas, segmentations and relevant offers.
- Real estate businesses will not only acquire new customers but also to optimize customer experience and longstanding relationships.
- Delivering relevant products that is specific to the consumer's needs thereby reducing unnecessary investment within the sale process (Krishna, 2015).

## 2.4 Consumer Behaviour Models

### 2.4.1 Overview consumer behaviour models

Behavioral models enhance our understanding and ability to explain the various types of components of consumer behaviour. Goodhope (2013) remarked that a model characterizes a hypothetical construction of theory, which forms the basis for predicting, manipulating and controlling the outcome of a specific problem, in this case the selling process.

McNeal (1973) described a model as an implied presentation of real phenomena. Williams (1982) holds the view that modeling occurs when a participant matches responses to cues provided by a model that may or may not influence consumer behaviour.

The problem posed in this research was to build a model that will be used to predict consumer-selling behavior. The parameters that should be factored for these models include economical, psychological and sociological that influence decisions in terms of selling their real estate property.

### 2.4.2 Classifications of consumer models

According to Goodhope (2013), consumer-buying models can be classified into the following:

- Quantitative or verbal: Quantitative can be viewed as numerical and symbols that can be transformed into function, equations and formulas. Verbal models are used to represent some object or situation that employ language as a means of expression.
- Physical or behavioral: Physical models look like the finished object they represent. Behavioral models focus on performance duplication rather than physical duplication.
- Prescriptive or analytical: Descriptive models are used to describe something mathematically that include mean, median, mode, range and standard deviation. Analytical on the other hand try to compare, challenge and interpret concepts.
- Decision process model: Decision models include a series of steps that a buyer potentially follows. Marketing opportunities can be tailored to fit each step of the decision process once the buyer's behaviour is analyzed.
- Theoretical model: This describes the influences that affect purchase decisions

### 2.4.3 Customer behaviour theoretical models

The review of decision models undertaken reflects the complexity of consumer choices that lead to a specific behaviour. There is a diverse range of variable and attributes that has been posited across models with their own justification in attempt to explain consumer behaviour. The major models for consumer decisions were proposed by Maslow's (1954) hierarchy needs model, Marshallian economic model (Marshall, 1890), Pavlov's Model (Kotler, 1965), Freud's Model (Kotler, 1965), Kotler's Behavioral Choice Model (Kotler, 1965) and Howard Seth Model (1960).

#### 2.4.3.1 Hierarchy of needs model

Maslow (1954) posited a hierarchy of human needs based on the theory that human beings are motivated by unsatisfied needs and certain lower needs are required to be satisfied before higher needs can be addressed. Human needs can be categorized into two groups: deficiency and growth needs. Deficiency needs at lower levels must be met before moving to the next higher-level need. The first four levels are:

1. Physiological: Hunger, thirst, and requirements for human survival.
2. Safety/ Security: personal security. Financial security, health and wellbeing.
3. Belongingness and love: Friendships, intimacy, family, being accepted
4. Esteem: to be competent, gain approval and recognition

Maslow and Lowery (1988) later recognized the growth need of self-actualization, which he referred to as 'What a man be, he must be'. This was described as the desire to accomplish everything that one can, to become the most one can.

Self-transcendence was another dimension of need that Maslow explored that was defined as connecting to something beyond ego or to help others find self-fulfillment.

Maslow's basic position is that as one becomes more self-actualized, they develop wisdom and position themselves to handle a variety of situations. According to Maslow and Lowery (1988), Self-actualized people are characterized by:

1. Being problem focused;
2. Focused on an appreciation of life;
3. Anxious about personal growth and;
4. The ability to have peak experiences.

Maslow's conclusion is that the hierarchy can be used to describe the various kinds of information that a consumer seeks. For instance, at the lowest levels, consumers seeking coping information to satisfy their basic needs. Consumers that are seeing how they can be safe and secure require helping information. Belongingness needs, especially around relationship development is known as enlightening information. Empowering information is sought by individuals at the esteem level i.e. how their egos can be developed. Finally, people in the growth levels of cognitive, aesthetic, and self-actualization seek edifying information.

#### 2.4.3.2 Micro models

##### a. The Marshallian economic model

The Marshallian economic model, according to Marshall (1890), states that buyers will spend income on goods and services that will provide them with the greatest fulfillment, depending on their tastes and price of the goods. In other words, a consumer would purchase a product within a price, income or wealth situation that will perfectly solve their utility maximization problem, whilst promoting their increased happiness. Mostert (2002) put in plain words that a consumer would spend their income on products that offer the greatest satisfaction that is conditional on their style and cost. In other words, the choice of goods that offers the greatest satisfaction versus the expenditure. However, economic factors in isolation cannot account for all variations in the purchasing process as it excludes brand and product preferences. It does however contribute to a small portion of a consumer's behaviour. A viewpoint of this model is that economic factors should be included in the comprehensive description of the buyer behavior since economic factors has a significant effect in all markets.

##### b. The Pavlov Model

The classical conditioning theory Pavlov's model plays a key role in marketing as buyers are conditioned to form impressions of various brands through associative learning process (Belch, 2012). Pavlov's model focused on four main concepts that include: impulse, suggestion, reaction and relapse. According to Kotler (1965), these concepts influence the marketing stimuli e.g. advertising, exclusive offers etc.

I. Impulses are a result of strong stimuli (needs and motives), which requires an action to take place by an individual. These actions may include primary (hunger, thirst) or acquired social relationships (fear, shame)
II. A suggestion, which is similar but weaker than impulses, is based on stimuli's and is also a determinant in the way an individual reacts.

III. Reaction is a response to suggestion and the repetition of the reaction is influenced by an individual's experience.

IV. Relapse is associated with strengthening of certain reactions on condition that the experience was according to the individual's expectation.

The Pavlovian model is a useful tool in aiding new product launches or creating promotional strategies.

**c. The Freud Model**

The Freud Model is based on psychoanalytic theory on the human being, addressing consumer behaviour being made in terms of biological and cultural elements. Kotler (1965) summarizes the model by stating that that Freud assumed that psychological forces that influences a consumer behaviour is unconscious which results in people not fully understanding their own motivations. This is in direct relation to a consumer's attitude, which may be positive, neutral or negative and is governed by the attitude strength.

**d. Kotler's Behavioral Choice Model**

According Kotler (1965), his model recognizes a consumer's product choice includes the following inputs: communication about available brands, its prices, quality, product availability, service, style, option and images. The channels in which this information reaches the consumer are advertising media, sales people, buyer's family and an individual's observation. The sales process must be aware that consumers and organizations have unique personalities and will vary from firm to firm.

**e. Howard Seth Model**

Sheth (1960) stated in his model that the first input into consumer decision-making consists of stimuli or information emanating from the environment. These inputs are made up of quality, prices, distinctiveness and availability of brand and related services. The second input includes Symbolic Stimuli, which is the information input from marketing and non-marketing sources, which also affects the problem recognition stage of the decision-making process. The third input is the information that the customer is exposed to, attention, perception, acceptance and retention. The fourth input looks at outcomes of the post choice consequences as a variable input.

## 2.5 Cluster Analysis: Statistical Technique

The aim of clustering is the organization of objects into groups, which share similarities among them. As depicted in Figure 2.3, this classification technique focuses on cases, data or objects (event or people) that are subdivided into groups (clusters) that are similar but not identical to one another and different from items in other clusters. In other words, this technique reveals associations, patterns, and relations within large volumes of data.  Clustering is defined as a collection of objects, which share similar characteristics between them and are dissimilar to objects that belong to other clusters (Balasko et al, 2006). Clustering is considered an unsupervised method, as it does not use previous class identifiers to identify the underlying structure in the collection of the data.

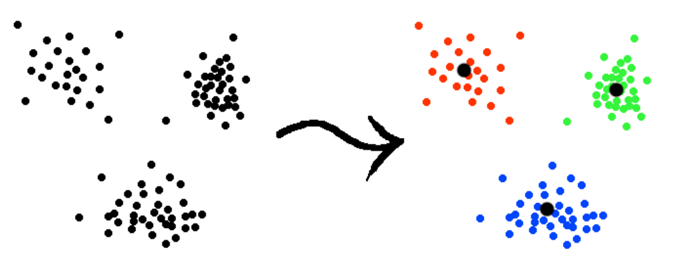


Figure 2.3. Examples of clustering

Authors Own source

Figure 2.3 represents three clusters, which contains data that can be easily identified and divided. According to (Estivill-Castro and Houle, 2001), the similarity criterion (in this case geometrical distance) indicates two or more objects belong to the same cluster if they are "close" according to a given distance. This method is known as distance-based clustering.

Estivill-Castro and Houle (2001) went on to state that conceptual clustering, which is another method of clustering that uses two or more objects can belong to the same cluster if one defines a common concept to all of the objects. In other words, objects are grouped per their fit to descriptive concepts, not according to simple similarity measures.

### 2.5.1 Key clustering concepts

- Supervised learning models: Supervised learning is based on training a data sample from data source with correct classification already assigned (Kalhori and Zeng, 2014). A supervised learning algorithm analyses the training data sample and produces an inferred function that can be used for mapping new examples.
- Unsupervised learning models: The aim of unsupervised learning or clustering is to discover groups of similar instances within the 'Unlabeled data'. In this approach, we have no information about the class label of data or how many classes exist (Kalhori, and Zeng, 2014).
- Constrained clustering: This is a form of semi-supervised clustering technique that allows the user to impose certain constraints. For example, the number of clusters. Constrained clustering algorithms will reject if the specified constraints are not met (Thi-Bich-Hanh et al, 2017).
- Heuristic clustering models: is a clustering approach that employs a real-world method that may not be guaranteed to be optimum or perfect, however will suffice to satisfy the immediate goals (Satchidananda et al, 2012).
- Hierarchical clustering methods: According to Ying and Karypis (2005), this method imposes a hierarchical structure on the data objects and their step-wise clusters. This can be in the form of a bottom up approach whereby each observation starts in its own cluster

and pairs of clusters are merged as you move up the hierarchy. Alternatively, it could form as a top down approach as all observations start in one cluster and splits occur as there is movement down the hierarchy.

- Partitioning clustering methods: According to Vijayarani and Jothi (2014), this method decomposes the data object sets into clusters where every pair of object clustered is either distinct (hard clustering) or has some members in common (soft/fuzzy clustering). It classifies the data by satisfying the following requirements: 1. Each group contains at least one point. 2. Each point belongs to at least one group.

### 2.5.2 Common clustering algorithms

- **K-Mean**: According to Gilliam and Voss (2013), K-mean clustering method is one of the simplest unsupervised learning algorithms that solves the clustering problem, in particular data mining. K-means clustering objective is to partition *n* observations into *k* clusters in which each observation belongs to the nearest mean.

  However, the output derived from this hard partitioning method is not always accurate in addition to the algorithm containing numerical issues. According to Trevino (2016), Figure 2.4 summarizes K-means by finding the best centroids by alternating between 1. Assigning data points to clusters based on current centroids and 2. Choosing centroids based on the current assignment of data points to clusters.

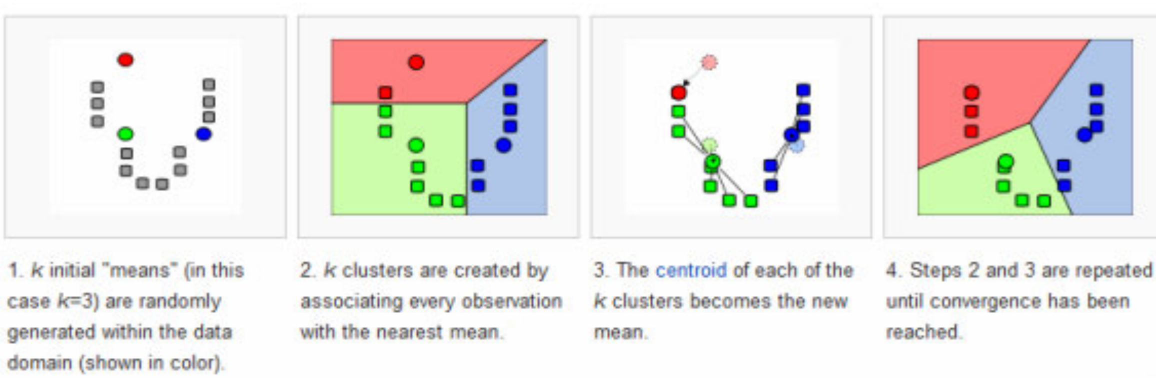



Figure 2.4 K-Mean Clustering

Source: Adapted from Trevino, A., 2016. *Introduction to K-means Clustering.* [Online] Available at: <www.datascience.com/blog/introduction-to-k-means-clustering-algorithm-learn-data-science-tutorials> [Accessed: 22 June 2017].

- **Bottom-up hierarchical algorithm:** According to Zhang et al (1996), this method begins the clustering by placing each object in different clusters and then merges them into more and more larger clusters until all of the objects are in the same cluster, based on the similarities shared among the clusters.
  This method builds the hierarchy from the individual data origins by increasingly merging clusters. According to Zhang et al (1996), Figure 2.5 below has six elements: [a] [b] [c] [d] [e] and [f]. Step one establishes which elements need to merge within the cluster (generally the two closest elements are chosen based on distance). As the clustering advances, rows and columns as the clusters merged and the distances updated i.e. at each step, the two most similar clusters are merged. The process stops when there is a single cluster of all examples.

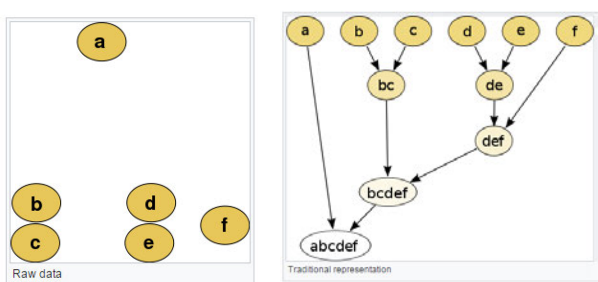



Figure 2.5 Bottom Up Hierarchical Algorithm

Source: Adapted from Zhang, T., Ramakrishnan, R. and Birch, M.L., 1996. An efficient data clustering method for very large databases. *SIGMOD*, 6(96), pp. 103–114. [Online] Available at: <https://www.cs.sfu.ca/CourseCentral/459/han/papers/zhang96.pdf >[Accessed: 22 June 2017].

- **DBScan algorithm**: According to Ester et al (1996), the density based clustering algorithm segments data based on particle density of some threshold. Clusters are formed by identifying particles that are most dense in a given plane. This is a common dimension reduction technique that is often employed on spatial /geo-spatial data. For example, this technique can be applied to define municipal boundaries based on the density of housing settlements in geographic landscape.
- **Regression clustering**: This algorithm offers clusters within multiple regression locations in which there is a dependent variable (y) and one or more independent variables (X's). According to Armstrong (2012), algorithms partitions the data into more than one clusters and performs multiple regression on the data within each cluster with the aim of estimating the relationships among variables. Regression analysis helps understand how the dependent variable changes when the independent variance is varied whilst other independent variables are held fixed.

In order to create a predictive model, there are many statistical and data mining algorithms that can be used. According to Harrell (2015), the general rule of thumb in deciding which model to use that address the business objective include:

- Clustering models are best suited for customer segmenatation.
- Classification algorithms are used for customer retention
- Regression algorithms are best applied to credit scoring or predicting the next outcome of time driven events.

## 2.6 The role of models in managerial decision making in marketing of products and services

Decision making is a systematic process that emanates out of an effort to derive a profitable outcome. As explained, purchase decisions apply across the board to individuals, households and organisations and buying behaviour are affected by the various inputs viz availability of funds, price, quality of products etc.

Models of consumer behaviour aids management decisions around marketing practice of planning their advertising and promotional appeal to ensure patronage and retention.

According to Kotler and Armstrong (2014), consumer behavior models aids marketing practice in the in the form of:

- Marketing planning and setting goals
- Aids with enhancing the sustenance of the firm's objectives and achievement of desired customer satisfaction.
- Evaluation of product and service offering to ensure conformity organizational goals.

The overall objective is to enable decision makers to take decisions that will render consumer satisfaction and company profit and growth.

With reference to the Howard Sheth Model, it demonstrates that consumers respond to stimuli in the environment during the course of the purchase decision. Consequently organisations make decisions around pre-determined goals and other external and internal factors with the aim of being committed to consumer need's satisfaction.

However, the Nicosa model exposes the relationship between the firm and the consumer in relation to the marketing practice which is to elicit purchase response from consumers. Marketing practitioners depend on the Nicosa model to evaluate their promotional efforts, sales campaigns etc. in order to act on prompt feedback.

Kotler Behavioural Choice model encourages realistic marketing decision making because the inputs are mostly features of a product and influences that drive buying behaviour.

These models help management make decisions relating to obtaining the best product quality, advertising reach, more streamlined distribution channels and the most appropriate pricing strategy that relates to the overall organisational goals. Hence, incorporating preferences, feelings and motives will guarantee overall corporate success.

### 2.7 Conclusion

Chapter2 is a review of theories, concepts and models that speaks to consumer behaviour. According to Soloman (2013), consumer behavior involves the psychological process that consumers undergo through their needs recognition, find means in which to satisfy those needs, understand surrounding information to make plans and finally to decide on how to implement those plans. This study has revealed basic concepts and theories on consumer behaviour from available literature with the overall goal to achieve a better understanding of consumer behaviour. The following chapters present the statistical techniques used to derive the propensity to sell model and describes rigorous tests required to meet the specific assumptions.

# CHAPTER THREE: Research Methodology

## 3.1 Introduction

This chapter focus was the nature of the research that provides a description of the methods adopted to carry out the study in addition to the research design. The contents of this chapter include the population, data collection procedures and instruments used in the study.

The research purpose was to determine if models can provide sufficient business intelligence that would help business strategy within the real estate industry. This chapter purpose was to present an analysis of the research process adopted in the study providing input into the statistical procedures involved in the research methodology.

According to Booth et al, (2003), everyone does research in one form or another that involves gathering information to answer a question that solves a problem. It is a process of collecting information and data for the intention of making business decisions. Research methodology may include research both present and past elements such as publications, interviews, surveys etc.

The study purpose was to understand the effectiveness that predictive models will have on business intelligence, with the aim of improving decision-making. Within the real estate environment, there is a need for insight on property owner's propensity to sell their property. This will allow real estate agencies to identify people that would potentially be willing to sell their property. This is for the purpose of obtaining housing stock or to be in the position to offer the property seller opportunities on other real estate properties to be purchased. This intelligence will aid real estate agents by reducing marketing effort and improving property sales.

Real estate data originates from multiple data sources that reside on different platforms and within different formats. This makes it challenging to consolidate data into one central location so that it can be analyzed. Another challenge faced relates to data being too complex due to its format being structured, unstructured and semi-structured, which makes data aggregation complicated. Misunderstanding of data would lead to a lack of business intelligence as reporting, graphs and dashboards are used in an incorrect manner. This study would unpack the various sources of data that is related to the real estate industry in order to create predictive models that will drive business intelligence and improve decision-making.

De Vaus (2001) states that research design is the strategy adopted that integrates the different components of the investigation into a logical manner. This approach ensures that the research problem is addressed. Research design forms the architectural plan for collections, measuring and data analysis, which is based on the defined research problem.

### 3.2 Research methodology and design

The research design in Figure 3.1 is the blue print for the collection, measurement and analysis of data.

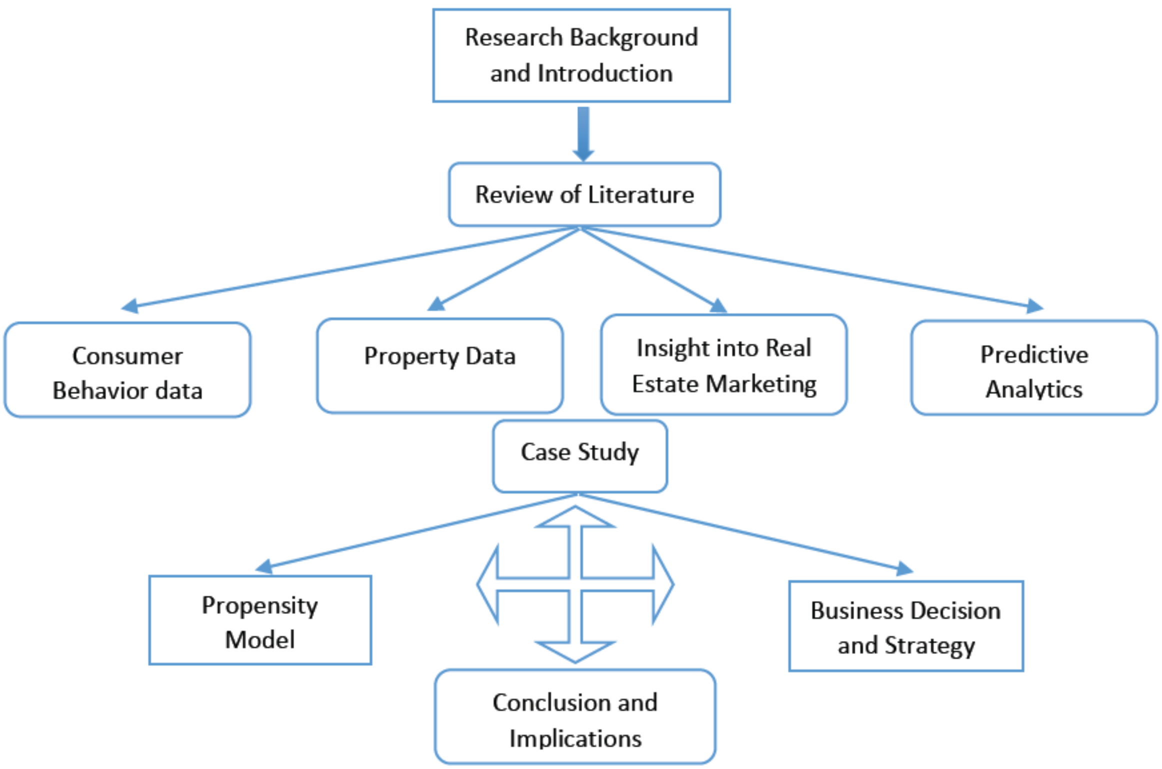



Figure 3.1 Research design process flow

Authors Own source

### 3.2.1 Research method

The main model created was based on three independent variables (household profile, wealth profile and property profile). The framework is aimed to investigate the relationship between the three independent variables and the dependent variable, which is the housing selling decision.

This study adopted a quantitative approach using a descriptive research design that would attempt to answer the research questions. The research objective was to ascertain the effectiveness that estate agents gain through leveraging predictive analytics. This was measured using objective means (e.g. the increase in product uptake, usage growths) thereby warranting a positivist approach.

The population data adopted within this research consists of secondary data as another party through primary data initiatives assimilated it. The population data consisted of 16126 individual property owners records. The data set is further split by demographic, psychographic and location based consumer information obtained from the following data sources:

- Deeds Office – Geographic location of property, Property size, Property value.
- Home Affairs – Spousal details, marital and deceased status.
- Credit Bureau – Credit risk profile, address and contact details, employment details and adverse indicators e.g. debt review.

The data would be a snapshot of suburbs within KwaZulu-Natal viz. Umhlanga, Mount Edgecombe and Phoenix.

The quantitative data collected uses a distribution approach in analyzing the frequency and percentage occurrence of each variable. Data would be analyzed using R, Microsoft Excel and visualization graphs.

### 3.2.2 Design

The study follows a retrospective design, which would be used to compare groups of people who share similar attributes but differ by certain characteristics. According to Babin and James (2010), a retrospective study, also known as a historic cohort study, is a study of individuals that share a common factor when compared to another group of equivalent individuals who are not exposed to those factors, in order to determine that factor's influence on the incidence of a condition.

Data was collected from existing sources and can be analyzed immediately by looking back with the benefit of better analyzing multiple outcomes. Design process includes:

1. Descriptions of what predictive models are.
2. Create a small sized database based on the population data provided that includes the various data categories.
3. An extensive literature review on the conversion of predictive models into business intelligence that would improve business decisions.
4. Define Big Data processing in terms of what data was processed, how and why would it be necessary to follow data processing standards.
5. Perform a descriptive analysis of the database from point 2 to determine consumer-selling behavior.
6. Define predictive models that could be used to measure consumer-selling behaviour by using behavioral clustering or propensity modeling.
7. Apply the models using reporting tools (e.g. dashboards) to represent on how it will improve business intelligence within real estate.

A propensity to sell model would be built for the real estate sector to demonstrate how data-driven results could affect marketing decisions. Founded from the model results, a validation and discussion marketing strategies would be carried out.

### 3.2.3 Quantitative method

Per Kumar (2005), there are two types of data namely quantitative and qualitative which can be further approached into a structured and unstructured manner. The choice of approach is dependent on the aim of the inquiry (investigation, validation or quantification) to either formulate policies or redefine processes. Quantitative data examines hypotheses that are composed of variables that are usually analyzed individually or grouped. The results of the analysis of the hypotheses are expressed numerically, and usually through the means of statistics (Creswell, 2003).

This study adopted structured research using a quantitative approach because the research process, which includes objectives, sample and data collection intended to create the model, is predetermined.

A quantitative research method was chosen in this study to test the hypothesis that was indicated in the study due to the following reasons:

- The selected positivist paradigm was better modeled using a quantitative approach.
- The aim of the study was to validate and test the hypothesis resulting from the research problem and examine the correlation between the variables and the research model.
- Due to the quantifiable measurements of the variables, which allowed for inferences to be drawn from large samples of populations.
- Variables specified in this research were quantifiable and measured.

Thus, a quantitative research methodology was best suited for this study and it was executed through assembling and evaluating data in terms of statistics. Quantitative methods used include secondary data analysis.

During the data construction in preparation of the model, an outline of both the independent and dependent variable with measures and size would be scoped. After the assembly of the master table, the next step was to split the data (60%) into training for the propensity model and another dataset (40%) for testing the model's accuracy by comparing the predictive and actual scores. 3716 of raw sales data from the year 2010 onwards was used for running the propensity models. The propensity to sell model would look at coefficients and the significant value of both the independent and dependent variables. With the significant values, it can be evaluated if the corresponding independent variable was significant and correlated for the model. If not, then the propensity models would be rerun until all the insignificants variables have been removed.

A logistic regression analysis would be used on significant values with the results from the test data to be used as an input into the production models in order to obtain a predictive score, which would be compared to the actual score. This would be the process to test the accuracy of the model. The final step was to draw preferable results for real estate agencies, by implementing the 'Propensity to sell real estate properties by property owners' model. It must also be noted that this study can be replicated within other industries by changing the dependent and independent variables as the technique and calculation will remain the same.

### 3.3 Location of the study

The province of KwaZulu-Natal has sought after property due to its diverse landscape, which consists of farms, golf estates, apartments, security estates, cluster homes and retirement villages. Properties along the north coast of Durban and Umhlanga Rocks represent a good split in terms of property value, income brackets, age groups and lifestyle index.

This study would be conducted on consumer behavior within the real estate industry geographically located in KwaZulu-Natal. The population data consisted of three suburbs including Umhlanga, Mount Edgecombe and Phoenix.

Reasons for choosing the data segments:

- Average value of the property differs substantially.
- Income inequalities, which is indicated by the Gini coefficient that measures the dispersion and variation in income.
- Average salary differs in the three locations.
- Consistent age group splits across the three suburbs.
- Good segmentation of the Living Standards Measures across the three suburbs.
- More differentiated ethnic groups, which is a source of a more diverse society.

It appeared that no research has been performed on a propensity to sell a property by real estate property owners based on data within the KwaZulu-Natal region. Based on data availability, the three coastal submarkets of this study were chosen. According to Property24 (2017), as at August 2017 there are currently 52325 properties for sale within its database, from KwaZulu-Natal, South Africa with an average value of R2 782 288.

Breaking these numbers further down into the suburbs chosen for this study, houses for sale within:

- Mount Edgecombe: 490 with an average value of R5 250 979.
- Umhlanga: 2471 with an average value of R 6 362 779.
- Phoenix: 363 with an average value of R1 012 178.

### 3.4 Target Population

According to Booth et al (2003), in statistics, population is used to describe the subjects of a particular study. This could be large or small in size, which share common characteristics e.g. age, gender i.e. everything or everyone who is the subject of a statistical observation.
Statistical populations are used in order to observe behaviors, trends, and patterns in a defined group of individuals that interact with the world around them, which allowed for conclusions to be drawn about the characteristics of the subjects for the study.
The population data adopted in this research included residential properties currently owned or previously owned within Mount Edgecombe, Phoenix and Umhlanga. The data was secondary, as another party and not the researcher through primary data research initiatives understood it.

The raw sales data from year 2010 onwards was made up of 3716 data entities before validation the data for missing fields, removing duplicates and incomplete data. Subsequently upon checking the data for accuracy, integrity and completeness, data cleansing techniques would need to be implemented. The data was received in a comma delimited text file, which was imported into a SQL database for easier manipulation. Duplicates were removed by applying heuristic business rules based on the property address segmented by owner.

The task of researcher in this stage is to check for errors and omissions in the data, and then to adjust the data so that it is complete, consistent and readable. The final data set being the total sample used for the study was 2848 observations.

The historical data was very messy with a fair amount of noise and missing values. The historical data from the property was used to update and clean the address fields and property size fields on population data. The cleaning and processing of the data was estimated to take several weeks as getting the data into the correct format in this historical dataset would take a fair amount of time. At this point exclusions are required to clean up the data for modeling purposes. The following exclusions have been applied:

- Properties that had a lifetime ownership of 0 months;
- A property size of 1 square meter; and,
- The purchase amount was less than R1 000.

### 3.4.1 Outliers

To improve the quality of the data being modelled, outliers were excluded from the analysis so that they do not skew the results. Excluding outliers would improve the data integrity as outliers could relate to data being incorrectly captured. The following variables had their outliers excluded:

1. The maximum lifetime ownership of private persons in the dataset is 1376 months, which equated to approximately 114 years.
   Lifetime ownership threshold has been set to exclude properties owned longer than 300 months (25 years). By choosing a ownership threshold of longer than 25 years skewed the results due to there being a high probability that the raw data was invalid.
2. 514 properties had more than 6 property owners sharing the property at the same time. These 514 properties were excluded from the analysis.
3. The maximum purchase price for a property was R980 000 000.00 which was registered in 2015. The data reflected that it had a R913 000 bonded amount and therefore appears that the purchase price was a data entry error.
   A total of 13 properties had a purchase amount of more than a billion Rand and have been excluded from the dataset.
4. Present value purchase price per $m^2$ is more than R50 000 000/sq. A further 20 properties were excluded.
5. The average age of the people who owned the property must be populated.

### 3.5. Research instruments

This study aimed to answer the question on how to use propensity model to predict an individual's behavior with regards to selling their property. For this purpose, the unit of analysis of the design would focus on individual's behavior, wealth indicator in terms of income and risk profiles and property trends in the areas of the study.

Thus, data was collected from the organizational CRM database, deeds office data, home affairs and credit bureau. With reviewing, analyzing and blending the data collected, a master table for propensity modeling was created using SQL coding.

#### 3.5.1 Secondary data

Unlike primary data, which is collected by the researcher to achieve a research objective, secondary data is data that was either collected by other researchers or research organizations that share the data with other researchers.

Secondary data analysis provides the researcher with an opportunity of saving time that would otherwise be spent collecting data. In particular, the case of quantitative data can provide larger and higher quality databases that would be unfeasible for any researchers to collect on their own.

According to Ghauri and Gronhaug (2010), the main advantage of secondary data is the time saving in terms of data gathering. Technology in the form of online databases has allowed the researcher to have access to data that would otherwise be confined to libraries and institutions. Another advantage of choosing secondary data is that it is the least expensive means of collecting data. For example, a component of the LSM data collected for this study was obtained from government data (e.g. Census survey) that is freely available for public access. This is especially useful for geographic studies. Lastly, the vastness, variety and veracity of secondary data generate new insights that can lead to unexpected discoveries. This allows the researcher to analyze the data and come up with relevant conclusions or simply verify previous results.

A potential downside to using secondary data collection is that it may not be appropriate in answering the research question. This could be due to the data lacking validity, which may lead to an alternative technique of data collection e.g. surveys or interviews. Another downside of using secondary data is the data quality. Government institutions usually have poor quality controls in place and therefore cannot guarantee the data quality.

## 3.6 Data collection

Quantitative analysis includes the collection of data with the purpose of utilizing statistical procedures and hypothesis testing with the aim of supporting or disproving claims (Creswell, 2009). A method of data collection is simply a technique that is used to collect research data by taking ungrouped data, which is not arranged, in any systematic order and grouping it so that it is presented in a form of frequency distribution.

### 3.6.1 Data flow

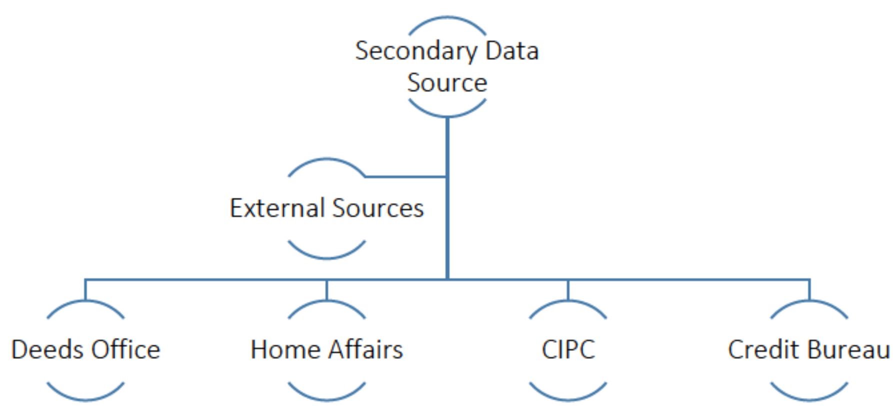



Figure 3.2 Data flow process

Authors Own source

According to Saad (2011), data flow utilizes a model or diagram to map the entire process of data movement as it passes from one component to the next within a systematic process flow, taking in consideration how it changes form during the process. As indicated by Figure 3.2, secondary data used within the study is sourced from external sources viz. Deeds Office (Property Details), Home Affairs (Personal Details), CIPC (Directorship) and Credit Bureau (Consumer).

### 3.6.2 Data categories

The data categories indicated by Figures 3.3, 3.4, 3.5 and 3.6 would be used in the research.

#### 3.6.2.1 Credit Bureau – Consumer Data

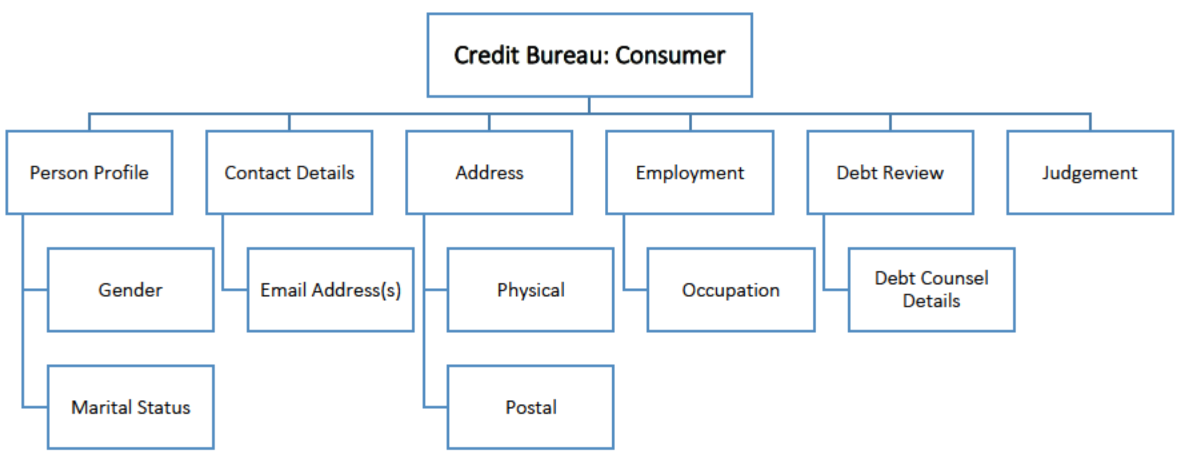



Figure 3.3 Credit Bureau – Consumer Data

Authors Own source

Data sourced from the Credit Bureau indicated by Figure 3.3 can be broken into various levels viz.:

- Person profile: Only gender and marital status from a person profile was applicable to the study. Personal information e.g. ID number and full name was excluded.
- Contact details: Email address could be used to identify the company that a consumer is employed at by interrogating the domain name within the email's address line. Telephone numbers has been excluded from the study.
- Address information: Was used to cross reference the location of the consumer to the title deed of the property.
- Employment: Employment in terms of occupation and employment status was used as an input into the wealth profile.
- Debt review and judgment: Was used to determine the risk profile as an input into the affordability of the consumer.

### 3.6.2.2 CIPC – Directorship Data

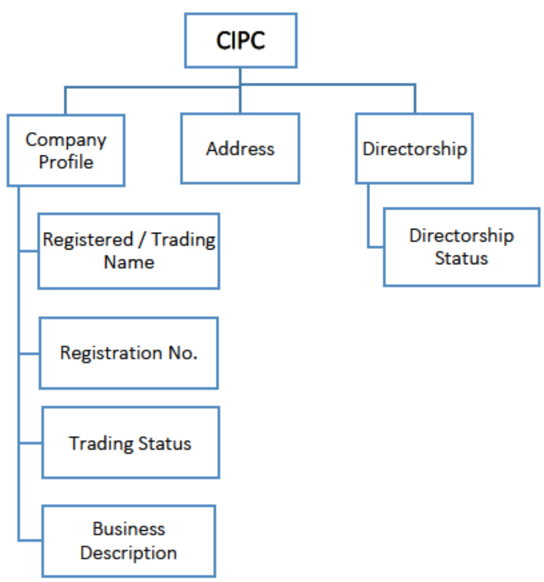



Figure 3.4: CIPC

Authors Own source

The CIPC data source indicated by Figure 3.4 consists of information, which includes current directors of the company, inactive directors of the company, company auditor details etc. For this study, the following information was used:

- Company profile: Business type and trading status indicated the type of business the consumer is employed at to determine the wealth profile, migration pattern and potential to sell their property. Business name and registration number was not applicable to the study.
- Address: Was used to understand the location of the business in relation to the residential address.
- Directorship: Was used as an indicator within the wealth profile based on the individual being a director of a business.

### 3.6.2.3 Deeds – Property Data

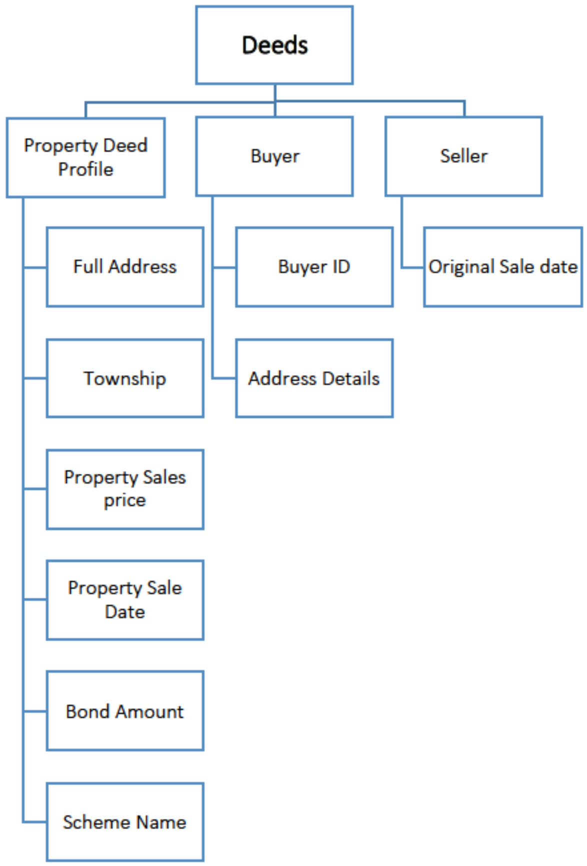



Figure 3.5 Deeds – Property details

Authors Own source

The Deed data indicated in Figure 3.5 consists of a variety of information that included: property bonded amount, transfer details, attorney details, identification of the plot of ground (referred to as ERF) and square meterage of the property. The following information was required for this study:

- Property Deed profile: This provided insight into the attributes of the consumer's property e.g. Date the property purchased, sales price and bonded amount of the property.
- Buyer: Buyer ID was required to link the property details to the Home Affairs and Consumer Data

- Seller: Seller information in terms of when the property was sold to the current owner was a good indication on how long the current owner was staying at the property.

#### 3.6.2.4 Home Affairs – Personal Data

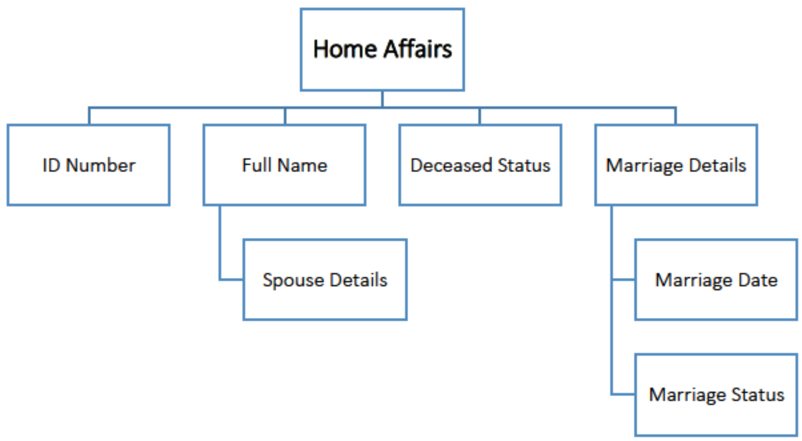



Figure 3.6 Home Affairs

Authors Own source

Home Affairs offered a diverse range of data relating to the consumer as indicated by Figure 3.6 e.g. identity number, passport number, adverse indicators, birth certification etc. For the purposes of this study, the information that will be relevant included:

- Deceased: Excludes consumers that have a deceased status.
- Marital details: Indicate years of marriage and current status e.g. single, divorced etc.

## 3.7. Data analysis:

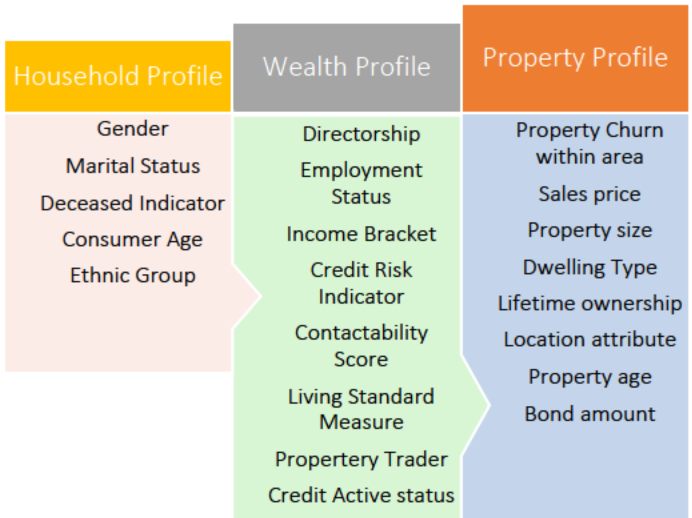



Figure 3.7: Consumer Profile

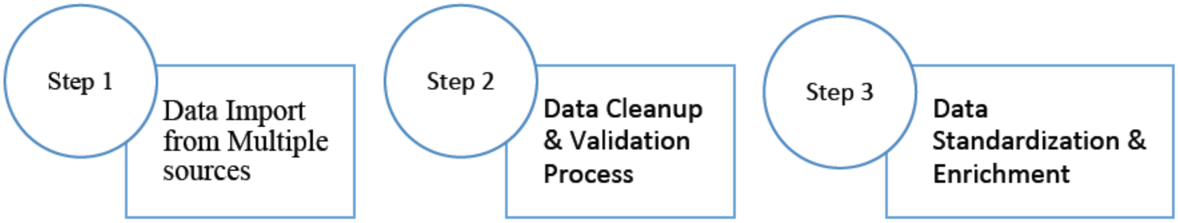
Authors Own source

As per Figure 3.7, three consumer profiles will be developed from the data categories viz.:

- Household profile: Insight into the consumer unit in terms of gender, consumer age of marriage, adult mortality, marital disruption, household sizes etc.
- Wealth profile: Income estimation, risk assessment, lifestyle measurement etc.
- Property profile: Salability rate of properties within a specific area, years of tenure at a property, price of the property and house type etc.

### 3.7.1 Data extract and cleanup process

According to Harrell (2015), not all data is accurate and is required to be cleaned for the studies. Data can be bad for any number of reasons including self-reporting errors, incomplete data aggregations and poor standardization methods. Therefore the following procedures are required




to prohibit data pollution.

Figure 3.8 Three-step Data Extract process

Authors Own source

Figure 3.8 is a high-level overview of the proposed workflow.

**Step 1: Data import**

The data import process is a process that uploads data from external sources with the aim of consolidating it into a central repository. The central repository in this instance is a Microsoft SQL database. The purpose was to organize, analyze and act upon unified data views in ways that is better aligned with specific objective and needs.

The external sources include data categories mentioned in 3.5.4.2:

- Consumer (CON)
- Commercial (COM)
- Deeds (DE)
- Home Affairs (HA)

**Step 2: Data cleanup**

The next step and one of the biggest challenges within this process was in the data cleanup and validation process. The data cleanup process indicated in Table 3.1 occurred for every record that comes into the database. Data would be kept in its raw state and consolidated until all data has been standardized.

Table 3.1: Data Cleanup Process

| No. | Data Cleaning Function | Description |
|---|---|---|
| **1** | Property Validation | Validation would sift through the property details and correct the spelling, address line layout, postal codes and street types. |
| **2** | Address Correction | Address correction included street names, suburb names, city names, province and postal code.<br>If an address was invalid it would be rejected and added to a rejected list. Rejections would then be analyzed and then re-imported. |
| **3** | Remove Bad Characters | The removal of any bad characters such as additional spaces, symbols, non-meaningful characters. |
| **4** | Property Details | Corrected the property details by using the latest or most used details for a property. |
| **5** | Remove Duplication | Removed any duplicate data. |

**Step 3: Data standardization and enrichment**

To maintain consistency, manage vagueness, poor data quality and lack of content, it was necessary to standardize and enrich the data from multiple platforms. This involved a process of transforming unstructured and incomplete data into valid, enriched information for reliable analysis and greater leverage. Enrichment used other data source e.g. South African Post Office and Google Maps to enrich the address data.

- **Standardization**

  - **Address validation**: With the address being the primary entity in which the system would be built upon, the address verification is one of the most essential tasks.
  - **Data consolidation**: By creating parent/child relationships, relationship links would be formed from the various data sources. Although the database would remain segregated on a data source level, consolidation of the aggregated data was required to be performed in-order to obtain the desired analytics.
  - **Validation:** Further data cleaning / validation would be performed at this stage which includes validation of property features.

- **Data Enrichment**
  - Once the data has been consolidated into a structured format, further enrichment will be performed on data relating to deeds, address and consumer information. E.g.
    - Google Maps was used to fill the gaps relating to address information;
    - Census data provided information relating to the consumer's lifestyle;
    - South African Post Office data assisted in filling the gaps on postal codes and suburb details.

## 3.8 Ethical Considerations

To obtain ethical clearance for this study the researcher first obtained a gatekeeper from the Managing Director, Mr. DW Muller, granting permission for the use of Metonymy (Pty) Ltd's data. Once the gatekeeper's letter was received the research proposal was submitted to the University of KwaZulu-Natal Ethical Committee for approval of the study. Both the gatekeepers letter from Metonymy (Pty) Ltd and the approval letter from the Ethical Committee can be found in the appendix of this research report.

### 3.9. Conceptual Framework – Scorecard Model

The aim of applying a scorecard model was to optimize the business activities within the real estate environment. According to Garrido et al (2014), predictive models allow to predict future behavior of people with scoring models being a special kind of predictive mode.

The scorecard to sell a property model was based on the likelihood of an individual to sell their property with household lifecycle being a primary determinant. The household lifecycles of an individual would be mapped to the property lifecycle behavior of an individual to determine if there was a strong correlation between the two.

In addition, the likelihood that a property will sell, as per Kassel (2016) is said to be influenced by the attractiveness of the suburb, the lifestyle of the suburb, the investment potential, the average lifespan of the property over time, and the demographics and past property acquisition of the current owner of the property.

For example, property buyers who intend to sell their existing property may or may not be correlated with single adults with no children; or a downscale in property may or may not be correlated with retirement, divorce or financial hard times. Additionally, lifestyle and investment opportunities may also appeal to a property seller. Such attributes would require suburb attractiveness index be addressed as a potential driver of a person's propensity to sell a property. Demographic attributes such as age, gender, occupation, earning potential, equity and existing property assets will be used as input attributes into a scorecard to sell.

#### 3.9.1 Key attributes of the propensity to sell model

According to Kryvobokov (2007), attributes are characteristics of an object (e.g. person), which are closely related to variables represented within data processing. Predictive analytics used to produce a model is most efficient when used to answer a specific inquiry. The data contains both company and private person owned properties. As there are additional variables we can obtain from the private persons data that is not available for the company data and vice versa. The two groups will be modeled separately.

There are variables that are common to both groups:

1. Lifecycle: This refers to the ownership of the property, which is defined by the length of time in months form the time that the property is bought to the time it is sold.
2. SELL: Using the 'Sale Date' from the deeds data will indicate when the property is sold. If the variable value = 1, then the property was sold else 0 indicated that it belongs to the current owner.
3. Purchase amount which is given as the present value purchase amount using the following formula.

(The present value formula using a CPI 5-year average rate is applied to the 'Purchase Amount' for each property owner

$$PV = R * (1 + \frac{rate}{12})^t$$

Where $PV = present\ value$

$R = purchase\ amount$

$rate = average\ 5\ year\ monthly\ inflation\ rate$

$t = number\ of\ months\ since\ the\ registration\ date$ )

### 3.9.2 Key attributes to measure include:

The property is to be modelled for the scorecard to sell a property, even though the property may have one or more property owners at a given point in time. Hence, private person variables are to be extracted as counts of the following variables per property:

1. Household profile: includes member groups of households through marriage, surnames, addresses or telephone numbers.
2. Property profile: includes when the property has sold overtime and the price of the property. To make the property prices comparable to the present time, a CPI index was applied to the data.

3. Wealth profile: current buyer's historical property behavior, demographic and directorship information.

The following list of significant attributes from these key measures was used to build a scorecard to sell a property:

#### 3.9.2.1 Household profile

1. Male vs female – A female consumer may have a stronger desire for bigger, safer and convenient homes than male consumers. This would thus influence their selling decision if they own a property.
2. Marital status:
    a. Single households
        i. Marital indictor (marital status obtained from the Home Affairs source or spousal relationship link contained within the deeds data)
        ii. Link the primary consumer's identity number to another person relationship who is younger than 18 years of age. This indicated if they are a single parent.
    b. Married
        i. Marital status obtained from the Home Affairs source or spousal relationship link contained within the deeds data
        ii. Link the primary consumer's identity number to another person relationship who is younger than 18 years of age. This indicated if they are parent who was potentially married.
        iii. Marriage date: Indicates years of marriage and was used to provide insight if the consumer was newly married.
        iv. Marriage date plus children under 18: Indicated if they required a bigger household.
    c. Divorced status

    i. Divorce date: Provided an indicator that the property would potentially be sold.
3. Deceased indicator
    a. Deceased date: Potentially indicates that the deceased estate be disposed of which involves the sale of the property.
4. Age of property owner/s
    a. Retired: Older couples with an empty nest.
    b. This indicated that there is a high probability of the property being sold.
    c. Younger people linked to the current size of the property as well as linked to the age of the bond was a good indicator if they would sell e.g. growing family, which represents a full nest.
5. Ethnic group – Indian; Black; white – Consumer behavior based on demographic.

#### 3.9.2.2 Property profile

1. Rate of sale of the properties within that suburb: postulates property churn that was grouped into properties with similar attributes e.g. number of bedrooms.
2. Security (movement from a standalone house to a sectional title property).
3. Title deed record: Provides information on the number of people registered on the title deed.
4. Sales Price of the property.
5. Size of the property.
6. Dwelling type – house, apartment, flat, farm, and gated community.
7. Lifetime ownership - Age of the property in terms of ownership.
8. Locations attributes – link it to Google API:
    a. Location close to schools and nurseries
    b. Health care
    c. Shopping Centre
    d. Food and restaurants
    e. Public transport

9. Average age that a property sold per suburb – properties are kept for a certain time period and then sold.
10. Average price at which a property was sold per suburb – properties are kept until they reached a certain market value (ceiling).
11. Relocation: Change in province linked to a change in employment.

#### 3.9.3.3 Wealth profile

1. Directorship details: Active directorship status within a company.
2. Linked back to the registration date of the company would indicate how old the company and this provided insight into the seniority status of the individual as a director.
3. Company, Sole Proprietor or Close Corporation.
4. Employment – indicated which sector the individual is employed at e.g. banking, retail etc.
5. Change in employment indicated relocation to another province.
6. Income bracket provided insight of the potential spending ability.
7. Propensity to churn based on a consumer's risk indicator indicated riskiness of a consumer to default on their debt.
8. LSM: Living standard measurement.
9. Property trader that bought and sold properties. These individuals have multiple properties in their portfolio.
10. Bonded amount versus the sales price indicates that a deposit was made against the property (i.e. liquid capital).
11. Credit active consumer was based on an application that was made in the prior 3 years using the date the consumer record was last updated.

### 3.10 Regression analysis

The cluster technique used within this study was regression analysis. The next step was to acquire the drivers to determine the propensity to sell by conducting regression models for each of the separate clusters. As previously mentioned, cluster analysis is a multidimensional statistical method that was used to identify objects into groups so that objects in the same cluster are more similar than two objects of different clusters. As per Armstrong (2012), with regression analysis, it provides and understanding of how the typical value of the dependent variable changes when any one of the independent variable is varied while other independent variables are fixed. Using this approach will aim to prove the differences in consumer behavior among the clusters and gain insight in the underlying factors that affect consumer behavior.

### 3.11 Conclusion

This chapter explained several aspects of the research methodology adopted in this research. The chapter reviewed alternative research paradigms and methodologies, and provided a justification of the research design and methodology. The details of data collection methods, data analysis processes, and data cleanup processes were discussed.

In this study, secondary data was used for data collection. There needs to be a deep understanding of the population data to build effective models that satisfactorily describes the data. With the creation of the statistical models, various assumptions must be met with the intention of satisfying the correctness of the model. Formal tests will be applied to test the validity of the assumptions made. The next chapter presents the detailed results of the research.

# CHAPTER FOUR: Results and Discussion

## 4.1 Introduction

This section discusses the development of the regression model constructed to segment the existing homeowner market of potential sellers based on trends inherent in the data. The data was coded and analyzed using a structured medium in the form of Microsoft Excel and a statistical program called R.

## 4.2 Source of data

To construct the model, property deeds data and home owner demographic profiling information were sampled from three popular suburbs in Durban; KwaZulu-Natal: Mount Edgecombe, Umhlanga and Phoenix. Title deed transactional information extending from 2010 onwards was considered, to include the most recent transactions resulting from recent economic factors (omitting the influence of 2008 financial crisis, changing crime levels, etc.).

## 4.3 Variable selection

This study ingests secondary data to yield the desired segmentation. As a result, the choice of variables was limited to the data source and may not measure all factors that may promote property sale.

The dependent variable (sold) marks properties that were sold and provides for a 2-year lag to cater for the time properties go on sale. This is required due to influential factors that lead individuals to release a property for sale. This additional layer improves the overall relationship within the model. For example, it takes up to 12 months from the time a property is sold till it appears on the deed registry due to delays experienced in capturing the data at the deeds office.

The variables considered for the model are listed below:

1. Age of property
2. Present value of the property
3. Size of property
4. Age of home owner
5. Gender of home owner
6. LSM of home owner
7. Household income
8. Credit Risk Score of home owner
9. Director status of home owner
10. Marital status of home owner
11. Deceased status of home owner
12. Property type of home owner

Of the variables listed above, only significant predictors were ultimately included.

## 4.4. Data cleaning and preparation

### 4.4.1 Outliers/Missing Data

Ahead of the data modeling, extreme values on property price (typically flagging farms, larger commercial properties) have been excluded from the dataset fed into the model. On the observed suburbs, Mount Edgecombe and Umhlanga, which are considered elite suburbs within the province, contain shopping centers and malls (no farms or industries). Phoenix, which is considered a middle-class suburb, contains shopping centers; malls and an industrial area. Title deeds and corresponding owners of such properties have been omitted from this research to maintain focus on residential properties.

Similarly, data records with incomplete or missing information necessary for prediction were excluded from the dataset. Often, the deeds office may have missing information such as marital status, ‘erf size’ or ‘house size’ etc.: all of such cases with any missing criteria within the list of selected variables have been omitted from the dataset.

### 4.4.2 Data manipulation

Since property values are subject to inflation, all monetary fields e.g. property values were projected to the present value using the formula below. The size of a property and the price of a property were used to calculate the present value per square meter. By expressing the price of a property relative to a square meter will inform the model on the relative price of the property. Furthermore, to improve linearity and achieve variance homogeneity, the price per square meter was log-transformed using the formulae below.

Rand value variables are

1. Natural log of the present value purchase amount

$$\ln(PV) = \ln\left(R * \left(1 + \frac{rate}{12}\right)^t\right)$$

Where, *rate* is assumed to be 5.0% representing South Africa’s average annual inflation rate.

2. Natural log of the present value purchase price per square meter

$$\ln\left(present\ value\ price\ per\ sqm\right) = \ln\left(\frac{presnet\ value\ of\ purchase\ amount}{property\ size} + 0.0001\right)$$

## 4.5 Transformation

It was important to identify the main predictive variables and distinguish between the independent and dependent variables. This was as per Armstrong (2012) in terms of determining the correlations between the variables that may skew the model i.e. find an identifiable signal to the noise ratio in the data.

According to the R Foundation (2015), R is an integrated software facility that is used for data manipulations, calculation and graphical display. Functionalities include:

- Data handling and storage facility.
- Operators designed for array calculation.
- Coherent and integrated tools for data analysis.
- Language which includes conditionals, loops, user defined recursive functions and input and output facilities.

R will be used to analyze the data due to the ability to handle independent variables that are in a nominal and numerical format. However, if there were many categories for a nominal attribute, which are not significant predictors for the model, they would remain part of the model. When the nominal values are dummy code into a binary variable with 1 if present, else 0 then attribute categories that are not significant can be manually removed from the model.

A description of the type of transformations in cluster, which is a collection of objects with similar characteristic, as per Balasko et al. (2006) is provided in this section. For example:

1. The age or age group attribute was calculated at the age of a first timer when he/she bought the property. It was calculated as his/her age less the number of years that he/she owned the property. The non-property buyers were given their actual age.
2. The following attributes are to be dummy coded into a 1 if the attribute was present, else 0.
    - Gender into male;
    - Ethnic into Indian, African, White and Other;
    - Directorship into directorship is active or directorship resigned; and,
    - Marital status code into single or married.
    - Credit application in the last 5 years.
    - South African or a foreigner
    - Deceased indicator
    - Credit Risk score indicator
    - LSM indicator

3. Dates are used to create the following attributes
    - Months with the income.
    - Months with the risk score.
    - Months with the mobile phone number.
    - Months with the home phone number.
    - Months with the work phone number.
    - Years with the occupation status.
    - Years of property ownership

    Missing values were replaced with zero values.

4. Ordinal attributes with nominal values are converted to numerical attributes
    - Alloy contactability
    - LSM indicator

Replacing missing values were performed in R because R can handle nominal attributes and not missing data. Due to this, the following attributes will have missing data:

- Living Standard Measure (LSM)
- Marital Status
- Number of years since the occupation was last updated
- Income categories
- Credit Risk Indicators

**4.6 Testing of Assumptions - Binary Logistic Regression:**

Validating the predictive model is critical by generating scores on a test sample and generalizes to the entire data population. Pearson's coefficient correlation measures the strength and existence of relationship between two variables and is defined as the covariance of the variables divided by the standard deviations of the variables (Keller, 2012). Pearson's pairwise correlation testing show that a few of the attributes have a strong positive linear relationship with selling or not selling a property. Examples of these attributes that influences this relationship include:

- Income group – The salary bracket that makes up the consumer's monthly earnings.
- LSM group – Living standard measurements differs significantly amongst consumer segments.
- Risk score – This indicates the riskiness of a consumer to meet their financial obligations.

Furthermore, the three attributes have a strong positive correlation among themselves. We will need to compensate for this correlation between independent variables in the model.

There are some necessary conditions for successfully constructing a binary logistic regression. The assumptions are listed and discussed in the Table 4.1:

Table 4.1 Conditions for constructing a Binary Logistic Regression

| **Assumption:** | **Propensity to Sell Model:** |
|---|---|
| Binary logistic regression requires a binary dependent variable. | The dependent variable of this model is if/if not a property was sold, thus being binary. |
| Binary logistic regression assumes that P(Y=1) is the probability of the event occurring, thereby the desired outcome must be coded as 1. | 1 represents a sold property, and 0 represents the property not sold. |
| The model must be correctly fit (avoid over- & under-fitting). | The study employed a train dataset to model and a test dataset to evaluate fit – this approach minimizes over- and under-fitting. |
| Each observation should be independent (error terms need to be independent) | All observations are independent of each other by design. |
| Linearity of independent variables and log odds. A possible solution is to categorize dependent variables from numeric. | To address this assumption, the age and property age variables were converted to categorical variables. |
| Recommended sample size at least 30. | The sample data used in this study is sourced from all property since 2010. The volumes, 2848 used for the study far exceed the minimum recommended base of 30. |

## 4.7 Data exploration

This section assesses some descriptive statistics to understand the distribution between sold properties and key variables that were considered for the model.

The bar charts below show the 100% stacked distribution between each variable and the proportion that was sold within the observed dataset.

### 4.7.1 Sale X Property Age

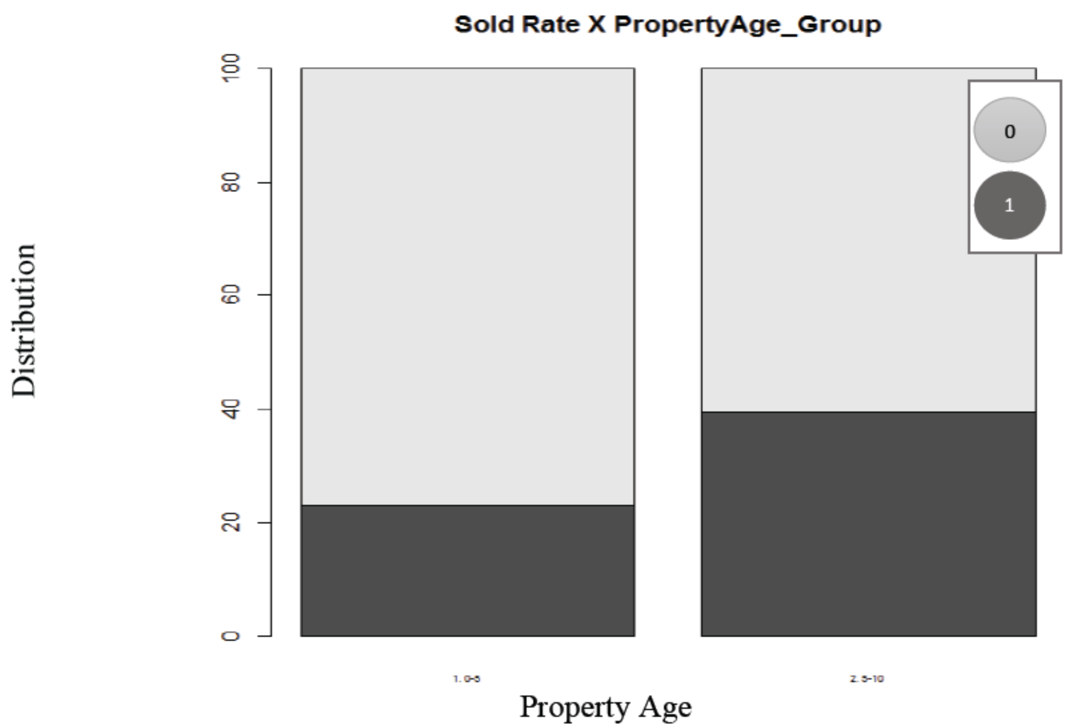



Figure 4.1 Property Age

Property age refers to the age of the building from the time it was built. From the bar graph in Figure 4.1, we see that older properties are more likely to sell than new properties. This ties back to Kotler and Armstrong (2012) thoughts that a consumer's decision is based on perception of attributes they consider important. Reasons for the behavior is that:

- Older properties are cheaper than newly developed units. This is evident when comparing the properties within Phoenix versus Mount Edgecombe.
- Older properties have larger plot sizes with bigger room sizes, double garages than the newly developed units. Umhlanga area, being a central business district has a large volume of apartment style units when compared to properties in Mount Edgecombe and Phoenix.
- Cultural norms and family values, as stated by Moslehpour et al. (2014) and Ling et al. (2015), plays a key role in a consumer's decision making. The information that the study has produced indicates that consumers within the Phoenix market tend to buy near family members, hence why the older property sales are higher.

### 4.7.2 Sale X Property Owner Age Group

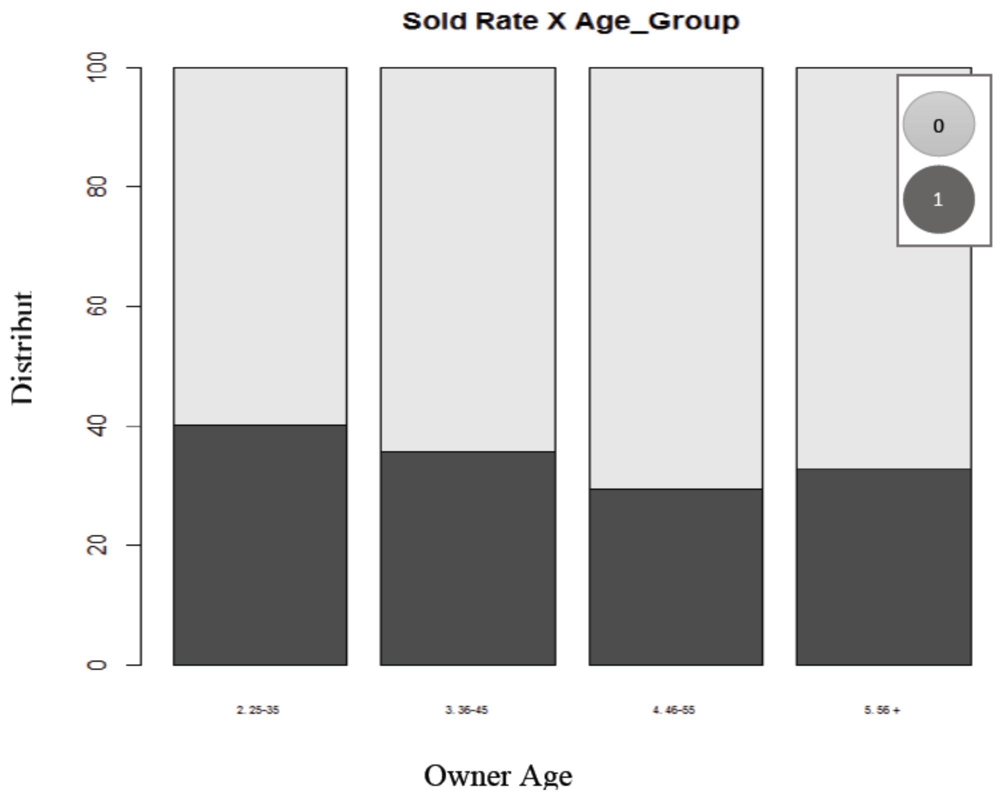



Figure 4.2 Property Owner Age Group

According to the bar graph in Figure 4.2, we see that younger people and older people are more likely to sell. The sale rate of younger property owners was higher than the selling decisions of older property owners. The potential reason that the study has indicated is that across the three suburbs neighborhood amenities or financial considerations will influence the younger consumers more. Older consumers would have a higher interest in neighbourhood attributes and living security whereas younger sellers should more importance to location attributes. This ties back to (Schiffman et al, 2012) thoughts that consumers will purchase properties according to their lifestyle habits. Quester et al. (2011) noted that social considerations is represented in consumption patterns, which have a deep impact on the value of housing attributes as well as the final selling decision. He went onto to state that the most important decision reflecting a family's social class is the decision on where to live in terms of type of home and neighborhood location. Perhaps, younger people have growing family needs whilst older people may be preparing for retirement.

### 4.7.3 Sale X Property Owner Gender

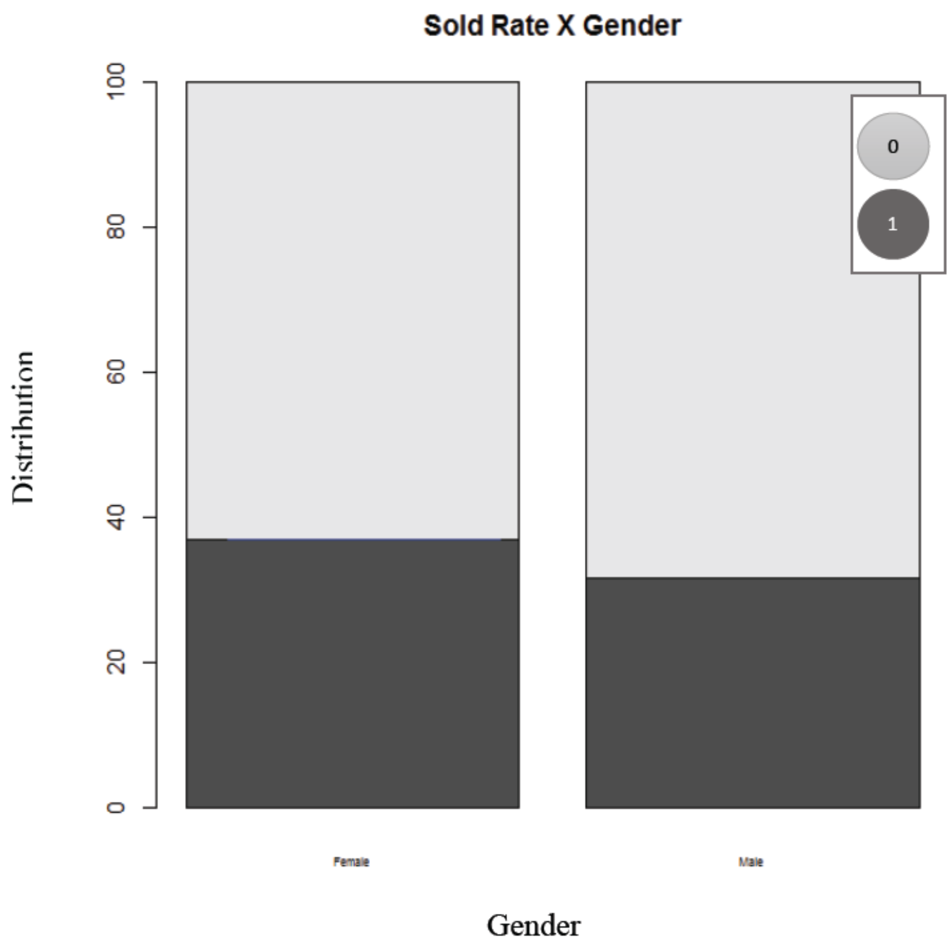



Figure 4.3 Property Owner Gender

Whilst gender profiles seem fairly balanced in Figure 4.3, female owned properties show to have a higher tendency to sell than males. Although financial contemplations will vary in terms of gender, the final selling decision was influenced by these disparities. Gender difference did not significantly influence a consumer's selling decision. Opoku and Abdul-Muhmin (2010) stated that women consider attributes like living space, aesthetics and exterior space more important than males did hence you would find that propensity to sell is more likely to be initiated by females when compared to males. However, the findings suggest gender differentiation does not factor as an important consideration. Therefore, should not factor as part of the business strategy for real estate companies.

### 4.7.4 Sale X Living Standard Measurement (LSM):

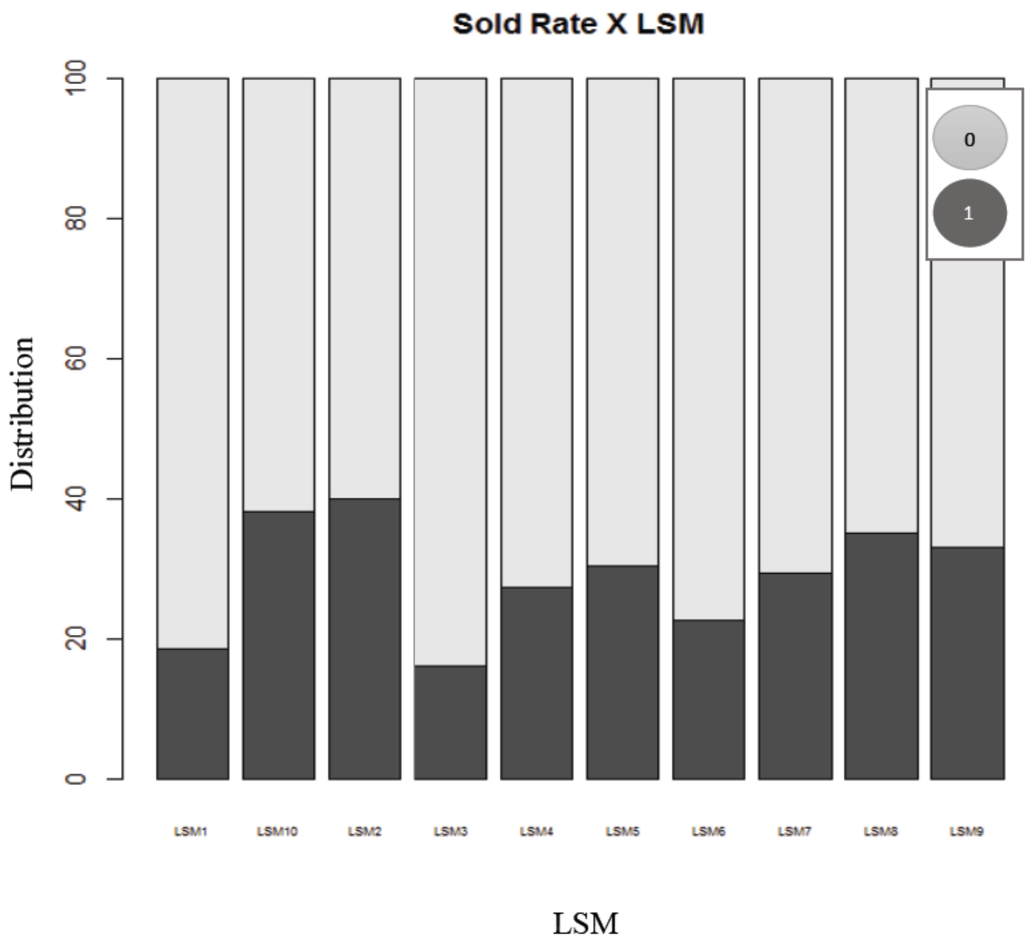



Figure 4.4 Living Standard Measurements

Figure 4.4 depicts that the higher LSMs (8, 9, 10), the middle LSMs (5, 6) as well as the lower LSM (2) have higher tendencies to sell. Based on Haupt (2001) definition of Living Standard Measure (LSM), we see that the results of the study groups people and households into ten distinct groups based on criteria such as their level of urbanization, ownership of vehicles and major electrical appliances. For instance, those with the least number of appliances, no vehicle and not living in urban areas will be classified as the low LSM. Perhaps due to the wealthier position of the higher LSM 8,9,10, places them able to sell their properties more easily. For example, by asking for a realistic selling price that churns the property faster or the property is well maintained which appeals to buyers.

LSM 2, 5 and 6 could be high due to the promotion of lower to middle LSMs to better properties over time. Keita (2012) stated, as part of the definition of consumer behavior in Chapter 2, a consumer's choice and purchase behaviour could be termed as utility maximization subject to the budget constraints i.e. obtaining value for money.

### 4.7.5 Sale X Income Band:

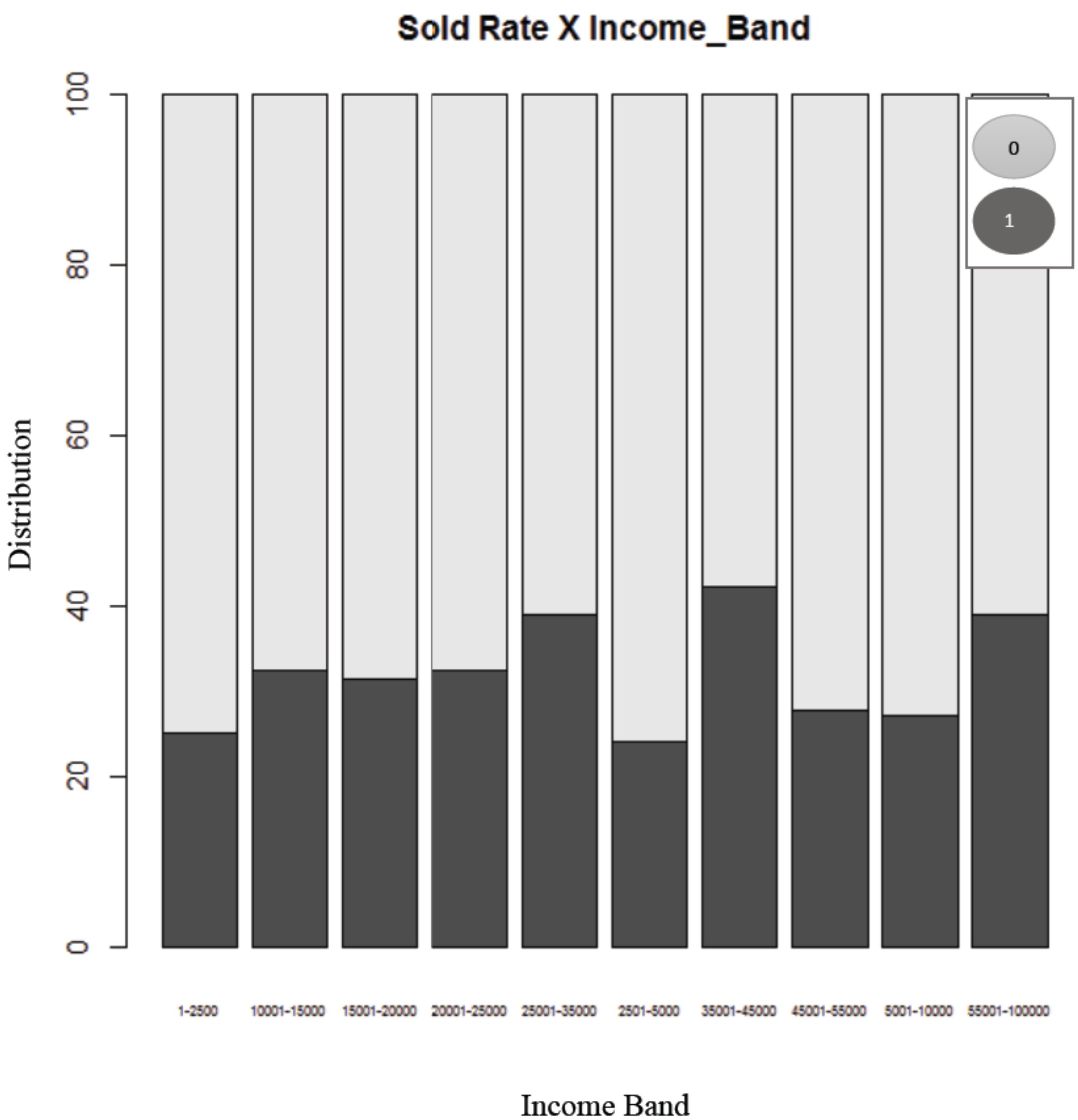



Figure 4.5 Income Band

Similar trends to the LSM are measured across Income Bands. As per Figure 4.5, the higher, certain middle and certain lower income bands are likely to sell their properties. The research found that income level has a considerable influence in consumer's property sold rate. The higher the income bracket, the higher the sale rate of the property. This point aligns to the past study of Kotler and Keller (2012) and Bramley (2012) that consumers with more disposable income can afford luxury goods and want based purchases. Haddad et al (2011) indicated that as consumers progress through their life cycle, they would prefer real estate properties that contain better attributes e.g. double garage. This could also be related to high-income groups who prefer better surrounding environments and convenient locations. In addition, consumers that want to move into a higher social status might choose a house or neighborhood that higher-class individuals would choose, which aligns to Quester et al (2011) thoughts.

### 4.7.6 Sale X Credit Risk Score

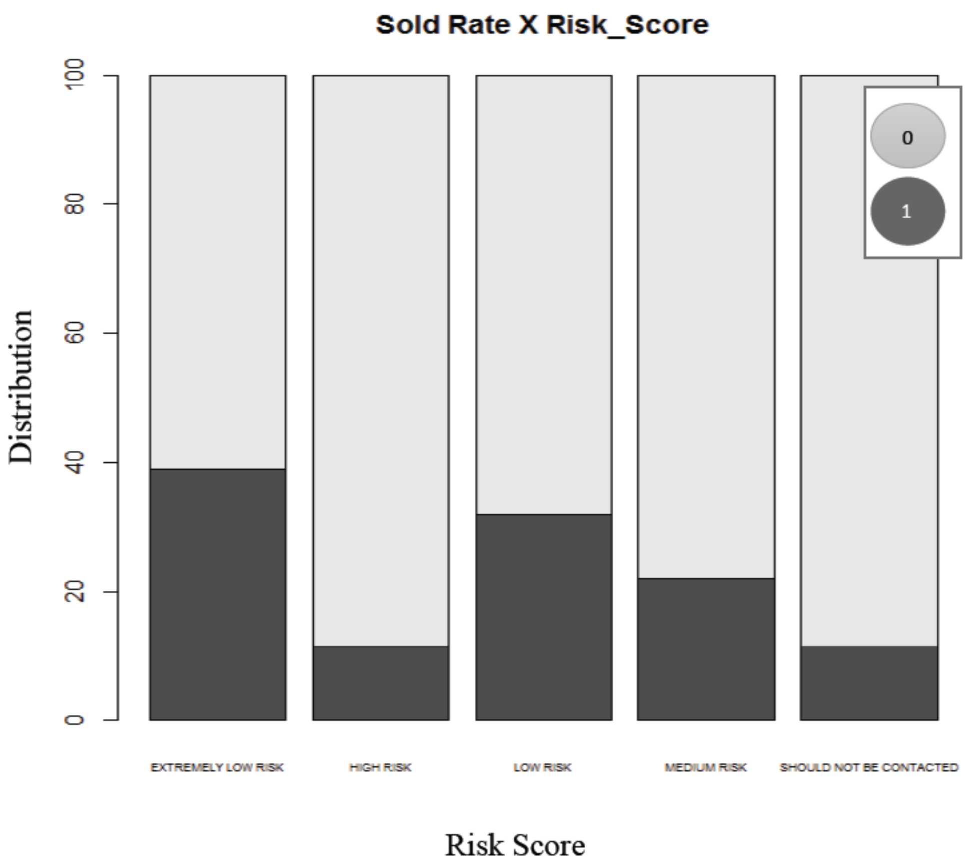



Figure 4.6 Credit Risk Score

Credit history is a record of the consumer's responsibility to repay their debts. As per Transaction Capital Credit Health (2013), poor repayments in terms of late payments or defaulting on debt repayment leads a high credit risk rating. The research has shown in Figure 4.6 that Lower credit risk homeowners are most likely to sell. Perhaps this is highly due to financial institutions funding low risk individuals to buy properties hence, often selling their existing residence. This aligns to Capitec (2017) statement that consumers with favourable risk scores, are usually granted credit more easily. Furthermore, the studies show a correlation amongst LSM, income earners and risks scores that higher LSM and income earners are generally prone to lower credit risk.

### 4.7.7 Sale X Director

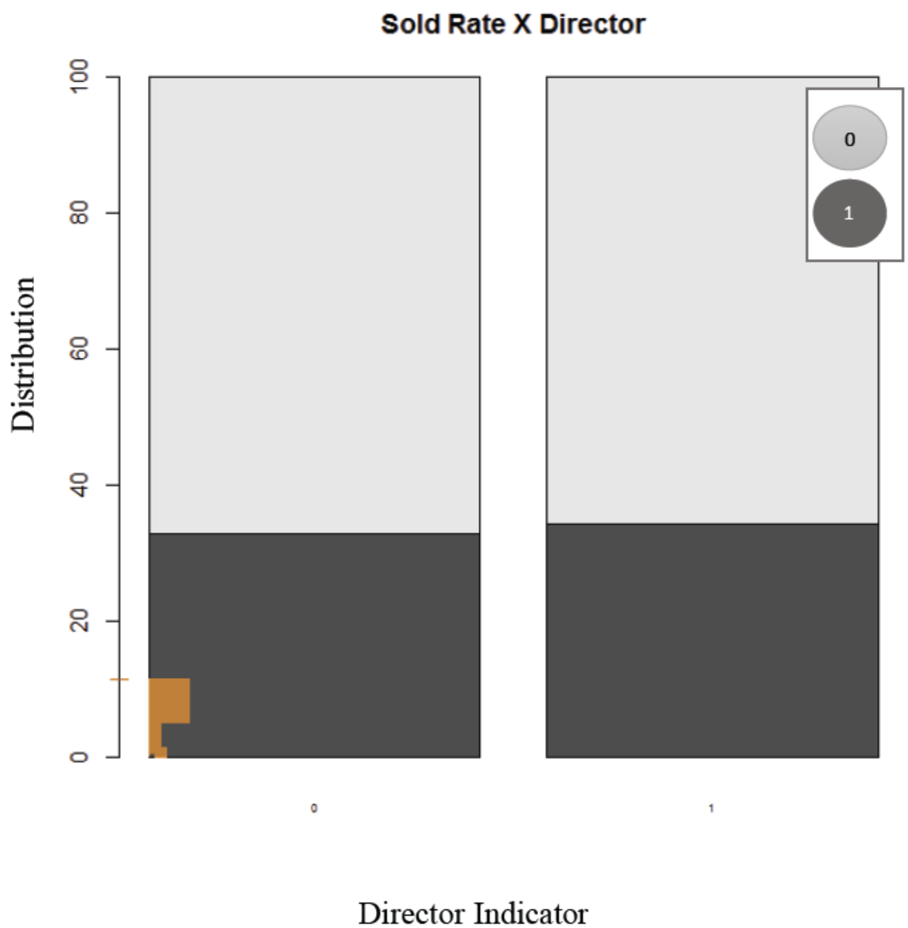



Figure 4.7 Director Indicator

Directorship status was used with the aim of being a good indicator within the wealth profile of the consumer i.e. being a director of a business potentially means that the consumer falls into a high LSM group. However, according to the directorship distribution in Figure 4.7, there seems to be no significant difference between directors and non-directors.

### 4.7.8 Sale X Deceased Indicator

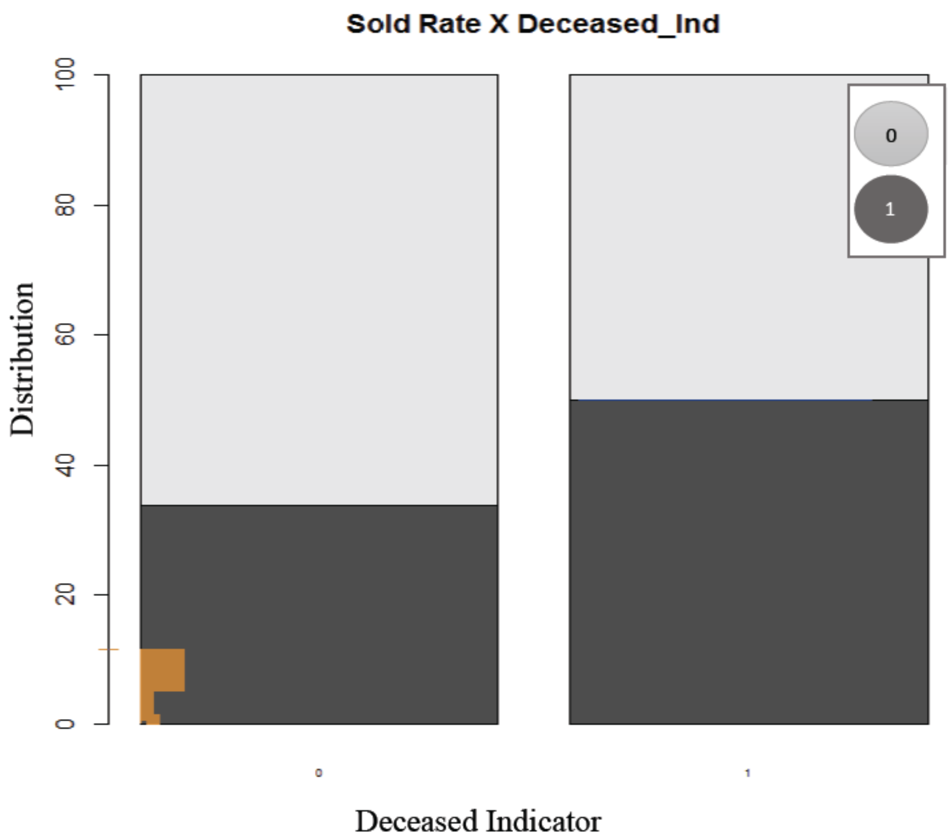



Figure 4.8 Deceased indicator

Although the given data at hand is minimal, the sale distribution in Figure 4.8 of the deceased proves to be a strong indicator of properties sold. As indicated by Adams (2011), people who handle a 'deceased estate will' are generally motivated to sell the property. This is due to property being old and therefore requires capital injection for maintenance or improvements. As a result, these types of properties are sold at a 'bargain' price.

### 4.7.9 Sale X Marital Status

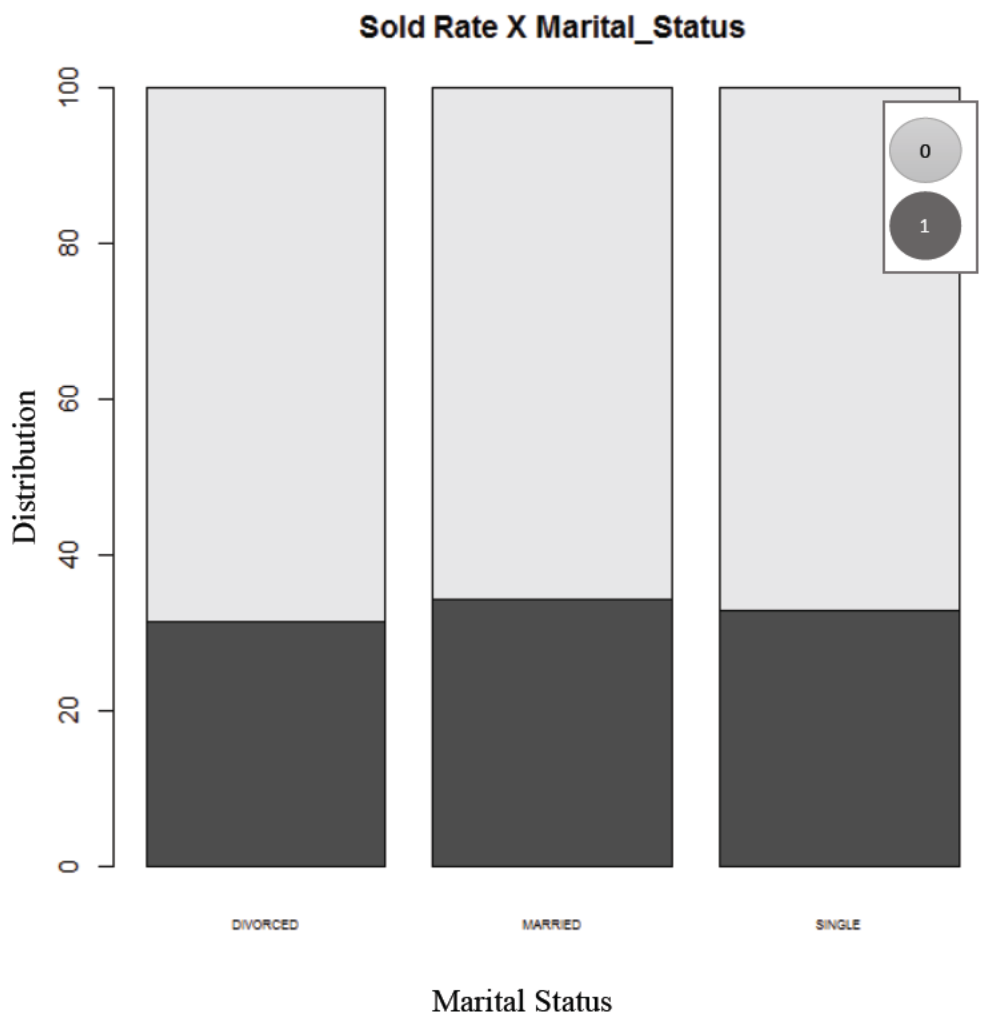



Figure 4.9 Marital Status

According to the information in Figure 4.9, the distribution of the property sale seems more-or-less balanced across the marital status groups. There was an expectation from the research, which indicated that the propensity to sell by married people was more significant than single consumers. This expectation was created by Opoku and Abdul-Muhmin (2010) who stated that married persons tend to be more concerned about environmental attributes being more important in terms of raising children. In addition, Pride and Ferrell (2011) indicated that family's progress through distinct phases called family life cycle. With this in mind, although there is a small indication based on the data, that married people have a higher sale rate than single people, there isn't enough information to make an informed decision.

#### 4.7.10 Sale X Property Type

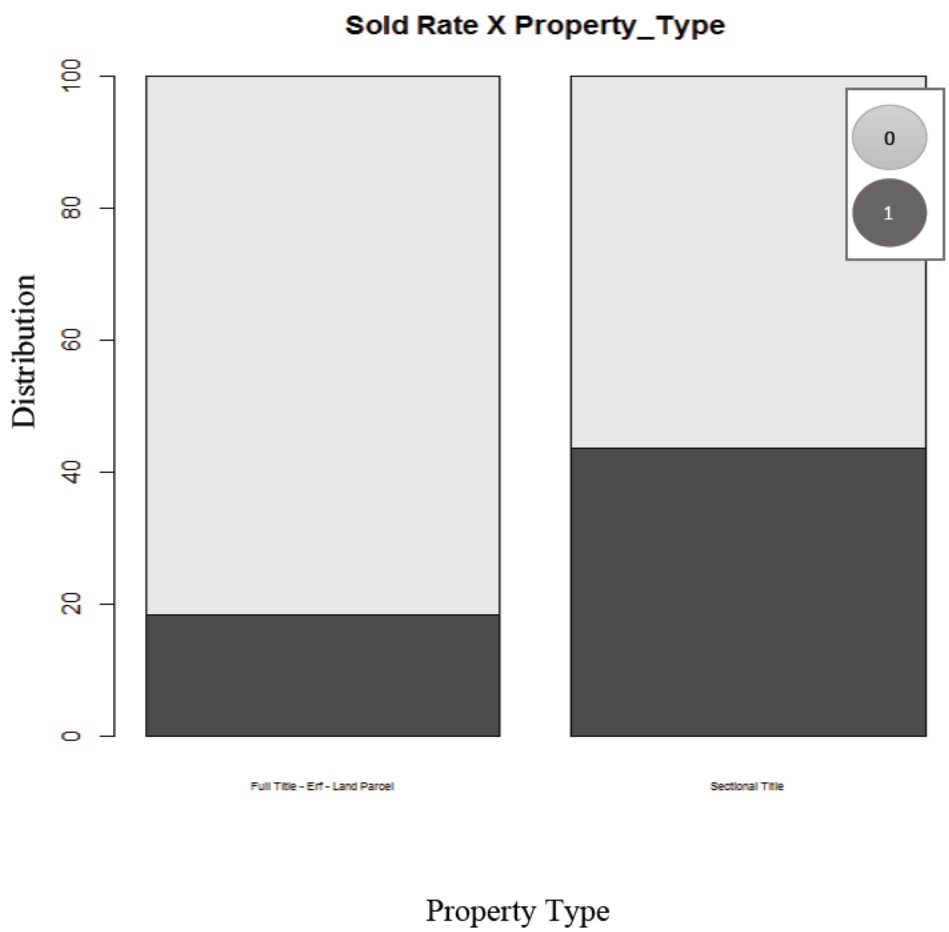



Figure 4.10 Property Ownership Type

The research has shown in Figure 4.10 that there is a higher sale distribution among section title properties when compared to full title properties. As per Ajayi (2012), one of the most important factors affecting a person choice of house is the location. This is perhaps due to the general life stage of people such properties attract e.g. ‘lock up the house and go’ mentality. It could also be due to sectional title properties offering the following benefits:

- Convenience factors due to minimal property maintenance e.g. smaller gardens.
- Heightened security by living in close proximity to your neighbors
- Affordability is better due to fixed monthly costs e.g. levy, garden maintenance
- A more communal way of life boasts close-knit communities and greater interaction with neighbours.

Finally, as per Gelfand (2004), housing location also affects the housing price. Due to Sectional title units being more cost effective than full title units, it may stand to reason that consumers will seek sectional title over full title due to the monthly cost saving.

## 4.8 The Logistic Regression Model

The regression model requires all variables to be in indicator format. For example, the gender variable, which has 2 levels: male and female was converted into 2 separate variables male and female with each of the form 0 or 1 (yes or no).
The sample dataset was split into a test and train dataset using the 80:20 rule (Pareto principle). The train subset was used to build the predictive model, whilst the test dataset was used to evaluate the model accuracy.

The model fit was of the form:

*Sold = β1*Ln_PV_Amount_per_sqm + β2*PropertyAge_Group + β3*Age_Group + β4*Gender + β5*Ethnic_Group + β6*LSM + β7*Risk_Score*

According to the model summary, the following variables have proven to be significant predictors of Sale: Property price per square meter, the age of the property, age, gender, LSM and credit risk scores of the home owners. The regression model further suggests that the price per square meter and age of the property are the most significant variables. Therefore, using a regression model within this study aligns to Harrell (2015) thoughts that regression algorithms are best applied to credit scoring or predicting the next outcome of time driven events. In addition, the variables that have been scored which act as significant predictors ties back to Fletcher's (2011) thoughts data modelling creates scenarios that can be tested which leads to the creation of new products or new ways of thinking.

The AIC score is a relative score (with no numeric meaning) used to select the best model of a few options. The final model has an AIC score of 2501.9. The model with the lowest AIC score was selected and insignificant variables were removed from the model.

Table 4.2 Generalised Linear Model Output

```
Call:
glm(formula = Sold ~ Ln_PV_Amount_per_sqm + PropertyAge_Group +
    Age_Group + Gender + Ethnic_Group + LSM + Risk_Score, family = binomial(link = "logit"),
    data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.0184  -0.9515   0.4990   0.8133   2.5345

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                        9.38766    0.93033  10.091  < 2e-16 ***
Ln_PV_Amount_per_sqm              -0.83301    0.05638 -14.776  < 2e-16 ***
PropertyAge_Group2. 5-10          -1.44359    0.11767 -12.268  < 2e-16 ***
Age_Group3. 36-45                  0.22294    0.19249   1.158  0.24678
Age_Group4. 46-55                  0.51735    0.19684   2.628  0.00858 **
Age_Group5. 56 +                   0.63741    0.19758   3.226  0.00125 **
GenderMale                         0.30090    0.10361   2.904  0.00368 **
Ethnic_GroupINDIAN                 0.47799    0.22040   2.169  0.03010 *
Ethnic_GroupOTHER                  0.08411    0.21486   0.391  0.69546
LSMLSM10                          -1.21968    0.67948  -1.795  0.07265 .
LSMLSM2                           -2.23774    1.65097  -1.355  0.17529
LSMLSM3                            0.38117    1.11263   0.343  0.73191
LSMLSM4                           -0.66584    0.82996  -0.802  0.42241
LSMLSM5                           -1.25087    0.78719  -1.589  0.11206
LSMLSM6                           -0.51973    0.68889  -0.754  0.45058
LSMLSM7                           -0.88870    0.68017  -1.307  0.19135
LSMLSM8                           -0.99817    0.68366  -1.460  0.14428
LSMLSM9                           -0.97316    0.68053  -1.430  0.15271
Risk_ScoreHIGH RISK                0.67803    0.61864   1.096  0.27307
Risk_ScoreLOW RISK                 0.14000    0.11801   1.186  0.23550
Risk_ScoreMEDIUM RISK              0.54834    0.18809   2.915  0.00355 **
Risk_ScoreSHOULD NOT BE CONTACTED  0.82822    0.47895   1.729  0.08377 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2975.5  on 2340  degrees of freedom
Residual deviance: 2457.9  on 2319  degrees of freedom
AIC: 2501.9

Number of Fisher Scoring iterations: 5
```

Table 4.2 indicates that all coefficients are statistically significant and therefore obtain the regression coefficients to fit the desired relationship. The 'LSM', 'Risk Score' and 'Age Group' deviance is less than the residual degrees of freedom and this indicates that the model is not misunderstood and that a saturated model which includes all the coefficients is better than a null model which only includes the intercepts.

## 4.9 Model validation

Given the model was constructed on the training dataset; the validation was performed on the test dataset. The explanatory variables observed on the test data set were run through the fitted regression model, and the predicted response was compared to the actual observed results. A summary of misclassification is expressed in the confusion matrix below:

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0  90  41
         1  93 280

               Accuracy : 0.7341
                 95% CI : (0.6933, 0.7722)
```

According to the confusion matrix, the model classification accuracy is 73.4%. The 95% confidence interval defines a range of values within this study that provides 95% certainty contains the population mean. Due to the large population data used, the mean is identified with more precision therefore the confidence interval is quite narrow.

This indicates that the regression model will roughly predict if a property is going on sale with an accuracy of 73.4%.

## 4.10 Customer segmentation

A key objective of this study was to employ a regression model to ultimately yield a customer segmentation model to group homeowners based on their likelihood to sell a property.

Given an acceptable regression model, the probability of selling a property can be estimated by:

$$Probability\ of\ selling\ a\ property = \frac{1}{(1 + \exp\left(-(model\ estimates)\right))}$$

Likelihood to sell a property is broken down into the following segments:

Table 4.3 Probability to sell a property

| Segment | Probability of selling a property |
|---|---|
| **Sell** | 0.80 – 1.00 |
| **Probably Sell** | 0.6 -0.80 |
| **Maybe Sell** | 0.4 -0.6 |
| **Probably not Sell** | 0.2 - 0.4 |
| **Not Sell** | 0 - 0.2 |

### 4.11 Contribution of the model to business strategy within the real estate industry

Below are the findings, which are tied back to the literature review in chapter 2, on how business strategy will benefit in terms of cost saving within the real estate industry.

As per Kotler and Keller (2012), it is important for business to understand the trigger points that initiate consumer interest within products. Bruner (1988) expounded that the problem recognition phase involves collaboration between individual desire and actual state of mind. Real estate agencies need to consider the driving factors that motivate a consumer to sell their property. It was mentioned that size of the home, up scaling to a better home, job transfer are factors that influences the consumer preference and ultimately feeds into the model that will improve business strategy.

This point ties back to Hoyer and MacInnis (2010), that marketers today recognize consumer behaviour and psychology traits as being their primary focus in terms of consumption of goods, services, time and ideas. Kotler and Keller (2012) stated that behavioral targeting is providing companies with the ability to target consumers and to find the best match between advertisements and prospects. The model has indicated that older properties are far more cheaper to obtain from a price point of view and therefore have a higher likelihood of being sold. Apart from a price point, older houses have larger yard space with bigger bedrooms for larger family units and therefore become more attractive to purchaser. This is the value that can be brought into business strategy through direct marketing using behavioral based advertising, email targeting or product recommendations.

Nandamuri (2012) mentioned that subdivision of markets into unique homogeneous subgroups of customers would influence business strategy by selecting a specific target market that can be met by a distinct marketing mix. As mentioned earlier, these demographics include age, gender, occupation and income. Delivering relevant products that is specific to the consumer's needs thereby reducing unnecessary investment within the sale process (Krishna, 2015).

Based on Verma's comment (2012), that an individual's decision making is based on information search and alternative evaluations, this behavior can be adopted into business strategy. For example, based on the findings that elderly people generally sell their properties at a certain age bracket, this can be a target market for Real Estate agents in two forms i.e. 1. Obtain stock in the form of a property that needs to be sold by the property owners and 2. Offer the elderly consumer alternative residence e.g. retirement village. This behaviour is evident within the model as it demonstrates that younger and older property owners are more likely to sell their property

Being able to understand consumer behavior that enhances target marketing helps improve on customer engagement. This potentially leads personalized interaction with the consumer, which is needed especially when it involves a high valued item like property. This will lead to customer satisfaction; especially around post sale behavior which aligns to Oliver (2015) statement that customized services and insights into customer preferences brings brand loyalty. According to Sunil (2015), scorecards help improve product design and marketing messages tailored for targeted marketing by adopting the right message to the right individual. This can be achieved by identifying patterns that leads to purchase/ selling behavior that will improve personas, segmentations and relevant offers. With LSM being a strong contributing variable in the model, it will be a good indicator for business strategies on how to identify 'hotter' leads e.g. eliminating lower LSM individuals from the marketing strategy. This will save costs on marketing material and resource utilization.

Although marital status within the study remained neutral in terms of the statistical results, it was expected that changes in household structure would lead to housing decision making being made jointly by the husband and wife, according to Anderson et al (2017). Pride and Ferrell (2011) stated that families progress through distinct phases called family life cycle which include young singles, young married without children, married with children, middle aged married with children, older married and older not married. With an expanded data population across more suburbs than the three selected for this study, there could potentially be a different outcome as compared to the results presented in this study that aligns to Anderson and Ferrell's comments.

Burnside (2015) suggested that the study of real estate would profit from a marketing and lead generation perspective, if elements of consumer behaviour concepts that include sociology and psychology were considered. These concepts would enable real estate analysts to predict consumer behavioral habits within the real estate realm.

The Marshallian economic model, according to Marshall (1890), states that buyers will spend income on goods and services that will provide them with the greatest fulfillment, depending on their tastes and price of the goods. This is evident especially with the LSM split and income brackets. As previously mentioned, consumers maximize utility whilst promoting their increased happiness. Mostert (2002) put in plain words that a consumer will spend their income on products that offer the greatest satisfaction that is conditional on their style and cost. In other words, the choice of goods that offers the greatest satisfactions versus the expenditure. Real Estate business strategy needs to understand this behavior in order to capitalize on marketing efforts. Finally, risk score is crucial for business strategies as indicated by the model. The model demonstrates that extremely low to medium risk individuals have a higher propensity to sell their property. This could be largely due to them qualifying for home loan financing as high risk individuals generally don't qualify due to the credit rating status.

### 4.12 Summary

This study has proved that it is possible to predict properties that are highly likely to go on sale given demographic and property detail with an acceptable degree of accuracy.

Earlier iterations of the model showed that directorship and marital status were insignificant. Property type and the deceased indicator were potentially significant but lacked robustness within the observed sample. And Income bands were highly correlated with LSM's and thus removed from the final model presented. However, the stages of a property owner's lifecycle demonstrates that it is an important determinant when a property owner sells their property e.g. retired people may want to downsize their housing living requirements. At the start of a marriage, we notice that couples own smaller properties however their requirements for larger houses become apparent when they have children. Therefore a property owner's age plays a significant role in the model prediction. In addition, LSM also played a significant role in the model prediction as it proved that property owners that fall in the middle to higher LSM brackets have a higher likelihood to sell their property.

# CHAPTER FIVE: Conclusions, Limitations and Recommendations

## 5.1 Introduction

Understanding consumer behaviour relating to residential property is an important segment within the property market in South Africa. The intention of this research was to construct a model to predict a consumer propensity to sell their property within KwaZulu-Natal coastal markets based on statistical attributes of a consumer's life cycle, consumer financial attributes and housing attributes that relates to factors that influenced the consumer's selling decision. The research objectives and research hypotheses were formulated in order to derive a regression model to predict a consumer's selling behavior to disseminate the results to real estate companies. The analysis of the data and discussion of the results were discussed in chapter 4. This chapter presents the fundamental findings of the study as well as the limitations and real-world use of the framework along with potential future research initiatives.

## 5.2 Key findings

The key findings in this study will be discussed in the context of the research hypothesis, which was indicated in Chapter 1. Statistical results in relation to the hypothesis were presented in Chapter 4. Comparing the results in brief with the literature discussed in Chapter 2 will draw conclusions on the hypothesis.

## 5.3 Research objectives

To address the research objectives discussed at the outset of the study:

- This study has proved that it is possible to predict consumer behavior on the likelihood to sell property with an acceptable level of accuracy using the data at hand.

- Given an accuracy of 73.4% prediction, industries such as the real estate market would benefit from such a model to shortlist prospective clients and perform data driven strategy to optimize their operations. This in-turn would result in positive time and financial efficiencies within the industry.

To address the test hypothesis of the study:

**Hypothesis 1:**

- ***H0a***: Models cannot be created to predict property owner's behavior when selling real estate located in KwaZulu-Natal submarkets.
- ***H1a***: Models can be created to predict property owner's behavior when selling real estate located in KwaZulu-Natal submarkets.

**Result:** We fail to reject H1a, as the study has proved that it is possible to predict sale of properties using a logistic regression with an accuracy of 73.4%.

Hypothesis 1 relates to the main research question whether a statistical model can predict a consumer's behaviour relating to the propensity to sell their property. The study has established a positive relationship, which predicts properties that are highly likely to go on sale given demographic, and property detail with an acceptable degree of accuracy. The model summary has proven that the following variables have been significant predictors of sale: property price per square meter, the age of the property, age, gender, LSM and credit risk scores of the home owners. For example, the model demonstrates that younger and older property owners are more likely to sell their property; middle to upper LSM property owners can afford to sell their property and extremely low to medium credit risk individuals also has a higher probability of selling their property, due to their home loan borrowing power.

The model has also pointed out that marital status was insignificant. However, this attribute could potentially add significant value to the overall model with an expanded property sample.

**Hypothesis 2:**

- ***H0b***: The variety of real estate data sourced, based on KwaZulu-Natal submarkets, cannot be consolidated into a usable format that adds value to business strategy.
- ***H1b***: The variety of real estate data sourced, based on KwaZulu-Natal submarkets, can be consolidated into a usable format that adds value to business strategy.

**Result:** We fail to reject H1b, as the study has employed property title deed data and homeowner demographic information to construct a regression model that can predict property sale with an accuracy of 73.4%

Hypothesis 2 conveys the research question of whether the data representative of the target population of all formal properties in the South African market and will the data add value in terms of the correlation of property sales and information on hand?

The present findings suggest that the data assembled from the variety of data sources has influenced a model to be created with a high percentile level of accuracy. Due to this, business strategies can be enhanced related to marketing activities. Target market is critical for businesses to reduce cost and have higher hit rate with respect to making a sale.

**Hypothesis 3:**

- ***H0c***: The consumer population identified within KwaZulu-Natal submarkets cannot be measured in terms of determining the propensity for a consumer to sell their property.
- ***H1c***: The consumer population identified within KwaZulu-Natal submarkets can be measured in terms of determining the propensity for a consumer to sell their property.

**Result:** We fail to reject H1c, as the study has proved that it is possible to predict sale of properties using a logistic regression with an accuracy of 73.4% based on 3 popular KwaZulu-Natal suburbs.

Hypotheses 3 ties back to the research question mentioned in chapter 1 whether the predictive accuracy of the model and the underlying drivers of the prediction are understood.

The study established that a significant positive relationship exists between consumer attitude and consumer selling behavior towards real estate. Consumer attributes that substantiated this relationship was evident in the consumer affordability in terms of income brackets, LSM, age of the consumer and the property lifespan (age of the property).

### 5.4 Implications of the findings

Consumer attributes were presented in chapter two's literature in terms of studies that have indicated the effects that it has in consumer behaviour (Opoku and Abdul-Muhmin, 2010). However, no specific analysis was achieved on specific demographics that influence a consumer's decision to sell an item. This study will indicate which attributes will affect home sellers behavior.

The contribution and implication of the study includes:

- This research contributes to the body of knowledge on consumer's residential housing selling decision making by making references to influences on consumer's behavior e.g. cultural.
- There has been limited scientific studies that identify or quantify housing attributes and consumer behavior that affects the selling decision in within a South African context in terms of the residential housing and the huge investment in this industry.
- Academic benefits: this study can be used as a platform for other researchers to build and expand upon.

Therefore, this study provides contributions to our understanding of consumer selling decision-making processed for the residential housing market in KwaZulu-Natal, South Africa.

### 5.5 Recommendations emerging from the study

Being a generalized regression model, the methodology used to design this framework can commercialize to several verticals that will benefit from this model. Apart from offering this model to real estate agencies, property investors, insurance and banks institutions can also apply this model to enhance their business strategies. For example, banking institutions would be able to determine from the onset what interest rate should be offered, based on the potential period that the home loan will run for. The model indicates that age of a property owner as well as the stage of their lifecycle is a significant contributor in determining when a property will be sold. Given this information, home loan institutions may offer higher interest rates if the belief is that young couples will sell their property within a shorter time frame than the standard 20 year period.

### 5.6 Recommendations for future studies

- With this model being a generalized regression model, which was developed to determine the sale propensity of real estate property specific to residential properties, the model can be extended to include a commercial property model.
- If possible the inclusion of societal and economic factors such as crime levels, property market index and/or additional household information (such as age of dependents etc.), number of people in the household and cultural factors (e.g. religion) that can add insight into the consumer's decision to sell the property.
- The model can be adapted for other industries e.g. motor vehicles industry. By understanding the stage in a consumer's life cycle e.g. marriage or predict when they are likely to have children, will aid motor vehicle retailers to determine when a consumer is most likely to sell their current vehicle to purchase a new one. This will provide motor vehicle dealerships the advantage to obtain new second-hand stock in addition to offering new vehicles.

## 5.7 Limitations of the research

- The study was conducted on 3 suburbs within South Africa. A larger study perhaps would be more representative of the entire country. Further research is required to determine how generalizable the context presented in this study to other segments of the South African residential property market with regards to diverse types of suburbs.
- Of the 3 suburbs selected, 2 are considered elite, whilst 1 is considered middle-class.
- This study was based on specific independent variables due to data availability. Perhaps with a larger suburb base, variables such as the deceased indicator may appear a significant predictor of property sale.
- The current model and the relationship between the monitored variables are subject to change as economic pressures may fluctuate in the future.
- A clear limitation within this study is the lack of empirical research that findings can be benchmarked against. However, given the data used and the statistical test applied to determine the suitability thereof, offers a high probability of confidence that the model is a good approximation of the data.
- Property attributes consist of features including house design, building quality, house type (double story), house finishing (air conditioning, finishes), age of the house, interior and exterior designs, which are expected to influence an individual's house purchase decision (Haddad et al, 2011). Gosling and McCunn (2013) defined living space with attributes that include living room size, kitchen size and its contents, floor area, number of bathrooms as well as bedrooms. Stevenson and Prout (2013) asserted that there is a strong relationship between living space features and a consumer's house decision-making and pricing. These factors were not considered in this study and would potentially be significant contributors to improving the accuracy of the model.

## 5.8 Conclusion

A large component of this study focusses on the property owner's behaviour regarding real estate selling and purchasing decisions. The study of consumer behaviour is challenging as it attempts to understand complex human beings and the reasons that trigger certain types of behaviour towards real estate decisions. Although individuals exhibit distinctive personality traits, we find that it is similar to other individuals who have been exposed to the same external influences of culture and society. Instead of these complexities being ignored real estate agents and markets need to embrace the reason behind consumer behavior that influences market choices.

The preditive model within this study unpacks the components of property owners that influence their behavior to sell their real estate property, why individuals place value on those components and why their preferences changes over time. The consumer behaviour literature would suggest that seller attributes like attitudes, lifestyle and tastes affecting consumer preferences needs to be considered as marketers cannot just rely on demographic data in which they infer information about consumers (Blackwell et al, 2012). Incorporating these consumer behaviour concepts into traditional real estate marketing approaches will improve understanding of individual decision making in a real estate context that will lead to better explanations, predications and success within the real estate market.

Since all of the research questions were answered, the study reached it objectives. The findings of this study indicate that understanding property owner's behaviour that drives real estate sales improves the marketing strategy of real estate businesses.

# REFERENCES


Adams, D., 2011. *Myths and logic of deceased estates.* [Online] Domain. Available at: https://www.domain.com.au/news/myths-and-logic-of-deceased-estates-20110318-1c0c1/ [Accessed: 24 Oct 2017].

Ajayi, C.A., 2012. Factors affecting the location of Real Estate. *Journal of Corporate Real Estate*, 14(4), pp. 214-225. [Online] Available at: Science direct [Accessed: 12 June 2017].

Ajzen, I., 1991. The Theory of planned behavior. *Journal of Organizational Behavior and Human Decision Processes*, 50(2), pp.179-211. [Online] Available at: Science direct [Accessed: 14 November 2017].

Anderson, C., Reynolds, L., Travis W., Gugerty, M.K., 2017. Husband and Wife Perspectives on Farm Household Decision-making Authority and Evidence on Intra-household Accord in Rural Tanzania. *World Development*, 90(15), pp.169-183. [Online] Available at: Ebscohost [Accessed: 01 June 2017].

Antoniou, C., Doukas, J. and Subrahmanyam, A., 2013. Cognitive Dissonance, Sentiment, and Momentum. *Journal of Financial and Quantitative Analysis*, 48(1), pp.245-275. [Online] Available at: Ebscohost [Accessed: 01 June 2017].

Armstrong, J.S., 2012. Illusions in Regression Analysis. *International Journal of Forecasting forthcoming*, 28(3), pp.689. [Online] Available at: Google Scholar [Accessed: 03 June 2017].

Babin, B.J. and James, K.W., 2010. A brief retrospective and introspective on value. *European Business Review*, 22(5), pp.471-478. [Online] Available at: Emerald Insight [Accessed 25 August 2017].

Balasko, B., Abonyi, J. and Balazs, F., 2006. *Fuzzy Clustering and Data Analysis Toolbox For Use with Matlab*. [pdf] Hungary: Veszprem University, pp.3-11. [Online] Available at: <ftp://ftp.unicauca.edu.co/Facultades/FIET/DEIC/Materias/computacion%20inteligente/parte%20II/semana12/clustering/mfiles/ClusteringToolbox/FuzzyClusteringToolbox.pdf> [Accessed: 22 June 2017].

Belch, G.E., 2012. *Advertising: An integrated marketing communications perspective.* 2nd ed. North Ryde: Mcgraw-Hill Australia Pty Ltd.

Betsch, T. and Haberstroh, S., 2012. *Routines of decision-making*. Psychology Press Mahwah. New Jersey: Taylor and Francis Group.

Blackwell, R.D., Engel, J.F. and David, T., 2012. *Consumer Behavior*. California: Thomson Brooks/Cole.

Blumberg, R. and Atre, S., 2003. Digging Into the Web: XML, Meta Data and Other Paths to Unstructured Data. *DM Review,* 38(2), pp.38. [Online] Available at: EbscoHost [Accessed: 08 October 2016].

Booth, W.C., Colomb, G.C. and Williams, J. M., 2003. *The craft of research*. 2nd ed. London: University of Chicago Press.

Bramley, G., 2012. Affordability, poverty and housing need: triangulating measures and standards. *Journal of Housing and Built Environment*, 27(2), pp.133-151. [Online] Available at: Google Scholar [Accessed: 03 June 2017].

Bruner, G.C., 1988. Problem Recognition: The Crucial First Stage of the Consumer Decision Process. *Journal of Consumer Marketing*, 5(1), pp.53-63. [Online] Available at: <https://www.researchgate.net/profile/Gordon_Bruner/publication/235253191_Problem_Recognition_The_Crucial_First_Stage_of_the_Consumer_Decision_Process/links/02e7e53b32d9c9e28c000000/Problem-Recognition-The-Crucial-First-Stage-of-the-Consumer-Decision-Process.pdf >[Accessed: 15 May 2017)].

Burnside, C., Eichenbaumz, M. and Rebelox. S., 2015. *Understanding Booms and Busts in Housing Markets*. [Pdf] Illinois: Northwestern University. pp.1-10. Available at: <http://www.kellogg.northwestern.edu/faculty/rebelo/htm/Booms-busts.pdf> [Accessed: 22 June 2017].

Capitec, 2017. *Credit Enquiries Affect your Score.* [Online] Available at: <https://www.capitecbank.co.za/bankbetterlivebetter/articles/how-credit-enquiries-affect-your-credit-score> [Accessed: 13 March 2017].

Creswell, J., 2003. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* 2nd ed. California: Sage Publications Ltd.

Daugherty, T. and Hoffman, E., 2014. eWOM and the importance of capturing consumer attention within social media. *Journal of Marketing Communications*, 20(1-2). [Online] Available at: Science Direct [Accessed: 17 June 2017].

Datascience, 2016. *Introduction-to-k-means-clustering-algorithm-learn-data-science-tutorials.* [Online] Available at: <www.datascience.com/blog/introduction-to-k-means-clustering-algorithm-learn-data-science-tutorials> [Accessed: 22 June 2017].

Dean, J. and Sharfman, M., 1996. Does Decision Process Matter? A Study of Strategic Decision-Making Effectiveness. *Academy of Management Journal*, 39(2), pp.368-396. [Online] Available at: Google Scholar [Accessed: 05 June 2017)].

De Vaus, D. A., 2001. Research Design in Social Research**.** London: SAGE.

Ester, M., Hans-Peter, K., Jiirg, S. and Xiaowei, X., 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Institute for Computer Science*, pp.226-231. [Online] Available at: <https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf> [Accessed: 22 June 2017].

Estivill-Castro, V.E. and Houle, M., 2001. Robust Distance-Based Clustering with Applications to Spatial Data Mining. *Journal of Algorithmica,* 30(2), pp.216-242. Available at: Google Scholar [Accessed: 05 June 2017].

Farzana, W., 2012. Consumers' Psychological factors association with brand equity of high-involvement product. *World journal of social sciences*, 2(5), p.90-101. [Online] Available at: <http://www.wbiaus.org/7.%20Wahida.pdf > [Accessed: 15 May 2017].

Festinger, L., 1957. *A Theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Fletcher, H., 2011. The 7 Best Uses for Predictive Analytics in Multichannel Marketing. *Target Marketing,* [Online] Available at: <http://www.targetmarketingmag.com/article/7-best-uses-predictive-analytics-modeling-multichannel-marketing/all> [Accessed: 18 November 2017].

French, N., 2015. Global Trends in Real Estate Finance: The Global Perspective. *Journal of Property Investment & Finance*, 33(1), pp.113-114. [Online] Available at: Emerald Insight [Accessed: 15 June 2017].

Garrido, M. M., Kelley, A. S., Paris, J., Roza, K., Meier, D. E., Morrison, R. S. and Aldridge, M. D., 2014. Methods for Constructing and Assessing Propensity Scores. *Journal of Health Services Research*, 49, pp.1701–1720. [Online] Available at: Google Scholar [Accessed: 23 Aug 2017].

Gelfand, A.E., Ecker, M.D., Knight, J.R. and Sirmans, C.F., 2004. The Dynamics of Location in Home Price. *Journal of Real Estate Finance and Economics*, 29, pp.149-166. [Online] Available at: EbscoHost [Accessed: 09-10-2017].

Ghauri, P.N. and Gronhaug, K., 2010. *Research methods in business studies: A practical guide*. England: Pearson Education.

Gilliam, D.A. and Voss, K., 2013. A proposed procedure for construct definition in marketing. *European Journal of Marketing*, 47(1-2), pp.5-26. [Online] Available at: Emerald Insight [Accessed: 23 June 2017].

Goodhope, O.O., 2013. Major Classic Consumer Buying Behaviour Models: Implications for Marketing Decision-Making. *Journal of Economics and Sustainable Development*, 4(4), pp.164-172. [Online] Available at: <http://iiste.org/Journals/index.php/JEDS/article/viewFile/4844/4922> [Accessed: 22 June 2017].

Gosling, S.D., Gifford, R. and McCunn, L., 2013. *The selection, creation, and perception of interior spaces: An environmental psychology approach*. Oxford: Berg.

Haddad, M., Judeh, M. and Haddad, S., 2011. Factors affecting buying behavior of an apartment and empirical investigation in Amman, Jordan. *Applied Sciences, Engineering and Technology*, 3(3), pp.234-239. [Online] Available at: Google Scholar [Accessed: 03 June 2017].

Harrell, F.E., 2015. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis.* United States: Springer Publishing.

Harris, R., 1998. *Introduction to Decision Making*. VirtualSalt, [Online]. Available at:

<http://www.virtualsalt.com/crebook5.htm> [Accessed: 05 June 2017].

Haupt, P., 2001. *The SAARF Universal Living Standards Measure (SU-LSM) - 12 years of continuous development.* [Online] Available at: <http://www.saarf.co.za/lsm/lsm-article.asp> [Accessed: 13 March 2017].

Hawkins, D.I., Mothersbaugh, D.L. and Mookerjee, A., 2011. *Consumer Behavior: Building Marketing Strategy.* 11th ed. New York: Tata McGraw-Hill.

Hawkins, D.I., Mothersbaugh, D.L. and Best, R.J., 2013. *Consumer Behavior: Building Marketing Strategy*. 12th ed. New York: Irwin McGraw-Hill.

Hoyer, W.D. and MacInnis, D.J., 2010. *Consumer behavior: A Marketers Tool.* 5th ed. London: Southwestern Cengage Learning.

Jones, R.P., Runyan, R.C. and Kim, Y.K., 2012. Moving single-brand retail shoppers from in-store to online: Is brand experience still important?. *Journal of Marketing at Retail*, 1(2), pp.17-30. [Online] Available at: Google Scholar [Accessed: 03 June 2017].

Kardes, F.R., CrOnnley, F.R. and Cline T.W., 2013. *Consumer Behaviour*, United State: Cengage Learning.

Kalechofsky, H., 2016. A Simple Framework for Building Predictive Models. *A Little Data Science Business Guide,* [Online] Available at: <http://www.msquared.com/wp-content/uploads/2017/01/A-Simple-Framework-for-Building-Predictive-Models.pdf> [Accessed: 18 Nov 2017].

Kalhori, S.R.N. and Zeng, X.J., 2014. Improvement the Accuracy of Six Applied Classification Algorithms through Integrated Supervised and Unsupervised Learning Approach. *Journal of Computer and Communications*, 2, pp.201-209. Available at: Emerald Insight [Accessed: 23 June 2017].

Kassel, M., (2016). *The Psychology of Buying and Selling a House*. [Online] Available at: < https://www.wsj.com/articles/the-psychology-of-buying-and-selling-a-house-1465783741> [Accessed: 10 March 2018].

Keita, L., 2012. Revealed Preference Theory, Rationality, and Neoclassical Economics, *Science* or Ideology Africa Development. *Council for the Development of Social Science Research in Africa,* 37(4), pp.73–116. [Online] Available at: <https://www.ajol.info/index.php/ad/article/viewFile/87486/77168> [Accessed: 02 June 2017].

Koklic, M. K. and Vida, I., 2009. A Strategic Household Purchase: Consumer House Buying Behavior. *Journal of Real Estate Practice and Education*, 7(1), pp.76-96. [Online] Available at: Emerald Insight [Accessed: 05 June 2017].

Kotler, P., 1965. Behavioral model for analyzing buyers. *Journal of marketing*, 20, pp.35- 45. [Online] Available at: Google Scholar [Accessed: 05 June 2017].

Kotler, P. and Armstrong, G., 2014. *Principles of Marketing*, 15th Ed, New York: Pearson.

Kotler, P., Armstrong, G. and Parment, A., 2012. *Principles of marketing*. Swedish Edition, Harlow: Prentice Hall.

Kotler, P. and Keller, K.L., 2012. *Marketing Management.* 14th Ed, Upper Saddle River: Prentice Hall.

Krishna, A., 2015. The Effect of Deal Knowledge on Consumer Purchase Behavior. *Journal of Marketing Research*, 31, pp.76-91. [Online] Available at: <https://ssrn.com/abstract=2552277> [Accessed: 02 June 2017].

Kryvobokov, M., 2007. What location attributes are the most important for market value?: Extraction of attributes from regression models. *Journal of Property Management*, 25 (3), pp.257-286. [Online] Available at: Emerald Insight [Access: 28 Aug 2017].

Kumar, R., 2014. *Research Methodology, A Step-by-step guide for beginners.* 4th Ed. California: Sage publications Ltd.

Lantos, G.P., 2015. *Consumer Behavior in action, real life application for marketers*. London: Routledge Taylor and Francis Group.

Ling, P., D'Alessandro, S. and Winzar, H., 2015. *Consumer behavior in action*. Victoria: Oxford University Press.

Loudon, D.L. and Della Bitta, A.J., 1993. *Consumer Behavior: Concepts and application Volume 1*. McGraw-Hill: Sydney.

Maclennan, D. and O'Sullivan, A., 2012. Housing markets, signals and search. *Journal of Property Research*, 29 (4), pp.324-34. [Online] Available at: Science Direct [Accessed: 17 June 2017].

Malik, S.U., 2012. Customer Satisfaction, Perceived Service Quality and Mediating Role of Perceived Value. *International Journal of Marketing Studies*, 4(1), pp. 69-73. [Online] Available at: Emerald Insight [Accessed: 27 June 2017].

Maslow, A.H., 1954. The Instinctoid Nature of Basic Needs1. *Journal of Personality*, 22, pp.326–347. [Online] Available at: Science Direct [Accessed: 17 June 2017].

Maslow, A. and Lowery, R., 1998. *Toward a Psychology of Being.* 3rd ed. New York: Wiley & Sons.

Marshall, A., 2009. *Principles, of Economics*. 8th ed. London: Macmillan.

McNeal, J.U., 1973. *Introduction to consumer Behaviour.* New York: John Wiley and Sons.

Mintzberg, H., 1979. *The Structuring of Organisations*. Englewood Cliffs, NJ: Prentice-Hall.

Mooya M.M., 2016. *Real Estate Markets and Neoclassical Economic Theory: A Heterodox Critique.* In: Real Estate Valuation Theory. Springer, Berlin, Heidelberg Moslehpour, M., Van Kien, P. and Danyfisla, I., 2014. Differences of customer purchase behavior toward organic rice in Indonesia and Taiwan. *International Journal of Quality and Service Sciences*, 6(4), pp.348 – 368. [Online] Available at: Emerald Insight [Accessed: 03 June 2017].

Mostert, P.G., 2002. *Buyer behaviour of South African Internet users*. Pretoria: University of Pretoria.

Nandamuri, P. P., 2012. Influence of Consumer Demographics on Attitude Towards Branded Products: An Exploratory Study on Consumer Durables in Rural Markets. *IUP Journal of Marketing Management*, 11 (3), pp.48-63. [Online] Available at: Business Source Complete [Accessed: 01 June 2017].

Nwanko, S., Hamelin, N. and Khaled, M., 2014. Consumer values, motivation and purchase intention for luxury goods. *Journal of Retailing and Consumer Services*, 21(5), pp.735–744. [Online] Available at: Sciencedirect [Accessed: 02 June 2017].

Oliver, R.L., 2015. *Satisfaction, A Behavioral perspective on the consumer*, 2nd ed. London: Routledge.

Opoku, R.A. and Abdul-Muhmin, A.G., 2010. Housing preferences and attribute importance among low-income consumers in Saudi Arabia. *Habitat International*, 34(2), pp.219-227. [Online] Available at: Google Scholar [Accessed: 03 June 2017].

Paul, A.L. and Hogan, S.K., 2015. On the couch understanding consumer shopping behavior. *Deloitte Insights,* [Online] Available at: <https://dupress.deloitte.com/dup-us-en/industry/retail-distribution/understanding-consumer-behavior-shopping-trends.html#endnote-4> (Accessed: 09 October 2017).

Pappu, R. and Quester, P.G., 2016. How does brand innovativeness affect brand loyalty?, *European Journal of Marketing*, 50 (2), pp.2-28. [Online] Available at: Emerald insight [Accessed 03 July 2017].

Peter, M. (2016) Creating value with Big Data analytics: making smarter marketing decisions, *International Journal of Market Research,* 58 (5), pp.761-764 [Online]. Available at: Ebsco Host (Accessed: 08 October 2016).

Pride, W.M. and Ferrell, O.C. 2011. Marketing. 16th ed. USA: Southwestern Cengage Learning.

Property24, 2017. *Property Values.* [Online] Available at <https://www.property24.com/property-values/kwazulu-natal/2> [Accessed: 24 August 2017].

Quester, P., Pettigrew, S. and Hawkins, D., 2011. *Consumer Behaviour –Implications for marketing strategy*. 6th Ed. McGraw-Hill: Sydney.

Saad, C., Bauer, B., 2011. The Model Analysis Framework An IDE for Static Model Analysis. *Journal of Industry Track of Software Language Engineering*, 4, [Online] Available at: Google Scholar [Accessed at: 24 August 2017].

Satchidananda, D., Bhabani, S.P.M., Nibir, N.B., 2012. Heuristic Frequent Term-Based Clustering of News Headlines. *Procedia Technology*, 6, pp.436-443. [Online] Available at: Science Direct [Accessed: 23 June 2017].

Schiffman, L.G., Hansen, H. and Kanuk, L.L., 2012. *Consumer Behaviour: A European Outlook*. USA: Financial Times/Prentice Hall.

Sent, E., 2004. The Legacy of Herbert Simon in Game Theory. *Journal of Economic Behavior and Organization*, 53, pp.303-317. [Online] Available at: Science Direct [Accessed: 02 June 2017].

Sheth, H.J, 1960. *Theory of Buyer Behaviour.* New York: John Wiley and Sons Inc.

Sidi, S. and Sharipah, N., 2011. Quality affordable housing: a theoretical framework for planning and design of quality housing. *Journal of Techno-Social,* 2(1). [Online] Available at: Emerald Insight [Accessed: 03 June 2017].

Soloman, M.R., 2013. *Consumer Behavior, Buying, Having, Being,* 3rd ed. Pearson.

Stevenson O. and Prout A., 2013. The psychology of Home Environments. *A call for research on Residential space*, 10, pp.135–157. [Online] Available at: Google Scholar [Accessed: 22 June 2017].

Sorescu, A., Frambach, R.T., Singh J., Rangaswamyd, A. and Bridges, C., 2011. Innovations in retail business models. *Journal of Retailing*, 87(1), pp. 3–16. [Online] Available at Emerald Insight [Accessed: 15 June 2017].

Sunil, 2015. Trends and practices of consumers buying online and offline: An analysis of factors influencing consumer’s buying. *International Journal of Commerce and Management*, 25(4), pp.442-455. [Online] Available at: Emerald Insight [Accessed: 02 June 2017].

Thi-Bich-Hanh, D., Khanh-Chuong, D. and Christel, V., 2017. Constrained clustering by constraint programming. *Artificial Intelligence*, 244, pp.70–94. Available at: Emerald Insight [Accessed: 23 June 2017].

Transaction Capital Credit Health, 2013. *What Does a Credit Score Mean?.* [Online] Available at: <https://www.credithealth.co.za/news-article/what-does-a-credit-score-mean/8/> [Accessed: 13 March 2017].

Trevino, A., 2016. *Introduction to K-means Clustering.* [Online] Available at: <www.datascience.com/blog/introduction-to-k-means-clustering-algorithm-learn-data-science-tutorials> [Accessed: 22 June 2017].

Verma, H.V., 2012. *Servicing marketing Text and cases*. India: Dorling Kindersley.

Vijayarani, S. and Jothi, P., 2014. Partitioning Clustering Algorithms for Data Stream Outlier Detection. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(4), pp.3975-3980. [Online] Available at: <https://pdfs.semanticscholar.org/c246/a01200ccf253d748c67a4176faed89ea687d.pdf >[Accessed: 22 June 2017].

Weintraub, E., 2017. *Reasons why Home Owners Sell.* [Online] Available at: https://www.thebalance.com/why-home-owners-sell-1799021 [Accessed: 10 March 2018].

Williams, T.G., 1982. *Consumer Behaviour: Fundamentals and Strategies.* New York: West Publishing Company.

Wu, K.S. and Teng, Y.M., 2011. Applying the extended theory of planned behavior to predict the intention of visiting a green hotel. *African Journal of Business Management*, 5(17), pp.7579-7587. [Online] Available at: Google Scholar [Accessed: 03 June 2017].

Wyman, D., Seldin, M. and Worzala, E., 2011. A new paradigm for real estate valuation?, *Journal of Property Investment & Finance*, 29 (4/5), pp.341-358. [Online] Available at Emerald Insight [Accessed: 15 June 2017].

Yin, R. K.. 2014. *Case Study Research Design and Method.* 5th ed. California: Sage Publications Ltd.

Ying, Z. and Karypis, G., 2005. Hierarchical Clustering Algorithms for Document Datasets. *Data Mining and Knowledge Discovery*. 10(2), pp.141-168. Available at: Emerald Insight [Accessed: 15 June 2017].

Zhang, T., Ramakrishnan, R., Birch, M.L., 1996. An efficient data clustering method for very large databases. *Paper of SIGMOD*, pp.103–114. [Online] Available at: <https://www.cs.sfu.ca/CourseCentral/459/han/papers/zhang96.pdf > [Accessed: 22 June 2017].

# APPENDIX 1: GATEKEEPER'S LETTER

metonymy

16 February 2016

To Whom it may Concern

**Approval to conduct research on the Metonymy Business Data Profiling program**

This letter serves to confirm our approval and agreement to support and participate in the research to be conducted by Mr HM Naidoo (Student number 961045758).

The topic, as proposed, is:

***Convert predictive models into business intelligence that drives strategy within the Real Estate Industry.***

We hereby grant access/ permission to the following data sources for the purpose of data aggregation, subject to consumer confidentiality being maintained:

1. Access to the Home Affairs data source excluding access to personal contact information.
2. Access to Deeds Data
3. Access to limited Credit Bureau information for the purpose of scoring individuals behaviour.

Upon completion of the study, a bound copy of the full research must be made available to Metonymy (PTY) Ltd.

[redacted] the undersigned.

Intelligence Based Profiling Solutions | www.metonymy.co.za

Blake House, 32 Flanders Drive, Mount Edgecombe, 4302, KwaZulu Natal, South Africa | Private Bag X27, Umhlanga Rocks, 4320
Tel: +27 31 538 4087 Fax: +27 31 538 4257 Email: info@metonymy.co.za

# APPENDIX 2: ETHICAL CLEARANCE

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

18 October 2017

**Mr Habendra Mischa Naidoo (961045758)**
**Graduate School of Business & Leadership**
**Westville Campus**

Dear Mr Naidoo,

**Protocol reference number: HSS/0169/017M**
**New project title:** Develop a Predictive Model that drives Business Strategy to determine Property Sales within the Real Estate Industry based in KwaZulu-Natal

**Approval Notification – Amendment Application**

This letter serves to notify you that your application and request for an amendment received on 05 September 2017 has now been approved as follows:

- Change in Title

**Any alterations to the approved research protocol i.e. Questionnaire/Interview Schedule, Informed Consent Form; Title of the Project, Location of the Study must be reviewed and approved through an amendment /modification prior to its implementation. In case you have further queries, please quote the above reference number.**

**PLEASE NOTE: Research data should be securely stored in the discipline/department for a period of 5 years.**

**The ethical clearance certificate is only valid for period of 3 years from the date of original issue. Thereafter Recertification must be applied for on an annual basis.**

Best wishes for the successful completion of your research protocol.

[redacted] **(Chair)**

**/ms**

Cc Supervisor: Dr Muhammad Hoque
Cc Academic Leader Research: Dr Emmanuel Mutambara
Cc School Administrator: Ms Zarina Bullyraj

**Humanities & Social Sciences Research Ethics Committee**
**Dr Shenuka Singh (Chair)**
**Westville Campus, Govan Mbeki Building**
**Postal Address:** Private Bag X54001, Durban 4000
**Telephone:** +27 (0) 31 260 3587/8350/4557 **Facsimile:** +27 (0) 31 260 4609 **Email:** ximbap@ukzn.ac.za / snymanm@ukzn.ac.za / mohunp@ukzn.ac.za
**Website:** www.ukzn.ac.za

1910 - 2010
100 YEARS OF ACADEMIC EXCELLENCE

Founding Campuses: Edgewood, Howard College, Medical School, Pietermaritzburg, Westville

## APPENDIX 3: TURNITIN REPORT

turnitin

# Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

| | |
|---|---|
| Submission author: | Mischa Naidoo |
| Assignment title: | Dissertation for turnuitin 2017 |
| Submission title: | 961045758 MBA Thesis 2017 TII |
| File name: | 961045758-_MBA_Thesis_Turnitin.. |
| File size: | 1.32M |
| Page count: | 97 |
| Word count: | 23,563 |
| Character count: | 131,943 |
| Submission date: | 22-Nov-2017 11:35PM (UTC+0200) |
| Submission ID: | 884259928 |

## 961045758 MBA Thesis 2017 TII

ORIGINALITY REPORT

| 5% | 4% | 0% | 1% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES