

Wilfrid Laurier University

Scholars Commons @ Laurier

Theses and Dissertations (Comprehensive)

2019

Spatial Big Data Analytics: The New Boundaries of Retail Location Decision-Making

Joseph M. Aversa

Wilfrid Laurier University, aver4310@mylaurier.ca

Follow this and additional works at: <https://scholars.wlu.ca/etd>



Part of the [Business Analytics Commons](#), and the [Real Estate Commons](#)

Recommended Citation

Aversa, Joseph M., "Spatial Big Data Analytics: The New Boundaries of Retail Location Decision-Making" (2019). *Theses and Dissertations (Comprehensive)*. 2138.

<https://scholars.wlu.ca/etd/2138>

This Dissertation is brought to you for free and open access by Scholars Commons @ Laurier. It has been accepted for inclusion in Theses and Dissertations (Comprehensive) by an authorized administrator of Scholars Commons @ Laurier. For more information, please contact scholarscommons@wlu.ca.

Spatial Big Data Analytics: The New Boundaries of Retail Location Decision-Making

DISSERTATION

Joseph Mattia Jr Aversa 2018

Submitted to the Department of Geography and Environmental Studies, Faculty of Arts
in partial fulfillment of the requirements for
Doctor of Philosophy in Geography
Wilfrid Laurier University

© Joseph Mattia Jr Aversa 2019

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis *may* be made electronically available to the public.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Dr. Sean Doherty and Dr. Tony Hernandez for their guidance and support throughout this process. The completion of this thesis would not have been possible without my wife (Tania), my children (Rosalie and Gemma) and my parents. Thank you for your constant encouragement and support.

ABSTRACT

This dissertation examines the current state and evolution of retail location decision-making (RLDM) in Canada. The major objectives are: (i) To explore the type and scale of location decisions that retail firms are currently undertaking; (ii) To identify the availability and use of technology and Spatial Big Data (SBD) within the decision-making process; (iii) To identify the awareness, availability, use, adoption and development of SBD; and, (iv) To assess the implications of SBD in RLDM. These objectives were investigated by using a three stage multi-method research process. First, an online survey of retail location decision makers across a range of sizes and sub-sectors was administered. Secondly, structured interviews were conducted with 24 retail location decision makers, and lastly, three in-depth cases studies were undertaken in order to highlight the changes to RLDM over the last decade and to develop a deeper understanding of RLDM.

This dissertation found that within the last decade RLDM changed in three main ways: (i) There has been an increase in the availability and use of technology and SBD within the decision-making process; (ii) The type and scale of location decisions that a firm undertakes remain relatively unchanged even with the growth of new data; and, (iii) The range of location research methods that are employed within retail firms is only just beginning to change given the presence of new data sources and data analytics technology.

Traditional practices still dominate the RLDM process. While the adoption of SBD applications is starting to appear within retail planning, they are not widespread. Traditional data sources, such as those highlighted in past studies by Hernandez and Emmons (2012) and Byrom *et al.* (2001) are still the most commonly used data sources. It was evident that at the

heart of SBD adoption is a data environment that promotes transparency and a clear corporate strategy. While most retailers are aware of the new SBD techniques that exist, they are not often adopted and routinized

1 TABLE OF CONTENTS

Author's Declaration	ii
Acknowledgements.....	iii
Abstract	iv
List of Figures	ix
List of Tables	x
1 Introduction	1
1.1 Research Aim and Objectives.....	4
1.2 Structure of Dissertation.....	5
2 Research Context	6
2.1 History of Retail Location Decision-Making	6
2.1.1 Wave 1 of Retail Location Decision Making.....	7
2.1.2 Wave 2 of Retail Location Decision Making:.....	16
2.1.3 Wave 3 of Retail Location Decision Making:.....	23
2.1.4 Wave 4 of Retail Location Decision Making:.....	27
2.2 Emerging Applications	41
2.2.1 Tracking Technologies.....	42
2.3 Innovation and technological Assimilation and Adoption	43
3 Research Design and Methodology	50
3.1.1 Questionnaire Design and Format	51
3.1.2 Sampling: Selecting Questionnaire Participants	53
3.1.3 Pre-Testing, Distribution and Maximizing Response Rate	54
3.2 Semi-Structured Interviews	55
3.2.1 Sampling: Selecting Participants for Semi-Structured interviews	57
3.2.2 Interview Design & Practices	59
3.2.3 Interview Data Coding and Analysis Techniques	62
3.3 Case Studies	64
3.3.1 Designing Case Studies.....	65
3.3.2 Selecting the Cases.....	67
4 Research Findings	69
4.1 Online Retailer Survey Results	69
4.1.1 Spatial Big Data Adoption by Retailers: Types, Importance, Perceptions, and Changes....	70

4.1.2	Techniques and Methods Adoptions by Retailers: Types, Importance, Perceptions and Changes	80
4.1.3	Business Culture: Opportunities and Challenges for Awareness, Availability, Use and Adoption of Spatial Big Data	88
4.2	Structured Interviews	90
4.2.1	Data Theme	90
4.2.2	Data Requirements for Method and Technique Implementation	95
4.2.3	Organizational Culture	102
4.3	Case studies	106
4.3.1	Case 1 – Developer/Brokerage/Leasing Company.....	106
4.3.2	Data and Technology Environment.....	108
4.3.3	Retail Location Decision-Making.....	111
4.4	Case 2 – Food Services	118
4.4.1	Data and Technology Environment.....	118
4.4.2	Retail Location Decision-Making.....	121
4.5	Case Study 3 – Financial Sector.....	126
4.5.1	Data and Technology Environment.....	128
4.5.2	Retail Location Decision-Making.....	131
4.6	Case Study Summary.....	136
5	Discussion.....	138
5.1	Online Questionnaire	138
5.2	Structured Interviews and Case Studies	140
5.2.1	Data Warehousing and Mining/Data Integration and Analysis/Data Interpretation	140
5.2.2	Corporate Buy-In.....	143
5.2.3	Compatibility with Current Systems	146
5.3	Best Practices	148
5.3.1	Integrated Approach – Best Practice	149
5.3.2	Reflexive Approach	157
5.3.3	Systemic Approach.....	159
5.4	A Framework for SBD Assimilation and Adoption in RLDM.....	161
6	Conclusion.....	166
6.1	Summary of Findings.....	167
6.2	Limitations.....	168
6.3	Future Research	169

References	171
------------	-----

LIST OF FIGURES

Figure 1: Utilization of analytical techniques in Retail Decision Making	22
Figure 2: Technology, Organization, and Environment Framework (TOE)	46
Figure 3: Integrated Approach: Ideal Corporate Data Environment for RLDM	150
Figure 4: Data Science Team	153
Figure 5: Reflexive Approach	158
Figure 6: Systemic Approach	161
Figure 7: Framework for SBD Assimilation	162
Figure 8: The Antecedents of SBD Assimilation Conclusion	165

LIST OF TABLES

Table 1: Trade Area Analysis Techniques	18
Table 2: Location Techniques by Usage (% respondents)	23
Table 3: Retail Channels	24
Table 4: Key Data Science Terms	29
Table 5: Potential Applications of Big Data	31
Table 6: Difference between Data Lakes and Data Warehouses	35
Table 7: RLDM Research Studies.....	40
Table 8: Time Geography Concepts	42
Table 9: A Summary of Key Research using Innovation Assimilation Theories.	45
Table 10: Technology, Organization, and Environment Framework (TOE)	45
Table 11: Innovation-Decision Process	47
Table 12: Attribute to Predict an Innovation Rate of Adoption.	47
Table 13: Roger's Adoption Categories	48
Table 14: Types of Content Questions Used in the Online Questionnaire.....	52
Table 15: Pre-Testing Purpose	54
Table 16: Evaluation of Different Interview Types	56
Table 17: Interview Respondents	59
Table 18: Interview Question Types	61
Table 19: Coding Categories.....	63
Table 20: Types of Case Study in Qualitative Research	66
Table 21: Distribution of Respondents by Sector, On-line Survey	69
Table 22: Retailers Frequently Used Data Sources (n=35), Online Survey	72
Table 23: Attitudinal Questions of the Importance of Geographic Data (n=29), Online Survey	74
Table 24: Byrom et.al, 2001 Attitudinal Statements: Geographical Data Issues, Online Survey	75
Table 25: Perceptions on External Spatial Big Data, On-line Survey	77
Table 26: Perceptions on Internal Spatial Big Data, On-line Survey	78
Table 27: How Spatial Big Data Usage has Changed in the Past Few Years (n=28), On-line Survey	80
Table 28: Changes in the Number of Physical Locations Counts over the last 5 years (n=34), On-line Survey	81
Table 29: Expected Future Change to Physical Locations Counts over the next 5 years (n=34), On-line Survey	81
Table 30: The Nature of Data Management Techniques and Strategies (n=29), On-line Survey	83
Table 31: Emerging Location Decision Making Techniques Definitions	85
Table 32: Retail Location Decision-making Techniques (n=28), Online Survey	86
Table 33: Utilization of Location-Based Applications on a Regular Basis (n=36), On-line Survey	87
Table 34: Difficulties with Data Integration in Retail Location Decision-making (n=27), On-line Survey	89
Table 35: Senior Management's role in Retail Location Decision-Making, On-line Survey	90
Table 36: Data Generating Methods	98
Table 37: Data Sources Frequently Used	109
Table 38: Corporate Strategies for increasing Sales and Customer Experience	113
Table 39: Data Sources Frequently Used by Spatial Research.....	120

Table 40: Support Structure	122
Table 41: Data Sources Frequently Used by Physical Distribution Planning.....	131

1 INTRODUCTION

The process of examining spatial features and their relationships have long contributed unique solutions to diverse problems, centred on the complex relationships between humans and their surrounding environments. Understanding and explaining these complex spatial arrangements are essential when making retail location decisions. Retailers are moving towards greater reliance on data driven decision-making by incorporating infrastructure that allows for the expansion of both the type and amount of data it works with, known as Big Data. Advancements in Big Data and Big Data Technologies have provided retailers an ability to obtain more granular level information on consumer behavior to support retail location decision-making (RLDM)

While studying the economic importance of location is not a new phenomenon (Jones and Simmons, 1987; Ghosh and McLafferty, 1987), the increasingly competitive retail marketplace and advances in information technology and e-commerce have placed added pressure on businesses to improve their understanding of the spatial nature of individual activity patterns (Cheng and Yu, 2005; Rogers, 2007). With paradigm changes in consumer behaviour being facilitated through the introduction of e-retail, questions begin to arise around the use of retail space and the need for space. Given the high risk nature of retail location investment RLDM research has a long history of focusing on appropriate retail location strategies for major retail organizations (Simkin et al., 1985; Clark and Wrigley, 1997; Hernandez et al., 1998; Hernandez et.al., 2004; Hernandez and Biasiotto, 2001; Reigadinha et al., 2017; Woods and Reynolds, 2012; Chacón-García, 2017; ELSamen and Hiyasat, 2017; Byrom, et al., 2001; Clarkson et al., 1996; Handa & Vohra, 2010; Theodoridis & Bennison,

2009). Techniques such as customer mapping, site selection, sales forecasting and other retail location portfolio management applications have been central in order to support RLDM. With advances in SBD there is a need to evaluate the role that new technologies play for retail organizations.

Store location decisions are frequently considered to be the single most important choice a retail business makes (Hernandez et al., 1998; Hernandez et.al., 2004). Retailers' have a variety of techniques at their disposal in order to help minimize the risks associated with retail decision making. Traditional approaches to location have been primarily based on the intuition of the decision maker (Hernandez et al., 1998). Due to companies committing both time and money into the RLDM process, the losses that a company can accumulate through poor decision support processes cannot be easily recovered (Simkin et al., 1985; Clark and Wrigley, 1997). These risks, in turn, place added pressure on retailers to develop rigorous strategic location decision-making practices that are aimed to create sustainable long-term success (Karande & Lombard, 2005).

An integral part of informing the decision-making process is the application of data analytics (the techniques, technologies, systems, methodologies, and applications that analyze critical business data) used to help businesses better understand their market and make timely business decisions (Chen, et. al, 2012). Given the ever-increasing scale of investment and subsequent increase in the risks associated with retail location decisions, it is not surprising that the use of data in RLDM has become a growing area of research interest. Big Data and associated analytics, such as, Social Media Data, Customer Surveillance Data, Predictive Analytics, Machine Learning, Social Media Analytics, Social Influence Analysis, Real Time Data

Visualization, Real-Time Demand Forecasts and Text Analytics, promise to offer comprehensive spatio-temporal data that can provide an increased level of sophistication and a more detailed understanding of consumer behaviour. This marks a new era of retail decision-making centered on data-driven decisions that will challenge the boundaries and nature of retail location decision-support. It may result in a transition from old and relatively simple theories and models to more complex methodologies (Kitchin, 2013; DeLyser and Sui, 2014; Goodchild, 2013).

Despite the rhetoric 'Big Data' little is known about how these new forms of data are truly leveraged, especially the new challenges presented by SBD and its role in retail locational decision-making (RLDM). Many key questions remain. For instance, are retailers merely collecting and attempting to use what is known as SBD because of a vague sense of its importance and/or the hype surrounding the burgeoning data science industry? Is there industry awareness, use and development of SBD practices? Is it possible to identify retailers that have realized the potential of SBD within their location decision-making activities? Are these new practices offering better insights? By investigating these questions we will be able to identify whether we are experiencing a revolution in RLDM or more simply just an evolution of previous practices. By developing an understanding of the impact that the so called 'big data revolution' (the new wave of RLDM) will have on retail geography we will provide a means for retail geographers to identify the opportunities and challenges related to the fusion of geography and big data within RLDM.

1.1 RESEARCH AIM AND OBJECTIVES

The aim of this doctoral research is to build on the theoretical and empirical foundations of RLDM studies within the field of retail geography. Emphasis will be placed on understanding the new wave of RLDM and its effects on the retail landscape. The main objectives are:

- i. To explore the type and scale of location decisions that retail firms are currently undertaking
- ii. To identify the availability and use of technology and SBD within the decision-making process
- iii. To identify the assimilation of SBD practices within retail firms
- iv. To create a conceptual framework outlining best practices for SBD adoption

This dissertation looks to bridge the gap in our understanding through exploring how RLDM and decision support environments have evolved in recent years, including the use of technology, SBD and data science-related methods. Overall, the intent is to provide a timely snapshot of location-based decision support activities amongst the leading retail and service firms operating in Canada and to critically assess the awareness, adoption, use and development of Big Data and associated data science approaches.

1.2 STRUCTURE OF DISSERTATION

The dissertation is divided into six chapters. Chapter Two outlines: (i) the history of RLDM and the theoretical approaches, models and data sources that are utilized by retail organizations when making decisions; (ii) the importance of understanding consumer behavior for retail planning purposes and the role that SBD plays in offering the potential for more granular level customer data; (iii) the literature that investigates the process of adopting new innovations and technologies within organizations.

Chapter Three discusses the methodology used to address the research objectives, including: (i) An online Questionnaire (ii) Semi-Structured Interviews (iii) Case Studies.

Chapter Four reports the findings from the three methods of data collection. Chapter Five interprets and describes the significance of the findings as it relates to existing literature. It further explains new understanding of how RLDM has evolved over the past decade. Furthermore best practices are identified and a conceptual framework is outlined to explain how SBD is adopted in RLDM. Finally, Chapter Six presents conclusions, limitations and future opportunities for research.

2 RESEARCH CONTEXT

2.1 HISTORY OF RETAIL LOCATION DECISION-MAKING

Retail location decision-making (RLDM) does not just include the process of locating new stores. It also includes the acquisition of stores and operating divisions, store development of both established formats and new formats, retail format and store closures, and the management of existing portfolios through refurbishments, relocations, re-fascias, and extensions (Byrom, et al., 2001; Hernandez, et al., 1998; Reynolds & Wood, 2010). Given the breadth of RLDM, it is common for retail organizations to use multiple methods and data sources when making decisions; no one method can be applied universally in every situation. Retail location decision-makers often leverage numerous methods in order to minimize weaknesses that may exist in any one technique (Ghosh & McLafferty, 1987).

Four distinct waves of RLDM can be identified. Each wave has a distinct set of theoretical approaches, models and data sources that leverage into the decision making process. The first wave (pre-1980's) is one that was primarily driven by ad-hoc decisions that were rooted in the experiences and institutional knowledge held by the decision makers. The second wave (beginning in the 1980's- 1990's) began with the introduction of Geographic Information Systems (GIS) technologies that allowed for easier implementation of more sophisticated forms of RLDM. This, in turn, created more formalized methods that began to displace ad-hoc decision-making related to location planning. It was during this time that the spatial modelling applications, developed decades earlier, saw a resurgence in their application within RLDM. This wave was largely defined by the introduction of individual licenses of GIS

software. The third wave 2000 to 2010 was defined by enterprise adoption of GIS and data-driven decision-making. The fourth and current wave (2010 to present day) is one that has largely been driven by new data sources known as Big Data. These types of data are characterized by the speed (velocity) in which they are collected, their large volume, as well as the variety of data sources available. This current wave focuses on new data mining and model optimization brought about by the development of Big Data. This wave is the most speculative in nature as the extent of new methodologies, theories, and applications of such detailed datasets is largely unknown within the retail sector.

2.1.1 Wave 1 of Retail Location Decision Making

Early RLDM research was largely rooted around four main theories: Central Place Theory (Christaller, 1933), Spatial Interaction models (Reilly, 1929; Huff, 1963), Bid Rent theory (Alonso, 1960), and Principle of minimum differentiation (Working and Hotelling, 1929). These theories are seen as the “four cornerstones” (Brown, 1993) to retail location theory, and as such, they have received considerable academic attention over the past century. Central Place theory (CPT) and Spatial Interaction Models (SIM) really focus on consumer spatial behavior while the Bid Rent Theory (BRT) and Principle of Minimum Market Differentiation (PMD) discuss strategies for identifying ideal retail locations. CPT and SIM provide valuable insight in the reactions of consumers towards selecting shopping destinations. By understanding these theories retailers are able to form better location decisions. These theories are outlined in terms of their ability to inform RLDM, even though several of these models were never formally employed within the retail sector. Although these were not regularly deployed formally, the

ideas and explanations that they provide about spatial arrangements were likely the driving forces that informed the prevailing ad hoc decision-making process at that time.

According to CPT, the primary purpose of a settlement is the delivery of goods and services for the surrounding market area; therefore, towns are centrally located and are referred to as central places (Jones & Simmons, 1993). These central places are attractive places for retail locations. Settlements that provide a larger degree or variation of goods and services are called higher-order central places. Lower-order central places have small market areas and provide goods and services that are purchased more frequently than higher-order goods and services. There are fewer higher-order places than lower-order places. This theory makes several assumptions about consumers' travel behaviour. It assumes that all consumers are identical, that the population is evenly distributed, and it assumes that there is an uninterrupted travel surface. This theory also relies on the concept that consumers will shop at the nearest centre and they will only participate in single purpose shopping trips (Clarke, 1968; Eaton & Lipsey, 1982). Ghosh and McLafferty (1987), went on to identified that multipurpose trips were viewed as more efficient as they reduce cost and time associated with travelling to a shopping destination. While both Christaller and Losch acknowledged the notion of multi-purpose trips, they did not make any formal attempt to account for these types of activities within their model (Brown, 1993).

CPT identifies that in an effort to minimize transportation costs, consumers will purchase goods and services from retail facilities that are closest to them (Christaller, 1933; Berry, 1964). This theory focuses on two basic concepts: market threshold and market range. Market threshold refers to the minimum population that is required for a good or service to be sold; while market

range refers to the maximum distance people will travel to purchase goods and services (Berry & Garrison, 1958). If range is greater than threshold, a business will be economically viable. This theory also makes two assumptions about human behaviour: First, Christaller stated that humans will always purchase goods from the closest place that offers the good, and second, if the demand for a good or service is high, it will be offered in close proximity to the population (Brown, 1993). If demand is reduced, the availability of the good is reduced as well.

Understanding market threshold and range allow retailers to make better location decisions as it helps to ensure that the retail location decision that a firm makes is in line with the distance their target market is willing to travel to acquire their goods.

In the 1950's there was a quantitative revolution in geography which resulted in a paradigm shift resulting in the development of new methodologies within the discipline of geography. One of these models was the SIM. SIM are based on the notion that consumers will consider the utility of retail location against the disincentive nature of distance when choosing a retail destination. These models offered an alternative to the previously discussed approach as they explain that the nearest offering is not the only factor considered when choosing a good or service (Clarke, 1998; Clarkson et al., 1996; Karande & Lombard, 2005). Gravity models have long been shown to be reliable in estimating the spatial boundaries of retail trade areas (Huff, 1963; Huff, 1964; Pooler, 1994). William A. Reilly was the first person to apply the Newtonian concept of gravity in physics to retail trade area analysis (1929). Reilly's (1929) model suggests two cities will attract retail trade from an intermediate town in direct proportion to the population of the two cities, and in an inverse proportion to the square of the distances from the two cities to the intermediate town. The purpose of Reilly's model was to

determine the relative retail “pulling power” of the two competing cities on a third town or city. Reilly’s work was then reformulated by Converse (1949) to determine the breaking point between the trade areas of two competing centres. Converse’s modification made it possible to calculate the approximate point at which two competing cities had equal trading influence. This modification has been used extensively to estimate trade areas of proposed shopping centres where the square footage of each shopping centre was substituted for attractiveness and travel time between centres for distance impedance (Ellwood, 1954). The value of these models depends on the ability to incorporate a variety of different measures for store attractiveness. By only including the size of a store and a measure of distance, it is difficult to explain changes in consumer interactions if a retail outlet adds attractions, such as, entertainment (movie theaters, restaurants, or an improved ambiance.

Widely considered the single most significant contributor to SIM, Huff’s ‘consumers choice model’ (1963, 1964) identified that consumers frequent competing shopping centres on the basis of their overall 'utility' or usefulness. This concept included a variety of variables that were associated with attractiveness as well as deterrents, such as travel times and competition (Stanley and Sewall; 1976). Therefore, each shopping centre had an estimated probability of being frequented by any individual that was directly related to its attractiveness, inversely related to its distance from the consumer and inversely related to the utility of competing centres (Brown, 1993). The Huff model (Huff, 1963) identifies that the attraction of a retail location can be expressed as the probability of a consumer in one market shopping at a specific retail location. The model identified this probability by identifying the square footage of the retail selling area of a shopping location, the sensitivity of changes in shopping probability to

changes in a selling area, the drive time between an area and the retail location, and the sensitivity of changes in shopping probability to changes in driving time (Jones & Simmons, 1993). Distance and attractiveness clearly act as proxies for the framework of CPT (Christaller, 1933) which were the value or importance of a retail facility and economic distance (Wee & Pearce; 1985). Due to consumer spatial behaviour being a result of many complex interactions, there have been several modifications to this probabilistic model. Examples of this include the introduction of image factors such as cleanliness, location, prices, friendliness, and variety (Stanley and Sewall; 1976; Wee and Pearce, 1985; Del Gatto and Mastinu, 2018; Sevtsuk and Kalvo, 2018). Jones and Simmons (1993) identified that the strength in these models depends on the ability to incorporate a variety of different indicators associated with store attractiveness.

Bid rent refers to the amount of rent an individual or company is willing to pay for a more central location, or the willingness to locate further from the central area in favour of lower rent (Narvaez, et al., 2013). BRT states that land users all compete for the most accessible land which is usually the centre of the city or central business district (CBD). The theory states that all businesses want to leverage the highly accessible sources of labour and customers that the CBD offers (Alonso, 1960). Therefore, the uses for any particular piece of land would usually go to the business that is willing to pay the most for it. Alonso (1960) stated that the amount businesses would be willing to bid would be based on their ability to attract the largest number of customers. In terms of retail, department stores and other large chained stores would be willing to pay the greatest amount to locate in the CBD. This was evident pre-WWII as the

development of the electric streetcar increased the level of accessibility within the downtown cores, acting as a catalyst for the development of retail strips within these central areas.

The Principle of Minimum Differentiation (PMD) or Hotelling's Law (Working & Hotelling, 1929) refers to the tendency of businesses or products to cluster in space. He observed that consumers prefer the nearest of two options when purchasing fixed priced goods with identical features. This idea forces both firms to locate in the middle of the market in order to eliminate the potential for its competitors to gain any additional portion of a market. Hotelling identified this process of retail clustering as socially wasteful (Brown, 1993); future adaptations of this model (Eaton & Lipsey, 1979) saw the clustering of retail as a result of retailers and consumers reducing uncertainty. Therefore, actions such as comparison shopping and multipurpose shopping are seen to be socially optimal for consumers while minimizing risk for retailers (Klemperer, 1992). Other research (Nelson et al., 1958) identified the Rule of Retail Compatibility, which states that "any two compatible businesses, in close proximity, will show an increase in business volume directly proportionate to the incidence of consumer interchange between them" (Nelson 1958 p. 66 as cited in Brown, 1993). The PMD model has been applied in a variety of applications including product development (Brown, 1993) and facility location problems (Gastner, 2011), often at micro-scale or intra-centre location (Brown, 1993). There have been a number of empirical studies that support the PMD's theory that sellers of the same or similar categories of merchandise tend to cluster closely together. Several researchers (Brown, 1993; Clarkson, et al., 1996) have identified that high-order retail categories, such as women's clothing stores and department stores, experience the most amount of clustering. On

the contrary, low-order retail sectors such as convenience stores and personal services, experience minimal clustering.

Before the 1980s, retail location techniques tended to favour ad-hoc approaches, with strong reliance on institutional knowledge and experience, despite the development of more formalized model-based approaches (Simkin et al., 1985). A big part of this had to do with a lack of data sources that were readily available as well as technologies with data handling capabilities. While some early work (Applebaum, 1966; Davies, 1977) outlined a selection of spatial analytical techniques that could be utilized, the majority of the RLDM processes were centered on instinct and intuition (Clarke, 1998). 'Gut-feel' approaches, tended to be highly subjective and inexpensive to operate (Hernandez et al., 1998). Due to its highly subjective nature, there is very little research that truly investigates the reliability or success of 'experience' as a viable site selection technique (Wood and Tasker, 2008). Simkin et al. (1985) further identified that while SIM existed (e.g., Reilly, 1931; Converse, 1949; Huff, 1963), many retailers considered these approaches to be complicated and onerous as they required extensive amounts of data, as well as rigorous hardware, software and personnel requirements (i.e., expertise needed to calibrate the models). While the institutional and experiential knowledge that retail decision makers hold are invaluable, Clarke (1998) outlines that these type of approaches have an apparent downside. First, this approach is highly subjective and relies heavily on the experience of the individual decision maker(s). Second, it can be very time consuming and expensive to visit the potential locations, especially when dealing with opening a network of stores (Clarke, 1998; Wood and Tasker, 2008).

Checklist analysis was common practice pre-1980 and has remained a popular technique (Clarke, 1998; Davies, 1977; Stanley & Sewall, 1976). The checklist method consists of evaluating a simple list of factors that are seen to be directly related to sales (Clarke, 1998) and can include variables such as population characteristics, accessibility, visibility, traffic flows, competition and cost (Stanley & Sewall, 1976). The variable selection process is generally an informal process as the location conditions are based on the decision maker's judgment or their experience evaluating the importance of these items (Stanley & Sewall, 1976). Several researchers (Simkin, 1990; Clarke, 1998; Hernandez & Bennison, 2000; Simkin et al., 1985) have identified the reliance of retailers on checklist and intuition-based approaches.

Although more formalized retail decision-making methods had been developed, retail decision-makers continued to rely heavily on ad-hoc methods mainly due to business culture, cost and time. To a lesser degree, analogue techniques were also used during this era of RLDM. These methods attempt to predict potential store success, often expressed as sales, by drawing comparisons with profitable stores within their corporate store network. While initial data and software requirements were not too onerous, analogue based methods can be dated back to the 1960's (Applebaum, 1966). Primary data sources for these models include customer surveys, competitive analysis and the collection of demographic and socio-economic information. Post 1980 these analogue based methods incorporate Geographic Information Systems (GIS) and loyalty cards (Clarkson et. al, 1996). The initial analogue models focused on the study of existing retail stores to identify potential future retail sites. Customers of these existing stores were interviewed to determine where they lived, allowing Applebaum to define primary trade areas for these stores. Several techniques (Rogers, 1984; and Applebaum, 1966)

have been developed based on constructing a formalized set of performance indicators in order to screen potential location options (Hernandez & Bennison, 2000).

The analogue method has also been a catalyst for the development of more advanced methods of site selection techniques. Multiple regression models are an example of this, as they use existing stores to determine the key variables that affect sales in order to forecast new development potential (Rogers, 1997 as cited in Hernandez & Bennison, 2000; Mendes and Themido, 2004; Chang and Hsieh, 2018). These models are generally tested and calibrated across a number of location scenarios to establish a benchmark tool for future development (Hernandez & Bennison, 2000). The use of multiple regression in location-based analysis has been well documented in branch-level banking location decision making (Simkin, 1990; Boufounou, 1995; Clawson, 2010; Doyle, et al., 1979; Simkin et al., 1985), specifically in planning new locations, evaluating performance, and providing marketing support (Boufounou, 1995). These models are generally analogous, using the company's historical data in order to determine potential success based on the demographic and socio-economic characteristics of the surrounding market. While these methods allow for some statistical rigor within the decision-making process, they do have several limitations. Primarily, the weaknesses are experienced because these models perform site evaluations in isolation without incorporating the effects of the company's entire location network and more importantly, competition. There is no gauge for cannibalization or competition between retail outlets (Clarke, 1998). Furthermore, these techniques present questions on whether or not location options or markets are ever genuinely homogeneous in nature. Rogers and Green (1979) identified that the selection of homogeneous analogue groupings by a retail location decision-maker may be

judgmental and arbitrary. Furthermore the selection of appropriate analogues for forecasting sales at a new site is dependent on the decision-makers understanding of the variables that produce fluctuations in levels of market penetration or of overall store sales (Rogers and Green, 1979).

2.1.2 Wave 2 of Retail Location Decision Making:

With the growth of information technologies through the 1980's and 1990's, RLDM began to experience growth in model sophistication. These new developments were primarily a result of growth in Geographic Information Systems (GIS) and the ability for these systems to process and manage spatial data, as well as increased data volumes. The benefits of GIS for site acquisition and store location research were discussed by Clark and Rowley (1995). Specifically they noted that GIS, at the very least, offered advanced visualization capabilities for spatially referenced information. This notation ultimately gave researchers an effective way to draw comparisons and to explore data on alternative locations (Clarke & Rowley, 1995). By the 1990's grocery retailers were adapting the analogue or checklist based approaches used in the past by incorporating GIS into their store selection process (Clarkson et al., 1996).

During the 1990's, GIS became more commonly used as a means for retailers to understand their market, customers and competitors. GIS can achieve this through the analysis of vital data including sales data, demographic data, and competitor data. The value of GIS in the retail planning process is its ability to support decisions in evaluating retail location performance (Drezner, 1998), find the optimum store territories (Handa & Vohra, 2010; Murad, 2008), identify the spatial implications of store closures and relocations (Brosekhan, et al.,

1995), find appropriate markets for launching new products (Gebauer & Laska, 2011), and find the best geographical fit for possible retail locations (Birkin et al., 2002; Marros, 2005). GIS contains built-in functions for geocoding locations, importing population-based databases, buffering and overlay that allowed for the growth of more sophisticated forms of spatial analysis and retail planning (Benoit & Clarke, 1997; Clarke, 1998). Buffer and overlay analysis allowed for greater ease in delineating market boundaries and extracting population information (Elliot, 1991; Ireland, 1994; as cited in Clark, 1998). Benoit & Clarke (1997) identified the process for conducting such analysis as: (i) estimating how far consumers are willing to travel to a store (either existing or potential) measured in time or distance (e.g., 10 minutes around a store or 5 km); (ii) delimit an area around that store (a buffer) that marks the limit of that time or distance factor in each direction outwards from the store; (iii) overlay and extract consumer spend data which lies within that buffer.

The importance of GIS in trade area analysis was highlighted by Birkin et al. (2002), addressing how it can be used for measuring potential demand within a geographical area and the definition and allocation of retail areas within a certain distance of the retail location. They also outlined the importance of retailers using GIS to produce high-quality maps, identifying their market boundaries. Several trade area techniques within GIS are highlighted in **Table 1**.

Table 1: Trade Area Analysis Techniques

Type	Description
Simple Ring	Created around stores using a specified radius
Data-Driven Ring	Created around stores using a radius that is proportional to store characteristics such as the total sales, square footage and Gross Leasable Area (GLA).
Equal Competition	Creates trade area boundaries halfway between each store and its neighbouring stores (Thiessen polygons).
Drive Time	Defines the areas that are accessible along the street network based on the specified maximum travel time or distance.
Gravity Model	Predicts the sales potential of an area based on distance, competition, attractiveness factors and consumer spending.
Threshold Ring	Creates rings that contain a specified population or household count

Reproduced from Murad, 2008.

Several researchers (Birkin et al., 2002, Lea et al., 2006; O'Dwyer and Coveney, 2006; Murad, 2008) have discussed the use of these techniques (Table 1) in RLDM. The facilitation of GIS to measure the availability and accessibility of supermarkets is discussed by O'Dwyer and Coveney (2006). Their research analyzes the location of supermarkets in relation to residential dwellings, car ownership and in terms of travel distance along the road network (O'Dwyer and Coveney, 2006). Murad (2008) uses GIS to define and analyze the market area of a mall located in Jeddah, Saudi Arabia. The research uses a drive time technique for implementing market analysis and especially for demand identification purposes. These models can easily prepare market profiles by extracting and aggregating data within the GIS software. While these techniques are efficiently computed, they do make several assumptions about the market and fail to address the complex relationships between customers and retailers. Benoit and Clarke (1997) and Clarke (1998) discuss the two significant limitations to be: (i) the problem of overlapping catchment areas; and, (ii) the fact that these forms of analysis assume spatial

monopolies. While the simple rings method, drive time/distance buffers or Thiessen (Voronoi) polygons are easy to conceptualize and use, they assume all households in the trade area will be patrons of the store and that no households outside the identified boundary will visit the store (Jones & Simmons, 1993). While methods representing spatial monopolies are commonly used (Jones & Simmons, 1993; Clark, 1998; Hernandez et al., 1998), they do oversimplify trade area dynamics as they do not account for the existence of competing stores.

Through the 1980's and 1990's, multiple retailers began investigating the use of SIM when participating in location-based analysis. While these theories date back to a pre-GIS era, their operationalization was not realized until GIS software packages became readily available (Okabe and Okunuki, 2001; Vlachopoulou et al., 2001; Zheng and Zhou, 2001). Defining the geographical market of retail facilities is a main factor that can be covered through GIS applications. "When performing gravity modelling and processing consumer data, users are capable of calculating how markets will change along with competition and consumer spending changes" (Murad, 2008, p.347). Supported by the availability of large-scale GIS and spatial interaction modelling systems (Birkin et al., 2010), large retail conglomerates experienced unprecedented growth through ambitious store development programs between the 1980's and 1990's (Wrigley, 1987). The growth of the location-allocation model was an example of this (strongly related to spatial interaction modelling).

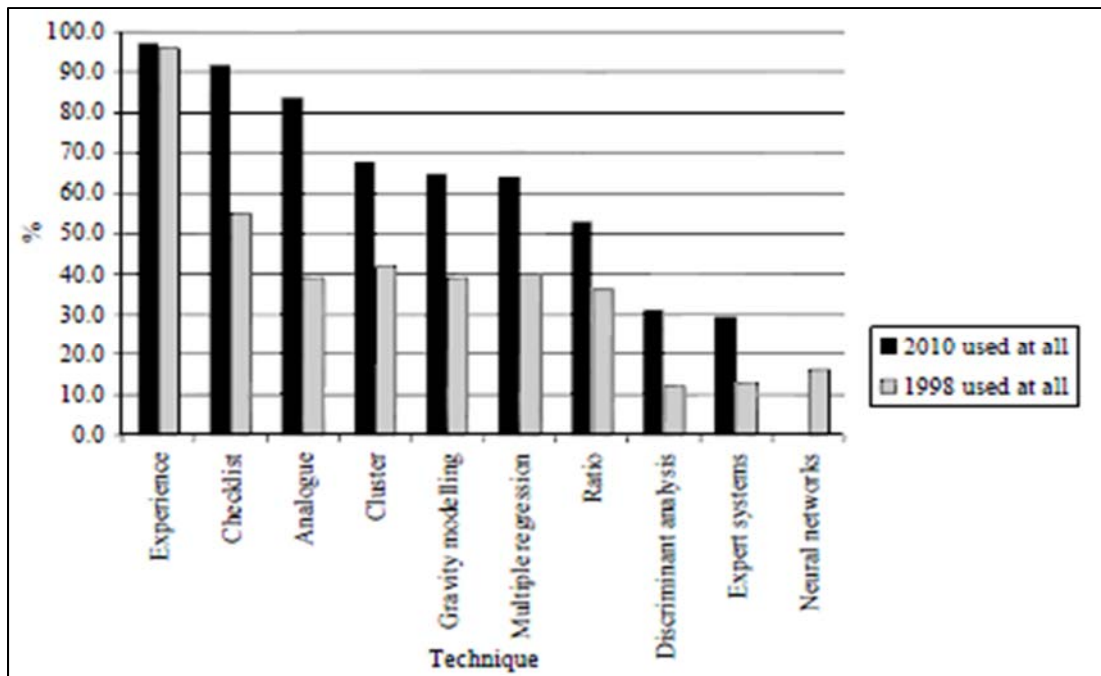
Location-allocation problems are concerned with the placement of an optimal number of facilities or supply points in a given geographic area (Beaumont, 1980; Goodchild, 1984; Drezner, 1998). The basic element of location-allocation is the interrelationship between supply points, demand points and locations (Suomalainen, 2006). The supply points are termed

‘facility,’ as it is to denote an object whose ideal spatial location we are trying to delineate in order to optimize the interactions with demand points (Goodchild, 1984). The locations are the physical spaces in which the facilities can be located; these eligible locations can either be discrete, continuous, or network in representation (Suomalainen, 2006). Location-allocation has two major elements: first is location modelling, which is the development of formal models used to identify the optimal place for a set of activities to take place based on certain parameters (Kemp, 2008); the second element is allocation modelling, which is the dispersion of the geographically distributed demand to the facility that would best service it, according to predefined parameters (Kemp, 2008).

GIS can work with many different types of data as long as they have a geographic element that can be represented by points, lines or polygons. The practicality associated with GIS technology began to experience growth with the arrival of geodemographic packages – that is a system that profiles market areas into customer segment types. In other words, geodemographics acted as a launch pad for GIS adoption within retail firms. It provided retailers with an ability to study the purchasing behaviour of target populations (Mitchell and McGoldrick; 1994; González-Benito, 2002 Gunter and Furnham, 2014), monitor competitive impact (Kaynak, and Harcar, 2005), and set performance targets based on individual stores’ trade area characteristics (O’Malley et al., 1997; Verhstsel, 2005). Despite the fact that geodemographic systems have been available since the late 1970’s in the UK, they did not thrive until the early 1980s with their availability through GIS packages. CACI became the first company to offer a geodemographic targeting system in the UK. CACI developed a segmentation system called A Classification of Residential Neighbourhoods (ACORN) (Clarke,

1998). The applicability of geodemographics to retail locational planning can be seen through the ability to profile the customer make-up of a potential store's catchment area and can be directly related to the product offering proposed by a given retailer, which is ultimately able to help identify new markets for future growth potential (Sleight and Leventhal, 1989; Batey and Brown, 1995; Birkin, 1995; Sleight, 1997). Examples of this include UK retailer Tesco using geodemographics for market analysis (Moore & Attewell, 1991), or IKEA's use of geodemographics to identify appropriate areas for its catalogue distribution (Handa & Vohra, 2010).

While formal techniques for location analysis have been available for over 50 years, these methods were not widely utilized in retail decision making (Byrom, et al. 2001; Hernandez et al., 1998; Hernandez & Bennison, 2000). With growth in low-cost computing and the availability of data, it is apparent that these techniques have diffused within RLDM. A survey conducted by Reynolds and Woods (2010) identified that during the 1980's and 1990's, the utilization of analytical techniques increased (refer to Figure 1). "For example, the use of analogue techniques more than doubled in the period, while that of ratio techniques increased by only 48 percent" (Reynolds & Wood, 2010, pg. 837).



Source: Adopted from Hernandez & Bennison, 2000.

Figure 1: Utilization of analytical techniques in Retail Decision Making

The data comparisons were derived from an earlier survey of UK retailers (operating more than 50,000 outlets across 8 retail sectors) undertaken in 1998, which investigated the use of RLDM techniques and how they utilize GIS. Table 2 identifies location techniques by level of usage. The study found that the simplest techniques, such as checklist, were used by two-thirds of all the retailers surveyed, analogue and ratio techniques by approximately one-third, and gravity models by approximately two-fifths (Hernandez & Bennison, 2000). While SIMs are more powerful in understanding the relationship between consumers and retail facilities, there are some drawbacks to these models. Firstly, these models are difficult to calibrate as they are data intensive (Clarke, 1998). Secondly, given the complex nature of consumer behaviour around product selection, it is unlikely that one model would be sufficient in describing

consumption patterns for different types of goods and services. This, in turn, requires skilled practitioners to make adjustments to the model based on the type of retail offering.

Table 2: Location Techniques by Usage (% respondents)

Technique	Of Which Used			
	Used	Used Regularly	Occasionally	Not Used
Experience	96	84	12	4
Checklist	55	33	22	45
Analogue	39	24	15	61
Ratio	36	15	21	64
Cluster	42	19	23	58
Multiple Regression	40	24	16	60
Gravity	39	27	12	61

Recreated from Hernandez & Bennison 2000

2.1.3 Wave 3 of Retail Location Decision Making:

The need for retailers to understand granular level details about consumers become an integral part of the RLDM process in the early 2000's (Brosekhan and Velayutham, 1995). A deeper knowledge of consumer behaviour helps to offer an understanding of how people think, feel, and select products or shopping destinations (Brosekhan and Velayutham, 1995). Consumer behaviour encompasses much more than what we buy; it focuses on the influences, either personal or group, that affect and shape a consumer's path-to-purchase. Traditional consumer theory identifies two main sources of information that are utilized by customers when purchasing goods and services; internal information and external information. Internal information is received from family, friends, work colleagues etc. and external information is

mainly derived from advertising and merchandising (Chaston, 2015). With the introduction of the internet offering customers a new channel to acquire information and new ways to interact with retailers, a new set of business challenges have arisen. With a clear push toward omni-channel retailing the customer now decides their preferred method for interacting (over the phone, online, in store) with a retailer. Table 3 provides an overview of the different retail channels.

Table 3: Retail Channels

Retail Channel	Description
Single-Channel Retailer	Sells merchandise through one channel only.
Multi-Channel Retailing	Sells merchandise or services through more than one channel or all widespread channels, whereby the customer cannot trigger channel interaction and the retailer does not control channel integration.
Cross-Channel Retailing	Sells merchandise or services through more than one channel or all widespread channels, whereby the customer can trigger partial channel interaction and/or the retailer controls partial channel integration.
Omni-Channel Retailing	Sells merchandise or services through all widespread channels, whereby the customer can trigger full channel interaction and/or the retailer controls full channel integration.

Beck and Rygl, 2014

With the growth in online retailing, there has been added pressure on traditional bricks-and-mortar retailers to understand these changes, in what can be referred to as Omni-channel consumer behaviour, in order to make informed location decisions and maximize their real estate decisions (Nicholson et. al, 2002). With this paradigm shift in the way that consumers shop, there has been a vast amount of research that has attempted to increase the understanding of consumers' attitudes towards online shopping and their intention to shop.

Several studies (Perea y Monsuwé et al., 2004; Moslehpour et al., 2018; Atulkar, and Kesari, 2018) identified that online shopping was not only affected by ease of use, usefulness, and enjoyment , but also by factors like consumer traits, situational factors, product characteristics, previous online shopping experience and trust in online shopping. This has led to the widespread collection of data to help make better-informed decisions. This data is comprised of a mix of data sources some of which are explicitly sought out and others which are a byproduct of running a retail business (Table 4).

Table 4: Data Generating Methods

Data Generating Methods	Definition	Example
Created Data	Is created because it would not exist unless a mechanism was put in place to collect that information.	Loyalty programs, Market research Surveys, Asking for Postal Codes
Provoked Data	Would not exist unless you invited people to express their views.	Product review, service review, etc.
Transaction Data	Generated every time a customer makes a purchase.	POS
Captured Data	Is information gathered passively from an individual's behaviour	GPS from mobile devices

Mar, 2005

An integral part of informing the decision-making process is the application of data analytics – that is the techniques, technologies, systems, methodologies, and applications that analyze critical business data, used to help businesses better understand their markets and make timely business decisions (Chen et al., 2012). Given the ever-increasing scale of investment and subsequent increase in the risks associated with retail location decisions, it is not surprising that the use of data in RLDM has become a growing area of interest (Byrom et al., 2001; Wood & Reynolds, 2012). While data collected via loyalty programs and point of sale is

not a new process of data acquisition, it is, however, one that has been changing. Initially, these forms of data collection were very static in the sense that retailers had an understanding of where you lived (address or postal code) and where you shopped.

It is difficult to discuss the advancements of GIS within RLDM without considering the growth in the availability of spatial data. A key element of any location planning analysis involving GIS or geodemographics is the storage and use of data. The power of GIS lies in its ability to manage data from a number of sources, and as a result, the development and subsequent diffusion of geodemographic databases were seen as being distinctly relevant for retailers (O'Malley et al., 1997).

The value of making more informed retail location decisions became more broadly accepted during this era (Theodoridis and Bennison, 2009; Clark and Rowley, 1995; Byrom et al., 2001) and subsequently, the use of spatial data began to play a key role. With widespread acceptance of loyalty programs by retailers (Byrom, et al., 2001; Clarkson et al., 1996; Handa & Vohra, 2010; Theodoridis & Bennison, 2009; Wood & Reynolds, 2012), data on customer expenditure could now be expressed spatially within a GIS. This acted as a powerful tool in the decision-making process regarding store development as well as marketing campaigns. Data obtained through these loyalty programs can be linked to geodemographic and lifestyle data and other aspects of consumer behaviour (Handa & Vohra, 2010). Several studies have investigated the use of data within the planning process (O'Malley et al., 1997; Byrom et al., 2001). Byrom et al. (2001) attempted to answer this question through the use of questionnaires sent to major retailers in the UK. The study was able to assess the current role and significance of geographic information in RLDM. Approximately two-thirds of the 104

respondents obtained census data, geodemographic data, and lifestyle data from external sources. Internally generated data was predominately competitor, transactional and operational data (Byrom et al., 2001). Furthermore, the study found that while 31 per cent of the respondents reported collecting loyalty card data the leveraging the data's geography in analysis was limited. While the retail respondents understand the data's potential for gaining a deeper understanding of customer behaviour, it was clear that knowing where customers were actually from was often thought to be of secondary importance (Byrom et al., 2001). Similarly, O'Malley et al. (1997) surveyed multiple retailers who collectively accounted for more than 60 per cent of the retail activity (sales) in the UK. This study found that while all of the retailers in this survey reported collecting point-of-sale data and loyalty data to support RLDM, there was little evidence of data integration into strategic decision making. This was viewed to be the result of insufficient user experience, or a general lack of awareness of the additional benefits to be derived (O'Malley et. al. 1997).

2.1.4 Wave 4 of Retail Location Decision Making:

With technological advancements, this information is changing into something that is much more dynamic – that is, retailers are now able to, not only know where you shop and where you live, they now have the potential to see the travel patterns associated with how you got there. Thus, this new era of retail decision making is one that will be driven by the use and management of Big Data (and increasingly SBD). It will strengthen processes related to purchase and sale tracking, which will offer insights on consumer behaviour. Retailers have incorporated systems (such as, Enterprise Resource Planning (ERP), Supply Chain Management

(SCM), Customer Relationship Management (CRM)) that are able to collect petabytes of detailed data on suppliers, customers and other aspects of businesses, increasing the amount of data by 10-fold to 1000-fold (Brynjolfsson et al., 2015).

We are currently in an era where there are fewer barriers to the collection, storage and outputting of data. Advancements in technology have revolutionized the way we collect, organize and use data (Graham & Shelton, 2013). While there is widespread support of the notion of SBD, there are no clear definitions (see. Goodchild, 2013; Graham & Shelton, 2013) or explanations that identify and demarcate the boundaries between “small data” and “Big Data.” Broadly, Big Data is comprised of three key characteristics—Volume, Velocity and Variety (Goodchild, 2013; Sagirolu & Sinanc, 2013). Volume refers to the fact that the data is bigger than previous or bigger than we can currently handle. Velocity refers to the fact that this type of data can be gathered in close to real-time from a Variety of sources (social media, geofencing, mobile devices and mobile applications) (see Boyd & Crawford, 2012; Goodchild, 2013). It’s estimated that up to 80% of Big Data is spatial in nature (Farmer and Pozdnoukhov, 2012; Folger, 2011, as cited in Leszczynski and Crampton, 2016), Mobile phones, automobiles, factory automation systems and other devices are equipped to generate streams of spatial activity data that can be leveraged in business decision making. “Sensors embedded in process machinery may be collecting operations data, while marketers scan social media or use location data from smartphones to understand teens’ buying quirks”(Brown et al., 2011, p. 38).

The awareness of both the significant challenges and opportunities associated with SBD is fundamentally questioning the traditions of RLDM and decision-making (Kitchen, 2013; Lee and Kang, 2015; Thatcher, 2014) and “this paradigm shifting is driven not just by data itself but

all other aspects that could be created, transformed, and/or adjusted by understanding, exploring, and utilizing data”. (Cao, 2017, p. 43:2). The field that is responsible for managing this transformation can be referred to as Data Science, that is, the field that includes everything that involves the “collection, management, processing, analysis, visualization, and interpretation of vast amounts of heterogeneous data” (Donoho, 2017, p 745). The potential of data science to facilitate spatial data-driven processes is increasingly being recognized (Lee and Kang, 2015; Thatcher, 2014) and it involves not only computing, informatics, GIScience and statistics, but also the social sciences and business (Cao, 2017). Cao (2017) outlines several key data science terms, often discussed interchangeably, these are highlighted in Table 5.

Table 5: Key Data Science Terms

Key terms	Description
Advanced Analytics	Refers to theories, technologies, tools, and processes that enable an in-depth understanding and discovery of actionable insights in big data, which cannot be achieved by traditional data analysis and processing theories, technologies, tools, and processes.
Big Data	Refers to data that are too large and/or complex to be effectively and/or efficiently handled by traditional data-related theories, technologies, and tools.
Spatial Data	Spatial Data is data or information that identifies the geographic location of features and boundaries on Earth, such as store locations, customer locations, and neighbourhood demographics. This is basically data that can be mapped and geographically analyzed
Spatial Big Data	‘Spatial Data’ + ‘Big Data’
Data Analysis	Refers to the processing of data by traditional (e.g., classic statistical, mathematical, or

	logical) theories, technologies, and tools for obtaining useful information and for practical purposes.
Data Analytics	Refers to the theories, technologies, tools, and processes that enable an in-depth understanding and discovery of actionable insight into data. Data analytics consists of descriptive analytics, predictive analytics, and prescriptive analytics.
Data Science	Is interdisciplinary field combining statistics and computer science in order to interpret data for the purpose of decision-making
Data Scientist	Refers to those people whose roles very much centre on data.
Descriptive Analytics	Refers to the type of data analytics that typically uses statistics to describe the data used to gain information, or for other useful purposes.
Predictive Analytics	Refers to the type of data analytics that makes predictions about unknown future events and discloses the reasons behind them, typically by advanced analytics.
Data Mining	Data mining is made up of a diverse and distinct set of methods that can be employed for identifying patterns. These include; summarization, classification, clustering, association, and trend analysis

Adopted from Cao, 2017

As senior executives move towards wanting more data-driven decision making (Brown et al., 2011; Brynjolfsson et al., 2015; Chaston, 2015; Chen, et al., 2012; Lavalley, et al., 2011; Sagioglu & Sinanc, 2013), organizations have added pressure in attempting to leverage data into their decision making process. A major caveat is that, for data-driven analysis to trigger new action across an organization, it needs to be closely linked to business strategy, ideally easy for end-users to understand and it needs to be embedded into organizational processes so

action can be taken at the right time (Lavallo et al., 2011). The applications of 'SBD' have been documented in five main areas (refer to Table 6).

Table 6: Potential Applications of Big Data

Topic Area	Usage
Healthcare	Clinical decision support systems, individual analytics applied for patient profile, personalized medicine, performance-based pricing for personnel, analyze disease patterns, improve public health
Public Sector	Creating transparency by accessible related data, discover needs, improve performance, customize actions for suitable products and services, decision making with automated systems to decrease risks, innovating new products and services
Retail	In-store behaviour analysis, variety and price optimization, product placement design, improve performance, labour inputs optimization, distribution and logistics optimization, web-based markets

Adapted from Sagioglu and Sinanc, 2013

A survey-based study conducted by Lavelle et al. (2011) set out to understand the challenges and opportunities associated with the use of business analytics and Big Data. The survey included approximately 3,000 business executives, managers and analysts working in different sized organizations. The study, which covered 108 countries and more than 30 industries, found that top performing organizations are twice as likely to apply analytics to activities. These top performers put analytics into the widest possible range of decisions and they were twice more likely to use analytics to guide future strategies (Lavallo et al., 2011). However, the way in which the data is used or how the data is leveraged is not clearly identified by Lavallo et al. (2011). Could traditional data or small data provide the same insights? With the

quantity of data being processed, it is important for decision-makers to highlight ways to deal with this volume, variety and velocity of information.

According to Kitchin (2013), Big Data sources can be broadly divided into three categories: directed, automated and volunteered. A directed data source refers to a digital form of surveillance which is usually focused on a person or place by a human operator (Kitchin, 2013). An automated data source is usually a direct function of a specific device or system. Kitchin identifies this as:

“traces from digital devices, such as smart phones that record and communicate the history of their own use; transactions and interactions across digital networks; clickstream data that records how people navigate through a website or an app; sensed data generated by a variety of sensors and actuators (that measure levels of light, humidity, temperature, gas, electrical resistivity, acoustics, air pressure, movement, speed, etc) embedded into objects or environments; scanning of machine-readable objects such as travel passes (e.g., Oyster card on the London Underground); passports or barcodes on parcels that register payment and movement through a system; machine-to-machine interactions across the Internet of things and capture systems, in which the means of performing a task captures data about that task” (Kitchin, 2013 p. 263).

There has been considerable research attention around the use of tracking technologies (see Doherty 2009; 2012). Near real-time data from mobile devices can provide detailed information about shoppers, providing knowledge around the complexities associated with their decision-making process. A retail-based example of this is a Toronto-based company Aislelabs, who provides indoor location technology in order to track customers. Through a

cloud-based service, Aisleabs collects information on the behaviour of their clients' customers regarding their traffic patterns, for example, producing heat maps that identify hotspots of activity based on walking paths and cross shopping data. This provides a wealth of new information because traditionally not much was known about a shopper's habits within a store until they paid for their goods. Through the use of cellular signals, foot traffic can help to inform decisions on setting lease prices, relocating amenities, and product positioning (Business New Network, 2015). Lastly, volunteered data is provided by a user. This can include "interactions across social media and the crowdsourcing of data wherein users generate data and then contribute them to a common system, such as OpenStreetMap" (Kitchin, 2013 p. 263).

Data mining has been able to provide a platform to manage substantial increases in the data flow that retailers are obtaining through loyalty programs and point-of-sale systems. Data mining has provided an opportunity to collect and manage valuable data, which has aided in understanding consumer patterns and trends. It allows for the extraction of vital data (geographic and personal) and the production of significant information that improves the decision-making process throughout an organization. Furthermore, it enables an organization to focus on the most important information in their consumer database, which allows managers to make more knowledgeable decisions by improving the prediction of future trends and behaviours (Chopoorian et al., 2001 as cited in Harmozi and Giles, 2004; Hebert et al., 2014; Raju et al., 2014; Turow et al., 2015). It has also offered unique opportunities for businesses to make more informed decisions by minimizing the amount of time required to extract and organize data. Data may be evolving, so it is important that the SBD mining techniques should be able to adapt to these changes (Jadhav, 2013).

Data Lakes are a relatively new technology that is meant to help with the data mining process. Data Lakes are storage repositories that holds large amount of raw data in its native format (structured, unstructured, semi-structured data) until it is needed (Khine et al., 2018). This ultimately eliminates data preparation because data are only classified, organized and analyzed when they are needed (Khine et al., 2018). Table 7 outlines the differences between data warehouses and data lakes which are contributing to their growth in popularity.

Table 7: Difference between Data Lakes and Data Warehouses

Difference	Data Lakes Description	Data Warehouse Description
Data Lakes Retain All Data	They retain data that is in use today but also data that may be used and even data that may never be used just because it might be used someday.	During the development of a data warehouse, a considerable amount of time is spent analyzing data sources, understanding business processes and profiling data. The result is a highly structured data model designed for reporting.
Data Lakes Support All Data Types	Data lakes support non-traditional data types. In the data lake, all data can be kept regardless of source and structure. It is kept in its raw form.	Data Warehouses generally consist of data extracted from transactional systems and consist of quantitative metrics and the attributes that describe them. Non-traditional data sources such as web server logs, sensor data, social network activity, text and images are largely ignored.
Data Lakes Support All Users	From operational users (80% of users looking at performance metric and reports) to users who perform in-depth analysis.	The data warehouse is usually ideal for operational users because it is well structured, easy to use and understand and it is purpose-built to answer their questions.
Data Lakes Adapt Easily to Changes	Since all data is stored in its raw form and is always accessible to anyone, users are empowered to go beyond the structure of the warehouse to explore data in novel ways and answer their questions at their pace.	Traditional Data Warehouses take very long to change. Considerable time is spent up front during development getting the warehouse's structure right.
Data Lakes Provide Faster Insights	Because data lakes contain all data and data types, because it enables users to access data before it has been transformed, cleansed and structured it enables users to get to their results faster than the traditional data warehouse approach.	The work typically done by the data warehouse development team may not be done for some or all of the data sources required to do an analysis.

Campbell, 2018

There are also institutional challenges around SBD. One major challenge identified by Brown et al. (2011) is the fact that the bulk of data are often stored in department-specific “silos,” delaying the use of such data. This is important because if organizations fail to leverage real-time aspects of data in their decision making, the value can be lost. Another challenge with such data is what Brown et al. (2011) refer to as “information hoarding”¹, that is business organizations that do not share data between departments. Examples of this can be seen within financial institutions, as they suffer from their own failure to share data among diverse lines of business, such as financial markets, money management, and lending (Brown et al., 2011). This can ultimately prevent these companies from forming a comprehensive view of their individual customers (Brown et al., 2011). This is not a new phenomenon, as several studies identified similar problems with the sharing of data and software licenses within organizations (Byrom et al., 2001; Hernandez et al., 1998; Reynolds & Wood, 2010; Wood & Reynolds, 2012).

Organizational, managerial and cultural issues have been identified as critical factors related to the widespread adoption of Big Data and SBD practices (Lavallo et al., 2011; McAfee et al., 2012; Vassakis et al., 2018; Erevellas et al., 2016; Kwon et al., 2014). Being able to visualize data differently becomes increasingly valuable as executives want better ways to communicate complex insights so that the decision-making process can move quicker (Lavallo et al., 2011). As the data is so big, it is very difficult to find user-friendly visualizations along with new techniques, and frameworks to bring data to life (Jadhav, 2013; Keim, 2013; Chen and Zhang, 2014; Jagadish et al., 2014). Other challenges exist around the ability to actually attain data and the lack of skilled professionals (data analysts) who can manage, organize and

synthesize the data (Davenport and Patil, 2012; Ahalt and Kelly, 2013; Lyon and Brenner, 2015; Akter et al., 2016). Large amounts of potentially useful data are being lost since new data is largely unstructured and untagged. The issue with unstructured and untagged data is that it is often in formats that are not compatible with current data warehousing systems making it impossible to extract data in order to perform analysis. Without a data lake this is impossible to overcome. This can ultimately result in potentially valuable unstructured and untagged data to be omitted from RLDM. In 2012, 18 percent of the digital universe in the US would be useful if tagged and analyzed (Gantz & Reinsel, 2013). Businesses need to proceed with caution as investments in analytics can be useless and costly unless employees can incorporate that data into complex decision-making (Provost & Fawcett, 2013).

There are also some social implications of Big Data, especially as related to privacy concerns. As we constantly provide vast amounts of data about our daily lives, it is seen to be easier than ever to identify individuals and create behavioural profiles. The basic task of ensuring data security and privacy become harder as information is multiplied and shared (Tene & Polonetsky, 2012; Jadhav, 2013). Data regarding “individuals’ health, location, electricity use, and online activity is exposed to scrutiny, raising concerns about profiling, discrimination, exclusion, and loss of control” (Tene & Polonetsky, 2012, p 65). This type of data can be collected and can be used by a variety of organizations from both the private and public sectors. “In some instances, such as the US government’s ... metadata alone were sufficient to build comprehensive portraits of individuals and groups” (Graham & Shelton, 2013, p 259). Data privacy and security is a major issue within the retail sector given the vast collection of spatial consumer data. There has been considerable consumer backlash to several major retail

data breaches. These include; Macy's, Kmart, Sears, Adidas, Best Buy, Panera Bread, Forever 21, Whole foods, Gamestop (Green, 2018).

With the increased automation of data collection and analysis, along with the development of algorithms that can extract and illustrate patterns in human behaviours (Boyd & Crawford, 2012), Big Data has raised key questions about research processes and the ways in which geographers should engage with information. There is a documented need to address questions on how to make sense of these vast amounts of raw information and how to evaluate the role of traditional forms of scientific theories and models in assessing these types of data (Bollier, 2010). Some have argued that there will no longer be a need for theories and models, as simple observations and measurements of large datasets will suffice (Anderson, 2008). Anderson (2008) identified that the new challenge is not to come up with new models but rather to find new ways to sort through the data in order to find meaningful correlations. While statements like this have generally been rejected (Boyd & Crawford, 2012; Graham & Shelton, 2013; Sagioglu & Sinanc, 2013), it does present a level of uncertainty around the idea of Big Data within a research context. Hal Varian, chief economist at Google, alternatively argued that theories allow you to extrapolate outside of what is actually observed. Therefore, the development of models that explain correlations in data allow for the development of new predictions (Bollier, 2010). Big Data can be seen to provide destabilizing amounts of knowledge and information that lack the regulating force of theories (Bollier, 2010). The increasing reliance on new data sources for measuring, models, algorithms, and information systems could mean that knowledge that is not so easily encapsulated within Big Data frameworks might become devalued (Graham & Shelton, 2013). With Big Data on its own not being self-explanatory, there

has been a growing dialogue around the need for models and theories to handle Big Data.

Furthermore, it is important also to identify whether or not existing data models can benefit or use the new data that is produced. For example, a limitation of SIMs was that they required an extensive amount of data (Simkin, 1989); therefore does Big Data now meet these data requirements and offer opportunities for the development of more robust models? No academic research looks at this in any detail.

Traditional location models are normative based approaches, meaning, they rely on a basic set of assumptions that attempt to outline what should happen as opposed to what is actually happening. The introduction of Big Data may offer an opportunity for traditional theories to be revamped to include a new set of assumptions around consumer activity. The presence of this new data should matter to the development or redevelopment of theories because the ability to identify real material affects lies within the data. In other words, it may offer an opportunity to verify the truth or accuracy of traditional retail geography theories. At a fundamental level, traditional models for RLDM are designed for smaller data requirements. While there has been progress in creating new data analytics software that manage and blend large datasets (such as, Altreyx), there is still very little known in regard to the implementation of these within retail planning.

Kitchin (2013) identifies a lack of preparedness within the discipline of human geography to manage this new era. His research argues that current methods being taught in universities and colleges are archaic with respect to their abilities to handle large volumes of data for analytical purposes. Other researchers (Goodchild, 2013; Graham & Shelton, 2013) highlight that the promises and risks of SBD practices in academia will influence the creation of new theories and

methodologies for geographers. The methods and training within human geography have not changed since the early 1990s when GIS began to be incorporated into data analytics (Kitchin, 2013). Therefore, the introduction of larger datasets places added importance on the adoption of a broader set of skills and methods (methodological pluralism) to come together to tackle a variety of issues (DeLyser & Sui, 2014; USGIF, 2018).

Table 8 identifies the current state of research in regards to RLDM. There is a clear need to evaluate the current state of RLDM as it relates specifically to this new wave of decision-making. There is a clear gap in the understanding of the role that the introduction of SBD and SBD analytics is playing within the retail sector.

Table 8: RLDM Research Studies

	Wood, and Reynolds, 2012	Reynolds, and Wood, 2010	Hernandez and Bennison, 2000	Hernandez and Emmons, 2012	Byrom, 2001	Byrom, et al., 2001
Country	UK	UK	UK	CND	UK	UK
Data	X	N/A	N/A	N/A	X	X
SBD	N/A	N/A	N/A	N/A	N/A	N/A
Techniques	X	X	X	X	X	X
Technology	N/A	N/A	N/A	X	N/A	N/A
Organization/Culture	X	X	X	N/A	N/A	N/A

2.2 EMERGING APPLICATIONS

A major area of promise with SBD is in the area of obtaining more granular level data about consumers. Smartphones and other mobile devices provide large streams of data tied to people, activities, and locations (McAfee, 2012). This data can be placed into two categories: (i) app data collected by the app companies directly from built-in sensor on the phones such as; GPS, accelerometer, magnetic field, gyroscope; and, (ii) spatio-temporal network-level data collected by the telecommunication companies such as user ID, location (GPS), device type, timestamps, type of service (Cheng et al., 2017).

Mobile data like this can assist with operationalizing fundamental models in human geography and time-geography (Hägerstrand's 1970), such as space–time prisms, space–time paths, and potential path areas (PPA) (Table 9) that explore how space can affect activity choices for individuals (Dijst and Kwan, 2005). These concepts help determine the locations where activities take place, the distance between the locations of activities, the available time for travel and travel velocities (Pred, 1984). While this work was initially developed in order to answer questions related to transportation planning (Miller, 1991; Miller, 1999; Anderson, 1971), many of the concepts of movement and influence on consumer choices are fundamental to the methods for identifying retail stores patronage (Couclelis, 2009; Scott and He, 2012; Shaw and Yu, 2009).

Table 9: Time Geography Concepts

Time Geography Themes	Description
Space-Time Prism	Space-time prisms model the ability of individuals to travel and participate in activities at different locations in an environment.
Space–Time Paths	The actual paths travels by individuals
Potential Path Areas	The region in space that is accessible to the moving object.

Miller, 2017

The benefits of collecting data and understanding consumer travel patterns have been well documented (Arranz-Lopez, 2017; Suel and Polak, 2017; Hagberg and Holmberg, 2017). Through the collection of loyalty program data, geo-fencing, and electronic point-of-sale, retailers now have the data and resources to leverage actual individual-level consumer SBD in retail decision-making.

2.2.1 Tracking Technologies

The use of modern tracking technologies provide new opportunities to examine the effects of individual travel patterns on consumer behaviour and operationalize concepts such as consumer space–time prisms, space–time paths, and potential path areas (PPA) with greater ease.

Several studies have looked at using tracking technologies to understand consumer/patron behaviour at a more granular level. For example, Yaeli et al. (2014) combined customer movement data (collected via Wi-Fi) with floor layout data and sales data to conduct

visual analysis on consumer movements. Lee et al. (2013) used data from smartphone sensors to examine the correlation between adjacent activities (cross-shopping) time of day. In a non-retail study, Ruiz et al. (2014) used smartphone detection over a 15 day period, utilizing a hospital's Wi-Fi network to classify visitors and to create various visualization tools (e.g., heat maps). Other studies have looked at utilizing GPS in order to infer travel demands (Sila-Nowicka et al., 2016; Gong et al., 2016). Sila-Nowicka et al. (2016), analyzed human mobility patterns from GPS trajectories and contextual information. Furthermore, Gong et al. (2016) used mobile data to infer trip purposes and uncovering travel patterns from taxi trajectory data. There have also been studies that have looked at mobile location data for deriving trade areas (Qu and Zhang; Han and Yamana, 2016; Wang et al., 2016).

2.3 INNOVATION AND TECHNOLOGICAL ASSIMILATION AND ADOPTION

SBD assimilation and adoption is a major technological challenge. It is imperative to look at technological innovation adoption in order to understand the propensity for an organization to adopt new technologies into their decision-making processes. There is an extensive body of research that attempts to model how technologies and innovations are adopted (Goodhue and Thompson, 1995; Taylor and Todd, 1995; Venkatesh, 2003).

Technology assimilation refers to the process or stages of innovation adoption. This is largely documented (Zhu et al., 2006; Wei et al., 2015; Hasgall and Ahituv, 2018) as being a process which moves from initial awareness of an innovation, to its full deployment. Zhu et al. (2006), identified that the assimilation of an innovation, (specifically within the adoption of e-

business), was linked to three stages; initiation, adoption, routinization. This initial stage requires awareness and evaluations of the potential innovation. This typically results in evaluating the benefits of the innovation to improve an organizational task, like RLDM (Zhu et al., 2006). The adoption stage is where the decision is made to actually use the innovation. If the decision is made to adopt the innovation, it will then be allocated resources (e.g. funding etc.). The third and final stage of assimilation deals with the wide scale adoption of the innovation by the organization.

The effectiveness of assimilating new technologies into an organization is related to the structure of the organization, its worker management and control methods, the amount of time and resource allocation and the worker's attitudes towards these new innovations (Hasgall and Ahituv, 2018). An antecedent or precursor to the adoption of an innovation is often explained through the TOE framework. Table 10 highlights key research that has used this approach to model the assimilation of an innovation.

Table 10: A Summary of Key Research using Innovation Assimilation Theories.

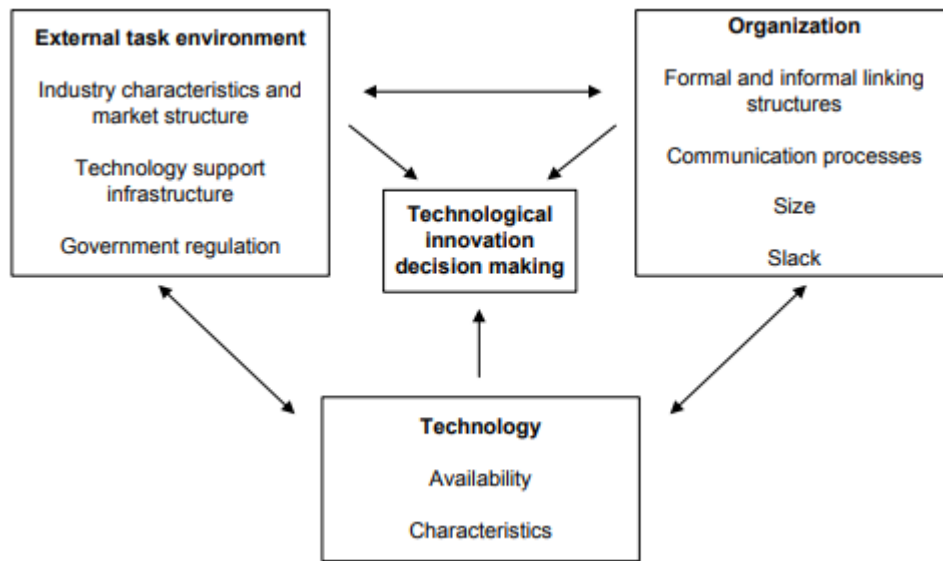
Study	Author(s)	Framework/Methods
The Process of Innovation Assimilation by Firms in Different Countries: A Technology Diffusion Perspective on E-Business	Zhu et al., 2006	Used TOE framework to investigate the diffusion of internet based e-buisness innovations by business firms.
The Assimilation of RFID technology by Chinese companies: A technology diffusion perspective	Wei et al., 2015	Used TOE framework to investigate the assimilation of RFID within Chinese firms.
Technological, organisational and environmental factors influencing managers' decision to adopt cloud computing in the UK	Gutierrez et al., 2015	Used TOE framework to investigate the assimilation of cloud compiuting application.
Understanding the determinants of business intelligence system adoption stages An empirical study of SMEs	Puklavec et al., 2018	The paper provides empirical insights about how technological, organizational, and environmental factors affect individual BIS adoption stages

Tornatzky and Fleischer (1990) developed the TOE framework in the 1990's (Figure 2). It focuses on three aspects of an organization that can influence whether or not a technical innovation will be assimilated. These are technological context, organizational context, and environmental context (Table 11).

Table 11: Technology, Organization, and Environment Framework (TOE)

Context	Descriptions
Technological Context	Describes both the internal and external technologies relevant to the firm. Includes: current practices and equipment internal to the firm and technologies
Organizational Context	Refers to descriptive measures about the organization such as scope, size, and managerial structure
Environmental Context	Is the arena in which a firm conducts its business—its industry, competitors,

Source: Tornatzky and Fleischer, 1990



Source: Tornatzky and Fleischer, 1990

Figure 2: Technology, Organization, and Environment Framework (TOE)

One of the most widely documented models is the Theory of Diffusion of Innovation created by Rogers (1995). His work states that before innovation adoption happens by an organization, it has to go through an innovation development process. The innovation-decision process, in the context of SBD adoption, is the process through which decision-making units progress from knowledge that an innovation exists, to an attitude formation towards the innovation, to a decision of whether to adopt or reject the innovation, to implementation of the new idea, and ultimately the confirmation of this decision (Rogers, 1995, as cited in Henderson et al., 2012). The five-stage process can be viewed in more detail in Table 12. Rogers further identifies five major attributes that can predict an innovation's rate of adoption within an organization, specifically: relative advantage, compatibility, complexity, trialability, and observability (Table 13).

Table 12: Innovation-Decision Process

Stages	Description
Knowledge	When the individual is exposed to the innovation's existence and gains an understanding of how it functions
Persuasion	When the individual forms a favourable or unfavourable attitude toward the innovation
Decision	When the individual engages in activities that lead to a choice to adopt or reject the innovation
Implementation	When the individual puts an innovation into use
Confirmation	When the individual seeks reinforcement for an innovation-decision already made but may reverse the decision if exposed to conflicting messages about it

Rogers, 1995 as cited in Henderson et al., 2012

Table 13: Attribute to Predict an Innovation Rate of Adoption.

Attributes	Descriptions
Relative Advantage	Is the degree to which an innovation is perceived as better than the idea it supersedes.
Compatibility	Is the degree to which an innovation is perceived as consistent with the existing values, past experiences, and needs of potential adopters.
Complexity	Is the degree to which an innovation is perceived as relatively difficult to understand and to use.
Trialability	Is the degree to which an innovation may be experimented with on a limited basis.
Observability	Is the degree to which the results of an innovation are visible to others.

Rogers, 1995 as cited in Henderson et al., 2012

Rogers (1995, as cited in Lai, 2017) further identified five adopter categories. These adopter categories identify the tendency of an organization to adopt an innovation earlier than others do. By compiling data from 508 innovation diffusion studies (as cited in Lai, 2017)

Rogers was able to identify the rate of adoption among individuals and organizations. The theory divides the adopters into 5 categories (Table 14): Venturesome Innovators, Respectable Early Adopters, Deliberate Early Majority, Skeptical Late Majority and Traditional Laggards (Longley et al., 2011 p.45).

Table 14: Roger's Adoption Categories

Category of Adoption	Description
Venturesome Innovators	Willing to accept risk and sometimes regarded as oddballs
Respectable Early Adopters	Regarded as opinion formers or role models
Deliberate Early Majority	Willing to consider adoption only after peers have adopted.
Skeptical Late Majority	Overwhelming pressure from peers needed before adoption occurs
Traditional Laggards	People-oriented to the past.

Source: Longley et al., 2011 p.45

Moore (1999) went on to identify that with high-tech innovations diffusion there are noticeable cracks or gaps [Moore, 1999 as cited in Martinaro and Liu, 2015). "Gaps occur in Rogers' theory between every adopter category, but at the same time the largest and most critical gap is formed between the early adopters and the early majority. Innovation has to cross chasms, in order to become successful in the markets" (Martinaro and Liu, 2015, pg. 140. This chasm occurs at the point where an innovation achieves majority adoption.

Montinaro and Liu (2015) further highlighted that Rogers (1995) proposes that IT-innovations can create uncertainty within an organization, which can ultimately lead to resistance. Therefore the larger the innovational change the greater the uncertainty it creates

and the more difficult it is put into practice (Montinaro and Liu, 2015). This becomes even more problematic if there is a need to change people's practices (Montinaro and Liu, 2015).

Theories related to the diffusion and assimilation of innovations are imperative to this dissertation as it will ultimately allow for the understanding of the current state of SBD on the adoption curve and the potential challenges (e.g. speed of adoption) they may face in assimilating SBD practices. By using innovation diffusion theory it will be possible to highlight issues which may exist for retail firms to move from being aware of what is available, to adoption and use of SBD analytics, all the way to the development of SBD practices.

3 RESEARCH DESIGN AND METHODOLOGY

The last major Canadian study focusing on data-driven retail decision making was more than a decade ago (Hernandez & Emmons, 2012). There is a need to revisit location decision-making considering present-day developments in SBD and data science. To accomplish the research objectives of this dissertation, a combined quantitative and qualitative approach was adopted utilizing a multistep survey model, similar to Wood and Reynolds (2012). Initially, an online questionnaire was conducted of major retailers in order to scope the current state of collection, storage, and use of data, techniques and technologies in RLDM. The second step was to conduct semi-structured interviews of approximately 24 analysts and managers responsible for retail location analytics. The final stage of data collection involved undertaking 3 case studies with major Canadian retail organizations.

3.1 ONLINE QUESTIONNAIRE

The online questionnaire focusses on retail location decision-makers who are responsible for identifying and acquiring retail location opportunities and managing store portfolios. The survey was designed to enhance the understanding of the nature and extent of location decision support activities undertaken within retail and service firms operating within Canada.

3.1.1 Questionnaire Design and Format

The first step in developing the online questionnaire was to develop a list of broad themes that were to be investigated. Four central themes for the survey were identified, which included: (i) *Personal* - to identify respondents and their decision-making experience; (ii) *Data Usage and Availability* – to audit the organization’s data environment; (iii) *Decision-making Techniques and Methods* – to ascertain the range of decision support tools used; and, (iv) *Organizational Culture* - to provide an organizational context for the decision-making activities.

In line with De Vaus (1995) as cited in Hay (2005), four content questions were developed for the online survey (Table 15). The survey questions consisted of both closed-ended questions (e.g., rating scales, forced choices, and dichotomous) and open-ended questions. The closed-ended questions were structured so that respondents had to select either specific categories, rankings (e.g., level of importance) or scales (e.g., measures of attitudinal intensity). A significant benefit of closed-ended questions is that their structure makes data coding and analysis straightforward (Reja et al., 2003). A challenge, however, with closed-ended questions is making sure that there is a relevant, extensive, and exhaustive list of possible answers. Additionally, closed-ended questions are limited to a set number of response options. Therefore, it relies on the assumption that the wording, categories, and concepts will have the same meaning for all respondents (Reja et al., 2003; Denscombe, 2009). As a means to try and lessen these limitations, four industry professionals from different retail sectors and with varying levels of experience) and two academics (with extensive survey design experience) were consulted to ensure that the questions were valid and relevant for the potential respondents. Open-ended questions were used much less frequently than closed-ended

questions in the online survey. Open-ended questions were used more extensively in the follow-up structured interviews. . The open-ended questions offered less structure, subsequently allowing respondents to recount their own understandings, experiences, or opinions in their own way. A major challenge with open-ended questions is the lack of consistency in the responses, making it difficult to draw comparisons (Reja et al., 2003; Hruschka et al., 2004; Turner III, 2010).

Table 15: Types of Content Questions Used in the Online Questionnaire

Type	Description
Attribute Questions	Aim to establish respondents' characteristics (e.g., Time spent in industry or with a particular organization)
Behaviour Questions	Aim to discover what people do (e.g., Techniques used in RLDM or data frequently utilized in decision making)
Attitudinal Questions	Aim to discover what people consider desirable or not desirable (e.g., Corporate challenges)
Belief Questions	Aim to establish what people believe to be true or false (e.g., Are you effective at SBD integration)

Hay, 2005

Note that for comparative purposes, sets of questions were strategically adopted from Byrom et al. (2001) and Hernandez and Emmons (2012). This was key to the assessment of not only the current state, but to gauge change over time in the retail data landscape. Additional questions were incorporated to capture the corresponding changes in the methods and techniques that are used to synthesise, analyze and capture SBD. Definitions for what constituted spatial data and SBD were provided to the survey participants in order to create consistency in responses. The definition provided to participants for spatial data was '*data or*

information that identifies the geographic location of features and boundaries on Earth, such as store locations, customer locations, and neighbourhood demographics. This is basically data that can be mapped and geographically analyzed' (Beal, 2018). Furthermore, SBD was defined as *'datasets that are so large or complex that traditional data processing applications are inadequate to deal with them'* (Marr, 2015).

3.1.2 Sampling: Selecting Questionnaire Participants

Following research ethics approval, a link to an online survey was distributed via the personal emails of 181 retail location decision-makers, defined as anyone responsible for locating/acquiring new stores, developing established or new formats, refurbishing or relocating stores, or closing stores. The survey focused on major Canadian retailer that typically made \$100 million (CND) in annual sales. These individuals came from 112 different retail businesses. The respondents were selected using a purposeful sampling method (Baxter and Eyles, 1997) which involved identifying and selecting individuals from retail organizations that were perceived to be responsible or participating in the process of RLDM. Potential respondents were identified through the industry network at the Centre for the Study of Commercial Activity (CSCA) at Ryerson University - a university not-for-profit research centre that focuses on the consumer service scape. A list of contacts was created using company websites and other online resources. For example, LinkedIn was used to expand the list of contacts by searching for individuals with comparable job titles.

3.1.3 Pre-Testing, Distribution and Maximizing Response Rate

Pre-testing was carried out to pilot the survey with a subsample of the respondents. This allowed for any issues to be identified and to make improvements to either individual questions or the questionnaire as a whole. Table 15 highlights the primary purpose of running the pre-test and the type of feedback that was solicited by test respondents.

Table 15: Pre-Testing Purpose

Pre-Test Category	Pre-Test Purpose
Individual Question Pre-Test	Are individual questions and question instructions understood?
	Would any question benefit from the addition of written prompts?
	Do respondents interpret questions as intended?
	Do any questions make respondents feel uncomfortable?
Questionnaire Pre-Test	How do respondents react to question ordering?
	Does it flow logically and intuitively?
	Does the questionnaire become repetitive?
	How long does it take to complete the questionnaire?

Hay, 2005

The pretesting process was the first step taken in an attempt to maximize the response rate. The following are additional strategies (adopted from Evans and Mathur, 2005; Hay, 2005) that were implemented to help increase the number of participants:

- A cover letter was included with the survey to introduce the research and the questionnaire
- The questionnaire was kept concise

- Follow-up emails were sent to early respondents thanking them for their response and to non-respondents prompting them to participate and complete the survey. This was done after the 1st, 2nd, 3rd, and 4th week after the questionnaire was circulated.

The email invitation informed potential respondents of the survey's objectives and the offer of a complimentary summary findings report, similar to other comparable studies from Canada (Hernandez and Emmons, 2012) and the UK (Byrom *et al.*, 2001). To maintain anonymity, all responses are reported in aggregated form by broad sector groupings.

3.2 SEMI-STRUCTURED INTERVIEWS

The second phase of this study involved undertaking semi-structured interviews. In total, 24 analysts, managers and executives responsible for retail location analytics were interviewed as part of the follow-up to the online surveys. Interviews were used at this stage because: (i) they offered an opportunity to fill knowledge gaps that the online questionnaire was unable to bridge; (ii) to investigate complex behaviours and motivations; and, (iii) to identify consensus and differences in industry experiences and opinions. Three types of interviews were considered and evaluated for inclusion before the choice of interview type was selected (Table 16).

Table 16: Evaluation of Different Interview Types

Type	Description	Benefits	Weaknesses
Structured Interviews	Follow a pre-determined and standardized list of questions. Questions asked in the same way and order.	Only the information required is collected. Easier for analysis because everyone is asked the same set of questions	Lack flexibility. Questions cannot be asked spontaneously or impromptu.
Unstructured Interviews	Are directed by the respondents rather than by predetermined questions.	Allows for more information and opinion. Tangents can lead to relevant and meaning information that was not initially considered.	Difficulty in comparing data
Semi-structured Interviews	Combines a pre-determined set of questions with the flexibility of unstructured questions. Therefore interviewers are able to explore responses further.	Large amount of detail generated. Fairly flexible and sensitive. Fairly reliable and easy to analyze.	Can't guarantee honesty of participants. Cause and effect cannot be inferred. Flexibility of interview may lessen reliability. Open-ended questions are difficult to analyze. Difficult to compare answers.

Hay, 2005

Semi-structured interviews were selected as they provided the flexibility of unstructured interviews while still providing a standardized list of questions in order to draw comparisons between respondents more easily, keeping it organized yet adaptable. Furthermore, they were chosen to allow for new questions to arise that were not initially considered. Semi-structured interviews were also used in comparable studies (Woods and Reynolds, 2012) and were viewed as a valuable and reliable method to gain information.

A primary concern when developing semi-structured interviews is attempting to ensure the credibility of the findings. Diefenbach (2009) identifies the following:

- (i) The quality of the data (e.g., what people say or don't say is only part of the picture).
- (ii) The quantity of data (e.g., need to meet sufficient people in order to make general comparisons).
- (iii) The time frame (e.g., their 'snapshot character' and the lack of longitudinal material).

Several methods and strategies were utilized and investigated to enhance the rigour of this research method. Sampling interview development practices and analysis strategies were investigated in order to enhance credibility. These methods were identified through several studies (Baxter and Eyles, 1997; Patton, 1990; Hay, 2005). These are outlined below.

3.2.1 Sampling: Selecting Participants for Semi-Structured interviews

The initial respondent list was drawn from volunteers (5 from the 45 participants) in the online survey. Due to the small number of volunteers multiple sampling techniques had to be used, as is common (Baxter and Eyles, 1997). This initial list of 5 was expanded using purposeful sampling methods. Patton (1990), highlights 15 types of purposeful sampling methods. The three methods that were utilized for this study were: (i) snowball sampling; (ii) criterion sampling; and, (iii) convenience sampling. Snowball sampling, often called chain sampling, identifies research subjects by having research subjects identify potential

participants. (Biernacki and Waldorf, 1981; Szolnoki and Hoffmann, 2013; Atkinson and Flint, 2001). All respondents were asked if they knew of anyone that would be willing to participate and they often made introductions. Criterion sampling is a method that involves selecting respondents that meet a set of criteria which are important to the research (Patton, 1990; Palinkas et al., 2015; Sandelowski, 1995). It was essential to the study to have a variety of respondents who held different positions across different sectors from organizations of varied size. As a result, individuals were sought out to add to the level of diversity in the respondent list. Convenience sampling is a technique that selects participants based on ease of access or convenient accessibility (Emerson, 2015; Etikan et al., 2016).

Table 17 identifies the list of the 24 respondents interviewed, by position and retail sector. The sampling goal was to ensure diversity in retail sector and positions. In order to gain a true picture of how retail location data is being adopted and the challenges and opportunities that exist in the world of SBD multiple retail sectors had to be surveyed, Furthermore, the perspectives and opinions of industry practitioners will differ based on the amount of time the respondent has spent in the industry and therefore it was important that junior analysts as well as senior managers were included in the interview processes.

Table 17: Interview Respondents

Respondents Position	Retail Sector
Partner	Retail Consultancy
VP of Real Estate Market Strategies	Grocery
Senior Director Branch Distribution	Financial
Market & Location Analytics Manager	Casual Dining
Senior Analyst	General Merchandise
Director of Market Development	Pharmacy and Personal Care
Senior Vice President of Real Estate Acquisition	General Merchandise
GIS Analyst	Developer/Leasing/Brokerage
Senior Director of Physical Planning	Financial
Senior Manager of Physical Distribution	Financial
Market Research Analyst	Developer/Leasing/Brokerage
Research Analyst in Asset Research	Developer/Leasing/Brokerage
Analyst – Analytics, Insights & Innovation	Developer/Leasing/Brokerage
Manager of Development and Portfolio Analytics	Casual Dining
Senior Director, Market Research	Developer/Leasing/Brokerage
Development Lead	Fast Food
Development Manager	Fast Food
Senior Market Analyst	Grocery
Senior Director of Physical Distribution Strategy	Financial
Development Lead	Fast Food
Director of Network Delivery	Financial
National Director Strategy and Insight	Fast Food
Marketing Analytics and Research Manager	Casual Dining
Spatial Research Analyst	Fast Food

3.2.2 Interview Design & Practices

The first step undertaken in the interview design was deciding whether or not to create a research “guide” or “schedule”. The primary purpose and intent of interview guides and schedules is to ensure that all issues of interest are investigated. Interview guides are used to outline a general list of issues that are to be covered by the interviews. Using guides offers a degree of flexibility when interviewing because questions are not as structured (Hay, 2005; Morehouse and Maykut, 2002). This allows conversations to follow more of a natural direction,

and for questions to be crafted impromptu, drawing on themes discussed (Hay, 2005; Morehouse and Maykut, 2002). A major disadvantage of using such guides exclusively is that you need to have good communication skills to be able to react quickly to the flow of the conversation (Hay, 2005; Morehouse and Maykut, 2002). In contrast, interview schedules use a list of carefully worded questions, allowing for much better comparisons to be drawn between participants. A significant disadvantage of this is that formally read questions can sound insincere (Hay, 2005; Morehouse and Maykut, 2002).

In the end, a mixed method was adopted as it capitalizes on the strengths of both guides and schedules while minimizing the weaknesses. In this method, fully worded questions can be placed in a guide while still being used as a topic area. The pre-determined questions act as a safeguard.

The interview questions were separated into primary and secondary questions. Primary questions were used as opening questions, which encouraged the respondents to initiate discussion on a new topic. Secondary questions were used as prompts that were meant to encourage respondents to follow-up or expand on issues (Hay, 2005, p. 83). Table 18 identifies the types of interview questions that were used.

Table 18: Interview Question Types.

Type of Question	Example	Type of Data and Benefits
Descriptive (knowledge)	What is the full name of your organization? What is your role within the organization?	Details on events, places, people and experiences.
Storytelling	What role do you play in the location planning process? Can you briefly outline the linkages between the departments?	Identifies a series of players, an ordering of events, or causative links. Encourages sustained input from the informant.
Opinion	How important is intuition in decision making?	Taps into people's ideology and assumptions. Encourages reflection on how events and experiences may have influenced opinions or perspectives.
Structural	How has the use of data evolved over your tenure in your organization?	Comparison of experience by place, time, gender and so forth. Encourages reflection on (dis)advantage.

Adopted from Hay, 2005

Interviews were manually documented to make the respondents feel more at ease. This was done on a recommendation by several of the participants. While audio recording would have been a more preferred method of collecting the responses, several respondents indicated their lack of comfort with being recorded due to the sensitivity of the topics. Manually documenting the interviews made the respondents feel less vulnerable and encouraged them to be more forthcoming. The interviews were typically conducted over the phone, however three were done in person and lasted between 30 and 90 minutes. Five of the interviews were carried out across multiple days due to interviewee availability.

3.2.3 Interview Data Coding and Analysis Techniques

Once all interviews were recorded, open coding was done to organize the findings and to extract meaningful themes. The benefit of coding interviews is that it makes it easier to extract and see patterns in responses allowing for common topics and trends to become apparent. This is a standard practice within qualitative research when interviews are used.

The following steps were taken during the coding process:

1. Coding Categories were defined
2. Labels were assigned to the coding categories
3. Classify relevant information into the relevant categories.
4. Test coding for reliability
5. Identify unreliable codes and correct coding as necessary.

The first step was to create a preliminary coding system based on three broad themes: (i) Data Usage (ii) Techniques and Methods (iii) Organizations Culture. The Codes were identified via by the preliminary survey and from previous research (Byrom, 2001; Hernandez and Emmons, 2012) that has identified the role of data in RLDM. Each code had a series of subcodes that were identified via the responses. Table 19 outlines the subthemes for each category.

Table 19: Coding Categories

Themes	Codes	Sub-Codes
Data Usage	Data Dimensions	Data Variety
		Data Volume
		Data Velocity
	Type of Data Growth	Consumer Data
		Data Automation
Techniques and Methods	The environment in which they operate	Sector
		Growth Strategy (e.g. Franchising)
	The nature of the location decision being made	Cost of Decision
		Experience
	Availability of Technology	Data Silos
		SBD Software
Organizations Culture	Corporate Challenges	Senior Management Buy In
		SBD Solutions
		Cost
	The SBD Challenge	Data Warehousing and Mining
		Data Integration and Analysis
		Data Interpretation

Careful consideration was given to the coding categories to ensure that the coding definitions were not too specific or abstract. The coding categories were developed to meet two basic characteristics, identified by Gordon (1992): (i) it is all-inclusive, and; and, (ii) it is mutually exclusive. In order to be all-inclusive, the coding had to include the entire range of relevant response categories and in order to be mutually exclusive, each category must be clear

enough so that responses could not logically fall into two categories (Gordon, 1992). The final coding needed to balance specificity against comparability.

Themes in the interview data were presented with summary statistics, including especially, the frequency count of their mention. Quotes from the interviewees were used extensively in order to add context to the themes discussed. All quotations are discussed in relation to, and contrasted with, the experiences and opinions of other respondents.

A major challenge to using semi-structured interviews was getting respondents to feel comfortable to be recorded. This would have made for a much more natural conversation allowing for the potential of greater dialogue. Furthermore another challenge was that the respondents use different terminology for the same thing. This made the coding process more challenging as it was not as simple as finding key terms. This challenge was overcome by creating a coding key that included all possible terms that would be referring to the same thing.

3.3 CASE STUDIES

Once some basic conclusions were identified on how SBD was used within retail businesses, the final stage of data collection was done through company visits and in depth interviews with several professionals from a variety of departments across their organization. Case studies allowed for more granular level data to be collected allowing for a more holistic view on a retail firms data environment, as it is not just focused on one departments perspectives. Furthermore this allowed for opportunities and challenges for SBD adoption to be more apparent. Case studies were chosen because of their strength in providing detailed

investigations and understandings of organizations process (Baxter and Jack, 2008). The three case studies chosen included: 1) a developer/brokerage/leasing firm, 2) a financial institution and 3) a fast food chain.

Case study research is a very common research practice in retail and business research (Wood, 2002; Ghemawat, and Khanna, 1998; Hindle, and Vidgen, 2018; Hargreaves et al., 2018; Gunasekaran et al., 2018; Jayakrishnan, et al., 2018; Wood et al., 2016 Christmann, et al., 2015; Okeahalam and Wood, 2009). Case studies are particularly useful for discovery, theory testing, classification, hypothesis development and identification of future research needs (Vissak, 2010). More importantly, because case studies do not necessarily need to rely on “previous literature or prior empirical evidence case study research can be used for theory-building even if little is known about the phenomenon” (Vissak, 2010, p. 371). Vissak (2010) also identified that contrasting to methods that rely on conducting statistical correlations and less on their core explanations, case studies can:

- (i) Discover causal relationships
- (ii) Understand how and why everything has happened in a certain way, and
- (iii) Create thick, interesting, and easily readable descriptions and rich understandings of phenomena in their natural settings.

3.3.1 Designing Case Studies

The first step in designing the case study methodology was to identify the type of case study to be adopted. There are three categories of case studies that exist (as defined by Yin

(1994): (i) exploratory; (ii) descriptive; and, (iii) explanatory. For this study, an exploratory case study was used. The central aim was to identify and explore the challenges and opportunities, and more importantly, the success factors associated with SBD adoption in RLDM. According to Yin (2003) this type of case study works well when pursuing answers that look to explain causal links in real-life interventions that are too intricate for survey or experimental strategies.

Furthermore, case studies can either be intrinsic, instrumental, or collective. Table 20 identifies the differences between these types. A collective case study was chosen mainly because it proved to be the best method for identifying the success factors and the hurdles for SBD adoption. A key aspect of the case study was to gain as many perspectives as possible. To truly understand SBD adoption it was important to understand the technological, organizational and personal perspectives that support implementing new techniques, technology and data in RLDM. The only way to accomplish this was to interview as many individuals from a variety of educational and work backgrounds and departments.

Table 20: Types of Case Study in Qualitative Research

Type	Description
Intrinsic	Is typically undertaken to learn about a unique phenomenon. The researcher should define the uniqueness of the phenomenon, which distinguishes it from all others.
Instrumental	Uses a particular case (some of which may be better than others) to gain a broader appreciation of an issue or phenomenon
Collective	Involves studying multiple cases simultaneously or sequentially in an attempt to generate a broader appreciation of a particular issue.

Cassell and Symon, 2004

Another vital decision that had to be made when developing the case studies was whether to use a single case study or multiple case studies. A multiple case study approach was selected for this study. While a single case study can provide valuable information, it is very difficult to extricate what is unique or common between retailers (Baxter and Jack, 2008; Meyer, 2001; Tellis, 1995). The use of multiple case studies provided a way to see firsthand how data and information are used and how it diffuses between different departments responsible for RLDM. Therefore, a multiple case study approach allowed for the explanation of differences within and between case subjects (Gable, 1994; Crowe et al., 2011; Cassell and Symon, 2004).

3.3.2 Selecting the Cases

A challenging issue in running these multiple case studies is identifying which retail organizations were going to be the case subjects as well as how many case subjects were going to be included. It was important to the research to have cases from different retail sectors that vary in size and scale. This allowed for best practices to be compared across the retail industry as a whole therefore allowing differences between sectors to be identified. The same sampling techniques from the semi-structured interviews were deployed: (i) snowball sampling; (ii) criterion sampling; and, (iii) convenience sampling. Through criterion sampling, key retailers were targeted who indicated high, medium and low success at SBD integration and adoption. Whether the company had high, medium, or low success was identified via the online questionnaire and semi-structured interviews. While these were not the only organizations that fell into the categories, they were selected through convenience sampling as they were most

willing to participate. Furthermore, the individuals targeted within these organizations were identified through snowball sampling methods. The main advantage of the snowball sampling method in this scenario is that individuals within these retail firms could be reached that otherwise would be difficult or near impossible to interview and sample.

These case studies were carried out through semi-structured interviews, unstructured interviews, formal and informal discussions, and when possible, on-site visits to the company. Interviews were manually recorded and they were informed that their anonymity would be preserved.

4 RESEARCH FINDINGS

4.1 ONLINE RETAILER SURVEY RESULTS

A total of 43 out of 181 individuals participated in the survey, representing a 23.8% overall response rate. Table 21 shows a breakdown of these respondents by three broad sector groupings: Food retailer, Non-food retailer, and Other. Food Retailers included any retail organization whose primary retail function involved selling groceries or food services, such as, grocery chains, restaurants and fast food chains. Non-Food included any retail firm whose primary purpose is to sell goods that are not food related which includes general merchandise, clothing and clothing accessories, home improvement, home furnishing, electronics and appliances, health and personal care, and hobby stores. The Other category included retail support services such as banking, leasing/brokerages, and retail consultants. Note that as questions in the survey were not all mandatory, sample sizes (n) by question varied from 27 to 43 participants, as indicated in the tabular results.

Table 21: Distribution of Respondents by Sector, On-line Survey

Retail Sector	# Sample	# Respondents	Response Rate (%)
Non-Food Retail	86	15	17.4
Food Retail	30	11	36.7
Other	65	17	26.2
Total	181	43	23.8

The job roles of respondents and the sizes of their companies varied. Larger retailers were targeted because they account for the largest amount of commercial activity in the country. The top 100 retail conglomerates in Canada account for approximately 75% of total non-automotive retail sales in the country (Daniel & Hernandez, 2017). Regarding respondent job roles, 21.6% of respondents were owners or executives, 37.2% senior managers, 21.0% middle management, 16.3% intermediate or senior analysts, 4.7% entry-level positions and 9.3% other. The vast majority of respondent's companies were very large retailers with over 250 locations (68%). The remainder had 100 to 250 locations (8.8%), 10 to 99 locations (20.6%), or less than 10 locations (2.9%).

4.1.1 Spatial Big Data Adoption by Retailers: Types, Importance, Perceptions, and Changes

Respondents reported a wide variety of internal and external data sources that they frequently collect or acquire for RLDM activities, as shown in Table 22. Internal data is collected within the organization while external data are sources that are not collected in-house but rather purchased or acquired from external providers. The RLDM departments used on average 11.4 different data sources. There were some clear differences between what organizations collected internally versus what they acquired from external providers. Demographic and socio-economic data was the most common externally sourced data of over 90% of respondents, whereas point of sale data was a common internal data source for over 60% of respondents, including, own-store sales, pedestrian counts, and customer transaction

data. Internal and external loyalty card program data were used by approximately 40% of the respondents. Loyalty card data is not a new method of data collection, but it is evolving and becoming more common in the retail market-place. Much of this is likely facilitated through the adoption of e-commerce because it provides the technological infrastructure needed to track customer activity.

It is also important to note that 20% of the respondents indicated frequently using social media data. This speaks to the potential integration of unstructured data into the decision-making process along with the adoption of new tools and frameworks that support the collection, analysis, and visualization of social media data.

Table 22: Retailers Frequently Used Data Sources (n=35), Online Survey

INTERNAL Data Set	% of Respondents	EXTERNAL Data Set	% of Respondents
Competitor Location Data	80.0	Census Data (e.g., Demographic and Socio-Economic)	82.9
Own Store Data (e.g., sales, sq. ft, store type)	68.6	Population Projection and Estimate Data	82.9
Traffic / Pedestrian counts	65.7	Adjusted Census Data (e.g., Demographic and Socio-Economic)	80.0
Customer Transaction (e.g., EPOS)	60.0	Business Location Data	85.2
Store Space Planning (e.g., merchandise mix)	60.0	Background Mapping Data (e.g., roads, rivers)	68.6
Customer Survey Data (e.g., exit intercepts)	48.6	Shopping Centre Location / Tenant Data	71.4
Planning Application	45.7	Geodemographic Data	57.1
Store Card Data (e.g., loyalty programs)	40.0	Daytime Population	57.1
Store Credit Card Data (e.g., VISA, Mastercard)	25.7	Consumer Expenditure Survey	45.7
Email/Electronic customer database	22.9	Traffic Data / Pedestrian Count	45.7
Social Media Data	20.0	Lifestyle / Values / Psychographic Data	42.9
Customer Surveillance Data (e.g., mobile tracking)	14.3	Employment Data (e.g., office employment)	40.0
Customer Mobile Application (e.g., retailer apps leveraging location data)	14.2	Planning Application / Land-use Data	40.0
Customer After-Sales (e.g., warranty)	5.7	Shopper Survey Data (e.g., mall intercept)	31.4
Crowdsourced Data	2.9	Loyalty Card Data (e.g., Air-Miles)	31.4

The respondents were asked a series of attitudinal questions to further gauge the *importance* of geographic data as a resource in RLDM (Table 23). Geographic data is viewed as a vital resource in the retail planning process by 93% of the respondents. It is also interesting to note that only 25% of the respondents indicated that their data has migrated to SBD. Perhaps reflecting that the trend towards SBD usage is still very much in the early stages.

When comparing this data to similar questions asked by Byrom *et al.* (2001) in a UK study in 2000, there were some clear changes in the corporate perception of Geographic Data (Table 24). Byrom *et al.* (2001), found that 68% of the respondents agreed with the statement that “*Geographic data are the key to many of our business requirements*” compared to 93% in this survey. This suggests that geographic data has grown to be a vital corporate resource in business decision making. Furthermore, there has been an increase in the awareness of the unique nature of geographic data with only 7.4% of the respondents agreeing “*Geographic data are no different from any other type of data*” compared to 27% in Byrom *et al.* 2001.

Table 23: Attitudinal Questions of the Importance of Geographic Data (n=29), Online Survey

Question	%		
	Agree	Neither Agree Nor Disagree	Disagree
<i>Geographic data are the key to many of our business requirements</i>	93.1	6.9	0.0
<i>Geographic data are unlikely to increase in importance over the next 5 years</i>	28.6	21.4	50.0
<i>Awareness of the geographic element of data is prevalent across our departments</i>	62.1	31.0	6.9
<i>Geographic data are no different from any other type of data</i>	7.4	25.9	66.7
<i>It often seems that we have too much data for our requirements</i>	17.9	42.9	39.3
<i>I leverage geographic data in my role</i>	89.3	10.7	0.0
<i>Geographic data is fully leveraged across my organization</i>	25.0	39.3	35.7
<i>Do you believe that your data has migrated from large datasets towards what has been defined as 'Big Data'?</i>	25.0	46.4	28.6

Table 24: Byrom et.al, 2001 Attitudinal Statements: Geographical Data Issues, Online Survey

Statement	%		
	Agree	Neither Agree Nor Disagree	Disagree
'Geographic data are the key to many of our business requirements'	68	27	5
'Geographic data are unlikely to increase in importance over the next 5 years'	27	19	54
'Awareness of the geographic element of data is prevalent across our department'	63	19	18
'Geographic data are no different from any other type of data'	20	35	45

Respondents were also asked to indicate whether they believed that the datasets they utilize are considered to be SBD as shown in Tables 25 and 26. It is interesting to note that while the majority of respondents (over 90%) indicated leveraging the geographic element of external data sources in their analysis, the majority do not believe that these data sources are considered Big Data, except for externally provided loyalty card data (58.3%), location data from cell phones (75.0%), crowdsourced data (70.0%) and social media data (85.0%). It is important to note that they were utilized the least by the respondents. Furthermore, while approximately 90% of the respondents indicated that geographic data is key to their business requirements, 75% of the respondents do not believe that their data has migrated to Big Data. All respondents, regardless of sector, fully utilizes the geographic element of Census Data (Demographic and Socio-Economic), Daytime Population, Consumer Expenditure Survey, Geodemographic Data, Lifestyle/Values/Psychographic Data, Background Mapping Data (e.g., roads, rivers), Business Location Data, Shopper Survey Data (e.g., mall intercept), Employment

Data (e.g., office employment) and Other government data (e.g., labour force). While the retail sectors are similar in accessing the spatial element of their data

Table 25: Perceptions on External Spatial Big Data, On-line Survey

External Data	Is the 'geographic' element of this data used in your analysis/decision-making?				Do you regard this data as <i>Big Data</i> in your organization?			
	%				%			
	ALL	Non-Food Retailers	Food Retailers	Other Retailers	ALL	Non-Food Retailers	Food Retailers	Other Retailers
Census Data (Demographic and Socio-Economic)	100.0	100.0	100.0	100.0	26.9	14.3	33.3	30.0
Daytime Population	100.0	100.0	100.0	100.0	8.7	0.0	11.1	11.1
Consumer Expenditure Survey	100.0	100.0	100.0	100.0	26.3	50.0	25.0	14.3
Geodemographic Data	100.0	100.0	100.0	100.0	21.7	25.0	22.2	20.0
Lifestyle / Values / Psychographic Data	100.0	100.0	100.0	100.0	20.0	25.0	0.0	33.3
Background Mapping Data (e.g., roads, rivers)	100.0	100.0	100.0	100.0	8.3	0.0	0.0	18.2
Business Location Data	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0
Shopper Survey Data (e.g., mall intercept)	100.0	100.0	100.0	100.0	21.4	0.0	33.3	14.3
Employment Data (e.g., office employment)	100.0	100.0	100.0	100.0	18.2	0.0	25.0	16.7
Other government data (e.g., labour force)	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0
Population Projection and Estimate Data	96.6	100.0	100.0	92.3	24.1	14.3	33.3	23.1
Adjusted Census Data (Demographic and Socio-Economic)	96.2	100.0	100.0	90.9	26.9	16.7	33.3	27.3
Shopping Centre Location / Tenant Data	96.0	100.0	100.0	90.9	8.0	0.0	0.0	18.2
Traffic Data / Pedestrian Count	95.7	100.0	100.0	88.9	17.4	0.0	11.1	33.3
Planning Application / Land-use Data	94.7	100.0	100.0	83.3	15.8	25.0	0.0	33.3
Loyalty Card Data (e.g., AirMiles)	91.7	100.0	80.0	100.0	58.3	100.0	60.0	25.0
Customer Surveillance Data	75.0	0.0	50.0	100.0	25.0	0.0	50.0	0.0
Location Data Collected From Cellphone Pings	75.0	100.0	0.0	100.0	75.0	0.0	100.0	100
Social Media	60.0	100.0	0.0	100.0	80.0	0.0	100.0	100.0
Crowdsourced Data	50.0	0.0	33.3	100.0	75.0	0.0	66.7	100.0

Table 26: Perceptions on Internal Spatial Big Data, On-line Survey

Internal Data	Is the 'geographic' element of this data used in your analysis/decision-making?				Do you regard this data as <i>Big Data</i> in your organization?			
	%				%			
	All	Non-Food Retailers	Food Retailer	Other	All	Non-Food Retailers	Food Retailer	Other
Customer Transaction (e.g., EPOS)	91.3	100.0	85.7	87.5	47.8	37.5	42.9	62.5
Customer After-Sales (e.g., warranty)	42.9	50.0	33.3	0.0	57.1	50.0	66.7	0.0
Customer Survey Data (e.g., exit intercepts)	90.0	66.7	100.0	100.0	30.0	50.0	12.5	33.3
Store Card Data (e.g., loyalty programs)	85.7	80.0	83.3	100.0	42.9	80.0	33.3	0.0
Store Credit Card Data (e.g., VISA, Mastercard)	81.8	100.0	75.0	50.0	45.5	40.0	50.0	50.0
Own Store Data (e.g. sales, sq. ft, store type)	91.3	85.7	88.9	100.0	39.1	42.9	33.3	42.9
Store Space Planning (e.g., merchandise mix)	91.7	100.0	85.7	100.0	16.7	25.0	14.3	0.0
Competitor Location Data	96.4	100.0	100.0	90.9	21.4	25.0	11.1	27.3
Traffic / Pedestrian Counts	95.8	100.0	100.0	87.5	8.3	0.0	0.0	25.0
Planning Application	81.8	100.0	83.3	0.0	18.2	0.0	16.7	100.0
Customer Surveillance Data (mobile tracking)	66.7	100.0	50.0	60.0	50.0	33.3	75.0	40.0
Social Media Data	55.6	66.7	25.0	100.0	77.8	33.3	100.0	100.0
Customer Mobile Apps (retailer apps leveraging location data)	44.4	50.0	0.0	100.0	66.7	50.0	100.0	50.0
Email/Electronic customer database	66.7	66.7	50.0	100.0	50.0	33.3	75.0	50.0
Crowdsourced Data	50.0	0.0	50.0	100.0	75.0	100.0	50.0	100.0

Non-food retailers are more likely to consider the external data sources that they use to be SBD. New modern data sources that are commonly associated with SBD, such as customer surveillance data, cell phone ping data, crowdsourced data, and social media data, are less frequently utilized by food retailers. This could indicate that Non-food retailers may have a smaller appetite for adopting in-house procedures and technologies capable of mining data that can be used in RLDM.

The perception of how SBD usage has changed in the past few years is presented in Table 27. Approximately, 96% of the respondents indicated moderate to significant increases in data volume, 89% indicated moderate to significant increases in data variety and 85% indicate moderate to significant increases in data velocity. When the increases in data volumes were cross-tabulated by sector, two key differences were apparent. First, there was moderate to significant increases in data volume (100%) and moderate to significant increases in data variety (100%) for Non-food retailers. Second, all Food retailers indicated moderate to significant increases in data volume, while 88% of the respondents indicated moderate to significant increases in data variety and 78% of the respondents indicated moderate increases in data velocity for the food sector. This indicates that advancements in corporate data environment are not likely uniform across all retail sectors.

Table 27: How Spatial Big Data Usage has Changed in the Past Few Years (n=28), On-line Survey

Change	Sector	%		
		Volume	Variety	Velocity
Significant Increases	<i>All</i>	46.4	32.1	28.6
	<i>Non-Food Retailers</i>	42.9	28.6	42.9
	<i>Food Retailers</i>	33.3	22.2	33.3
	<i>Other</i>	58.3	41.7	16.7
Moderate Increases	<i>All</i>	50	57.1	57.1
	<i>Non-Food Retailers</i>	57.1	71.4	42.9
	<i>Food Retailers</i>	66.7	66.7	44.4
	<i>Other</i>	33.3	41.7	75.9
No Increase	<i>All</i>	3.6	7.1	10.7
	<i>Non-Food Retailers</i>	0	0	14.2
	<i>Food Retailers</i>	0	11.1	22.2
	<i>Other</i>	8.4	8.3	0
Decrease	<i>All</i>	0	3.7	3.6
	<i>Non-Food Retailers</i>	0	0	0
	<i>Food Retailers</i>	0	0	0
	<i>Other</i>	0	8.3	7.4

4.1.2 Techniques and Methods Adoptions by Retailers: Types, Importance, Perceptions and Changes

To understand how data is utilized, current locational decision-making practices were examined, as shown in Table 28 and Table 29. Approximately two-thirds of the respondents (68%) indicated that they had experienced an ‘increase’ or ‘significant increase’ in the number of retail locations operated over the last five years and 64% of respondents expected ‘increases’ or ‘significant increases’ in retail property counts over the next five years. Of interest is the

23.5% of the respondents that indicated that they will experience a 'decrease' in their retail property counts. While collectively there appears to be an increase in the total store counts for most organizations, there were some important sector differences. Most notably, while most sectors have already experienced decreases in retail locations, food retailers did not indicate any decrease. This likely has to do with the nature of both the grocery and food service businesses not being affected by e-retail in the same way as other retail sectors.

Table 28: Changes in the Number of Physical Locations Counts over the last 5 years (n=34), On-line Survey

Type of Change	%			
	<i>All</i>	<i>Non-Food Retailers</i>	<i>Food Retailers</i>	<i>Other</i>
Significantly Increased	17.7	27.3	27.3	0.0
Increased	50.0	54.6	54.6	41.7
Remained the same	11.8	9.1	18.2	8.3
Decreased	17.7	9.1	0.0	41.7
Significantly Decreased	2.8	0.0	0.0	8.3

Table 29: Expected Future Change to Physical Locations Counts over the next 5 years (n=34), On-line Survey

Type of Change	%			
	All	Non-Food Retailers	Food Retailers	Other
Significantly Increased	14.7	36.4	9.1	0.0
Increased	50.0	45.5	81.8	25.0
Remained the same	11.8	18.2	0.0	16.7
Decreased	23.5	0.0	9.1	58.3
Significantly Decreased	0.0	0.0	0.0	0.0

The nature of the data management techniques and strategies used within respondent's organization are shown in Table 30. Over 88% of the respondents indicated that their data analytics function was managed in-house. It was also evident that RLDM departments rely on data-driven decision-making. Over 80% of the respondents indicated that their company is oriented towards decisions that are supported by data analytics, and that metrics drive their decisions. However, 70% of the respondents indicated that experience is the most important factor when making decisions in the retail industry. Additionally, whilst many respondents indicated they are orientated towards data-driven decision-making, about one-third indicated that their techniques were archaic and outdated. Together, this suggests that while data analytics are important, they must be interpreted using an experienced lens to extract true meaning and organizational context.

Table 30: The Nature of Data Management Techniques and Strategies (n=29), On-line Survey

Statement	%		
	Agree	Neither Agree Nor Disagree	Disagree
Our company is oriented towards decisions that are supported by data analytics	84.6	7.7	7.7
Metrics drive our decision making	80.8	11.5	7.7
We manage our analytics in house	88.5	7.7	3.9
The recommendations our department makes are rarely accepted by senior management	7.7	11.5	80.8
We fully leverage 'Geographic Big Data' in our Organization	30.8	38.5	30.8
'Geographic Big Data' are a vital part of our department's decision-making processes	50.0	26.9	23.1
Analysts understand the techniques they are using	84.6	11.5	3.9
Our decisions are based on detailed analysis and research	80.8	11.5	7.7
Multiple techniques are employed for any single decision	73.1	19.2	7.7
Experience is the most important factor when making decisions in the retail industry	70.4	22.2	7.4
Model accuracy is let down by inaccurate source data	61.5	30.8	7.7
We often do not have the time to undertake in-depth analysis	44.4	29.6	25.9
We employee Data Scientists who are responsible for supporting our location decisions	15.4	30.8	53.8
We outsource our data to a third part	15.4	23.1	61.5
Analysts understand the techniques they are using	65.4	30.8	3.9

New employees are required with academic backgrounds in mathematics and data science	30.8	30.8	38.5
Our techniques for location decision making are archaic and outdated and need to be postdated	33.3	25.9	40.7

The relative use of established and emerging RLDM techniques is shown in Table 32 (Table 31 provides definition for some of the emerging techniques). New modern forms of location decision-making techniques such as machine learning, social media analytics, social influence analysis, sentiment analysis, real-time demand forecasting, text analysis and live visual analytics, are not often used, as over 50% of respondents indicated rarely or never using these types of techniques. Some organizations are just starting to implement the use of modern techniques such as association rule learning, real-time demand forecasting and live visual analytics in their decision-making process. However, the majority of respondents still frequently rely on traditional techniques including experience (96.4%), checklists (85.7%), analogues (88.5%), multiple regression (59.2%) and spatial interaction models (55.5%). Similar to Woods and Reynolds (2010), and Hernandez and Emmons (2012), the use of experience was the most frequently used location technique. Furthermore, comparable to Hernandez and Emmons (2012), the more sophisticated techniques were less frequently used. With both spatial interaction and regression models used the least frequently.

Table 31: Emerging Location Decision Making Techniques Definitions

Technique	Definition (from Gandomi and Haider, 2014)	Research Literature
Predictive Analytics	Comprised of a variety of techniques that predict future outcomes based on historical and current data. They are primarily based on statistical methods.	Waller and Fawcett, 2013
Machine Learning	A form of predictive analytics where computers automatically develop new knowledge.	Jordan and Mitchell, 2015
Social Media Analytics: Content-Based Analytics	Focuses on the data posted by users on social media platforms, such as customer feedback, product reviews, images, and videos.	Batrinca and Treleaven, 2014
Social Media Analytics: Structure-Based Analytics	Is concerned with synthesizing the structural attributes of a social network and extracting intelligence from the relationships among the participating entities.	He <i>et al.</i> , 2015
Social Influence Analysis	Refers to techniques that are concerned with modelling and evaluating the influence of actors and connections in a social network.	Kim and Srivastava, 2007
Sentiment Analysis	Techniques analyze opinionated text, which contains people's opinions toward entities such as products, organizations, individuals, and events. Also known as opinion mining	Ravi and Ravi, 2015
Association Rule Learning	Used to analyze and predict customer behaviour by looking at past behaviours	Prasad, 2011
Real Time Data Visualization	Visualizing customer location data and distributions in real-time	Sagiroglu and Sinanc, 2013
Real-Time Demand Forecasts	Forecasting behaviours in real-time	Beheshti-Kashi <i>et al.</i> , 2014
Text Analytics	Refers to techniques that extract information from textual data (e.g. Social network feeds, emails, blogs, news).	Lim <i>et al.</i> , 2013

Table 32: Retail Location Decision-making Techniques (n=28), Online Survey

Type	Technique	%		
		Always/Of ten	Sometimes	Rarely/Never
Established	<i>Experience</i>	89.3	7.1	3.6
	<i>Checklist</i>	53.6	32.1	14.3
	<i>Analogue</i>	65.4	23.1	11.5
	<i>Multiple Regression</i>	29.6	29.6	40.7
	<i>Gravity Models/Spatial Interaction</i>	37.0	18.5	44.4
Emerging	<i>Predictive Analytics</i>	40.7	22.2	37.0
	<i>Machine Learning</i>	14.8	18.5	66.7
	<i>Social Media Analytics: Content-Based Analytics</i>	3.7	18.5	77.8
	<i>Social Media Analytics: Structure-Based Analytics</i>	3.7	14.8	81.5
	<i>Social Influence Analysis</i>	7.4	7.4	85.2
	<i>Sentiment Analysis</i>	14.8	18.5	66.7
	<i>Association Rule Learning</i>	14.8	18.5	66.7
	<i>Real-Time Data Visualization</i>	30.8	19.2	50.0
	<i>Real-Time Demand Forecasts</i>	22.2	22.2	55.6
	<i>Text Analytics</i>	3.7	11.1	85.2

There appeared to be a significant amount of variety related to the adoption of retail location applications, as seen in Table 33. Respondents indicated that competitor analysis (92%), market mapping (89%), cannibalization (89%), trade area identification (86%), and site screening and site selection (83%) were operationalized most frequently (Table 34). These were likely most frequently used because they work well with existing GIS technology. There are also some notable sector differences, for example, Food Retailers perform logistics planning more frequently compared to the other sectors. This could be a result of food retailers handling perishable products making it imperative for them to streamline logistic planning. Also, all food sector respondents (100%) indicated that they conduct cannibalization analysis on a frequent

basis. Similar to Hernandez and Emmons (2012), competitor analysis, market mapping, site screening and trade analysis are still applications undertaken on a regular basis. A major difference is the increased frequency of cannibalization analysis being implemented by the respondents. This could potentially be linked to the growth in e-commerce, reducing the need for retail locations as well as a greater territorial reach in shoppers.

Table 33: Utilization of Location-Based Applications on a Regular Basis (n=36), On-line Survey

Application Type	%			
	All	Non-Food Retailers	Food Retailers	Other
Competitor analysis	91.7	81.8	100.0	93.3
Market mapping	88.9	81.8	90.0	93.3
Cannibalization	88.9	81.8	100.0	86.7
Trade area identification	86.1	63.6	90.0	100.0
Site screening and selection	83.3	63.6	100.0	86.7
Network planning	66.7	45.5	80.0	73.3
Setting sales targets	58.3	36.4	90.0	53.3
Monitoring outlet performance	55.6	54.6	60.0	53.3
Store portfolio segmentation and planning	52.8	45.5	60.0	53.3
Customer profiling	50.0	36.4	50.0	60.0
Acquisition and merger planning	33.3	27.3	60.0	20.0
Customer database planning	25.0	18.2	30.0	26.7
Merchandising mix analysis	22.2	36.4	30.0	6.7
Promotional/media analysis	13.9	0.0	30.0	13.3
Targeting direct mail	13.9	0.0	40.0	6.7
Logistics planning	11.1	18.2	20.0	0.0

4.1.3 Business Culture: Opportunities and Challenges for Awareness, Availability, Use and Adoption of Spatial Big Data

The top reasons why respondents experienced difficulties integrating data into their decision-making are outlined in Table 34. A lack of understanding from senior management on how analytics/SBD can be used to improve decision-making proved to be a major concern among the respondents. This was apparent as over half of the respondents identified cost (55.6%) as a major challenge. Other significant challenges included: the ability to transfer data findings into meaningful solutions (44.4%); the ability to obtain data (37%) and software (30%); concerns that perceived costs outweigh the benefits (30%); and, information hoarding (30%). Most interesting was the fact that, despite the growth of SBD, one-third of respondents indicated that obtaining data was a major challenge in data integration. This speaks to the significant differences between organizations in terms of their receptiveness, suitability, and capacity to handle more complex data. Food retailers were more likely to identify cost, buy-in from senior management, software availability and perceived return-on-investment as being major difficulties with data integration, when compared to other sectors.

Table 344: Difficulties with Data Integration in Retail Location Decision-making (n=27), On-line Survey

Difficulties	% in Agreement			
	All	Non-Food Retailers	Food Retailers	Other
Cost	55.6	42.9	75.0	50.0
Ability to translate data findings into meaningful solutions	44.4	28.6	12.5	75.0
Ability to obtain data	37.0	28.6	50.0	33.3
Finding Talent (Lack of internal skills to manage and handle data)	29.6	28.6	25.0	33.3
Buy in from senior management	29.6	14.3	50.0	25.0
Software Availability	29.6	14.3	50.0	25.0
Perceived costs outweigh the benefits	29.6	14.3	62.5	16.7
Information Hoarding (data is not shared between departments)	29.6	42.9	37.5	16.7
Do not have the capacity of using Big Datasets	25.9	28.6	25.0	25.0
Lack of understanding of how to incorporate the data to improve business decisions	22.2	14.3	25.0	25.0
Concerns with data integrity	18.5	28.6	25.0	8.3
Data are not accessible across your organization	14.8	0.0	12.5	25.0
There are no tools and expertise available to work with unstructured real-time data	11.1	0.0	12.5	16.7

Table 35 highlights the role that management plays within RLDM. Table 31 clearly shows upper management support for data-driven decision-making however, the differences among the respondents centred on the role that SBD plays. Less than half of the respondents (44%) indicated that senior managers fully buy-in to SBD analytics and the majority do not agree that SBD is a corporate resource (46%). While the vast majority of senior managers support data-driven decision-making (81%), retailers appear to be conservative in their adoption of SBD and

SBD analytics. Data hoarding was also found to be an issue, with only 23% of respondents indicating that they disagree with the statement that ‘data resources are tightly controlled in department silos.’

Table 35: Senior Management’s role in Retail Location Decision-Making, On-line Survey

Senior Management’s Role	%		
	<i>Agree</i>	<i>Neither Agree Nor Disagree</i>	<i>Disagree</i>
Our Senior Management fully buy into Big Data analytics	44.4	40.7	14.8
Management support data-driven decision-making	81.5	14.8	3.7
Big Data is viewed as a corporate resource	46.2	38.5	15.4
Data resources are tightly controlled in department silos	38.7	38.5	23.1

4.2 STRUCTURED INTERVIEWS

The results of the structured interviews are organized by question theme. The major themes are broken down into three broad categories: Data Theme, Data Requirements for Method and Technique Implementation and Organizational Culture.

4.2.1 Data Theme

All 24 interviewees indicated that there was growth in the *volume of data* available within the organization. The interviewees documented that this increase in data volume was coming from both external and internal data sources. Customer transactions, for example, are increasingly leaving a digital imprint and therefore organizations are looking at ways to capture

this information. Nineteen interviewees indicated that the major increase in customer data was a direct result of internally generated loyalty card data and growth in online shopping. The 5 interviewees that did not indicate growth in customer data were all from the food services sector.

Seventeen interviewees reported significant increases in the sophistication that is – using new or different techniques, to improve the collection of customer information, though some indicated this to be a major challenge. All of the food service interviewees (7) indicated that this was a challenging process.

“More customer data would be great but it’s difficult.” Fast Food

Retailer

“No customer loyalty data ... we are developing a loyalty program

to access our own data.” Casual Dining Retailer

These food service interviewees were asked why it was so difficult to get customer information. The challenge was commonly linked to the nature of their business. For instance, 2 of the casual dining interviewees indicated that it was a result of the fact that they lacked one single point of contact (like a centralized cashier) and had to rely on service staff who are more focused on service quality.

“It’s harder when your full service ... there is no single point of

guest interaction. You rely on the server’s... their main focus is

getting food to the table and not screwing up an order.” Casual

Dining Retailer

These companies were also less likely to have loyalty programs in place. While every single retailer indicated that the development of an internal loyalty program or the improvement of an existing program as a major priority, 6 of the 7 food service retailers indicated that their business environment did not lend itself to effective integration of these types of programs.

“...food service is not ubiquitous like GAS. Before everything else

food quality, experience and service matters more than

points...therefore loyalty becomes a secondary motivation when

deciding to visit one restaurant over another.” Casual Dining

Retailer

Half of the interviewees (12) reported seeking more granular level data in order to gain significant insight into consumer spatial behaviour. While traditional transactional data was identified as valuable by virtually all interviewees (23), it doesn't always provide an exact measure of individual behaviour as many of the methods of data collection have significant flaws. A major challenge of loyalty programs, regardless of whether they were internally operated or out-sourced, was a lack of participation as highlighted by 21 interviewees. For example, loyalty programs do not typically capture all customers as it is rare to have 100% consumer adoption and usage. Fourteen of the interviewees indicated that while they have

long been collecting customer data (some for over 50 years) at the Point of Sale (POS), they are now undergoing a shift in the data being collected, specifically as it pertains to variety. For instance, 11 interviewees reported they are looking at the potential of triangulating their POS data with data collected through a variety of technologies that are able to track customer movement both inside and outside of the retail locations. Four-fifths of the Developer/Leasing/Brokerage interviewees indicated that they have already implemented tracking technologies including advanced traffic counters, digital signage with retina recognition, internal positioning systems, as well as, mobile phone tracking.

“Digital signage in our shopping centres have eye-tracking and facial recognition used to get demographic data.”

Developer/Leasing/Brokerage

While it appeared that the development sector was the most active in this area, they were not the only ones that indicated they were adopting new methods of tracking customers. Three of the five fast food interviewees (representing the largest retail organizations in terms of store count) were looking at adopting a variety of consumer tracking technologies such as facial recognition software and mobile phone data. It is important to note that 7 interviewees indicated it was difficult to extract insights from customer data, including all the Financial Sector respondents, the Pharmacy and Personal Care Respondent and one Developer/Leasing/Brokerage respondent.

"We are not using new techniques...the tools have not changed, analysis has not changed and does not look like it is changing... Risk mitigation will not improve...we will paralyze the business if we use more data." Pharmacy and Personal Care Retailer

"We are now looking at cell phone pings to track people during the day" Fast Food Retailer

While this real-time data was viewed as vital for some (9 interviewees), many (14 interviewees) indicated that they were not concerned with real-time data analysis. Six of these interviewees indicated that while there may be benefits for marketing purposes, it is not realistic to leverage such data when making long-term retail location commitments.

"Hard to use real-time data. We are making 20-year decisions and this is not happening on yesterday's data." Grocery Retailer

"Day to day data is not important ... period data is important."

Pharmacy and Personal Care Retailer

A catalyst for the increases in the variety of data sources was linked to the automation of consumer data collection and the ability for data to be generated from a diverse amount of digital devices, as indicated by 19 interviewees. This automation included the following: customer surveillance data, cell phone ping data, crowdsourced data, and social media data.

However, few interviewees reported effective integration, adoption or utilization of this data into RLDM. Only 3 interviewees indicated that they were able to leverage these data sources and none of the interviewees indicated that they were effective at this in RLDM process. All the respondents that identified the increase in variety coming from the automatically collecting data (19) identified that it was very difficult to establish a single depiction of each customer across multiple sources of customer information (point-of-sale, loyalty program, social media, etc.). Therefore, a major challenge appears to be finding a way to identify the various ways the same customers are represented across different methods of collection.

“An inability to do this presents problems with filtering and compressing data in a usable way”. Developer/Leasing/Brokerage

4.2.2 Data Requirements for Method and Technique Implementation

The interviewees were asked a series of questions around the nature of the data used within the methods and techniques they deployed in their RLDM process. There was a great deal of variability in the adoption of RLDM techniques that were able to accommodate these new dimensions of data (Variety, Volume, Velocity). The level of decision-making sophistication being adopted was linked to a number factors including; (i) the business environment in which they operate; (ii) the nature of the location decision being made; and, (iii) availability of technology.

4.2.2.1 The Business Environment in Which They Operate

The likelihood of adopting more advanced methods for decision-making was linked to internal and external factors related to the business environment. For instance, 3 of the fast food interviewees and 2 of the casual dining respondents, for whom franchising was the dominant growth strategy for these firms, indicated that when franchisees were involved in the retail location decision process, the use of more sophisticated techniques and data become less prevalent in comparison to retailers that were opening and operating corporately owned stores. Companies that relied on franchising as a corporate growth strategy encountered a unique set of challenges. These 5 interviewees all indicated that the franchisees' willingness to invest in a property would hold precedence over any other form of retail decision-making. In other words, if a potential franchisee was adamant about a specific location, it was common practice for the corporate real estate departments to move forward with the location acquisitions without relying on data-driven analytics. Moreover, 15 of the interviewees worked for organizations with multiple brands, of which, 10 indicated that they had to rely on more data-driven decisions because of needing to manage multiple brands. In other words they indicated that because of the complexities of selecting the right brand for the right market, they had to lean heavily on data. Twelve of these fifteen interviewees indicated that they feared cannibalization amongst their brands.

*"We need to apply advanced analytics and detailed customer data
in order to make sure that our brands are reaching the right
customers" Grocery Retailer*

Another common environmental factor influencing the use of RLDM techniques expressed by 15 interviewees was the fact that they pay close attention to how competitors make business decisions. Specifically, they stated that they keep tabs on the data inputs and the use of technologies within decision making from comparable retailers in the same retail sectors. To put it simply, if there is a perception that a company's competitors are engaged in more advanced data-driven decision-making, then it becomes a greater priority.

Based on the interviews it appears that some sectors are better equipped to collect data. All financial sector interviewees (5) indicated having access to the largest volume and variety of internally collected data when compared to the other respondents. Participants were asked which data generating methods were utilized when generating data sources. Specifically, participants were asked whether data was generated by: (i) created data; (ii) provoked data; (iii) transaction data, and, (iv) captured data methods (Table 36). The financial sector interviewees were the only ones that indicated that they actively generated data from all 4 of these methods. Four out of the five interviewees indicated that the main reasons for generating data from all these sources were based on their own personal feeling that having a large data repository could help with all aspects of their business including marketing, real-estate, and operations. It is important to note that the collection of data did not necessarily indicate that the data generated was being utilized on a regular basis. These financial sector interviewees only regularly used created data and transaction data for RLDM, rarely integrating provoked and captured data. Outside of the financial sector respondents, no other participants indicated that they collected data from these 4 methods of collection. The next closest to this was the Developer/Leasing/Brokerage interviewees as 4 participants indicated that they generate data

by created, transactions, and captured methods of collection. Three interviewees did, however, indicate that while transaction data was collected, it was not provided daily nor were they able to see individual purchases but rather they were provided with aggregates over an extended period of time (typically monthly).

Table 36: Data Generating Methods

Data Generating Methods	Definition	Example
Created Data	Is created because it would not exist unless a mechanism was put in place to collect that information.	Loyalty programs, Market research Surveys, Asking for Postal Codes
Provoked Data	Would not exist unless you invited people to express their views.	Product review, service review, etc.
Transaction Data	Generated every time a customer makes a purchase.	POS
Captured Data	Is information gathered passively from an individual's behaviour	GPS from mobile devices

Mar,2015

An organization having access to data does not necessarily indicate that they are using the data for decision-making. The financial sector interviewees (5) indicated that they only regularly used created data and transaction data for RLDM and that they rarely integrate provoked and captured data. Furthermore, a total of 16 of the interviewees indicated that a lot of the data that was used was too detailed and not useful and thus SBD adoption was not present in their departments - that is, research departments or real estate departments were not responsible for SBD or SBD analytics integration. They indicated that this is left in the hands

of marketing departments whose objectives were more in line with using advanced internal data sources and targeting customer groups and individual customers.

4.2.2.2 The Nature Of The Location Decision Being Made

It was a common trend that the more complex a retail location decision is, the greater a reliance is placed on data-driven decision making. It is important to note that the interviewees, more often than not, defined complexity as the cost of the decision, that is the higher the cost, the greater the complexity and risk. Twenty interviewees indicated that experience was an effective tool for driving high-level ideas from a creativity perspective and aids in validating decisions on the ground, but cannot be used in order to make more intricate decisions. For example, understanding complex relationships regarding customer interaction is not possible through experience alone. The following are several statements that reinforce this idea.

“Big Level decision are never made without data analytics...”

Financial Sector

“Experience alone does not dig deep enough... it’s impossible to understand implications of store size... through experience.”

Grocery Retailer

“Some things cannot be categorized in data so we rely on the experience of our brokers to aid in the decision process” Casual

Dining Retailer

“Intuition and Experience will never disappear... we start off with a qualitative thought and we try to support it and quantify it with data.” Financial Sector

Irrespective of the growth in the availability of data, experience continues to be an important factor when making retail location decisions, as was apparent in the results of the online survey. Twenty-one interviewees indicated that while there is strong growth in data-driven decision-making techniques, experience is still vital in order to extract true meaning and organizational context. Seven interviewees even argued that experience was more important than ever given the data deluge. They stated that there was a greater need to filter through data to find the data that is relevant for RLDM.

“Greater reliance is being placed on departmental experience in order to filter down the data in a more meaningful way... It is hard to filter through the noise.” Grocery Retailer

Conversely, some of the interviewees were looking to break away from having a human element in the decision-making process altogether. Four of the food service companies, retailers who historically never or very rarely relied on data-driven decision making, wanted to remove the experience element completely. These retailers actually indicated that the fear of lawsuits from franchises, as a result of cannibalization, is the biggest catalyst for no longer wanting to rely on intuition or experience alone.

“We have made decisions solely based on intuition and it has not bode well...Experience is still a vital resource but we do not want to make decisions without the validation of data.” Fast Food Retailer

4.2.2.3 Availability of Technology

An organizations ability to integrate more data is one that relies heavily on the available technologies within the organizations. The SBD gap was largely linked to issues with data mining related technology. Sixteen interviewees indicated that they were challenged by a lack of computer processing capacity to discover patterns in SBD sets. These issues presented problems with adopting machine learning, statistics, and database systems. It was unanimously expressed that the only way to leverage more granular level data into the decision-making process would be to develop or incorporate better software capable of handling SBD. Twenty-one interviewees further indicated a clear need to fuse the new data that is being collected with technological infrastructure that is capable of allowing for more granular level data to be integrated, ultimately improving the quality of decisions being made. The adoption of software capable of handling nontraditional forms of data (unstructured and semi-structured data) has not been realized by many RLDM departments (14 interviewees). Twenty interviewees indicated an inability to handle these new data sources.

“There is a need for more data but retailers need to become IT companies in order to effectively integrate the data” Grocery Retailer

Furthermore, the ability to get the data appears to be a significant challenge for all respondents. All of the interviewees indicated data silos to be a problem, mainly due to the fact that cloud computing and other data storage is not accessible across the organization.

“We know the data is out there (through social media and other sources)...we lack the infrastructure to accommodate that data.” Grocery Retailer

There is a significant lack of awareness between departments of the data sources that are available. While some (3 interviewees) are actively breaking down these barriers, the majority are not. “We are looking at introducing a data lake for all institutional data to be accessed by everyone within the organization” Developer/Leasing/Brokerage

4.2.3 Organizational Culture

There were some clear, corporately driven obstacles to SBD adoption that were highlighted by the interview participants. One major impediment stated by all of the interviewees was a lack of understanding of how to use SBD analytics to improve RLDM. There is clear confusion amongst the interviewees in regards to the RLDM solutions that SBD can

offer. As a result of the increased rhetoric that SBD can present unmatched opportunities to extract greater insights there appears to be strong motivations for retailers to collect customer data. It is evident that retailers are getting lost in a sea of data (indicated by 21 interviewees) and that without more clearly definable deliverables, it is ultimately providing no valuable output. Most interviewees indicated that while that data may be novel and 'cool,' it doesn't actually provide any decision-making benefit.

"But just because we can measure, monitor and access everything doesn't mean we should." Pharmacy and Personal Care Retailer

"I am not really sure what the benefit is...Constantly collecting new data but it adds little value" Financial Sector

"We buy data to prove the mall is not dying. Not actually changing how we make decisions." Developer/Leasing/Brokerage

*" People are slow to adopt solutions... It's difficult because it's overwhelming...What does it (data) mean?...How do we use it?...New data has brought new problems.
Developer/Leasing/Brokerage*

Another corporate challenge seems to be related to a lack of executive or senior manager buy-in for SBD adoption. In the 14 organizations where senior management is onboard with adopting SBD and SBD analytics, there appears to be a greater appetite to incorporate new methods, ideas and product innovations. This really seems to be a top-down issue. The departments responsible for RLDM don't appear to be well versed or capable of looking at non-traditional data sources to make evidence-based decisions. Wherever new methods are being adopted, senior managers are attempting to leverage new employee skillsets by bringing in individuals with knowledge of machine learning methods, predictive modelling and data retools that can handle both structured and unstructured data. This was apparent in 10 of the organizations surveyed as they indicated that they either recently hired or were actively searching for employees with a strong data science background. In some situations (3 organizations) much of this is actually outsourced to third party companies. In the 12 instances where upper management have not looked at advancing their decision-making capabilities to include SBD practices, 7 reported competing priorities as being a significant challenge. This seemed to be strongly related to tight budgets creating an inability to accommodate any innovation in terms of software, data or even expertise. Furthermore, 9 interviewees expressed serious displeasure with the current state of decision-making at their organization. This was tied to a lack of advancement and progress in the tools used for RLDM.

"Just keep doing what you are doing ... May not be right ... the process of using data has not changed even though data and markets changing ... if you ask me it is ridiculous." Pharmacy and Personal Care Retailer

Interviewees (20) indicated that the perceived costs of SBD and SBD analytics overshadowed the projected benefits it would provide.

“Politics vs numbers. Definitely there are still senior people who make decisions. The last 3-5 years people are starting to trust the process a bit more.” Grocery Retailer

Eleven interviewees indicated that they did not feel like more data nor new models would be able to increase the ability of their department or organization to make better decisions. One general merchandise retailer even indicated that the success of a retail location was solely driven by demographics and incorporating any other variables or new methodologies would be a complete waste of time.

4.3 CASE STUDIES

This section provides an in-depth look at three organizations in order to (i) develop an understanding of SBD in RLDM, (ii) understand the diffusion process for SBD and SBD analytics and, (iii) identify the success factors for effective awareness, availability, use and adoption of SBD. Each case study included a different number of interviewees. These are outlined below. The data on each organization was carried out through the use of semi-structured interviews, unstructured interviews, formal and informal discussions, and when possible, on-site visits to the company.

4.3.1 Case 1 – Developer/Brokerage/Leasing Company

The first case study was of a large Canadian development/brokerage/leasing company. The data collect for Case 1 was obtained through conversations with 4 individuals. Three individuals were within the Analytics, Insights & Innovation (AI&I) (Director, and 2 Analysts) department and 1 individual was from Strategic Marketing (Manager). Data was collected through multiple in-person meetings and phone conversations.

Enterprisewide data-centric RLDM is controlled and supported by the Analytics, Insights & Innovation (AI&I) team which is housed under Asset Management. This department consists of a small team who are responsible for spearheading the organization's data science initiative. The newest member of the team is the only employee with both educational and work related experience in data science. In essence, this department acts as a support team, helping to minimize risk when any location-based decision is being undertaken (from leasing through to

marketing). The department is responsible for data-driven consulting when any decision has to be made outside of acquisitions. Acquisitions are made by a separate team, with research capabilities. The acquisitions team is mainly responsible for looking at new developments or the purchasing of existing retail (mainly shopping centres) or office space.

AI&I is currently in a transition phase as they are attempting to overhaul the entire data operation. This is proving to be a major challenge as the organization actively expressed concerns over talent issues.

“New people or full-on training had to happen to keep up with the current situation. Or overhaul and hire new people”.

“Everyone does one job. Limit the number of hats people wear... this means nothing gets done effectively.”

All three members of this team indicated that they were overworked because they were still responsible for the day-to-day tasks (what they were initially hired to do) but have now added new initiatives to help extract untapped potential either from their current data or by acquiring new data sources.

It was identified that the amount of reliance on data-driven decision-making largely depends on the type of decision being made. For example, the AI&I team noted that within the Leasing department about 80% of decisions are largely based on gut feel while only 20% rely on data analytics. The Leasing department largely relies on personal relationships with potential retailers therefore they were not likely to ask AI&I to support a decision. Furthermore, retailers

often conduct their own analysis and have usually already made a location decision before approaching the Leasing department to work out the details of the deal.

“...the Leasing department was not obligated to consult with us (AI&I) when working out lease agreements with potential tenants...They only use for Ad-Hoc requests as they see fit”

It was indicated that the major issue with this is the fact that there is a real lack of automation, making it hard to rely on data when decisions need to turn around faster than the data processing time. It was identified that this was changing within this organization as there is an active push from Upper Management (c-level individuals), indicating that this needs to stop.

“Our President gets it and is willing to invest in all our new initiatives”

4.3.2 Data and Technology Environment

The initiative to overhaul the data operations stemmed from pressure placed on by the Leasing department because they, in turn, are under pressure from the retailers. Retailers are now requesting a larger volume and variety of data so that they can have a more holistic view of the potential performance that a given shopping centre may provide.

“This helps sell retailers when you say you have all this data and did all this market analysis”

“Helps to keep the mall relevant given the challenges with e-commerce”

Table 37 identifies a list of the data sources mentioned by participants as being used regularly at this organization.

Table 37: Data Sources Frequently Used

Description	Source	Accessibility	Volume
Visitor Demographics	Aislelabs	API	TB
Visitor Path Behaviour	Aislelabs	API	TB
Traffic Counter	A1 Solutions	API	TB
Retail Sales	Oasis/JDE	Sql Server	GB
Survey Results	Questback/Aislelabs	Flat files	MB
Demographics by Geography	Enviroics	Flat files	MB
Financial	Bloomberg	API	GB
Traffic	TomTom	API	GB
Real Time Events	N/A	API	GB
Weather	N/A	Aislelabs API	GB
Digital Signage Cameras	Quividi/Pattison	Export	TB
Floorplan Distances	MappedIn	Flat files	MB
Interactive Directory Search	MappedIn	API	TB
Social Media	Hootsuite	Flat files (potential API)	MB
Website Users	Google Analytics	API	GB
Email/Newsletter	Internal	Flat files	MB
Property Level Financials	Oasis/JDE	Sql Server	GB
Occupancy, Vacancy, Rollover Data	Oasis/JDE	Sql Server	GB
Closed / Forecasted Leasing Deals	Oasis/JDE	Sql Server	GB

With this added pressure from retailers, the type of data analysis being conducted is starting to evolve and change within this organization. Initially, this department was solely responsible for doing fundamental trade area profiles as a means to highlight the demographic and socio-economic composition of markets. This was almost exclusively used to support the leasing team. Furthermore, the customer data collected within these organizations did not

provide a very detailed perspective on consumer behaviour within individual shopping centres. Traffic counters situated in doorways were the only method utilized to collect indoor traffic data even though it was made apparent that these were not an effective method for data collection. These counters lacked the ability to provide any detail on traffic flow. Also, real-time data counts were not available and therefore it was very difficult to understand temporal changes in consumer travel patterns. This hurdle acted as a motive for change mainly because it was viewed as an absolute necessity to have that level of data (at a minimum) to remain competitive.

“Major pressure coming from retailers wanting to know granular level mall information before they make leasing commitments”

This mentality also triggered the introduction of digital signage with the ability to collect retina and facial recognition data, providing patron demographic information. Formally, on-site surveys were done to obtain customer data, but this has now transitioned from paper-based to something done via smartphones.

“We had to rely on mall intercept surveys to identify demographic data on consumers. We are now implementing facial recognition software to gain a better understanding of who our customers are”

If a customer logs on to the free Wi-Fi service available within their shopping centres, the mall can push surveys and questionnaires to customers. This organization offers incentives to participants to help with response rates by sending emails while people are actually in the shopping centre; the shopping centre owner/manager can communicate with the patrons,

however, the retail tenants are still not able to directly access this information. Significant strides are being taken to achieve this level of communication and customer engagement with the program roll out scheduled for 2019.

They have also installed an indoor positioning system (IPS) in most of their shopping centre locations. This was done by an external technology company (based out of Toronto) who is responsible for both the installation of the sensors and for providing the automatic traffic count and flow data. Through Bluetooth enabled smartphones, these beacon sensors can locate patrons within a predefined time-frame. The travelling signals (pings) being produced are frequently measured to calculate the distance between the patron and the beacon sensor. The data is made available in two ways: (i) specific data requests made to the technology provider; and, (ii) real-time summary statistics viewed on a dashboard via a cloud server. Therefore, the raw data that is produced by the sensor is not actually managed in-house. A team of data engineers (employed at the technology company) process the data into a readable format that allows for individual travel routes to be determined. The sensors are placed near shopping centre and tenant entrances which helps determine customer flows - that is where shoppers enter and leave the shopping centre and which stores are being visited.

4.3.3 Retail Location Decision-Making

The AI&I department has several tasks that they are responsible for. One task is creating what is referred to as pitch books. These are put together for the leasing team in order to assist when pitching a new tenant. These pitch books are comprised of market area data (socio-

economic and demographic), shopping centre sales data, and shopping centre traffic count data that is generated from indoor sensors. This unit largely relies on traditional GIS based approaches (drive distance and drive time) to create market plans. The organization is also responsible for forecasting site performance and performing market research – that is, gathering information about the market areas of their shopping centres and their customers. The AI&I team is also responsible for ad-hoc requests for demographic data from other business units within the organization. Sales forecasting is not undertaken by the AI&I team, but rather the Acquisitions department via an in-house regression model that is frequently tweaked and updated. While the data has changed, offering advanced location intelligence potential, much of this potential has not been realized. New data sources, collected in real time (data in motion), are primarily being used for telling interactive stories mainly for senior executives, leasing and marketing departments. It has not been effectively integrated within the forecasting models being employed.

The ultimate vision for the organization was made quite apparent - they are looking to increase profits at each shopping centre, while enhancing the overall consumer experience. The consumer experience element is viewed as the key driver for increasing customer interaction. Table 38 identifies the three corporate strategies that they are looking to deploy.

“We are looking to use experience focused marketing to change how shopping happens”

Table 38: Corporate Strategies for increasing Sales and Customer Experience

Goal	Description	How
Change Consumer Behaviour	Increase visitation/dwell time; Increase spending; Increase market share.	Understand the customer, create personalized experiences, track purchasing, use real-time traffic, target marketing and advertise efforts.
Boost Operational Efficiency	Supply chain management; On-site labour/duty scheduling; Integrated building systems.	Track traffic flow, uncover labour inefficiencies, embrace Internet of Things, indoor/outdoor positioning technology for complete digital experience.
Support Leasing Decisions	Optimize rent structure; Optimize tenant mix; Optimize development initiatives.	Understand the unique spatial context of each centre, track the industry and market, model the environment, forecast behaviour.

This organization is in the process of building robust customer databases that link demographics and shopping behaviours. They are doing this by compiling Wi-Fi login information, newsletter sign-ups, and supplementary census demographic data. This ultimately helps create an advanced segmentation system in order to target distinct customer profile.

“We are now targeting distinct customer segments by using data from all our data sources”

Along with understanding what happens within their shopping centres, they are interested in understanding what happens outside their centres and are actively looking to map the gaps

in their current customer base. Credit card data and external loyalty (such as AirMiles) are providing the opportunities to identify these gaps.

From an operationalization efficiency perspective, they are looking to use shopping centre traffic data in order to optimize shopping centre staffing. Furthermore, they are looking at historical customer flow data (day, week, month, and year) along with weather data to be able to predict shopping centre traffic. This was indicated to be beneficial for customer service, specialty leasing/events, and promotional offers.

The data is also seen to help enhance Leasing/Development decisions. The data will be able to identify and utilize traffic volume in the tenant space and the amount of cross-shopping that exists with other tenants. Therefore, this would ultimately produce what they refer to as an Influence Index, which is based on sales, traffic and cross-shopping. This metric will indicate the influence that each tenant has on the overall shopping centre, broken down by gross leasable area. From the customer perspective, they are looking to calculate what is referred to as an experience value, which is measured by the amount of time spent in the shopping centre along with the amount of money spent. They can see this through the traffic data, which will show, in essence, the longer the dwell time and higher dollars spent means the shopper had an optimal experience. This requires data sharing agreements between the landlord and tenants with regard to accessing actual sales data from POS systems.

"If we get tenant buy-in and have them share there POS data with us we can see the relationship between dwell time and purchases"

4.3.3.1 Success Rate and Challenges

While there is a clear strategic plan in place, the organization has experienced difficulty with execution. The primary problems are related to data mining and warehousing. While the company has been very effective at obtaining new data sources, the data warehouse is not able to handle all these new data sources (such as, traffic data, POS data from multiple tenants, facial recognition data, etc.).

“Backend data infrastructure is a big problem...it is not prepared to handle what we want to do”

Currently, data is not stored in one central place. Data silos are a significant problem as they are not aware, nor have access to, certain data sources within the organization. These are often stored locally on an employee's machine. Furthermore there is no documentation available to understand how data is stored. In other words, they are missing metadata identifying the basic database schema, outlining how the database is constructed. This metadata is in the head of the database developer and the organization risks losing this knowledge if this employee was to leave the organization. This is not only an issue related to data but also access to certain software packages. For example, the organization has a license for Tableau but it is currently only accessible by one individual.

“No one knows who might have access in the company. People might have it but its individual.”

In order to combat this, they are in the process of creating a data lake in order to offer a centralized place for all data (regardless of type) to be stored and accessed by any part of the organization. They have outsourced the development of the data lake to an external company. With corporate buy-in in place, they were hoping to have the data lake up and running by the summer of 2018. Unfortunately, for a variety of reasons, this has been significantly delayed. With substantial push-back, due to push-back from departments and employees, the process has been delayed considerably. This resistance seems to largely stem from fear of having to learn a new system.

“Employees are scared of having to learn a new process...they are comfortable with the traditional methods”

A major issue is the fact that a data lake is a relatively new concept and as such, is quite complicated to implement effectively. Internally, there is a significant gap in personnel who understand the technology. While corporate support exists, this gap in knowledge has made it hard to get employee or departmental support. While some departments/people want to help, others would rather not. It was believed that the push-back was a result of fear. It was identified that the fear was either related to having to learn something new, or the fear that they would be replaced by a data engineer/scientist. The organization has a committee that has to approve any enterprise-wide investment, which has held up the implementation because

they ultimately need to sign-off on the funding. There was no sense that it would be declined because it is a c-level initiate but organizational bureaucracy has slowed the process.

Accessing data is also a challenge because no tools exist within the organization to allow analysts and data users to go straight to the data warehouse. They are currently in the process of building API's that will make this easier. They expressed a clear need to have data come into their software packages (e.g., automatically into Tableau), and it is not happening.

"Data is hard to access. We know it exists but it's hard to get our hands on it"

"Getting data reports on consumer traffic on selected sites (shopping centres) is not possible, making it difficult to conduct real-time decision making within the shopping centres."

Without these tools, it has made it difficult to illustrate the value to other business units. It was often indicated that the data science within this organization was 'out of touch', as in the fact that IT and other business units don't understand it. While time-consuming, the analytics team has booked one-on-one meetings to explain benefits of the data lake and API's to solve the organizations data problems. Therefore there is a strong push to educate people so they can see the value. If they can't see the value, it is hard to obtain the needed buy-in from colleagues. Buy-in is becoming easier within this organization because the push towards a more sophisticated RLDM environment centered on SBD and data science is coming from the President of the organization. This has made continued resistance from certain individuals or

business units difficult. The Analytics, Insights & Innovation unit is responsible for championing the role out the changes to their data and analytics environment.

4.4 CASE 2 – FOOD SERVICES

The second case study was of a large fast food chain. The data collect for this case was obtained through conversations with 3 individuals. Two individuals from the Spatial Research department (National Director, and an Analyst) and 1 individual from Business Insights (Consultant). Data was collected through multiple phone conversations.

All geographic data and spatial analytics at this fast-food chain is undertaken by the Spatial Research Team (SRT). This team is one part of a division known as the Strategy and Insights Division (S&I). Along with Spatial Research, Consumer Insights and Business Insights are also part of the S&I division. The spatial research team has three position levels: analysts, a manager and a director. The director is responsible for the entire S&I team. The analysts and manager are all geographers, trained in spatial analysis, with no data science experience or education.

4.4.1 Data and Technology Environment

This company proved to be a very data-poor organization. The poor quality of their data operation was a result of several factors. Primarily, this organization suffered significantly as a result of data silos. Table 39 highlights the issues with regards to data not being accessible

across the organization as all their data sources, minus their sales data, is only locally accessible by the SRT. Secondly, the organization has historically had challenges in acquiring customer level data. Unlike some retailers that have constant flows of customer transaction data, this retailer has had an inability to link their voluminous transaction data to individual consumers.

“We don’t have a great data picture of our customers...no idea who our customers are...no info on repeat customers or individual purchase habits”

They have recently implemented ways to collect more customer data but these have not proved to be lucrative in generating data. They predominately use a Customer Satisfaction Survey to identify the postal codes of their clientele. Unfortunately, participation was documented as being very minimal and therefore likely to be biased, making it difficult to effectively integrate into any form of analysis.

“The amount of customers who fill out our customer survey is negligible...it is not incentive driven”

Participation in this survey is voluntary as participants are informed of the survey details on their receipts. Another way this organization has attempted to get more customer data is through their mobile app.

“We are currently looking at ways to access our mobile customer data in order to better understand how customers interact with us”

The primary function of their app is to get all customers to place orders and browse menu options. The app only works through customer registration whereby the user needs to identify their name and postal code. When asked about the percentage of purchases carried out on their mobile app, it was stated that this level of granularity was not collected or accessible across the organization nor was it integrated.

“This data is still in its infancy and is not fully integrated”

Table 39: Data Sources Frequently Used by Spatial Research

Description	Source	Accessibility	Volume
Demographic Data Profiles	External	Local	GB
National traffic data	External	Local	GB
Competitor data	External	Local	GB
Business Data	External	Local	GB
Expenditure Data (NHS)	External	Local	GB
Sales & Transaction data	Internal	Enterprise	TB
Market Classification	Internal	Local	GB

Beyond this, there have been only modest advancements in their data environment. For example, while they indicated that they still rely heavily on their traditional data sources (mainly demographic data), they have made incremental changes in the quality of those data sources. They indicated that there has been some investment into data upgrades in regards to

demographics data profiles and day-time population. They acquire this data from a major demographic data provider.

From a technology perspective, this organization has invested in data software packages that help with streamlining their research operation. Specifically, the organization has invested in Alteryx. Alteryx is a data blending software platform that is primarily suited to handle a larger volume and a diverse variety of data. This is not necessarily the benefit of this product for this organization because their data sources are neither uniquely large or in an unstructured format. The significant benefit is that this software package has allowed for the automation of their sales forecasting models. With the automation, the sales forecasting models can be completed quicker, allowing for shorter lead times and faster decisions. Models are also able to be calibrated much faster in Alteryx because the data flow is more efficient. All data is funnelled through Alteryx, and therefore all mapping and analysis are done within one data software package. Team members seldom use any other software package.

“I only open Excel and Access to convert new data files into an Alteryx format...all analysis is done with Alteryx... no need for multiple software packages”

4.4.2 Retail Location Decision-Making

The SRT, in essence, acts as a support team that is mostly responsible for providing spatially driven data analysis for a variety of internal business units. The divisions regularly supported were identified as: Restaurant Development, Supply Chain Management, Marketing,

and the non-spatial divisions of the Strategy and Insight Team (SIT). Table 40 highlights the support structure for each of these divisions and what is provided to those divisions. Mainly this department conducts market-based analysis for the individual business units. More often than not they use basic distance buffers to create demographic snapshots of the immediate market area surrounding their locations or proposed locations. They also rely on statistical models (primarily, in-house built regression models) to quantify potential acquisitions, restaurant reinvestments, and to identify market penetration.

Table 40: Support Structure

Business Unit	Description
Restaurant Development	Provide high-level market overviews and insights for restaurant development and network optimization; Create, define and apply statistical models to quantify market variables for market level decision support.
Supply Chain Management	Provide support for market deployment, scheduling, and market insights for decision support regarding new channel strategy initiatives; Support the tracking and reporting of new channel strategy initiatives such as restaurant reinvestments and implementation of new interior & exterior technology to service guests.
Marketing	Create specific market overviews (demographic, competitive, restaurant data) to identify our existing market penetration and where/when to market specific products or initiatives; Support marketing programs related to channel strategy by leveraging segmentation system for decision support in the strategy and execution of new opportunities.
Strategy & Insights	Assist other LOB (line of businesses—menu line of business etc.) for reporting and tracking by identifying and measuring KPI's that measure the market and business; Provide decision support for market testing by identifying tailored control groups to measure how our new/existing opportunities trend at specific sites behave and react compared to their markets.

Out of these business units, the most frequently supported team was documented to be the Restaurant Development team.

“We work closest and most often with the Restaurant Development team”

This has a large part to do with the fact that this organization is going through a high growth phase and is expanding quickly; as a result, they are usually employing analytics for site selection purposes. This included trade area analysis, site selection, and cannibalization analysis. The first step is conducting analytics to identify demand potential and service gaps in the market by looking at the current restaurant network. The real estate team then goes out to find potential sites. These sites are brought back to the SRT who then conducts preliminary sales forecasting analysis to assess suitability. Once the development team decides they want to move forward, a full market plan is created, which includes a sales model and in-depth cannibalization analysis model. This development cycle (from site selection to restaurant opening) is documented to take anywhere between 18-24 months. The amount of time is largely linked to the fact that this retailer owns the majority of their locations.

“We are not looking at short-term leases. The majority of our properties are owned and act as long-term investments”

4.4.2.1 Success Rate and Challenges

While the data environment has progressed and changed, there have been minimal advancements with regards to location planning, as the core decision-making techniques used have not changed in the past 10 years. Just as before, they rely on sales forecasting models (that has been improved by data and software quality) and market analysis techniques. The biggest difference has been the expansion of the SRT as it was initially run and operated by only one individual. In the past few years, it has grown to include 5 individuals, all with very similar educational and training backgrounds.

The SRT faced several challenges. While the SRT has worked to help minimize risk and to support any decision that could use spatial data much of the work is not actually being used. The ultimate decision-makers would almost always rely on their intuition and experience to drive any decision forward regardless of what the data analytics might say. This has changed substantially due to several key corporately documented failures costing the company millions of dollars.

“Screwing up one sales estimate can cost several million dollars...

the role of location supported decision making is now seen as

essential... Almost entirely data-driven decision making now”

With that said, intuition was indicated to be a very vital component to how decision are still being made. These were largely linked to the role of the site visit. It was indicated that in some situations, the data does not have the ability to provide a full picture and in some cases can also be misleading.

“Site visit is important...can’t see barriers and limitations without visiting. Real estate rep is important for this.”

The real estate representatives still prove to be the biggest challenge for complete SBD adoption. It was evident that these reps still prefer to rely on relationships that they have built and their knowledge of how the real-estate market, and more specifically, how this fast food chain operates and what they regard as key success factors.

“it’s really difficult to get real-estate reps to buy in to analytics...its like ‘Moneyball’ [book and film discussing the changing culture of data-driven decision-making in sports management and the push back the early adopters experienced]. Reps thought they knew this”

Another major challenge identified was getting corporate support to invest in new data sources and technologies to help improve their RLDM. While the real estate team can see the potential of more granular level consumer data, it has not been an easy sell to senior (c-level) decision-makers. They indicated that a challenge has been the fact that these individuals do not fully understand the methodologies and, therefore, it is difficult to get financial support. They lack employees capable of conveying details about their data models, as well as, the results without the use of industry jargon.

“People don’t fully understand the methodologies and this poses a money problem. The statistical models are complex so hard sells are needed.”

“They (senior decision-makers) have operations and business background. Hard to convert and explain data in a meaningful way.”

The business units, as opposed to the SRT, drive the support environment. Therefore, individuals that are not necessarily equipped with the skills and/or the knowledge of what can be done to help support decision-making along with the potential benefits that those solutions could provide, dictate the level of support. So even though the SRT claims to be “smart enough to know what we don’t know and what we do know” in relation to data-driven decision-making, other units may challenge the value of such decision support.

4.5 CASE STUDY 3 – FINANCIAL SECTOR

The final case study was of a large financial company. The data collected for this Case was obtained through conversations with 6 individuals. Three individuals were within the Physical Distribution Planning department (PDP) (Former Director, Current Director, 1 senior Analysts, 1 Analyst, and 1 intern) and 1 individual from Data & Analytics (D&A) department. Data was collected through multiple in-person meetings and phone conversations.

The Data & Analytics Department (D&A), which has approximately 450 employees, is responsible for the entire data operation at this organization, which includes data-housing,

data-management, data-privacy, and data acquisition. This team is under the Technologies and Operations Unit (TOU), which is similar to IT Services at other institutions.

The D&A has two major functions, the data storage side and the data usage side. The data storage side is largely responsible for the data warehouse and the data lake at the organization. More specifically, they take care of the data architecture, data development and data support. These responsibilities include developing data models, the policies that govern and decide what data is collected, as well as the storing, and arranging data. The data usage component is largely responsible for data integration. This is a multi-faceted group, which includes a Lab Team (LT), Reusable Services Team (RST), and Location Intelligence Team(LIT). The LIT is predominantly responsible for conducting applied research and has the flexibility to be innovative and apply data analysis for a variety of different applications. This may or may not include locational data and location-based planning. The RST, often called the Business Integration/Insight Team (BIIT), works closely with the LIT as they often pick up some of the data applications/projects that come from the LIT and attempt to integrate these projects into various business units' day-to-day decision-making operations. The LIT is a relatively new part of the D&A division, joining the team approximately one year ago. Their main responsibility is to provide custom location-focused data analytics to help all business units. For example, one of their projects was to create a credit risk index. By leveraging customer address data, they can get credit risk scores at the sub- regional level.

The PDP division is the only division that is directly responsible for physical branch-location decision-making, that is, branch or ATM investment decisions. This includes any renovation, relocation, new branch location, and branch closures. This organization is under the

Retail Distribution Strategy and Performance (RDSP), which is under the Regional Banking division (RB). This consists of six planners (physical distribution analytics and planning), two data analysts, and a strategic director. The planners work closely with regional leaders such as regional VPs and operating officers who report to the regional presidents.

4.5.1 Data and Technology Environment

This is a very data-rich organization, having access to over 20 petabytes of data. They stated that their organization collects more data than any other retail organization in Canada.

“Outside of government we collect the most data.”

“We rival Google and Amazon in the volume of data we collect.”

The massive volume and the level of detail on consumers is so extremely fine-grained, making it a fertile environment for innovation. They reported negligible use of location data (only used to make branch and ATM decisions) even though 92% of their internally collected data has location information.

“92% of our collected data has location information...Traditionally not used for much more than location decisions for branches and ATM’s”

The interviewees identified three major reasons for not adopting more location-based data analytics. These are: (i) they did not know what data existed; (ii) they were unaware of what they could access; and, (iii) they would gain access to some data but never actually integrate it into their decision-making. In order to combat these issues, the organization has recently gone through a complete data overhaul. They have implemented a data lake, which has been active within the organization for over 9 months. This is a major catalyst for breaking down data inefficiencies that the organization has experienced due to data silos. This has created a single touch point for data sharing across the organization, as it has made all the data within the organization accessible to all the business units.

“With the introduction of the data lake, data sharing is becoming easier. Allowing for greater flow and access to data and information”

Data stewards are responsible for administering the permissions in order to enable access to the data lake. This is an ongoing process in the sense that they do not wait for business units to approach them, but instead they (Data & Analytics Unit) reach out to inform the units of what exists and of its potential uses. Therefore, they are active in ensuring that all members that require access, or could benefit from access, have access.

The interviews were largely focused on the PDP department because historically this was the only business unit that was responsible for location decision-making. Over the last two years, they have made substantial progress in the quality and variety of data (Table 41) that

they rely on when making decisions. They have newly acquired data from a major data provider based out of Toronto in regards to the census profile and daytime population. They have also incorporated a segmentation system that allows them to easily create market profiles for specific areas across the country.

“We can now create market profiles much easier and more efficiently since we purchased our segmentation system.”

From a technology perspective, they have actively been expanding the geographic and information technology packages that they have access to. While they formally only relied on MapInfo, they have now included Alteryx and Tableau. They use MapInfo and Alteryx for data-process, visualization, and all GIS applications, and Tableau for reporting and visualization. They purchased Tableau to help create better visualizations as a means of displaying the data and finding ways to make the data more understandable to decision-makers.

“We are moving away from the traditional GIS packages and starting to bring in new software packages with greater data capabilities (size and variety)”

Table 41: Data Sources Frequently Used by Physical Distribution Planning

Description	Source	Accessibility	Volume
Personal banking client information	Internal	Enterprise	TB
Business banking client info	Internal	Enterprise	TB
Retail branch & ATM locations	Internal	Enterprise	GB
Demographic Data	External	Enterprise	GB
Daytime Population	External	Enterprise	GB
Wealthscapes (assets, liabilities and wealth)	External	Enterprise	TB
Business Locations	External	Enterprise	GB
Shopping centres Data	External	Enterprise	GB
Business counts (Stats Can	External	Enterprise	GB
Geodem segmentation data	External	Enterprise	GB
Streetpro	External	Enterprise	GB

4.5.2 Retail Location Decision-Making

The PDP largely acts as a support team; they work closely with corporate real estate, finance, as well as external partners (e.g., real estate brokerages). Formally, a big part of that dealt with store openings and closures but this has now shifted with added attention on understanding and repurposing their current retail locations.

“The nature of banking has changed...we are looking at repurposing our physical spaces in order to adapt to remain relevant to our changing consumer.”

In other words, a big part of their business now focuses on refurbishments, relocations and re-fascia in order to reposition themselves as a leader in an evolving retail sector.

From a RLDM perspective, the past two years have brought about significant changes in the way the organization carries out location-based analytics. Formally, the PDP was the only department responsible for location analytics at this organization. They mainly, and in large part continue to, make branch location-decisions using very established methodologies with GIS software (specifically MapInfo). Specifically they rely on basic trade area analysis, experience and analogue based approaches (e.g., multiple regression, demographic data benchmarks) They selected new real estate opportunities by constructing market plans that consisted of trade area profiles, indicating market potential. They rarely carried out more advanced techniques, like predictive modelling, to aid their decision-making.

“We still rely on our traditional approaches (methodologies) to make real-estate decisions”

The resistance from the team leader(s) was the primary reason that decision-making processes took so long to evolve. Change in senior leadership fostered change in the decision-making process they employed.

“...our former Director was not open to change...(he) didn’t really care to adapt”

Location intelligence is now something that they are incorporating into all aspects of their business, adopting a more advanced approach for retail analysis. Furthermore, they traditionally lacked flexibility in the way decisions were made as their process of supporting decisions was standardized and did not allow for customization.

“Our team is under a cultural transformation...is taking on a broader role...centered on accountability and responsibility.”

This degree of customization is allowing flexibility. They are hoping that this greater flexibility will reinforce the perceived value of their team internally therefore increasing the frequency of adopting location-based data -driven decisions in favour of intuition based ones.

They are developing trajectory databases that track consumer behaviour in space and time. These trajectory databases will enable advanced data analytics identifying similarities, commonalities, differences and overlap in consumer behaviour. They are also undertaking territory analytics (the only retailer in this study doing this actively) by leveraging mobile data. This creates temporal behavioural clusters based on reoccurring location destinations. This is distinctively different from traditional residence based analytics or the identification of causal or consequential events. With a greater focus on data-driven decision-making, this team still values the role that intuition and institutional knowledge bring to the location planning perspective.

“intuition drives high-level ideas from a creativity perspective and helps validate the decision on the ground... Some things cannot be categorized in data....In order to find the needle in the haystack

better insights and analysis are required to find new opportunities...Intuition will never disappear. We start with a qualitative thought and we try to support it and quantify it with data.”

4.5.2.1 Success Rate and Challenges

The organization indicated that former data practices made it challenging to adopt SBD methodologies. While the organization was extremely data-rich, they experience significant limitations in access.

“We (the organization) have access to a lot of data but we only tap into a small portion of that.”

A major challenge was data awareness. The team was not fully aware of what was available across the organization, which made it impossible to integrate new data into their decision-making processes.

“Knowing what’s out there and how to use it.”

Before the data overhaul took place, new data sources were rarely brought in. Data acquisitions only happened if its use was clear and if it would provide a quantifiable increase in the quality of a decision.

“It was very difficult to bring in new data...there was always a lot of skepticism around the value of new data”

Although new data sources are considered shiny new objects and create a lot of attention, senior management were not quick to buy-in, and therefore, justifying new data acquisitions to senior managers was a major challenge. At a department level (e.g., PDP) there might be value, but the increased level of insight was difficult to get across to Senior Directors, who traditionally would not support data-driven decisions if it went against intuition/institutional experience. This is no longer the case as there is complete management support for growth in data-driven initiatives. Along with having a supportive management team, from c-level individuals to business unit directors, a major benefit at this organization is the richness of data that is available.

“Data processes exist and they work...We can get at data with relative ease”

Not only are they data-rich (richness comparable to Amazon and Google data), the data is easy to access. This makes new innovative applications possible. This turnaround was largely related to change in management changing the culture at the institution.

“This change had everything to do with changes in Directors... they now see the value of location-based data”

Acquiring talent was a challenge at this organization. It is difficult to find location intelligence people with strong data science backgrounds. While it's easy to find GIS specialists, it was said that it's not as easy to find data scientists.

"...few data scientists with a location intelligence training exist."

4.6 CASE STUDY SUMMARY

These case subjects were selected for a number of reasons. Initially these were targeted after gaining an understanding of where the companies were at in terms of the awareness, availability, use, adoption and development of SBD. This was identified through the online questionnaire and the semi-structured interviews. These 3 organizations proved to be very different from one another in terms of their SBD adoption curve and therefore made excellent case subjects for identifying the success factors as well as challenges. While other retailers were approached these 3 organizations were easier to gain access to because I was able to develop a good rapport with at least one an internal person. These individuals were responsible for making introductions with a variety of individuals that would have otherwise been unreachable. This was a key element of my research because the only way a holistic view of their SBD operation was to talk to as many individuals as possible from as many departments as possible. Not only did the 3 cases organizations represent different levels of SBD diffusion they were also from different retail sectors. This allowed for potential sector differences to be explored. These three cases were ultimately chosen because they represented 3 different areas of SBD diffusion and 3 different sectors.

5 DISCUSSION

5.1 ONLINE QUESTIONNAIRE

From these results, it is evident that the retail environment has undergone some significant changes. Most notable are changes in the collection and usage of SBD. The majority of respondents (well over 80%) indicated overwhelming increases in the variety, volume and velocity of data sources – the “3 V’s” of Big Data. This change presents new technological challenges, as well as opportunities for retailers to incorporate advanced analytical tools to reveal meaningful new information, and better insights when making decisions. The diversity in data sources may well pose the biggest challenge for retailers as traditional data warehouses may not be equipped to handle and integrate SBD.

Along with increases in the adoption of new SBD sources (such as, customer surveillance data, location data collected from mobile phone pings, social media, and crowdsourced data) there is evidence that certain traditional data sources are migrating to what might be considered SBD. One example of this would be loyalty data. Early loyalty programs provided nothing more than POS data linked to an individual and their home address, therefore retailers simply knew where their customers lived and where they shopped. With mobile applications readily collecting location information from their users through cellular or near-field sensors, retailers now have the potential to track a customer outside and inside the retail locations. Therefore, traditional loyalty data may be starting to migrate to (what can be defined as) SBD as

retailers integrate technology and more sophisticated data collection approaches to their loyalty programs.

Retailers' attempts to gain greater access to more granular level customer data proved to be a significant data-oriented advancement. Developments in sensory-based technology, traffic data, store space planning data, and daytime population data are starting to increase the breadth of customer data that retailers have access to. With the adoption of new technologies, such as Indoor Position System (IPS), there is the potential for greater detailed tracking of consumers and potential consumers.

In addition to changes in SBD, the range of location research methods that are employed within retail firms are also changing significantly. New emerging techniques are starting to transform the ways that retail location decisions are made. With more than a third of the respondents indicating some use of advanced forecasting techniques (e.g. Machine Learning, Social Media Analytics, Social Influence Analysis, Sentiment Analysis, Real-Time Data/Demand Visualization), it is clear that retailers are starting to add techniques that will allow for more significant insights into individual consumer behaviours. With technology redefining consumer interaction with retailers, the future of bricks and mortar locations has been significantly threatened. An individual's path to purchase is no longer linear as the growth in omni-channel retail has offered consumers more ways to engage in complex shopping behaviours.

This is likely a key reason why established methods such as experience, checklists, analogue, and multiple regression techniques still play a key role in decision-making. As was

identified by Hernandez and Emmons (2012) and Woods and Reynolds (2010), reliance on experience is still the single most important element in RLDM.

Large data sets, or Big Data, has always been a challenge for as long as people have been able to collect information at some level. In other words, whatever data sets might appear to be large and unmanageable today, might not be in the not so distant future. Similar to Sivarajah et al. (2017), respondents identified that the intricacies of SBD are rarely related to the volume of data being generated but more to do with the type of data. Therefore, certain small data sets, that might be unstructured, can pose greater challenges than large structured data.

5.2 STRUCTURED INTERVIEWS AND CASE STUDIES

This section discusses some of the finding from the structured interview questions pertaining to SBD assimilation. The following sections will discuss assimilation in regards to (i) Data Warehousing and Mining/Data Integration and Analysis/Data Interpretation, (ii) Corporate Buy-in and , (iii) Compatibility with current system.

5.2.1 Data Warehousing and Mining/Data Integration and Analysis/Data Interpretation

The dialogue with the respondents indicated that challenges appeared to address 3 major themes related to data operations; (i) Data Warehousing and Mining (ii) Data Integration and Analysis (iii) Data Interpretation. Most respondents indicated their lack of effectiveness at

data warehousing and mining. Generally, the complexities associated with SBD made it difficult for organizations to compile and organize data into one common database, which therefore makes the process of extracting meaningful data for analysis challenging. This raises questions around the increasing role and need of data blending practices, that is - the process of combining data from multiple sources into functional dataset (Bazeley, 2009). The growth in software packages such as Tableau and Alteryx can be partially attributed to their ability to create single views of data even when multiple data sources exist.

A clear challenge associated with SBD is directly linked to a firm's effectiveness at integrating data (new and old) into the decision-making process. This is brought about by diverse types of data sources and data formats. It does not appear that a lack of data is an issue for any retailer, as all respondents identified having an abundance of data, the real problem seems to be related to data integration. Twenty-two respondents highlighted the fact that their data warehouses, which were designed to handle structured data, are not equipped to integrate SBD. This challenge, in the variety of data, is related to integrating unstructured data and semi-structured data which do not conform to existing data models. This makes it increasingly difficult to create new knowledge that ultimately improves decision-making. Similar to what was identified in Marr (2015), it is clear that most companies are 'data-rich but insight-poor.' This really has to do with a lack of understanding on how to use data by most retail organizations.

“It’s a challenge because customer data is defined by multiple actions...not just one point of interactions. This makes data variety the major challenge” Financial Sector

With regards to data analysis, there appear to be significant barriers for retailers to start to perform analysis and modelling involving new data formats. If the ultimate goal of collecting SBD is to deliver better insights and value in decision-making, it is essential to find and refine methods that are capable of supporting new data. A problem with performing data-analysis largely had to do with a lack of experienced personnel (Data Scientists) within the RLDM departments, capable of using and developing machine learning techniques and improving data collection procedures. While some decision-making departments are starting to hire data-scientists to tackle some of these data-driven challenges, the transition has not been smooth. One of the major challenges documented stems from the fact that these data-scientists lack industry-specific knowledge which makes it difficult for them to communicate their findings effectively to senior management.

“A disconnect from the data scientist as they don’t have industry knowledge.” Developer/Leasing/Brokerage

Data interpretation has always proven to be a challenge with spatial data analysis, as documented in several studies (Stone et al., 2003; Kohavi, et al., 2004; Jagadish, et al., 2014). It is important to make the results and conclusions generated from data analysis understandable and interpretable for senior decision-makers. A major challenge exists in retail corporations

understanding of how computing technological solutions have evolved to allow SBD sources to be retrieved, aggregated, examined, and interpreted. The real power of SBD adoption has to do with the process of interpreting and gaining new insights from new, untapped data sources. One respondent went into detail about the difficulties in the packaging of data and analysis in an understandable way.

“We talk about our mechanisms for data reporting on a regular basis...like an iceberg, all the data processing and data-driven analysis happens beneath the surface...upper management only cares about what is above the surface...we are always trying to find new ways to package our findings and recommendations in clear easily digestible ways” Financial Sector

Data visualization proves to be a significant challenge with data interpretation. Data visualization methods allow decision-makers to have easily understandable visuals, as the more complex the data becomes, the more difficult it is to offer easily interpretable information for decision-making.

5.2.2 Corporate Buy-In

To efficiently establish SBD decision-making practices, initiatives need to be directly linked to corporate strategy. There needs to be a level of understanding and synergy between the initiatives and the subsequent value that they provide to the organization. It was apparent

that the most successful organizations had corporate buy-in for the collection of new SBD sources as well as any subsequent technologies that may be required. If the promotion of a data environment was not coming from the top down, it seemed less likely to gain any corporate traction. In the case of the Financial Sector, it has experienced substantial growth in the adoption of practices seen to create value from the extensive in-house data sources that are available. This is done mainly because there are clear strategic objectives centred on implementing a data-driven business. When retailers decide to become data-centric organizations, monetary buy becomes a lot easier as well. Difficulties in justifying expenditure on geospatial technology and data seem to be less prevalent when there is strong executive support.

If an organization is able to prove that their team, or the organization at large, has a capacity to apply SBD Analytics to solve essential business problems, corporate buy-in becomes much easier. From a data perspective, organizations need to create value from in-house (internally collected) data before purchasing or collecting new data sources. With that said, some organizations have an inability to maximize the value of their in-house data sources because they lack adequate technologies required to do so. The organizations that are consistently successful at SBD initiatives are the ones that have created decision-making models that can leverage the benefits of the data and advanced analytics in a repeatable way. This allows for the streamlining of the decision-making process. The successful data-driven businesses are able to align the organization's strategies, the data systems, processing, and analysis in order to make better location decisions.

It was clear from the respondents that the organizations without a clear corporate data-champion, pushing towards integrating SBD decision-making, experienced severe challenges. In other words, these organizations lacked individuals with the set of competencies required for SBD adoption. Three types of talent deficiencies exist. The first refers to an organization that does not even realize the potential in hiring an individual with a specific set of skills that can help implement and refine spatial data decision-making. This is more difficult to correct as there is a lack of awareness in data-driven potential. The second is where the corporate talent strategy is not in-line with the business' strategy. Therefore, the company becomes unable to meet the talent requirements that are necessary for the adequate implementation of SBD initiatives. The third pertains to the scale of hiring practices. The most successful firms have multiple employees, or entire departments, that are comprised of a talented group of individuals equipped with the tools and knowledge to handle and incorporate SBD and SBD Analytics. The ability to gain traction in SBD initiatives is linked to whether there is one data-scientist versus a team of data-scientists responsible for growing data-driven decision-making. More so, when entire departments are developed to tackle the challenge of SBD, the results are more profound.

From a RLDM perspective, this resource issue is not easy to overcome. There has been growth in educational programs and training that focus on preparing people for careers in business intelligence (BI), that is, positions that use technology-driven processes for analyzing data and presenting actionable information to help inform business decisions (Chen et al., 2012). However, there is not an abundance of individuals with the capabilities for is coined

Location Intelligence (LI), which uses a set of tools to relate geographic contexts to business data. Therefore, even if an organization wants to expand its LI implementation, it's often hindered by a significant gap in the market as very few geographers or GIS practitioners are equipped with capabilities to do more complex data processes and analytics.

This does, however, produce an opportunity for geographers, and individuals with GIS training, to familiarize themselves with the benefits of BI frameworks and the potential solutions they can provide. Awareness is essential for adoption to happen. That is, awareness in data-availability as well as awareness of the techniques and technologies available to help inform decision-making.

5.2.3 Compatibility with Current Systems

While promises an increase in the quality of decision-making, the process of integration is not always seamless. Current practices often obstruct the development of new methodologies and techniques for location-based decision-making. IT infrastructure that is archaic, or unadaptable, proved to be a hurdle for evolving a retailer's data operation to include more novel (machine learning, predictive analytics leveraging multiple unstructured data sources, social media analytics etc.) RLDM methods. An internal champion who is active in investing in data technology solutions is essential. Without a leader it is likely that barriers, such as information hoarding, will take place.

Talent issues also proved to be a challenge because people felt threatened by the potential shift in how decisions were being carried out. With traditional practices being

threatened in the presence of new methodologies and techniques, employee pushback seemed to be driven by fear. Fear of losing their job in favour of someone with a data science background, and fear of having to learn something new. It seemed to really threaten their level of comfort within their current positions.

There is a level of attraction and a lure to utilize and acquire new data sources given the rhetoric around the benefits of SBD. While this appeal exists, the reality is that most organizations are overwhelmed by their current data sources (evident in both the structured interviews and online survey). The level of access to already collected corporate data would dictate the level of success as it relates to data integration. Gaining access to, or sharing, corporate data sources is a major challenge as most data is stored in silos where it becomes next to impossible to access in order to interrogate the data. The challenge becomes more complex when organizations attempt to add new data sources without a clear strategy, or more importantly, without fully understanding what they have in-house and how new data sources might complement their current data. The most successful organizations are ones that are able to capitalize on their internal data sources. This usually comes in the form of integrating 'data lakes.'

The importance of breaking the organizational boundaries related to data silos is crucial for success. Having a central place that individuals can extract and utilize data is essential for SBD adoption. It became clear in the case studies that an organization's inability to access data results in a significant loss in potential adoption and progress within data-driven decision-making. Throughout the interviews, there was a clear evidence that data lakes are important to successful SBD practices. Data lakes are data repositories that are not bound to structured data

warehouses, such as the relational database model. The critical value of a data lake is that it provides a place for unstructured and semi-structured data to be stored. The barriers associated with traditional warehousing are removed with data lakes as all data can now be stored, regardless of whether or not it is of immediate use.

While this versatility in data integration is a key advantage when attempting to incorporate more complex data sources, it is not the only benefit. A quintessential strength is its ability to break down data silos. When a company incorporates a data lake, it does not necessarily mean they will inevitably adopt SBD practices when making location-decisions. Rather, it indicates a willingness to evolve the data environment to include the processing and storing of non-relational data. Without a data lake, adoption of SBD processes are completely impossible. There is no way to store and easily access unstructured data log files, internet clickstream records, sensor data, images, and social media posts. Adopting a data lake signifies a paradigm shift in the cultural identity of an organization, as their implementation requires significant investment in time, money and talent.

5.3 BEST PRACTICES

From the data collected, it is clear that there are three distinct types of data environments that exist. In this section the three types will be discussed in detail with clear indications of the benefits and challenges that each provides to their respective organizations when attempting to adopt SBD in RLDM.

5.3.1 Integrated Approach – Best Practice

Figure 3 identifies what an ideal SBD environment (Integrated Approach) looks like. Not all retail organizations are well situated to transform their RLDM practices. In order to transform current processes and to adopt new practices effectively, they need to start with key business objectives. Organizations that start with data, but do not have a clear strategy for adoption, have an inability to integrate the data effectively. These strategies can either be organization-wide or ones that focus on specific departments. Location decision-makers need to be clear about what they want to achieve. If data acquisition happens without a clear plan or objective, it is next to impossible to find ways to integrate the SBD as the data will serve no purpose. Therefore, all data acquisitions and the adoption of new technologies need to be centred on meaningful decision-making strategies, as it will reduce the likelihood of getting overwhelmed by the data and the hype it brings.

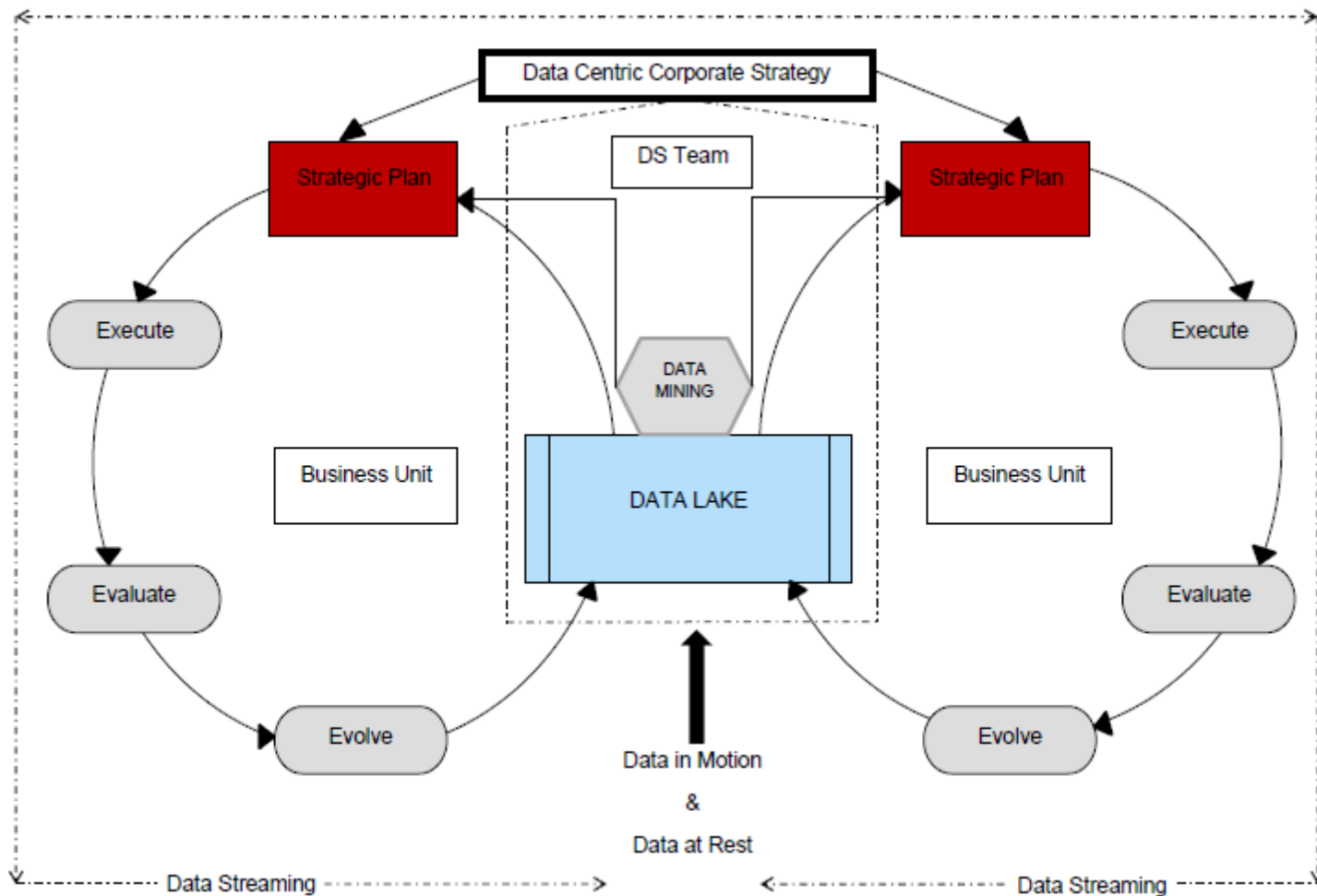


Figure 3: Integrated Approach: Ideal Corporate Data Environment for RLDM

Once a strategy is developed (Data Centric Corporate Strategy), data acquisition can happen. This might occur by acquiring new data or leveraging the untapped potential in existing data. This minimizes the likelihood of organizations being overwhelmed by the amount of data they have (that is- buying data or collecting data just for the sake of having it). This complicates the data environment, often overwhelming the data infrastructure (IT) teams as well as the analysts. Data sources cannot function effectively without the presence of technologies built to

handle the data. Therefore, data acquisition cannot happen if the data environment is not ready to receive it. This was the case for several respondents who purchased data they interpreted as being beneficial but had no mechanisms in place to actually store or view the data. This forces such companies to pay an external company to clean and provide data visualization for them.

Through the surveys and interviews it became clear that a major issue with data integration was the presence of a more sophisticated data warehousing system capable of retaining all types of data, support all data types, and supporting all users. The lack of access and awareness of institutional data that could be available proved to be a major challenge. Before advancements in an SBD processes can actually occur there is a need for a free flowing data environment with respects to data sharing, data access, and data awareness. The only way that this can actually happen is if all data sources, regardless of type, size, and complexity were centrally stored and accessible by all business units equally. The success of a centralized data warehousing system is contingent on the willingness for individual employees and the individual business units within the retail firm to no longer store data within their own units. If any retail organization continues to hoard data all efforts in attempting to move towards greater sophistication in all aspects of RLDM will be futile. Once again this reinforces the need for a clear corporate strategy making it mandatory that all data silos need to be absolved. Without corporate pressure the hurdles of integrating new and innovative systems for data warehousing, specifically, the resistance from individual business units to change their practices for RLDM will be too great. This largely will be related to internal uncertainties to the business units to put advanced data methods into practice. This is why at the heart of an ideal SBD

environment is the data lake (Figure 4). Data lakes are required in order to tap into existing data sources, creating a data lake that is accessible by all business units is essential. Without the removal of data barriers, brought about by data hoarding and departmental data silos, it is difficult to mine the locational data to help inform business decisions.

The only way to do this effectively is by developing a Data Science Team (DST) (Figure 5). This would ideally be an interdisciplinary team with employees coming from a variety of backgrounds including, statistics, and computer science etc. with responsibilities including SBD maintenance and data mining. An important task for the DST would be the development of a data environment capable of handling all forms of data regardless of the data source. The DST would be responsible for the following:

- i. Securing technical and financial resources to maintain core datasets and acquire new data
- ii. Creating a plan to address challenges in SBD talent recruitment and retention
- iii. Developing an evaluation framework to determine the effectiveness of SBD talent over time
- iv. Planning for and employing an adequate number of permanent staff to maintain growth in SBD.
- v. Creating and implementing a communications plan, aligned with corporate strategic goals, and ensure communications are tailored to each business unit

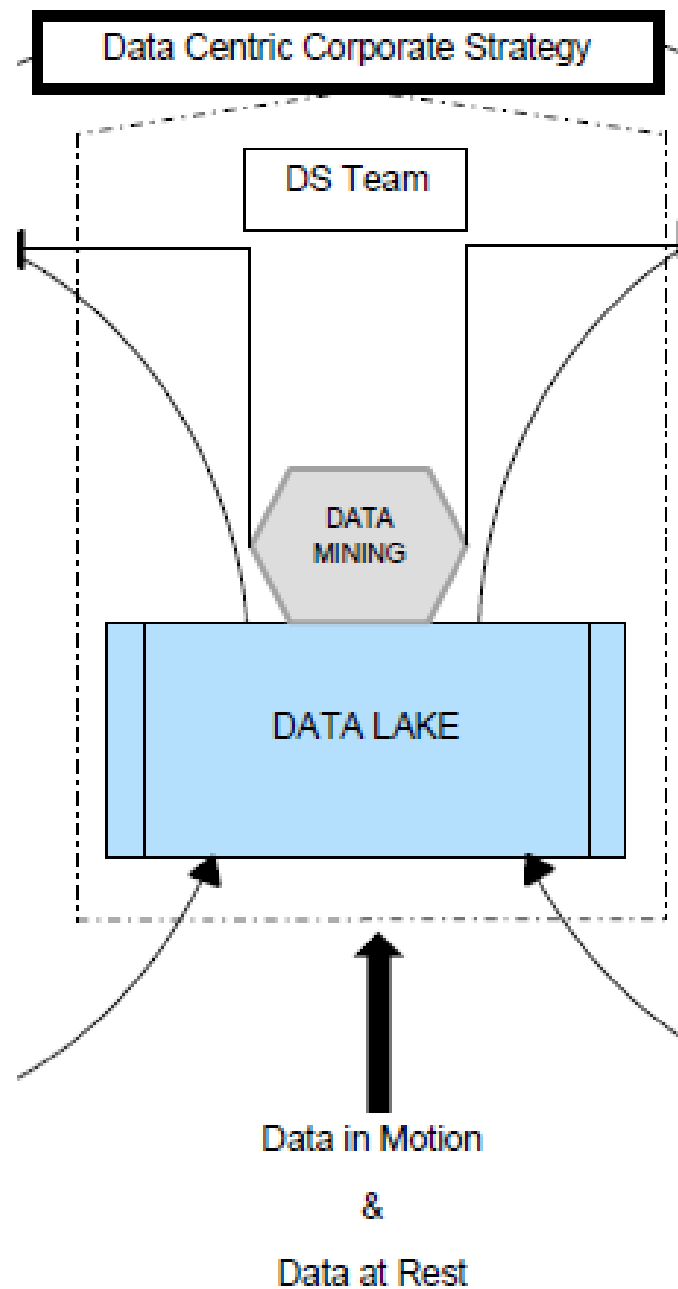


Figure 4: Data Science Team

Once the challenge of housing complex data sets in a centralized location is overcome, the organizations are then faced with the issue of tapping into that data in order to offer meaningful solutions to RLDM. This is where the data mining unit becomes crucial (noted above

the data lake Figure 3 and 4). The most successful companies were the companies that had the greatest degree of flexibility in their data processes. Data processing posed a significant challenge if data could not be requested from the data warehouse or lake with relative ease. If the data discovery, and data controls are not improved to a point that data is flowing to the units that require it, data mining becomes near impossible. If significant barriers within the data process exist, it is difficult to analyze data and to detect or predict patterns. The best-case scenario would be to have a data mining team responsible for supporting individual business units who would make ad-hoc requests and who are engaged in a constant process of looking to discover new applications for their data sets.

It was evident that these data mining challenges would be solved by having a DST team acting as business wide consultants who are tasked with developing data mining applications for individual business units. This would encourage greater efficiencies as relevant applications can be shared between other business units. If mining processes are centralized to the individual business units, it is less likely that the data mining applications will be shared. Therefore, this would encourage silos to develop. This is usually where data mining applications experience the most amount of difficulty. Furthermore, if SBD mining is done within the individual business unit's organizational learning becomes near impossible.

Without a strong corporate strategy pushing for the adoption of location analytics, it is less likely that the individual business units will be as receptive to adopting new SBD practices. In order for the individual business units to be successful, they need to have strategic plans in place. An example of a strategic plan could be a retailer wanting to increase its trade area size, or increase customer dwelling time. Once this is in place, the retailer can look to execute a

data-driven project in order to fulfill the objective set forth through the strategic plan. Keeping with this example, the organization could opt to purchase sensory-based technologies that utilize a patron's Wi-Fi signal. This could enable an opportunity to communicate with the patron in real-time in order to increase their dwelling time within the space. Once SBD project is executed, the success and failures of these practices need to be evaluated. This was identified as a hurdle in effective SBD adoption as it was difficult to identify whether the quality of a decision was actually improving with the newly deployed data techniques. The evaluation process also helps with managing expectations. When SBD initiatives are discussed and there is potential for deployment, employees and senior management were often under the impression that this would indicate immediate improvement in decision quality. Unfortunately, the reality is that this is far more complex and more difficult to measure. Furthermore, evaluating these initiatives encourages evolution in data-driven decision making as adjustments in order to maximize both the return and the effectiveness of the SBD practices being executed. The evolution component helped identify new data sources and approaches they should be incorporating and what data sources are not effective.

Without an entire team capable of handling data mining initiatives, business units will lack the talent (work force). The individual business units do not employ data scientist capable of handling the data mining initiatives and if the expectation is for current employees to learn how to do deep and intricate data analysis the learning curve is often too steep. Most of the employees within the individual business units neither have the time (responsible for making decision relatively quickly so they don't have time to adopt new processes), nor the ability (lack the technological know-how) to engage in SBD initiatives effectively. When the units are

responsible for the adoption and advancement of new data practices (which is the case for the Systemic Approach), it provides a challenge in gaining traction and in changing the way RLDM happens. Furthermore, the individual business' units experience the challenge of training current employees or hiring new talent. It is clear that the business units are not always the best versed in data progress within the retail organizations as a whole. This makes it difficult to have any growth in potential data initiatives to improve decision-making. A major issue is that business units do not communicate easily with other departments nor do they share data or expertise.

The companies with the greatest amount of sophistication in their RLDM are the ones that are tearing down data barriers. They are the ones who have invested heavily in a data environment that is centrally accessible by all employees and departments. It was clear that obtaining support for heavy investment into data infrastructure was not an easy endeavour. The only real way to gain traction was to have top-down support from the organization. If pressure was coming from the bottom-up (analysts up to management) it always proved to be futile. Traditional institutional views on reliance and importance of data-driven decision-making was a catalyst for evaluating the propensity of success. The transition to more sophisticated SBD driven RLDM is much easier, and the learning curves are far less steep, if investment in data practices were not a new phenomenon. Unfortunately, not all sectors had historically prosperous data environments, and as such, there were clear winners and losers when looking at adoption.

5.3.2 Reflexive Approach

Figure 5 identifies a Reflexive Approach to data-driven RLDM. This environment is not one that is built on a strong data centered corporate strategy. In these environments there is typically minimal top down pressure to innovate or improve the current state of RLDM to include more data-driven decision-making. Therefore, any growth or support for using SBD would need to come from within the individual business units. In these types of environments it was evident that SBD awareness was a lot less common. There was less departmental discussion and communication of new methods, techniques and data sources that could enhance the quality of their RLDM processes. In this environment RLDM is performed as a reflex, often without conscious thought as they heavy rely on previous practices and experience to support all decisions.

Enterprise-wide GIS licenses are also uncommon at this level and typically traditional types of RLDM techniques are carried out by a select few with individual desktop licenses. This happens mainly as a result of the fact that these organizations are more likely to have silos that form around the individual business units. Within these organizations individuals might not even be aware that GIS is being leveraged by other departments (Business Units) within the organization so therefore don't even recognize the need for enterprise level licensing. Furthermore all of the data used by these business units are typically stored within local servers. Therefore corporate awareness of available data sources is practically non-existent. This is also amplified by the fact that there is minimal or no formal communication between departments. In these types of organizations, communication between departments is

extremely rare. This is mainly due to either business units not understanding what other units are responsible for or out of fear of losing work to other departments or other individuals.

These business units are capable of experiencing some innovation within their RLDM process but it is challenging. Any innovation tends to come from an internal champion who is aware of available technologies, methods and data that can be used to evolve the way that their RLDM happens. This is, however, rare within the Reflexive Approach because of the lack of corporate pressure for advancing data-driven decision-making there is minimal or no financial support allocated to pay for new data, technologies or even talent, as it is not considered corporate priorities.

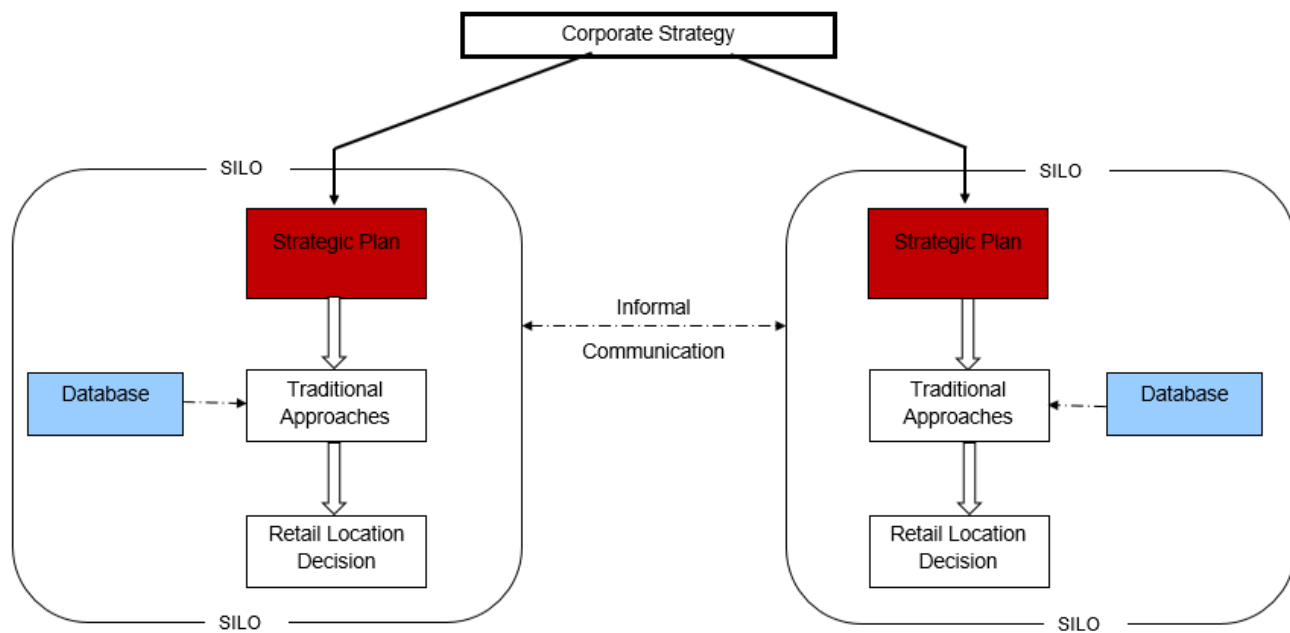


Figure 5: Reflexive Approach

5.3.3 Systemic Approach

Figure 6 identifies a Systemic Approach to data-driven RLDM. These environments are typically effective in integrating/using and developing traditional approaches to RLDM. This includes the use of more traditional spatial data (census data, competitor location data, POS data, loyalty program data etc.), and GIS to conduct competitor analysis, market mapping, cannibalization, trade area analysis, site screening and selection, network planning, customer profiling and acquisition and merger planning. Their ability to engage in data-driven decision-making is fostered by a relatively stronger data centered corporate strategy. In these environments there is more top down pressure to innovate or improve their RLDM processes to include more data-driven decision-making. Therefore, any growth or support for using SBD is more likely to come from not only the individual business units but also from C-level individuals within the organizations.

Since there is emphasis on the data centered corporate approach to decision-making, it is common that these types of organizations have a centralized data warehouse accessible to most business units. These warehouses rely on traditional relational database models (only capable of handling structured data) in order to house the organization's data. This includes data related to operations (Sales/SQFT, Inventory) or even point of sale (loyalty program data). The data warehouse is typically controlled by the IT department within the organization. They are typically responsible for providing access to the database, as well as, its overall maintenance. While there are Enterprise level data sources available, individual business units still have the ability to acquire, manage and update their own data. The degree of data sophistication that exist within those individual data units as well as the corresponding RLDM

techniques available for use are largely dependent on the employees within those departments. Within these types of organizations the corporate pressure to adopt new practices while existing is typically not strong enough to overhaul their approach to RLDM. Therefore if the individual business units responsible for location decisions don't prioritize the acquisition of new practices, procedures and data SBD adoption of any kind is not likely. Even within certain organizations where an individual business unit wants to adopt SBD and SBD analytics they tend to encounter significant challenges when integrating more sophisticated techniques. A major issue for these organizations is that they experience major talent gaps because the traditional skillsets of the employees within those individual business units does not include new SBD techniques. These units typically do not employ data scientists and therefore they likely need to hire employees with different skillsets. Even the organizations that have attempted to overcome this talent issue by hiring individuals with data science backgrounds (typically one person) are unable to effectively integrate and develop SBD practices within reasonable timeframes.

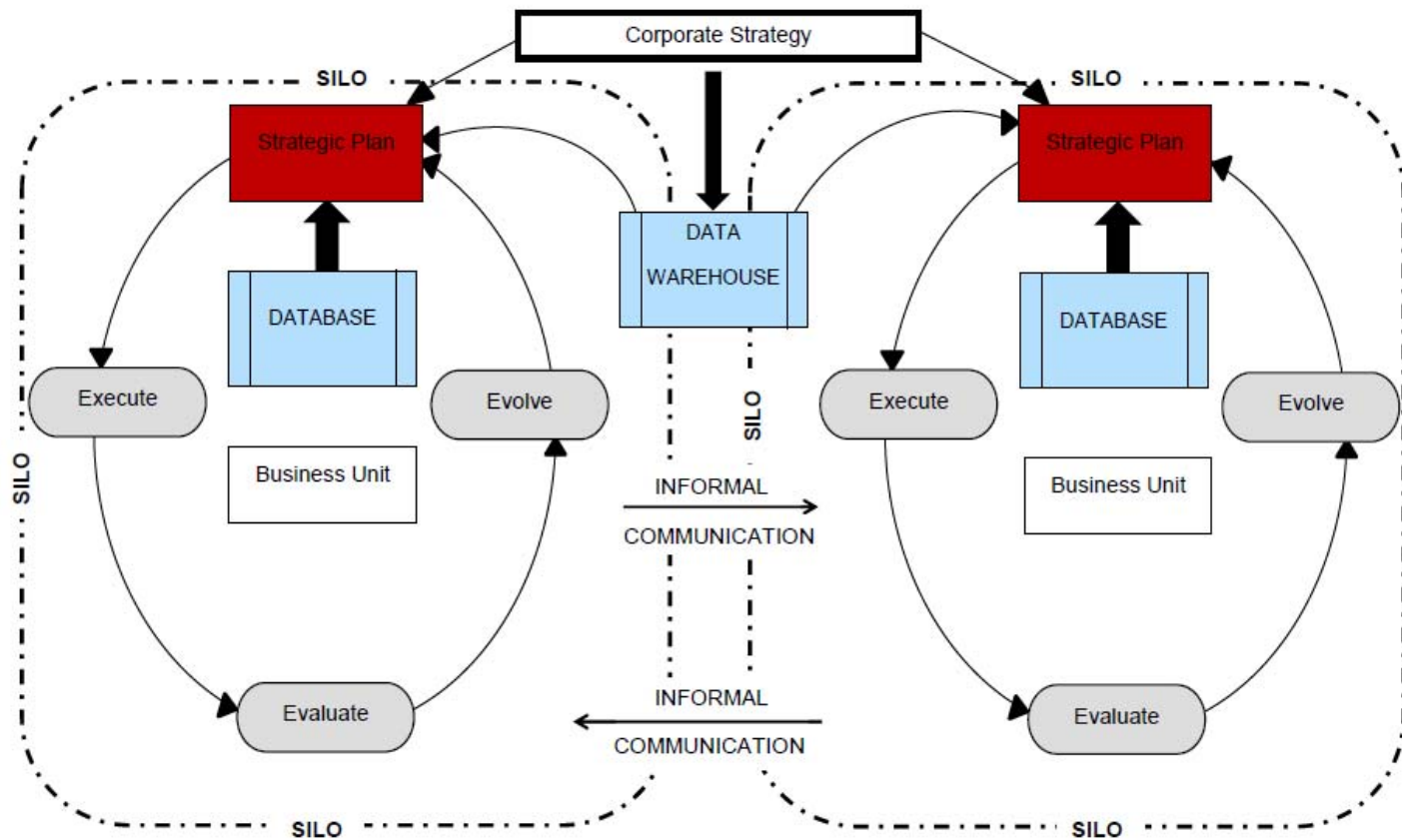


Figure 6: Systemic Approach

5.4 A FRAMEWORK FOR SBD ASSIMILATION AND ADOPTION IN RLDM

SBD practices were investigated in order to identify how and if they are being assimilated into RLDM practices. It was found that SBD assimilation for RLDM moved through a four stage process: (i) awareness, (ii) evaluation (iii) adoption, and, (iv) routinization (Figure 7).

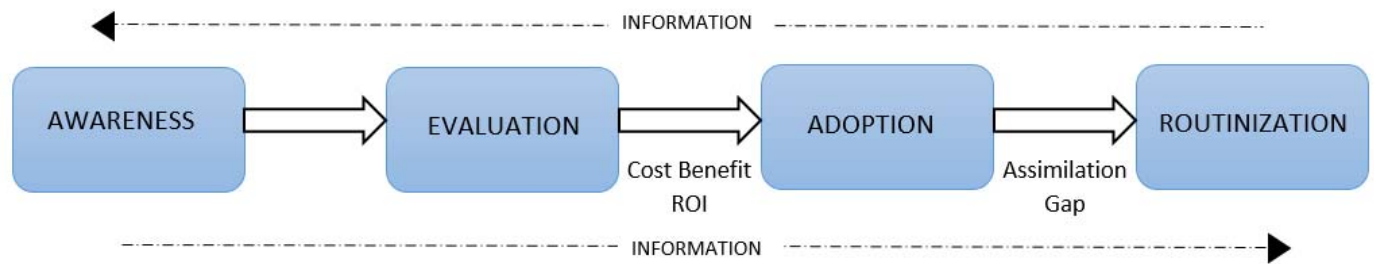


Figure 7: Framework for SBD Assimilation

Before any SBD practices can be assimilated in RLDM processes there has to be an awareness of what technologies and data are available. The level of awareness needs to ideally be organization-wide or else corporate resistance becomes more common (as illustrated in Figure 3, 5 and 7). Therefore it needs to be hierarchical in nature, coming from the top and moving down into the business units. Once retail organizations become aware of what's available they then move through a process of evaluating the benefits (cost benefit analysis) that the SBD practices can offer in order to improve the way that RLDM takes place. If the technology or data source is seen to have significant potential at increasing the quality of retail location decisions it then moves into the third stage; adoption. At the adoption stage a corporate decision is made to actually acquire the data or technology. It was at this point that the retail organizations would receive funding to buy software (such as Alterxy) or to set up new data collecting methods (Ibeacons). The final stage of assimilation, and one that proved to be the most difficult was the routinization of the SBD processes into day to day RLDM. SBD was seen to have high implementation complexity (data warehouse, new technologies, methods, and personnel) and therefore, for most firms SBD implementation never actually occurred.

The antecedents of SBD assimilation proved to be related to technological, organizational and environmental themes (Figure 8). From a technological perspective, adoption was largely linked to Rogers' (1995) attributes for predicting technology adoption, specifically, relative advantage, compatibility, and complexity. Through the structured interviews and case work it was clear that SBD assimilation was largely affected by these three characteristics. For any retail firm to adopt a new data source or technology to handle SBD the relative advantage of the SBD innovation over existing practices would need to be clear. Any hesitation with this and a chasm would exist between the awareness and evaluation stage and adoption stage of SBD assimilation. Therefore, the new SBD innovation would undeniably need to supersede the previous technique. From a compatibility perspective, if the innovation was not consistent with the existing values of the retail organization (e.g. data centered corporate strategy), current practices (e.g. data warehouse versatile enough to handle all sorts of data) and the needs of the business unit responsible for RLDM the adoption and routinization of SBD innovations would not happen. If the current infrastructure is not in place to adopt new SBD innovations it is not likely to gain traction in order to move towards adoption. Finally it was clear that SBD innovations that were seen to be overly complex to understand and use would not progress from the evaluation stage of assimilation to adoption.

From an organizational perspective the likelihood of SBD assimilation in RLDM are related to the scope of RLDM done within the retail firm, the size of the retail organization, and the skillset and knowledge of the employees. Retail firms that had historically engaged in data centered RLDM for a variety of location decisions were more prone to adopting new SBD

practices. It was also found that the larger retail organizations (measured by number of store locations) had the greatest appetite for adoption, as they were more likely to have the resources, mainly monetary, to assimilate SBD. Without having the funds to invest in new technologies and methods, it was not possible to gain greater sophistication in analysis. The skill level of the managers and employees responsible for RLDM also indicated whether a retailer was likely to adopt new practices. If the skillsets of these teams were not in line with the new SBD innovations there would be severe hesitation in bringing on new techniques and data sources.

From an environmental perspective the assimilation of SBD is dependent on both the retail sector and the competitive environment. Retailers were very cognizant of what their competitors were doing from a data analytics perspective. Knowledge of what competitors were doing, in large part, came from industry symposiums and conferences where technology adoption is heavily discussed. Therefore, it was not uncommon for a retail organization to actively seek certain tools and techniques that were being discussed by industry competition. In other words, retailers will actively follow other organizations that indicate that they have successfully implemented certain SBD practices. For example, when looking at developers and property managers of large shopping centres, the use of sensory-based technology has been widely adopted however, the likelihood of effective integration is not realized equally amongst all parties. Retail sector differences also encouraged or discouraged the likelihood of adoption. Some retail sectors have the privilege of rich-data which easily streams into a centralized place. Some organizations have already built infrastructures that allow for a seamless collection of customer data; the finance sector is a very good example of this. Some sectors don't have the

same privileges, mainly due to not having easy mechanisms in place for data collection; an example of this is the food services sector. Other things, such as whether retail locations are corporately owned or franchised also plays a role in the likelihood of adopting a strong data-centric decision-making platform. When decisions have multiple players with multiple perspectives, it is more complicated to find unity in best practices.

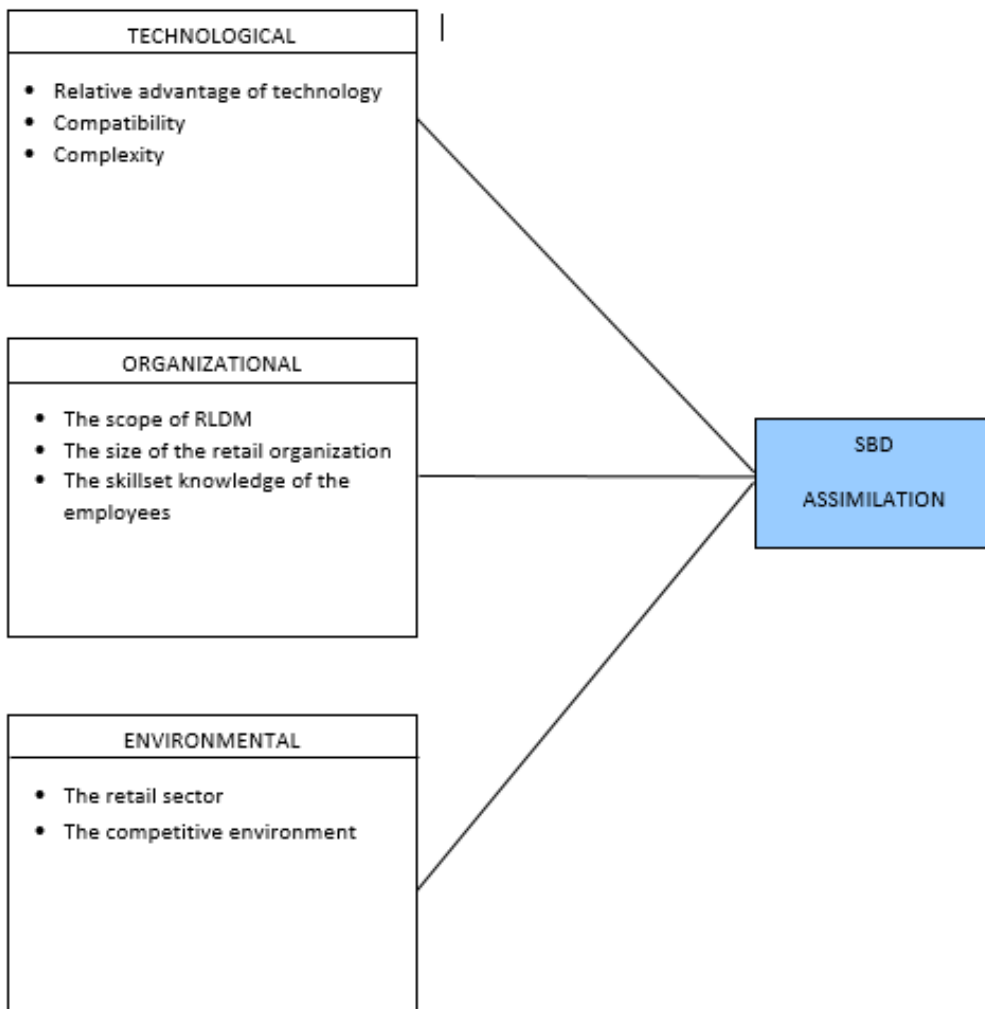


Figure 8: The Antecedents of SBD Assimilation Conclusion

6 CONCLUSION

As retail businesses move towards demanding more SBD-driven decision-making there is an added pressure to leverage data into the decision-making process. Generating data-driven results to inform the strategies and processes for RLDM represents a major challenge for retailers. This dissertation has identified that information hoarding, a lack of understanding from senior management on how technologies/methods/data can be used for RLDM, and a lack of skilled data analysts who can manage and synthesize the data, to be significant hurdles for SBD adoption. Retailers at the forefront of SBD analytics have dedicated time and money, including the hiring of new personnel with data science skill sets, in order to foster the development of new methodologies. With SBD and SBD analytics technologies being the most influential IT innovations in the last decade (Wang and Hajli, 2016) attention needs to be given to both the technical challenges as well as the organizational strategic views that could obstruct the advancement of SBD analytics.

While this research suggests that the adoption of more formalized SBD practices are starting to affect RLDM there are still questions around how this increases the quality of the decision a retail firm makes, even with advancements in analytical methodologies traditional techniques still hold an important place in RLDM. There is growing concern that SBD will dilute this value of experience and institutional knowledge in RLDM. SBD adoption is in an infancy stage within the retail sector and as we move through this fourth wave of RLDM the ultimate effect that SBD adoption will have on RLDM techniques is still uncertain.

6.1 SUMMARY OF FINDINGS

When considering the first objective (Objective 1: To explore the type and scale of location decisions that retail firms are currently undertaking), traditional practices still dominate the RLDM process. While the adoptions of SBD applications are starting to appear within retail planning, they are not widespread. Retailers still largely rely on traditional approaches (data and techniques) when engaged RLDM. Although experience remains the single most readily used method in RLDM it is clear that the decision-making techniques that are leveraged in decision-making are growing when compared to the past.

In regards to the second objective (Objective 2: To identify the availability and use of technology and SBD within the decision-making process), the retail sector has experienced changes in: (1) The availability and use of technology and geospatial data within the decision-making process, (2) The type and scale of location decisions that a firm undertakes, (3) The range of location research methods that are deployed within retail firms. Although there have been changes in the retail landscape, a number of aspects of location decision-making remain unchanged. Firstly, retail decision-makers still rely heavily on traditional data sources, such as those highlighted in past studies by Hernandez and Emmons (2012) and Byrom *et al.* (2001). For instance, Census Data (Demographic and Socio-Economic), Population Projection and Estimate Data, and Own Store Data (e.g. store sales, customer counts) are still used more frequently than other more novel data sources (e.g., Customer Surveillance Data, Social Media Data, Customer Mobile Application, Crowdsourced Data).

Finally, SBD assimilation (awareness, adoption and routinization) (Objective 3), is influenced by: The Retail Sector, The Organization, and the Technologies Available. At the heart of SBD adoption is a data environment that promotes transparency and a clear corporate strategy. While most retailers are aware of the SBD techniques that exist (Predictive Analytics, Machine Learning, Social Media Analytics: Content-Based Analytics, Social Media Analytics: Structure-Based Analytics, Social Influence Analysis, Sentiment Analysis, Association Rule Learning, Real Time Data Visualization, Real-Time Demand Forecasts, Text Analytics) few are leveraging these techniques in RLDM. Accessing the data needed to conduct these methods was documented as the greatest hurdle for doing this. While accessing more granular level customer data via Store Card Data (e.g., loyalty programs) Store Credit Card Data, Email/Electronic customer database, Social Media Data, Customer Surveillance Data (mobile tracking), Customer Mobile Application (retailer apps leveraging location data), Crowdsourced Data) was of high importance to retailers, few have managed to find ways to either effectively collect or integrate this type of data.

6.2 LIMITATIONS

It is difficult for any one of the research methods deployed to completely unravel the complexities of organizational decision-making, hence the use of all three methods. Case studies alone present three major limitations to conducting qualitative research. If you have too few case subjects there can be very little basis for scientific generalization (Yin, 1993; Zainal, 2007). By adding the online questionnaire and semi-structured interviews this limitation was

minimized as there was an abundance of data to pull from in order to generalize SBD practices within the retail environment in Canada. Another limitation had to do with the interviewees and respondents of all three survey methods. These individuals largely worked in real estate or research departments (about 75%) making it challenging to get a full picture of RLDM, as some location decisions are made in other departments. Through the case studies it was possible to talk to people from different departments but the number was still quite small in comparison.

There was a fair bit of contradictions in how people within the same departments or even the same institutions viewed their SBD operations which meant follow-up interviews had to be conducted in order to verify true practices. This was minimized by talking to several individuals in order to tease out similar responses or to confirm whether or not certain RLDM procedures were actually done. Furthermore, it was difficult to obtain a holistic picture of which retail sectors are early adopters versus laggards in their use of SBD practices for RLDM as multiple retailers of the same sector were not utilized for the case study. While this was investigated through the semi-structured interviews this might be a worthwhile area for greater exploration.

6.3 FUTURE RESEARCH

While this dissertation looks at the role that certain retail sectors play in the adoption, awareness and development of SBD practices it could be beneficial to dig deeper into the RLDM practices of one individual sector. In order to do this multiple case subjects from the same retail sector will need to be interviewed. Even though some similarities and differences between

retail sectors were discussed in this dissertation it did not delve deep into the minutia of how the characteristics of certain retail sectors might make SBD adoption easier or more difficult when making retail location decisions. Furthermore, it may also be beneficial to look at how the retail sector might differ from other types of private and even public organizations. By investigating how RLDM is conducted outside the retail world, it can be identified whether the retail industry is a laggard or innovator in the adoption, development and awareness in SBD practices for the use of location decisions.

REFERENCES

- Ahalt, S., & Kelly, K. (2013). The big data talent gap. *UNC Kenan-Flagler Business School White Paper*, 1-15.
- Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment?. *International Journal of Production Economics*, 182, 113-131.
- Allahviranloo, M., & Recker, W. (2015). Mining activity pattern trajectories and allocating activities in the network. *Transportation*, 42(4), 561-579.
- Alonso, W. (1960). A theory of the urban land market. *Papers in Regional Science*, 6(1), 149-157.
- Applebaum, W. (1966). Methods for determining store trade areas, market penetration, and potential sales. *Journal of Marketing Research*, 127-141.
- Ammenwerth, E., Iller, C., & Mahler, C. (2006). IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. *BMC medical informatics and decision making*, 6(1), 3.
- Anderson, C. (2008) 'The end of theory, will the data deluge makes the scientific method obsolete?', Edge, [Online] Available at: http://www.edge.org/3rd_culture/anderson08/anderson08_index.html
- Anderson, J. (1971). Space-time budgets and activity studies in urban geography and planning. *Environment and Planning A*, 3(4), 353-368.
- Arranz-López, A., Soria-Lara, J. A., López-Escolano, C., & Campos, Á. P. (2017). Retail Mobility Environments: A methodological framework for integrating retail activity and non-motorised accessibility in Zaragoza, Spain. *Journal of transport geography*, 58, 92-103.
- Atkinson, R., & Flint, J. (2001). Accessing hidden and hard-to-reach populations: Snowball research strategies. *Social research update*, 33(1), 1-4.
- Atulkar, S., & Kesari, B. (2018). Role of consumer traits and situational factors on impulse buying: does gender matter?. *International Journal of Retail & Distribution Management*, 46(4), 386-405.
- Batey, P., & Brown, P. (1995). From human ecology to customer targeting: the evolution of geodemographics. *GIS for business and service planning*, 77-103.

Baxter, J., & Eyles, J. (1997). Evaluating qualitative research in social geography: establishing 'rigour' in interview analysis. *Transactions of the Institute of British geographers*, 22(4), 505-525.

Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The qualitative report*, 13(4), 544-559.

Bazeley, P. (2009). Integrating data analyses in mixed methods research. *Journal of Mixed Methods Research*, 3(3), 203-207.

Beaumont, J. R. (1980). Spatial Interaction Models and the Location-Allocation Problem. *Journal of Regional Science*, 20(1), 37-50. Beck, N., & Rygl, D. (2015). Categorization of multiple channel retailing in Multi-, Cross-, and Omni-Channel Retailing for retailers and retailing. *Journal of Retailing and Consumer Services*, 27, 170-178.

Benoit, D., & Clarke, G. P. (1997). Assessing GIS for retail location planning. *Journal of retailing and consumer services*, 4(4), 239-258.

Berry, B. J., & Garrison, W. L. (1958). Recent developments of central place theory. *Papers in Regional Science*, 4(1), 107-120.

Berry, B. J. (1964). Cities as systems within systems of cities. *Papers in regional science*, 13(1), 147-163.

Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research*, 10(2), 141-163.

Birkin, M., Clarke, G., & Clarke, M. (2010). Refining and Operationalizing Entropy-Maximizing Models for Business Applications. *Geographical Analysis*, 42(4), 422-445.

Birkin, M., Clarke, G., & Clarke, M. P. (2002). *Retail geography and intelligent network planning*. John Wiley & Sons.

Bizer, C., Boncz, P., Brodie, M. L., & Erling, O. (2012). The meaningful use of big data: four perspectives--four challenges. *ACM SIGMOD Record*, 40(4), 56-60.

Bollier, D., & Firestone, C. M. (2010). *The promise and peril of big data* (p. 1). Washington, DC: Aspen Institute, Communications and Society Program.

Boufounou, P. V. (1995). Evaluating bank branch location and performance: A case study. *European Journal of Operational Research*, 87(2), 389-402.

Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662–679.

Brown, S. (1993). Retail location theory: evolution and evaluation. *International Review of Retail, Distribution & Consumer Research*, 3(January), 185.

Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'. *McKinsey Quarterly*, 4(1), 24-35.

Brosekhan, a. A. B., Velayutham, C. M., & Phil, M. (1995). Consumer Buying Behaviour – A Literature Review. *Journal of Business and Management*, 8–16.

Brynjolfsson, E., Hitt, L., & Kim, H. (2015). Strength in Numbers : How Does Data-Driven Decisionmaking Affect Firm Performance. *Ssrn Electronic*, (August), 1–28.

Burke, R. R. (2002). Technology and the Customer Interface: What Consumers Want in the Physical and Virtual Store. *Journal of the Academy of Marketing Science*, 30(4), 411–432.

Byrom, J. (2001). The role of loyalty card data within local marketing initiatives. *International Journal of Retail & Distribution Management*, 29(7), 333–342.

Byrom, J. W., Bennison, D. J., Hernández, T., & Hooper, P. D. (2001). The use of geographical data and information in retail locational planning. *Journal of Targeting, Measurement and Analysis for Marketing*, 9(3), 219-229.

Campbell, C. (2018). Top Five Differences between Data Lakes and Data Warehouses. Retrieved August 17, 2018, from <https://www.blue-granite.com/blog/bid/402596/top-five-differences-between-data-lakes-and-data-warehouses>

Carolan, M. (2018). Big data and food retail: Nudging out citizens by creating dependent consumers. *Geoforum*, 90, 142-150.

Cassell, C., & Symon, G. (Eds.). (2004). *Essential guide to qualitative methods in organizational research*. Sage.

Chaston, I. (2015). *Internet marketing and big data exploitation*. Palgrave Macmillan.

- Chacón-García, J. (2017). Geomarketing techniques to locate retail companies in regulated markets. *Australasian Marketing Journal (AMJ)*, 25(3), 185-193.
- Chang, H. J., & Hsieh, C. M. (2018). A new model for selecting sites for chain stores in China. *International Journal of Industrial and Systems Engineering*, 28(3), 346-359.
- Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314-347.
- Chen, M. S., Han, J., & Yu, P. S. (1996). Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and data engineering*, 8(6), 866-883.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data To Big Impact. *Mis Quarterly*, 36(4), 1165–1188.
- Cheng, E. W. L., Li, H., & Yu, L. (2005). The analytic network process (ANP) approach to location selection: a shopping mall illustration. *Construction Innovation (Sage Publications,Ltd.)*, 5(2), 83–97.
- Cheng, X., Fang, L., Hong, X., & Yang, L. (2017). Exploiting mobile big data: Sources, features, and applications. *IEEE Network*, 31(1), 72-79.
- Christaller, W. (1933), Central Places in Southern Germany, Jena, Fischer
- Christmann H, Alexander A, Wood SM (2015) Exploring Brand Identity and Entrepreneurship as Drivers of Small Specialist Retailer Internationalisation: A German Case Study, *International Review of Retail, Distribution and Consumer Research* 26 (2) pp. 137-153 Taylor & Francis
- Clarke, G. (1998). Changing methods of location planning for retail companies. *GeoJournal*, 45(4), 289.
- Clark, G. L., & Wrigley, N. (1997). Exit, the firm and sunk costs: reconceptualizing the corporate geography of disinvestment and plant closure. *Progress in Human Geography*, 21(3), 338-358.
- Clark, W. A. (1968). Consumer travel patterns and the concept of range. *Annals of the Association of American Geographers*, 58(2), 386-396.
- Clarke, I., & Rowley, J. (1995). A case for spatial decision-support systems in retail location planning. *International Journal of Retail & Distribution Management*, 23(3), 4-10.

- Clarkson, R. M., Clarke-Hill, C. M., & Robinson, T. (1996). UK supermarket location assessment. *International Journal of Retail & Distribution Management*, 24(6), 22-33.
- Clawson, C. J. (2010). Locations , Performance to Standards , and Local Conditions. *Strategies*, 38(1), 8–14.
- Converse, P. D. (1949). New laws of retail gravitation. *Journal of marketing*, 14(3), 379-384.
- Couclelis, H. (2009). Rethinking time geography in the information age. *Environment and planning A*, 41(7), 1556-1575.
- Crowe, S., Cresswell, K., Robertson, A., Hubby, G., Avery, A., & Sheikh, A. (2011). The case study approach. *BMC medical research methodology*, 11(1), 100.
- Curry, M. R. (1997). The digital individual and the private realm. *Annals of the Association of American geographers*, 87(4), 681-699.
- Davenport, T. H., & Patil, D. J. (2012). Data scientist. *Harvard business review*, 90(5), 70-76.
- Davies, R. L. (1977) "Store location and store assessment research: the integration of some new and traditional techniques." *Transactions of the Institute of British Geographers* (1977): 141-157.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.
- DeLyser, D., & Sui, D. (2014). Crossing the qualitative-quantitative chasm III: Enduring methods, open geography, participatory research, and the fourth paradigm. *Progress in Human Geography*, 38(2), 294-307.
- Del Gatto, M., & Mastinu, C. S. (2018). A Huff model with firm heterogeneity and selection. Application to the Italian retail sector. *Spatial Economic Analysis*, 1-15.
- Denscombe, M. (2009). Item non-response rates: a comparison of online and paper questionnaires. *International Journal of Social Research Methodology*, 12(4), 281-291.
- Dijst, M., & Kwan, M. P. (2005). Accessibility and quality of life: time-geographic perspectives. *Social Dimensions of Sustainable Transport: Transatlantic Perspectives*, 109-126.

Doherty, S. T. (2009). Multisensor Monitoring of Human Activity and Physiology in the Built Environment. In Transportation Research Board 88th Annual Meeting (No. 09-2660).

Doherty, S. T. (2012). Exploring blood glucose variation over geographical space. *Diabetes technology & therapeutics*, 14(3), 276-284.

Doyle, P., Fenwick, I., & Savage, G. P. (1979). Management Planning and Control in Multi-Branch Banking. *The Journal of the Operational Research Society*, 30(2), 105–111.

Drezner, T. (1998). Location of multiple retail facilities with limited budget constraints - in continuous space. *Journal of Retailing and Consumer Services*, 5(3), 173–184.

Eaton, B. C., & Lipsey, R. G. (1982). An economic theory of central places. *The economic journal*, 92(365), 56-72.

Ellwood, L. W. (1954). Estimating potential volume of proposed shopping centers. *The Appraisal Journal*, 22(4), 583.

ELSamen, A. A. A., & Hiyasat, R. I. (2017). Beyond the random location of shopping malls: A GIS perspective in Amman, Jordan. *Journal of Retailing and Consumer Services*, 34, 30-37.

Emerson, R. W. (2015). Convenience sampling, random sampling, and snowball sampling: How does sampling affect the validity of research?. *Journal of Visual Impairment & Blindness (Online)*, 109(2), 164.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1-4.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897-904.

Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet research*, 15(2), 195-219.

Gable, G. G. (1994). Integrating case study and survey research methods: an example in information systems. *European journal of information systems*, 3(2), 112-126.

Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future*, 2007(2012), 1-16.

Gastner, M. T. (2011). Scaling and entropy in p-median facility location along a line. *Physical Review E*, 84(3), 036112.

Gebauer, H., & Laska, M. N. (2011). Convenience stores surrounding urban schools: An assessment of healthy food availability, advertising and product placement. *Journal of Urban Health*, 88(4), 616–622.

Ghemawat, P., & Khanna, T. (1998). The nature of diversified business groups: A research design and two case studies. *The Journal of Industrial Economics*, 46(1), 35-61.

Ghosh, A., & McLafferty, S. L. (1987). Location strategies for retail and service firms. Lexington, MA: Lexington Books.

Gong, L., Liu, X., Wu, L., & Liu, Y. (2016). Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*, 43(2), 103-114.

González-Benito, Ó. (2002). Geodemographic and socioeconomic characterization of the retail attraction of leading hypermarket chains in Spain. *The International Review of Retail, Distribution and Consumer Research*, 12(1), 81-103.

Goodchild, M. F. (1984). ILACS: A Location Allocation Model for Retail Site Selection. *Journal of Retailing*. 6(1), 84-100

Goodchild, M. F. (2013). The quality of big (geo) data. *Dialogues in Human Geography*, 3(3), 280–284.

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS quarterly*, 213-236.

Gorden R. Coding interview responses In: Basic interviewing skills. Long Grove (IL): Waveland Press; 1992.

Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, 3(3), 255–261.

Green, D. (2018, August 08). If you shopped at these 15 stores in the last year, your data might have been stolen. Retrieved August 9, 2018, from <https://www.businessinsider.com/data-breaches-2018-4#sears-3>

Gutierrez, A., Boukrami, E., & Lumsden, R. (2015). Technological, organisational and environmental factors influencing managers' decision to adopt cloud computing in the UK. *Journal of Enterprise Information Management*, 28(6), 788-807.

Gunasekaran, A., Yusuf, Y. Y., Adeleye, E. O., & Papadopoulos, T. (2018). Agile manufacturing practices: the role of big data and business analytics with multiple case studies. *International Journal of Production Research*, 56(1-2), 385-397.

Gunter, B., & Furnham, A. (2014). *Consumer Profiles (RLE Consumer Behaviour): An Introduction to Psychographics*. Routledge.

Hagberg, J., & Holmberg, U. (2017). Travel modes in grocery shopping. *International Journal of Retail & Distribution Management*, 45(9), 991-1010.

Hägerstrand, T. (1970, December). What about people in regional science?. In *Papers of the Regional Science Association* (Vol. 24, No. 1, pp. 6-21). Springer-Verlag.

Han, J., & Yamana, H. (2016). A study on individual mobility patterns based on individuals' familiarity to visited areas. *International Journal of Pervasive Computing and Communications*, 12(1), 23-48.

Handa, M., & Vohra, A. (2010). Geographic Information Systems: A Decision Support Tool for Business. *Drishtikon : A Management Journal*, 1(2), 123–141.

Hargreaves, I., Roth, D., Karim, M. R., Nayebi, M., & Ruhe, G. (2018). Effective Customer Relationship Management at ATB Financial: A Case Study on Industry-Academia Collaboration in Data Analytics. In *Highlighting the Importance of Big Data Management and Analysis for Various Applications* (pp. 45-59). Springer, Cham. Hasgall, A., & Ahituv, N. (2018). Implementing continuous adaptation to technology innovation in complex adaptive organizations. *The Journal of High Technology Management Research*.

Hay, I. (2005). *Qualitative research methods in human geography*. South Melbourne, Vic.: Oxford University Press.

Henderson, C., Dancy, M., & Niewiadomska-Bugaj, M. (2012). Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process?. *Physical Review Special Topics-Physics Education Research*, 8(2), 020104.

Hebert, D., Anderson, B., Olinsky, A., & Hardin, J. M. (2014). Time series data mining: a retail application. *International journal of business analytics (IJBAN)*, 1(4), 51-68.

Hernandez, T. and Emmons, M. Driving Decisions: the use of GIS by retail organizations in Canada, *Proceedings of the Applied Geography Conferences*, Vol 35, pp. 240-248, 2012.

Hernandez, T., Bennison, D., & Cornelius, S. (1998). The organizational context of retail locational planning. *GeoJournal*, 45(4), 299–308.

Hernandez, T., & Bennison, D. (2000). The Art and science of retail location decisions (UK). *International Journal of Retail & Distribution Management*, 28(8-9), 2013.

Hindle, G. A., & Vidgen, R. (2018). Developing a business analytics methodology: A case study in the foodbank sector. *European Journal of Operational Research*, 268(3), 836-851.

Hormozi, A. M., & Giles, S. (2004). Data mining: A competitive weapon for banking and retail industries. *Information systems management*, 21(2), 62-71.

Hruschka, D. J., Schwartz, D., St. John, D. C., Picone-Decaro, E., Jenkins, R. A., & Carey, J. W. (2004). Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field methods*, 16(3), 307-331.

Hu, P. J., Chau, P. Y., Sheng, O. R. L., & Tam, K. Y. (1999). Examining the technology acceptance model using physician acceptance of telemedicine technology. *Journal of management information systems*, 16(2), 91-112.

Huff, D.L. (1964). Defining and Estimating a Trading Area. *Journal of Marketing*, July 28, pp. 34-38.

Huff, D.L. (1963). Probabilistic Analysis of Shopping Center Trade Areas. *Land Economics*, 39, pp.40-46

Jadhav, D. K. (2013). Big data: the new challenges in data mining. *Int J Innov Res ComputSci & Technol*, 1(2), 39-42.

Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, 57(7), 86-94.

Jayakrishnan, M., Mohamad, A. K., & Yusof, M. M. (2018). Assimilation of Business Intelligence (BI) and Big Data Analytics (BDA) Towards Establishing Organizational Strategic Performance Management Diagnostics Framework: A Case Study. *Journal of Digital Information Management*, 16(1).

Jones, K., & Simmons, J. W. (1993). Location, location, location. Nelson Canada.

- Karande, K., & Lombard, J. R. (2005). Location strategies of broad-line retailers: an empirical investigation. *Journal of Business Research*, 58(5), 687-695.
- Kaynak, E., & Harcar, T. D. (2005). American consumers' attitudes towards commercial banks: A comparison of local and national bank customers by use of geodemographic segmentation. *International Journal of bank marketing*, 23(1), 73-89.
- Keim, D., Qu, H., & Ma, K. L. (2013). Big-data visualization. *IEEE Computer Graphics and Applications*, 33(4), 20-21.
- Kemp, K. (Ed.). (2007). *Encyclopedia of geographic information science*. SAGE publications.
- Khine, P. P., & Wang, Z. S. (2018). Data lake: a new ideology in big data era. In *ITM Web of Conferences* (Vol. 17, p. 03025). EDP Sciences.
- Klemperer, P. (1992). Equilibrium Product Lines: Competing Head-to-Head May Equilibrium be Less Competitive. *The American Economic Review*, 82(4), 740–755.
- Kitchin, R. (2013). Big Data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3), 262–267.
- Kohavi, R., Mason, L., Parekh, R., & Zheng, Z. (2004). Lessons and challenges from mining retail e-commerce data. *Machine Learning*, 57(1-2), 83-113.
- Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, 34(3), 387-394.
- Lai, P. C. (2017). The Literature Review of Technology Adoption Models and Theories For The Novelty Technology. *JISTEM-Journal of Information Systems and Technology Management*, 14(1), 21-38.
- Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*, 52(2), 21–32.
- Lee, D. H. (2013). Smartphones, mobile social space, and new sociality in Korea. *Mobile Media & Communication*, 1(3), 269-284.

Lee, J. G., & Kang, M. (2015). Geospatial big data: challenges and opportunities. *Big Data Research*, 2(2), 74-81.

Leszczynski, A., & Crampton, J. (2016). Introduction: Spatial big data and everyday life. *Big Data & Society*, 3(2), 2053951716661366.

Lloyd, A., & Cheshire, J. (2017). Deriving retail centre locations and catchments from geo-tagged Twitter data. *Computers, Environment and Urban Systems*, 61, 108-118.

Longley, P. (2011). *Geographic information systems & science*. Hoboken, NJ: Wiley

Lyon, L., & Brenner, A. (2015). Bridging the data talent gap: Positioning the iSchool as an agent for change. *International journal of digital curation*, 10(1).

Marros, R. J. (2005). Using a Geographic Information System (GIS) to Visualize and Analyze Spatial Location in a Retail Environment. *University of Minnesota, Winona*. 1–10.

Matinaro, V., & Liu, Y. (2015). Virtual design and construction: innovation process and diffusion in Finnish construction business. *International Journal of Innovation and Learning*, 18(2), 133-150.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 60-68.

Mendes, A. B., & Themido, I. H. (2004). Multi-outlet retail site location assessment. *International Transactions in Operational Research*, 11(1), 1-18.

Meyer, C. B. (2001). A case in case study methodology. *Field methods*, 13(4), 329-352.

Miller, H. J. (1991). Modelling accessibility using space-time prism concepts within geographical information systems. *International Journal of Geographical Information System*, 5(3), 287-301.

Miller, H. J. (1999). Measuring space-time accessibility benefits within transportation networks: basic theory and computational procedures. *Geographical analysis*, 31(1), 1-26.

Mitchell, V. W., & McGoldrick, P. J. (1994). The role of geodemographics in segmenting and targeting consumer markets: A Delphi study. *European Journal of marketing*, 28(5), 54-72.

Moore, S., & Attewell, G. (1991). To be and where not to be. *OR Insight*, 4(1), 21-24.

Morehouse, R. E., & Maykut, P. (2002). *Beginning qualitative research: A philosophical and practical guide*. Routledge.

Moslehpour, M., Thanh, H. L. T., & Van Kien, P. (2018, January). Technology Perception, Personality Traits and Online Purchase Intention of Taiwanese Consumers. In *International Conference of the Thailand Econometrics Society* (pp. 392-407). Springer, Cham.

Murad, A. (2008). Creating a GIS application for defining retail catchment area at Jeddah City. *International Journal of Services, Technology and Management*, 9(1), 79–92.

Narvaez, L., Penn, A., & Griffiths, S. (2013). Spatial configuration and bid rent theory: How urban space shapes the urban economy. *Proceedings of the Ninth International Space Syntax Symposium*, 89:1–19.

Nelson, R. L., & Curtis, A. (1958). *The selection of retail locations*. New York: FW Dodge Corporation.

Nicholson, M., Clarke, I., & Blakemore, M. (2002). “One brand, three ways to shop” : situational variables and multichannel consumer behaviour. *The International Review of Retail, Distribution and Consumer Research*, 3969(July 2015), 131–148.

O’Dwyer, J. & Coveney, L. (2006). Scoping supermarket availability and accessibility by socio-economic status in Adelaide. *Health Promotion Journal of Australia*, 17(3), 180.

O’Malley, L., Patterson, M., & Evans, M. (1997). Retailer use of geodemographic and other data sources: an empirical investigation. *International Journal of Retail & Distribution Management*, 25(6), 188–196. Okabe, A., & Okunuki, K. I. (2001). A computational method for estimating the demand of retail stores on a street network and its implementation in GIS. *Transactions in GIS*, 5(3), 209-220.

Okeahalam CC, Wood S (2009) Financing internationalisation: a case study of an African retail transnational corporation, *Journal of Economic Geography* 9 (4) pp. 511-537

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533-544.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. SAGE Publications, inc.

Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International journal of electronic commerce*, 7(3), 101-134.

Perea y Monsuwé, T., Dellaert, B. G. C., & de Ruyter, K. (2004). What drives consumers to shop online? A literature review. *International Journal of Service Industry Management*, 15(1), 102–121.

Piatetsky-Shapiro, G. (1996). *Advances in knowledge discovery and data mining* (Vol. 21). U. M. Fayyad, P. Smyth, & R. Uthurusamy (Eds.). Menlo Park: AAAI press.

Pikkarainen, T., Pikkarainen, K., Karjaluoto, H., & Pahnla, S. (2004). Consumer acceptance of online banking: an extension of the technology acceptance model. *Internet research*, 14(3), 224-235.

Piotrowicz, W., & Cuthbertson, R. (2014). Introduction to the special issue information technology in retail: Toward omnichannel retailing. *International Journal of Electronic Commerce*, 18(4), 5-16.

Pooler, J. (1994). An extended family of spatial interaction models. *Progress in Human Geography*, 18(1), 17-39.

Pred, A. (1984). Place as historically contingent process: Structuration and the time-geography of becoming places. *Annals of the association of american geographers*, 74(2), 279-297.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data*, 1(1), 51-59.

Puklavec, B., Oliveira, T., & Popovič, A. (2018). Understanding the determinants of business intelligence system adoption stages: An empirical study of SMEs. *Industrial Management & Data Systems*, 118(1), 236-261.

Raju, P. S., Bai, D. V. R., & Chaitanya, G. K. (2014). Data mining: techniques for enhancing customer relationship management in banking and retail industries. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(1), 2650-2657.

Qu, Y., & Zhang, J. (2013, May). Trade area analysis using user generated mobile location data. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 1053-1064). ACM.

Ramanathan, U., Subramanian, N., & Parrott, G. (2017). Role of social media in retail network operations and marketing to enhance customer satisfaction. *International Journal of Operations & Production Management*, 37(1), 105-123.

Reigadinha, T., Godinho, P., & Dias, J. (2017). Portuguese food retailers—Exploring three classic theories of retail location. *Journal of Retailing and Consumer Services*, 34, 102-116.

Reilly, W. J. (1929). *Methods for the Study of Retail Relationships*, Research Monograph, No. 4, of the Bureau of Business Research, University of Texas.

Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. *Developments in applied statistics*, 19(1), 159-177.

Reynolds, J., & Wood, S. (2010). Location decision making in retail firms: evolution and challenge. *International Journal of Retail & Distribution Management*, 38(11), 828–845.

Rogers, D. S. (1984). *Store location and store assessment research*. John Wiley & Sons Inc.

Rogers, D. (2007). Retail location analysis in practice. *Research Review*, 12(2), 73–78. Retrieved from http://www.docunator.com/bigdata/2/1365585232_6cd4b48411/retail_location_analysis.pdf

Rogers Everett, M. (1995). *Diffusion of innovations*. New York, 12.

Ruiz-Ruiz, A. J., Blunck, H., Prentow, T. S., Stisen, A., & Kjærgaard, M. B. (2014). Analysis methods for extracting knowledge from large-scale wifi monitoring to inform building facility planning. In *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on* (pp. 130-138). IEEE.

Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on* (pp. 42-47). IEEE.

Sandelowski, M. (1995). Sample size in qualitative research. *Research in nursing & health*, 18(2), 179-d183.

Scott, D. M., & He, S. Y. (2012). Modeling constrained destination choice for shopping: a GIS-based, time-geographic approach. *Journal of Transport Geography*, 23, 60-71.

Sevtsuk, A., & Kalvo, R. (2018). Patronage of urban commercial clusters: A network-based extension of the Huff model for balancing location and size. *Environment and Planning B: Urban Analytics and City Science*, 45(3), 508-528.

Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., & Fotheringham, A. S. (2016). Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, 30(5), 881-906.

Simkin, L. P., Doyle, P. P., & Saunders, D. J. (1985). How retailers put site location techniques into operation: An assessment of major multiples' practice. *International Journal of Retail & Distribution Management*, 13(3), 21–26.

Simkin, L. P. (1989). SLAM: store location assessment model—theory and practice. *Omega*, 17(1), 53-58.

Simkin, L. P. (1990). Evaluating a store location. *International Journal of Retail & Distribution Management*, 18(4). <https://doi.org/10.1108/09590559010140345>

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263-286.

Shaw, S. L., & Yu, H. (2009). A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. *Journal of Transport Geography*, 17(2), 141-149.

Sleight, P. (1997). Targeting customers: How to use geodemographic and lifestyle data in your business. *Henley-on-Thames: NTC Publications*.

Sleight, P., & Leventhal, B. (1989). Applications of geodemographics to research and marketing. *journal of the Market Research Society*.31(1), 75-101.

Stanley, T. J., & Sewall, M. A. (1976). Image inputs to a probabilistic model: predicting retail potential. *Journal of Marketing*, 40(3), 48-53.

Stone, M., Bearman, D., Butscher, S. A., Gilbert, D., Crick, P., & Moffett, T. (2003). The effect of retail customer loyalty schemes—Detailed measurement or transforming marketing?. *Journal of Targeting, Measurement and Analysis for Marketing*, 12(3), 305-318.

Suel, E., & Polak, J. W. (2017). Development of joint models for channel, store, and travel mode choice: Grocery shopping in London. *Transportation Research Part A: Policy and Practice*, 99, 147-162.

Suomalainen, E. (2006). Multicriteria analysis and visualization of location-allocation problems. *Master's thesis, Department of Engineering Physics and Mathematics, University of Helsinki*.

Szolnoki, G., & Hoffmann, D. (2013). Online, face-to-face and telephone surveys—Comparing different sampling methods in wine consumer research. *Wine Economics and Policy*, 2(2), 57-66.

Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models. *Information systems research*, 6(2), 144-176.

Tellis, W. M. (1997). Application of a case study methodology. *The qualitative report*, 3(3), 1-19.

Tene, O., & Polonetsky, J. (2012). Big data for all: Privacy and user control in the age of analytics. *Nw. J. Tech. & Intell. Prop.*, 11, xxvii.

Thatcher, J. (2014). Big data, big questions| Living on fumes: Digital footprints, data fumes, and the limitations of spatial big data. *International Journal of Communication*, 8, 19.

Theodoridis, C., & Bennisson, D. (2009). Complexity theory and retail location strategy. *The International Review of Retail, Distribution and Consumer Research*, 19(4), 389-403.

Tornatzky, L. G., & Fleischer, M. (1990). *The processes of technological innovation*. Lexington, MA: Lexington Books.

Turner III, D. W. (2010). Qualitative interview design: A practical guide for novice investigators. *The qualitative report*, 15(3), 754-760. Turow, J., McGuigan, L., & Maris, E. R. (2015). Making data mining a natural part of life: Physical retailing, customer surveillance and the 21st century social imaginary. *European Journal of Cultural Studies*, 18(4-5), 464-478. USGIF, 2018 STATE AND FUTURE OF GEOINT REPORT. (2018). Trajectory Magazine. doi:http://trajectorymagazine.com/wp-content/uploads/2018/02/SFoG_2018.pdf

Vassakis, K., Petrakis, E., & Kopanakis, I. (2018). Big Data Analytics: Applications, Prospects and Challenges. In *Mobile Big Data* (pp. 3-20). Springer, Cham.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.

Verhetsel, A. (2005). Effects of neighbourhood characteristics on store performance supermarkets versus hypermarkets. *Journal of Retailing and Consumer Services*, 12(2), 141-150.

Verhoef, P. C., Kannan, P. K., & Inman, J. J. (2015). From multi-channel retailing to omni-channel retailing: introduction to the special issue on multi-channel retailing. *Journal of retailing*, 91(2), 174-181.

Vigneron, F., & Johnson, L. W. (2017). Measuring perceptions of brand luxury. In *Advances in Luxury Brand Management* (pp. 199-234). Palgrave Macmillan, Cham.

Vissak, T. (2010). Recommendations for using the case study method in international business research. *The Qualitative Report*, 15(2), 370-388.

Vlachopoulou, M., Silleos, G., & Manthou, V. (2001). Geographic information systems in warehouse site selection decisions. *International journal of production economics*, 71(1-3), 205-212.

Wang, Y., Jiang, W., Liu, S., Ye, X., & Wang, T. (2016). Evaluating trade areas using social media data with a calibrated huff model. *ISPRS International Journal of Geo-Information*, 5(7), 112.

Wee, C. H., & Pearce, M. R. (1985). Patronage behavior toward shopping areas: a proposed model based on Huff's model of retail gravitation. *ACR North American Advances* 12 (1). Wei, J., Lowry, P. B., & Seedorf, S. (2015). The assimilation of RFID technology by Chinese companies: a technology diffusion perspective. *Information & Management*, 52(6), 628-642.

Wood SM (2002) Organisational restructuring, knowledge and spatial scale: the case of the US department store industry, *Tijdschrift voor Economische en Sociale Geographie* 93 (1) pp. 8-33

Wood SM, Wrigley N, Coe NM (2016) Capital Discipline and Financial Market Relations in Retail Globalization: Insights from the Case of Tesco plc, *Journal of Economic Geography* 17 (1) pp. 31-57 Oxford University Press

Wood, S., & Reynolds, J. (2012). Leveraging locational insights within retail store development? Assessing the use of location planners' knowledge in retail marketing. *Geoforum*, 43(6), 1076-1087.

- Wood, S., & Tasker, A. (2008). The importance of context in store forecasting: the site visit in retail location decision-making. *Journal of Targeting, Measurement and Analysis for Marketing*, 16(2), 139-155.
- Working, H., & Hotelling, H. (1929). Applications of the theory of error to the interpretation of trends. *Journal of the American Statistical Association*, 24(165A), 73-85.
- Wrigley, N. (1987). The concentration of capital in UK grocery retailing. *Environment and Planning A*, 19(10), 1283-1288.
- Yaeli, A., Bak, P., Feigenblat, G., Nadler, S., Roitman, H., Saadoun, G., ... & Sandbank, T. (2014). Understanding customer behavior using indoor location analysis and visualization. *IBM Journal of Research and Development*, 58(5/6), 3-1.
- Yin, R. K. (2003). Applications of case study research. Applied social research methods series. Thousand Oaks: Sage Publications. Yu, KH (2013). Institutionalization in the context of institutional pluralism: Politics as a generative process, *Organization Studies*, 34(1), 105-131.
- Yin, R. K. (2011). *Applications of case study research*. Sage.
- Zeng, T. Q., & Zhou, Q. (2001). Optimal spatial decision making using GIS: a prototype of a real estate geographical information system (REGIS). *International Journal of Geographical Information Science*, 15(4), 307-321.
- Zerbino, P., Aloini, D., Dulmin, R., & Mininno, V. (2018). Big Data-enabled Customer Relationship Management: A holistic approach. *Information Processing & Management*.
- Zhu, K., Kraemer, K. L., & Xu, S. (2006). The process of innovation assimilation by firms in different countries: a technology diffusion perspective on e-business. *Management science*, 52(10), 1557-1576.

