Electronic Theses, Projects, and Dissertations   Office of Graduate Studies

12-2022

BIG DATA FOR COMPREHENSIVE ANALYSIS OF REAL ESTATE MARKET

YongLin Xiao

BIG DATA FOR COMPREHENSIVE ANALYSIS OF

REAL ESTATE MARKET

_____

A Project

Presented to the

Faculty of

California State University,

San Bernardino

_____

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Information System and Technology

_____

by

YongLin Xiao

December 2022

BIG DATA FOR COMPREHENSIVE ANALYSIS OF

REAL ESTATE MARKET

_____

A Project

Presented to the

Faculty of

California State University,

San Bernardino

_____

by

YongLin Xiao

December 2022

Approved by:

Dr. Conrad Shayo, Committee Chair

Dr. Jesus Canelon, Committee Member

Dr. Conrad Shayo, Dept. Chair, Information and Decision Sciences

ABSTRACT

This Culminating Experience Project explored the application of big data in the real estate industry in order to address the problem of analyzing the accurate property estimates value. The research questions were: (Q1): What are the benefits and advantages of utilizing big data in the real estate market? (Q2): What are the trends in the application of big data in the real estate market? (Q3): What are the challenges in applying big data in the real estate market? (Q4): What are the methods and processes of applying big data in appraisal of assets in the real estate market? To answer these four questions, this study used qualitative and quantitative methodology, content analysis conducted on data collected through Google Scholar, and One Search for industry reports, conference papers, and select literature about big data adoption trends in the real estate industry. The findings were as follows: (Q1): The benefits of using big data analytics are to help clients to make the right decisions and advice, have higher efficiency for appraisals, better risk evaluation of risk in the real estate industry simplifying applications in valuations and pricing. (Q2): there is anecdotal evidence that real estate has already started adopting big data. Adoption is most likely to be beneficial for first mover industry players at the top of the industry pyramid including investment banks, commercial banks, and mortgage banks that hold the highest interest in the real estate industry. (Q3): complexity of big data solutions and the costs of implementation are a major challenge while smaller players such as real estate agents and brokers do not

find utility or justification for the huge investment in big data. (Q4): the development of algorithms remains as the main process of applying big data solutions as there are no off-the-shelf big data solutions for the real estate industry. Adoption of Machine learning (ML) and Artificial Intelligence (AI) in real estate would help buyers and sellers to learn from data and make informed decisions. The conclusions of the culminating experience project are Real Estate Industry has a low adoption of big data solutions because many of the players in the industry have not yet learned how to translate big data to business objectives. Areas of further studies include the development of models and algorithms for use by the real estate industry.

TABLE OF CONTENTS

CHAPTER FOUR: DATA ANALYSIS AND FINDING

CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Background Information

Real estate remains to be the biggest asset class around the world.

During the covid-19 pandemic, housing prices increased by 17.5 percent in

2021(Federal Housing Finance Agency, 2022). The pricing in this industry is

mostly qualitative as the realtors consider various real estate-oriented feature,

including location, number of rooms, square footage, and number of baths, which

affect the way a certain property is priced (Coleman et al., 2022). Such factors

render predicting real estate prices extremely complicated. People place

divergent values on different property features, thus making it challenging to

establish a model that can precisely reflect the human decision-making

processes. Data generates more than 2.5 quintillion bytes every day. Big data is

used in every possible field and walk of life, such as in management, healthcare,

business, and other ventures. (Munawar et al., 2020). Most evidently, tools and

systems have been created in the past years in a more efficient, collective, and

manageable way. Big data helps to define the huge and interlinked databases

that have been established and the many processes that focus on distilling

creative knowledge from the digital data that is in existence (Coleman et al.,

2022). However, the potential of big data in real estate is in gradual progress. Scholars are still finding the best way to produce good results in the real estate big data industry.

Although big data is new and exciting, it characteristically remains a complex subject. The concept refers to large and hard to manage data that is collected when business conducts its day-to-day operations. The data is collected in high volumes and at significant velocity. The massive data from the real estate industry can effectively produce the prediction of growth of consumer demands and productivity, and data can provide the basis and sources for business decision-making and strategy formulation (Li, 2021). For instance, if a customer wants to purchase a property, he can use the area sold data record to see if the listing price is reasonable or not and help to make the decisions. However, Big data in the real estate industry is still new and needs to be discovered.

Great focus has been on the expected advantages of mining data that will create novel understanding of human behaviors, habits, and lifestyles. The details acquired from big data can alter the way that individuals think about different segments of research. For most of the individual customers, real estate holds emotional connection or response (Meier et al., 2019). Emotions play a significant role in determining real estate transactions. Buyers with an emotional attachment to a particular property will place a higher bid despite its market value. Furthermore, the 2008 housing crisis revealed how herd behavior can

2

create housing bubble as many investors believed that housing prices would always go up. In the end, the market crashed as emotions made investors panic selling their housing units. However, real estate as a sector has always been data driven. The question would be how big data can provide anything novel to the sector to be steeped in data collection and proper management (Winson-Geideman & Krause., 2016). Therefore, there is the need to investigate the conventional forms of real estate data, their connections, and the expected drawbacks.

Real estate data can be categorized as financial, transactional, and physical. The financial data includes information details related to financing, including bond issuance and stock. According to Winson-Geideman and Krause (2016), transactional data is related to information on real estate purchases, mortgages, leases, and the general returns to the developments. The physical data incorporates information such as the actual real estate land or the property itself. The other important fact to note is that with the launch of the geographic information systems (GIS), novel kinds of data could be appropriately used in the real estate sector (Winson-Geideman & Krause., 2016). GIS provides room for extra-locational spatial details to be garnered, quantified, and analyzed. With extra-locational details, spatial analysis along with business geography emerged as viable real estate processes and have eventually been seen as the norm across real estate analysis. As GIS widens the capability to measure and scrutinize data, big data brings in the human and psychological aspects (Winson-

Geideman & Krause., 2016). GIS can enable prediction of demand of houses in a particular area based on past trends. Hence, investors can build sufficient inventory of houses to match the anticipated demand.

One issue that remains clear is that the new data collection models, data management systems, along with the data analysis organizations represent a broader set of phenomena and firms. In the real estate sector, a huge segment of resources is used for the purposes of prediction. Realization of what is likely to occur in the coming days is powerful information for the decision-makers of the real estate market (Winson-Geideman & Krause., 2016). With the presence of core data, prediction methods are mostly statistical as they utilize past trends to forecast what is likely to happen in the future. The integration of peripheral data introduces the capacity to improve prediction through observation and analysis of causal relationships. Such factors can be seen as structural prediction and provide the capacity for more precise and timelier details on which to base real estate choices (Kok et al., 2017). The utility of big data is not just on the volumes of collected information but on how connections can be generated from the different data sets.

In the past, real estate companies only focused on their privately held data to formulate crucial insights that they could use to compete in the market. However, many companies have seen the advantages that arise from linking data from different companies and regions. Another thing to note is that there is presently lack of research on the role of big data in improving clients'

comprehension of commercial and industrial markets, including real estate developments (Winson-Geideman & Krause., 2016). While the residential markets are mostly the simplest to collect huge datasets, other notions across the real estate study can benefit from improved research into the potentials of big data. The application of big data in the real estate sector seems to be fledgling (Kok et al., 2017). However, it can be challenging to identify the organizations that are utilizing big data and how tactics are being approached because of the concerns about sharing proprietary business information (Winson-Geideman & Krause., 2016). The few illustrations that have gone public do indicate the detailed use of big data in improving efficiency along with profitability of the real estate organizations and ventures.

In the real estate industry, valuation and appraisal are major processes that utilize big data. Property development depends on forecasting of anticipated costs and returns. The managers in the industry use the forecasts of supply along with demand to estimate the costs to be incurred and returns that are expected (Wei et al., 2022). With forecasting playing a crucial role in directing investing decisions, artificial intelligence tactics for residential valuation have been recommended in research for a while now (Wei et al., 2022; Kang et al., 2021). The emphasis on data mining from huge property transactions has become very popular. A wide range of methods can be applied for data mining. Nural networks have been recommended as appropriate for the estimation of hedonic price index based on cross-sectional transactional data (Grybauskas et

al., 2021). The fact remains that there is a need to explore and identify the trends and challenges facing the real estate market. The focus should be on the application of big data to obtain the most accurate property value in real estate.

Problem Statement

Appraisals in the real estate market are typically founded on capitalization of the net income of a property using cap rate that is acquired from neighboring transactions. Transactions of comparable assets are never similar in either time or building features (Kang et al., 2021). The appraisers are focused on past valuations or transactional prices of a property. Such imprecise measures of property values tend to be problematic in the current hyper-efficient capital markets. Moreover, appraisal-oriented price indexes that are still utilized by most investors as the decision-making benchmarks are imprecise as well (Kang et al., 2021). The fact is that investors lack the precise pricing data that is required to underwrite the properties, especially in evaluating the risks of real estate assets. Furthermore, one thing is that the real estate industry experiences noticeable costs from the appraisal bureaucracy. To address the error of appraisals, the real estate sector has established an optional property index series, including transaction-oriented indexes (Kok et al., 2017). However, such measures fail to address the primary issue of imprecise measurement of value of properties.

While the inefficacies of property appraisals exist, the real estate sector has experienced a notable rise in data collected along with the machine learning capabilities. Timely estimations of property values are crucial for investors in the real estate industry since it helps them make informed underwriting choices (Oluwunmi et al., 2019). The changes emanating in real data have pushed the whole sector to generate ways of establishing higher efficacy for appraisals along with dynamic monitoring. Moreover, external threats and prospects have pushed entities to change to big data. While the firms are making new efforts towards progress, the academic circles are also considering many attainments across big data (Oluwunmi et al., 2019). The tendency has invigorated scholars to learn by correspondence from other segments and integrate novel technologies with conventional industries to guide in building information dominance. Furthermore, just as asset transactions along with mortgage incidence rise, the costs linked to the appraisals still increase. The conventional single valuation approach is not only inappropriate but also idiosyncratic (Pérez-Rave et al., 2019). Particularly, it relies on the expertise capacity and experience that the evaluator has. With developments in computer technology and the rising application of such advancements across the real estate sector, scholars are not fulfilled with the linear hedonic pricing systems that have shown to be of low steadiness and accuracy as identified by Wei et al. (2022). Hence, an accurate system needs to be formulated.

Significance of the Study

Studying is very important for several reasons. The first one is that the precise and timely estimations of property values are important for real estate investors and lenders since it helps them to make informed decisions. One fact is that systematic errors or biases during estimations and assessments can have negative impacts on the provision of equity or debt (Wei et al., 2022). Moreover, investors, regulators, and other parties in the public domain depend on appraised values to evaluate returns in the real estate market. Automated property valuations offer more accurate reflections of the flow and real estate stock. The other fact is that the presence of instant and accurate property value may expand the opportunities for financial innovation in the real estate industry (Kok et al., 2017). The finding will also show that big data is having a heightening effect on the assessment and comprehension of real estate markets. The rationale is that big data incorporates vast amounts of semi-structured and unstructured critical information that can be mined. The collection and mining of this big data plays a significant role in helping draw inferences that are more scientific and useful compared to the small data analysis.

The research is also important to policymakers in the real estate industry as they will understand the proper application of big data, especially in appraisals and making valuations of properties. Big data offers big opportunities and potentials for sectors that require data analysis (Pérez-Rave et al., 2019). The results will show that there is a rising need for analytic instruments that are

applied in conducting sophisticated predictive big data analysis. At the same time, the variability of standard remedies in big data scrutiny is rising. Many companies have believed that big data business intelligence would introduce profit and other benefits in the coming years. Many large firms already utilize big data for business intelligence and obtaining external data (Pérez-Rave et al., 2019). Another anticipation is that many other entities will begin to benefit from their data by selling it or offering the content based on big data. Stakeholders in the real estate market will understand that big data technology can be used at divergent phases of project planning, implementation, and analysis. For entities that cannot or are not able to introduce their automated evaluation system, they can also realize that they could collaborate with real estate technology organizations. The entities apply big data together with predictive analytics to offer accurate property estimates to clients.

Research Questions

To address the issue raised in the problem statement, this culminating experience project will answer the following questions:

1.   What are the benefits and advantages of utilizing big data in the real estate market, include how the industry applies ML and AI?

2.   What are the trends in the application of big data in the real estate market?

3.   What are the challenges in applying big data in the real estate market?

4.   What are the methods and processes of applying big data in appraisal of assets in the real estate market?

Organization of the Study

Chapter One introduces the reader to the study. It captures the background information, the problem statement, research questions, and significance of the study. Chapter Two encompasses the literature review, it captures information on adoption trends, benefits, challenges, and processes in big data. Chapter Three presents the methodology. It presents information about all the processes and procedures used in acquiring, processing, presenting, and interpreting data for this experience project. Chapter Four will include data analysis and finding. Chapter Five will a discussion of the findings, conclusion, and the areas for further study.

CHAPTER TWO

LITERATURE REVIEW

Overview

The following chapter provides a systematic review of the literature. The section offers a broad view of the subject of use of big data while doing comprehensive analysis of the real estate market. One Search found 1,651 results between 2018 to 2022 after filtering. When sorted by relevance and open access, the result decreased to 463. This project selected the first 50 results and found 26 relevant articles after reading the abstracts. On the other hand, Google scholar also searches the same terms. The initial results found 18,700. Sorted by relevance and date, it produced 2,220 results. This project selected the first 50 results and found 40 relevant articles. There were two articles found to be duplicates of One Search results, so it was removed. Therefore, this project will be based on roughly 68 articles. This study will focus on the scholar article found in Google Scholar and One Search.

Table 1. Summary of Articles Related to Research Questions

Database	Search Words	# Of relevant publications	# Selected	Authors
ScienceDirect; Google Scholar;	Content Analysis	30	4	Assarroudi et al., 2018;

Pfau Library's OneSearch				Kleinheksel et al., 2020; Kuckartz, 2019; Selvi, 2019;
ScienceDirect; Google Scholar; Pfau Library's OneSearch	Big data Real Estate	43	26	Ali, B., & Siniak, N., 2020; Al-Qawasmi, J, 2022; Barkham et al., 2022; Battisti et al., 2019; Chen, X., & Lu, 2018; Cheryshenko& Pomernyuk, 2021; Coleman, W et al., 2022; Conway, J, 2018; Du & Zhang., 2014 Fields & Rogers., 2021; Grybauskas., 2021; Kok et al., 2017; Locurcio & Morano,. 2020;

				Meier et al., 2019;
				Obinna & Udo., 2022;
				Oluwunmi et al., 2019;
				Pai & Wang., 2020;
				Pérez & González., 2019
				Rabby & Hassan.,2022;
				Rossini., 2000;
				Treleaven., 2021;
				Ullah & Al-Turjman., 2021
				Wei et al.,2022
				Winson & Krause., 2016;
				Munawar & Sepasgozar., 2020;
				Li., 2021;
				Bazeley., 2006;
				Yousuf & Zainal., 2020

The recent developments in technology along with the improved data collection tactics have introduced several transformations toward the evaluation of real estate from applied and theoretical facets. Pérez-Rave et al. (2019) conducted a study on machine learning as an approach to big data regression analysis across real estate prices for predictive purposes. According to Pérez-Rave et al. (2019), large volumes of data are being gathered, transitioned, and investigated to predictively evaluate market trends and assess customer sentiments by entities (Pérez-Rave et al., 2019). The application of big data is at the cutting edge and has the capability of changing decision-making processes in the real estate market. Nevertheless, big data tends to vary hugely in both scales and category by sector and much wider discussion neglects matters crucial to the analysis of real estate property. Therefore, there is a need to understand what big data is, why and how it is used in a property setting, and how it is transforming the real estate market. Wei et al. (2022) focused on the price model-based model for real estate appraisal in the period of big data. Wei et al. (2022) state that a crucial point to be noted is that the veracity of big data is with its uncertainty, including the level of its accuracy and the dependability. Organizations that operate in the highly competitive segment ought to understand that the accumulation of big data does not provide anticipated positive business results. The implication is that a contemporary institution

should not set the action of processing present data to improve the quality of its financial and economic analysis of results (Wei et al., 2022). In this regard, it is crucial to utilize the whole available data for segmentation, construct a huge number of models for small entities, and address specific issues utilizing information technology.

Grybauskas et al. (2021) also focused on big data for the real estate market through predictive analytics. According to Grybauskas et al. (2021), big data introduces many benefits in the real estate industry, including more accuracy of property values, improved transaction success rates, along with proper decision-making. In the present day, websites have emerged where estimates and predictions can be made for properties (Grybauskas et al., 2021). Such values are founded on transaction prices in the past periods to provide an estimated price. However, they at times overestimate the property prices and provide incorrect details. Using big data, many unstructured data can be simple to get online. The issue is how to process the data sets to formulate valuable insights. The fact is that the uneven data quality would lower the correctness of the appraisal outcomes (Grybauskas et al., 2021). In this regard, the combination of different methods and divergent data is very important. Manual appraisals have also become outdated with the emergence of big data. Kok et al. (2017) studied big data in real estate through the transition from manual appraisal to automatic valuation. They acknowledged that utilizing the technology in an attempt to process appraisals is more accurate. Some researchers have also

recommended the use of improved and better systems instead of forecasting tactics (Kok et al., 2017). However, the issue is that several uncertainties impact such datasets, including the way to utilize an effective manner to process the data and acquire more accurate value for every property.

Meir et al. (2019) embarked on the application of data mining development in assessing real estate investment plan. Meir et al. (2019) acknowledge that big data is consistently being established through streaming data or complex event processing. Meier et al. (2019) state that there exist two fundamental challenges linked with big data streams, including the analysis of data in real-time for immediate application along with batching as well as storage for future analysis. The variety of big data is the different kinds of data that is garnered, including the structured and unstructured ones. The structured data conforms to the set regulations. The unstructured data does not in any way conform to the regulations (Meier et al., 2019). There is also the veracity of big data that has much to do with the uncertainty it holds. Data cleaning ought to be reliable and researchers have to make sure that it is relevant to the purpose being sought. One thing that comes out clear about these dynamics is that big data and associated techniques have continued to evolve and the awareness about it has expanded. Tabesh et al. (2019) focused on the implementation of big data strategies across the managerial facet indicating that the process of analyzing this data is known as analytics and calls for divergent skills and instruments compared to the conventional data analysis. Moreover, the analysis of big data

streams can offer the additional edge required to compete in case the approach is well-structured. Using the opportunities across big data analytics can enable the managers to become more agile.

Oluwunmi et al. (2019) provided a review depicting the effect of big data to encompass digitization of records, information on user tastes, sensor details on urban development, along with sensor information on movement. Oluwunmi et al. (2019) acknowledge in the review that the rapid expansion of big data and its promising utilizations have resulted in passionate pursuit of big data technology across different segments. In effect, a series of standout achievements have also emerged. As a typical information-intensive sector, big data is playing a notable role in the real estate appraisal segment (Oluwunmi et al., 2019). The introduction of big data to study is not in any way a paradigm shifts in knowledge acquisition since it is a computational change in thought as well as research. Regardless of size, researchers still believe that data and research constraints introduce more questions. However, Oluwunmi et al. (2019) indicate that it should be noted that big data does not always provide more insightful results just due to its size. The problem is usually prone to different methodological issues, including input errors. Furthermore, it is subject to problems not generally linked to conventional data. Another thing is that big data is combined with other data sets that can compound the issues (Winson-Geideman & Krause., 2016). Data curation is crucial, and most important. Furthermore, big data cannot be seen as

the answer to every aspect as it should be reflective of the purpose being sought.

<p style="text-align:center">Benefits of Big Data in Real Estate</p>

The drastic transformations arising from data analytics have pushed the whole real estate sector to find ways of enhancing more efficiency for appraisals and dynamic monitoring. Kang et al. (2021) embarked on understanding house price expansion using machine learning and big data. They indicate that adoption of data-oriented strategies plays a significant role in helping real estate broaden its market. Valuations can be done more accurately based on the data that has been acquired or shared by different players. Property managers and investors would have improved control of their portfolios and associations with clients. Being a very huge industry, the real estate players depend on data to make crucial decisions (Kang et al., 2021). The data can be property valuations in a particular area, the number of homes, the percentage of owners, and other related aspects. Wei et al. (2022) adds that having access to this kind of data supports business decision-making processes. However, the major issue experienced is that accurate and up to date data is challenging to obtain. Real estate investors are knowledgeable about how big data analytics can help in the improvement of operational performance of their portfolios.

Cheryshenko and Pomernyuk (2021) targeted the incorporation of big data in the decision-making procedures within the real estate sector. According to

Cheryshenko and Pomernyuk (2021), the significant expansion in data access and the probability of collecting each kind of information measurable gives room for development and novel revenue models in the real estate market. According to Cheryshenko and Pomernyuk (2021), the present applications of big data across the industry already entails developers, agencies, and property managers that can grow their domains with the application of technology for business model building among others. Big data makes it possible to manage assets in a rational and individualized manner, thus improving investors and clients' fulfilment. Firms in the real estate industry are also able to diversify investments and have novel access to individual information. The tools also alter the way of thinking and open new revenue channels for realty entities (Cheryshenko & Pomernyuk, 2021). However, it should be noted that the use of big data and analytics on an entity should be scrutinized first, especially regarding the type of available data and its relevance in a wide array of possibilities. In other words, the kind of data to be used should be one that creates value for divergent players in the real estate market.

Big data offers strong support for rational way of development in the real estate market. The approach has great benefits to the realty entities since it helps them implement and enhance diversified investment through data mining for potential value. According to Pai and Wang (2020), the digital personal details and the revolutionary changes in the manner of thinking render innovative investments the novel revenue growth chances for realty entities in the era of big

data. One thing that should be made clear is that harnessing of digital data following revolution has emerged as more complex than before. Winson-Geideman and Krause (2016) state that the conventional statistical tools in real estate had constraints in addressing the data sets. Through leveraging big data tools combined with other technologies, the entities can gather, summarize, aggregate, and analyze the datasets in an attempt to conduct different functions. In the real estate industry, data analysis has a significant role in picking up the correct deal at a minimum risk (Wei et al., 2022). The huge volumes of historical and digital data gathered by real estate players and users can be utilized by big data tools, incorporated with advanced artificial intelligence technologies to generate benefits.

Big data assists the real estate market to enhance quicker workflow by advising effective and profitable projects to clients. Grybauskas et al. (2021) indicate that the risk of selecting the wrong property is also minimized by reliance on big data. By using predictive as well as sentiment analysis of big data, the customized as well as filtered assets for customers offers a competitive edge. Price declines are more common across the industry because of the high risk involved in the market. In this regard, big data can help mitigate such risk by being used to predict future prices of properties. Marketing strategies can also be improved for heightened customer reach and engagement when big data is used (Grybauskas et al., 2021). By analyzing the data in real-time, conducting regional price scrutiny, and predicting future price fluctuations, the real estate firms can

collaborate and provide individualized insurance coverage for each client in real-time by quoting the future value. The real-time details ought to be represented in an authentic and profound way for better understanding and engagement of clients (Wei et al., 2022). The capabilities explain the reason several companies, including, Zillow are using artificial intelligence-enabled big data analytical instruments for improved enhancement of customer engagements to generate profits.

Big data in the real estate market also has the potential toward revenue optimization. The fundamental goal of any sector is to have a return on investments to optimize revenue in the long run. Barkham et al. (2022) focused on the application of urban big data across the city management and real estate markets arguing that with big data-enabled analytical instruments, incorporated with more advanced technologies, there is the chance of enhancing overall growth across different functionalities. Moreover, by being able to minimize the risk in investments through the application of big data capabilities, the future of an entity is heightened (Barkham et al., 2022). The thing that is clear here is that future digital intelligence would change the business strategies and practices associated with real estate, which would happen through a shift in responsibility, risk, and sufficiently addressing customers' needs (Pai & Wang, 2020). The applications of big data across innovative investment have become common at foreign real estate firms. The functionalities have enabled the innovative entities not only to fulfil the demands of their clients by improving quality of service, but

also enhance the realty seals. Such companies can decide to focus their attention on the satisfaction of their clients.

Challenges in Using Big Data in Real Estate

When data is gathered, organizing, and analyzing it is a critical phase in unlocking its knowledge. Software and cloud-based platforms are now being utilized to facilitate the role (Kok et al., 2017). Through visualization, filtration, analysis, or simulation of upcoming events, the sector can evaluate the market trends, financial properties, and design decisions. Their capabilities make it possible to forecast the potential future results. However, Kok et al. (2017) indicates that there is huge and fierce competition among the technologies to handle the market. Conventional software is challenged by new cloud-based platforms that enhance automated data aggregation into huge valuable databases. Although forecasting using big data is soon to become a game changer in the real estate market, each entity in the industry struggles with the problem of reputation. The conventional model of risk analysis is still identified as standard. There is also the issue of integrity associated with the datasets. Rossini (2000) expounded on the utilization of expert systems and artificial intelligence for real estate predictions. If huge datasets are aggregated over time, it is mostly challenging to assess the quality of this data (Rossini, 2000). The filtering of data can be useful but problematic to some extent. The other issue is that the precision of the forecast relies heavily on the amount of processed data.

22

Developing the channel to catch up with the ever-increasing size of data sets is even more difficult.

   The primary challenge in the application of big data in the real estate market is about how it is interpreted. According to Ali and Siniak (2020), there are recent trends in the use of big urban data and their effect on real estate markets. The authors acknowledge that the interpretation should be done intelligently, safely, and from an aggregation point. The other challenge that emerges with this issue is the understanding of exactly what to do with the datasets, the way to best analyze it, and the insights to be extrapolated. Some companies do not utilize big data analytics due to the fact that it requires a costly and cumbersome infrastructure. Moreover, the contradictions between privacy protections and big data remain irreconcilable (Ali & Siniak, 2020). The fact is that big data processing technology is not easily available for many companies in the market. Furthermore, the characteristics of real estate entities brings more difficulties into the use of big data. Big data entails much more than just many numbers, thus making it more complex and disordered. Oluwunmi et al. (2019) add that the gathering, storage, and processing of this huge unstructured data requires unusually advanced technology. It is a challenge for any real estate corporation to capture critical information from the big numbers. In fact, it would be a wise decision for real estate firms to have big data handled by professional third parties. Divergent kinds of entities play different roles in the era of big data (Wei et al., 2022). The estate entities can be able to center on the applications of

information in an attempt to deepen development, leaving the big data

processing to vendors that are deemed excellent with technologies.

Coleman et al. (2022) studied the use of machine learning to assess real

estate prices through the use of location big data. Coleman et al. (2022) identify

that the real estate segment remains to be the mainstay of real economy with

some features of fictitious economy, including complexity, greater risk, and

metastability. Such features raise the uncertainties in the use of big data.

Considering that a fictitious economy system is extremely sensitive to

anticipations, the publicity as well as sharing of big data can influence investment

demand due to the alterations on individuals' expectations of realty (Coleman et

al., 2022). Owing to the metastability, this impact would damage the stability of

real estate market, thus influencing the economic development of a nation as

well. Coleman et al. (2022) add that there is also the aspect of structural

imbalances and information asymmetry when handling big data. The platform

ought to be formulated for realty data storage and assist in building a big real

estate database (Du et al., 2014). One thing that becomes clear is that more

detailed business and collective management are the growing trends in the real

estate market. The way to function effectively remains to be a big data problem.

The real estate firms are required to construct a big data warehouse and conduct

integration management through proper data mining and analysis for forecasts.

Big Data Models Used to Conduct Appraisals and Valuations

Various developed models are used for conducting valuations and appraisals in the real estate market. Most of these are the automated valuation models (AVMs). According to Kok et al. (2017), an AVM refers to a service that makes use of mathematical models to offer the estimated value of an asset or property at a particular point in time. The AVMs can assume different approaches to establish a property's value. Most AVMs in the present day tend to utilize regression or hedonic models. Some more contemporary algorithms utilize adaptive estimation models or neural, self-learning network systems (Kok et al., 2017). Regarding application of big data, Hedonic AVMs are well acknowledged within the real estate market. Hedonic pricing systems demand information on basic property characteristics including the building year, size, as well as the level of capital improvements made (Kok et al., 2017). The models typically encompass a search engine that can compare the attributes of the asset with comparable characteristics using a radius search pattern. Most importantly, Hedonic AVMs are deemed common since they are founded on easy regression models that are simple to implement and comprehend (Pérez-Rave et al., 2019). However, there is a drawback in that regression models are universal ones meaning that they create a single predictive formula that is consistent across the whole range of variables. Additionally, a single model might not be appropriate in making predictions of property value in an accurate manner.

A machine learning approach has also been widely used in the real estate market in analyzing big data as it has a great potential for prediction. Pérez-Rave

et al. (2019) indicate that machine learning utilizes at least two subsamples with one for training and another for the purpose of validation. The integration of hedonic regression and machine learning can give valuable chances for analysis of real estate prices while still enhancing the joint execution of inferential as well as predictive analysis, on the basis of the obtained models. However, it should be noted that this integration is not in any way elementary (Pérez-Rave et al., 2019). The rationale for this is that when utilizing regression from a machine learning point to analysis of big data, the inferential handling is still constrained. Some of the reasons for this is that the estimates of coefficients are not consistent. They do not accomplish the critical prerequisite of estimation issues. The other reason is that regularization processes are conducted in an attempt to alleviate overfitting in non-training samples (Pérez-Rave et al., 2019). The practice can result in bias due to the omission of variables. Furthermore, considering the high correlations among variables, the models can be unstable. The acquired predictors vary based on the sample, which would mean that the emerging models are inconsistent. Kok et al. (2017) identify that it is evident that due to such shortcomings in the models, it is important to integrate other strategies to help in making decisions, while fulfilling the criteria of importance and improvement of knowledge. In looking for hedonic approach remedies based on machine learning for big data that enhances meeting the needs of inference and forecast, a crucial tool is the selection of variables. This entails selecting a

subset of variables through certain criteria to get the variables that are most crucial for comprehending a phenomenon.

Summary

The chapter has provided a systematic review of literature materials. This literature review helped me to reach research questions 1 to 3. From analysis, it has become clear that big data has a crucial role in the real estate market. The application of big data has been making appraisals and predictions of property values. Big data has emerged as a critical segment in helping players in the market realize the value of investments. However, the utility in the real estate market has not come without challenges, especially the high cost of processing big data. Chapter 3 below will provide the methodology used to conduct this study.

CHAPTER THREE

METHODOLOGY

Introduction

The focus of this culminating experience project will be to develop a

methodology that will answer the research questions related to the problem

statement posed. In line with that goal, this section of the study presents the

methodology. Among the factors considered as the design of the study, data

collection, data analysis, and the reliability and validity of the research methods.

Research Design

The research adopts a qualitative and quantitative design. The

methodology involves the collection and analysis of data covering opinions,

views, observations, text, and other non-numeric information (Kleinheksel,

Rockich-Winston, Tawfik & Wyatt, 2020). In considering the qualitative design,

the research captured that the real estate industry involves a lot of qualitative

information in informing decisions. Such information includes location, historical

information about the property, materials, weather information, design, and other

information that directly affects the value of the property. Other factors include

property size, the market value of properties in the region, the number of rooms,

baths, garages, and parking spaces, the number of years the property has been

in the market, and other such information. Both qualitative and quantitative

research helps in capturing all the information that affects decision-making in the real estate industry.

On the other hand, this study will also use the quantitative approach to collect numerical data and implement it in RStudio. Quantitative techniques allow people to analyze data deeply and see the relationship between variables and factors (Bazeley, 2016). In this project, I will manually collect the data from Redfin, and the data will insert into Excel (CVS) and import to R studio to analyze the price in the city of Brea and the city of Walnut. The purpose of this analysis is to compare the differences between listing price and sold price and what is the price difference between these two cities. However, this study will be more focused on a qualitative study. The quantitative approach will only be used to collect the numerical sets and produce fewer graphs and tables by using RStudio.

The industry has been considering the potential presented by big data for the sector (Al-Qawasmi, 2022). The ideas of the practitioners are captured in various textual sources, including conference papers and research reports. However, the information is barely harmonized. Therefore, the industry continues to experience a limited understanding of how it could adopt and apply big data, machine learning, and artificial intelligence to improve price prediction. An exploratory research design helps the study to reach far and wide to acquire information and help the sector understand the applications of big data to the

industry, the application barriers, and what the industry can do to implement the use of big data.

Qualitative and quantitative research designs are useful for the type of study attempting to conduct for the real estate market. With the focus on benefits, challenges, processes, and trends in application of big data in the real estate industry, the use of qualitative and exploratory designs helps gather descriptive information that will help ensure that the research questions are fully answered and that the aims of the study are met. Most importantly, the approach ensures that all the information pertaining to the subject is presented comprehensively in an organized manner that enhances its utility to industry practitioners.

Data and Data Collection

The study aims to understand the benefits of using big data in the real estate industry. The second goal is to understand the challenges to the application of big data, machine learning, and artificial intelligence in the real estate industry. Additionally, the trends and processes of applying big data in real estate will be explored, also will provide a manual dataset to analyze by RStudio. All the aims of the research are captured as requiring qualitative information, besides manually collecting datasets using a quantitative research method, more precisely, the data required is textual information from records of how both the industry and researchers have explored the applications of big data in the real estate industry.

The data required for this study is secondary from publications. The data shall be gathered by conducting both global searches on the worldwide web and systematic searches of information from Google Scholar. The goal is to acquire both grey literature and white literature about the study. Grey literature mainly includes industry reports and commentaries about the uses of big data, machine learning, and artificial intelligence in the real estate industry. The source also includes working papers and research reports that may not be peer reviewed but which remain highly relevant to this emerging area of big data applications. On the part of white literature, the focus will be on acquiring research reports published in peer reviewed journal articles in the last five years and which focus specifically on the applications of big data in real estate. The data will be acquired by searching various key terms and phrases, including big data, artificial intelligence, machine learning, and real estate. Other related terms include artificial intelligence in real estate, real estate decision trees, data and property indices, and challenges in big data applications for real estate (Yang, Yin & Li, 2021). The conception of terms and phrases considered herein captures the key aspects that the study seeks to cover.

As mentioned earlier, this research will also include a quantitative component which is essential especially in the valuation and pricing of real estate. The experience project will manually collect the data from Redfin, and the data will be inserted into Excel and imported to R studio to analyze the price in the city of Brea and the city of Walnut. The purpose of this analysis is to compare

the differences between listing price and sold price and why house prices are different between these two cities. Table 1.1 attached below will introduce an overview of these two cities.

Table 2. Structure of Dataset in RStudio

```
str(Re_walnut)
...

'data.frame':    20 obs. of  9 variables:
 $ Month     : chr  "2-Jul" "1-Jul" "29-Jun" "30-Jun" ...
 $ List.Price: int  1120000 800000 1239900 895000 1109000 998000 958888 898000 1100000 849900 ...
 $ Sold.price: int  1180000 820000 1071001 975000 1135000 985000 1020000 955000 1180000 880000 ...
 $ Per.Sqf   : int  539 490 306 531 461 484 376 455 598 450 ...
 $ Sqft      : int  2188 1679 3500 1836 2460 2060 2712 2100 1974 1954 ...
 $ Sqft.lot  : chr  "8857" "5002" "1.87acres" "9401" ...
 $ Bedroom   : int  4 3 4 4 4 3 5 4 4 4 ...
 $ Bath      : num  3 2 5 2 3 3 3 3 3 3.5 ...
 $ Address   : chr  "20512 Julliard Dr, Walnut, CA 91789" "21533 Magnolia St, Walnut, CA 91789" "19830 E
Saddle Ridge Ln, Walnut, CA 91789" "20309 Shadow Mountain Rd, Walnut, CA 91789" ...
```

```
{r}
str(RE_brea)
...

'data.frame':    20 obs. of  9 variables:
 $ Month     : chr  "9-Jul" "8-Jul" "7-Jul" "7-Jul" ...
 $ List.Price: int  949900 975000 900000 999999 920000 1080000 1200000 875000 910000 850000 ...
 $ Sold.price: int  960000 1055000 885000 1130000 920000 1080000 1200000 930000 899900 890000 ...
 $ Per.Sqf   : int  359 418 407 523 368 491 650 421 498 491 ...
 $ Sqft      : int  2676 2526 2176 2160 2500 2198 1723 2207 1827 1814 ...
 $ Sqft.lot  : chr  "0.32acres" "5115" "5984" "5700" ...
 $ Bedroom   : int  5 4 3 3 5 4 2 4 3 3 ...
 $ Bath      : num  4 2.5 2.5 2.5 4 2.5 2.5 3 2.5 2.5 ...
 $ Address   : chr  "1440 Bexley Ln, Brea, CA 92821" "2101 Windbreak Cir, Brea, CA 92821" "1116 Niguel Canyon
Way, Brea, CA 92821" "859 Zion Canyon Way, Brea, CA 92821" ...
```

There are 9 variables in this Table 1.1, which will include Listing price, price per square foot, number of bedrooms, address, and so on. This dataset collects 20 houses in each city; The property size is between 1429 square feet to 3500 square feet with 2 to 5 bedrooms.

Data Analysis

The study will apply content analysis methods in data analysis and using RStudio to analyze sold price from redfin. The concept refers to any technique for making inferences by systematically and objectively identifying special

32

characteristics of messages (Kleinheksel, Rockich-Winston, Tawfik & Wyatt, 2020). The second definition describes content analysis as an interpretive and naturalistic approach. The approach is also a research technique for objective systematics and descriptive description of the content of communication. Content analysis finds use cases in all instances where the goal is to differentiate communication content because it helps reveal patterns in the communication content. A key characteristic is that it helps identify the dominant views about a particular subject of interest by providing a tool for the comparison of different competing views present in data (Selvi, 2019). Using content analysis in exploring big data applications in real estate will help identify the most important views about the benefits, challenges, trends, and processes of its use in real estate.

Method and Process

Two major levels of content analysis are dependent on the specificity of information and the depth of analysis that is required. The first is thematic analysis while the second level is key word analysis (Kuckartz, 2019). Thematic analysis broadly identifies the key themes from content. Key word analysis isolates the key information under each theme and may be used to determine subtle differences in information presented under each theme. For this research, the focus will be on thematic analysis (Kuckartz, 2019). The reason is that the

research will be focusing on identifying the key thematic areas under each the research question.

Content analysis is also described to include two types. The first is conceptual analysis while the second approach is relational analysis. Conceptual content analysis focuses on identifying the occurrence of the main concept under study (Kuckartz, 2019). In this case, the concept is big data applications in real estate. The concept of big data includes both machine learning and artificial intelligence. Since the goal is to improve decision-making in real estate valuation and transactions, the concept is presented in the form of a decision model or decision tree that tracks the information of consideration from the point where a customer gets the intention to purchase property to the point whether they decide to purchase or not purchase the property. Through the path is information and data that the customer considers. The information is acquired through a customer purchase journey that involves looking for details from various sources. Often, the journey begins online before the customer can proceed to acquire information from real estate agents and then from authorities responsible for the custody of real estate information. Hence, by exploring the concept, the analysis will be looking at all the information that applies in real estate decision-making and how that information can be modeled.

The other type of content analysis is relational analysis. The approach considers the relationships between concepts. For instance, it is not possible to assess big data and fail to look at decision modeling, machine learning, and

artificial intelligence. In real estate decision-making, it is not possible to discuss real estate decision-making without looking at information from the financial sector, employment, and incomes of persons because all the information plays a critical role in informing the decisions of the consumers and also how the real estate agents promote the products to consumers (Kuckartz, 2019). In this research, relational analysis will include cognitive mapping and proximity analysis.

The implementation of the content analysis methods outlined above is designed in several steps. The first step is deciding how many concepts to code for, which will be guided by the research questions. The second step is to determine the themes and reduce the content to the themes of analysis (Assarroudi, Heshmati Nabavi, Armat, Ebadi & Vaismoradi, 2018). The third step was to determine the coding rules by focusing on whether the data is coded for occurrence of a theme or for the frequency of a theme. The fourth one is to determine the patterns in the themes and the last is to explore the relationships between themes. Based on this level of analysis, it becomes possible to make inferences about the subject of research in a repeatable and confirmable manner.

Machine Learning (ML) is the application of Artificial Intelligence to learn and improve themselves without programming (Selig, 2022). This application is a great way to process the amount of big data in the real estate industry. ML is a powerful tool to reduce the difficulty of modeling and can handle uneven-quality

data without giving a specific mathematical model (Wei et al., 2022). The most modern methods of machine learning algorithms are Neural Network (ANN), Support Vector Machines (SVM), Deep Learning (DL), and Ensemble Learning (EL) (Wei et al., 2022). However, each method of machine learning has their intrinsic advantages. And these four methods will be discussed next.
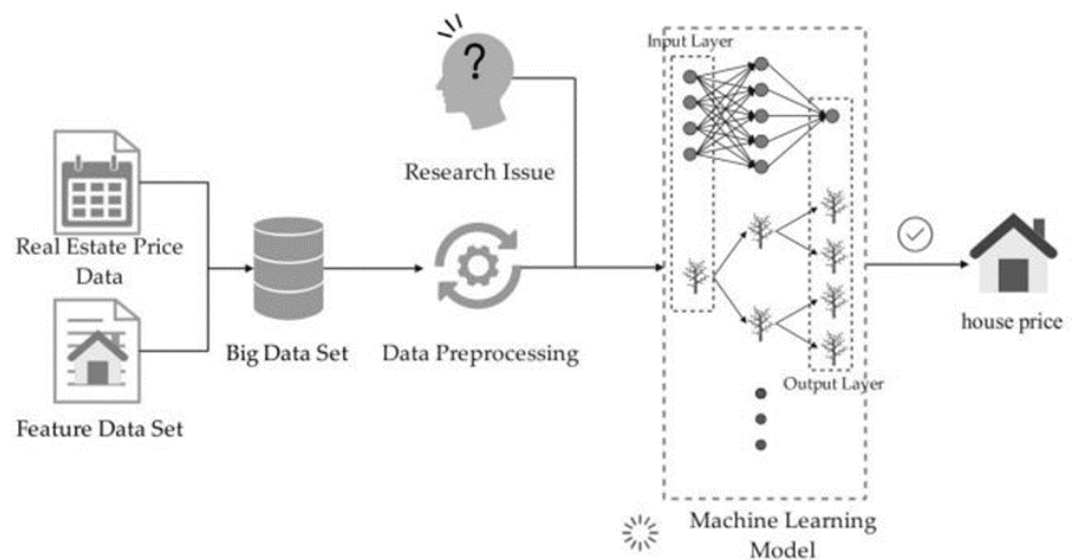
Figure 1. Flow Chart of Applying Machine Learning (Wei et al, 2022)

The artificial neural network (ANN): Neural network is a multivariate analytical tool that has become a major tool for the real estate industry, and the concept of the ANN is a learning algorithm that has a similar function to the human brain (Rossini,2000). The ANN algorithm is the longest-serving, and it is the most frequent method in the real estate industry (Wei et al., 2022). However, a Neural network is a modern algorithm in the industry.

- Support vector machine (SVM): This method is focused on small-sample learning. It is based on fewer support vectors instead of using

the dimension of the sample, and SVM is a great method to avoid

dimension disaster (Wei et al, 2022). This method is also very widely

used in forecasting housing prices. For instance, SVM was a great

method to predict changes in Shanghai real estate market prices, and

they achieved some good results (Shen et al,2013).

- Deep Learning (DL): DL deems an extension of the traditional

neural network model. The advantage of DL is that DL avoids the

complicated manual engineering process (Wei et al., 2022). The weak

point is that this method needs excessive data demand and

computation; the DL prefers to use recommendation algorithms,

medicine, and image recognition, and so on, which is not

recommended use by the real estate industry (Wei et al., 2022).

Therefore, this complex algorithm is not achieved well in the real estate

industry.

- Ensemble Learning (EL): EL cannot be referred to as single

machine learning algorithm. This method uses multiple weak classifiers

to gain a strong classifier and improve the accuracy of the results (Wei

et al, 2022). Overall DL strategies can be divided in three types, which

is boosting, bagging and stacking. However, EL is not very popular in

the real estate industry.

☐ Overall, Machine Learning (ML) is a complex algorithm, and different

methods support different data types and aims. ML improves valuation accuracy,

no matter what method types are used (Wei et al., 2022). However, ML is an effective way to get the accuracy of house appraisals.

Reliability and Validity

Qualitative study designs often are affected by the challenges of validity and reliability. The systematic approach to content analysis helps enhance the stability, reproducibility, and accuracy of study results (Selvi, 2019). The reason is that the approach to the collection of data is methodically outlined, making it possible for other researchers to replicate the study. Further, the coding criteria based on concepts and relationships between concepts apply theory to the research approach. The approach makes it possible to understand the methodology and results and at the same time for other readers of the research report to critique the findings (Kleinheksel, Rockich-Winston, Tawfik & Wyatt, 2020). Such facts about the content analysis approach ensure high reliability and validity of the study findings.

Quantitative methods play a big role in the statistics field. This method helps to gather data and conduct the necessary analysis (Yousuf, 2020). In this research, the quantitative research method applies to collect numerical data, such as prices, size, number of rooms, and so on. However, the quantitative method can apply in many fields, such as the financial industry, healthcare, and more.

The study takes further steps to enhance reliability and validity. Particularly, the paper enhances reliability by outlining the methodical approach to the collection of data and setting out the plan for collecting white literature data through a systematic review of only the peer-reviewed literature for the period of the last five years. Drawing information from the systematic review ensures that the study gathers data that is of high quality by focusing on the highest standard of evidence-based research. On the other hand, using quantitative research methods to collect the data ensures the reality of the study.

Summary

In summary, the focus of the study is to explore the applications of big data in the real estate industry. Also, using quantitative methods collect data and go thought RStudio to extract valuable information. In another way, the section presented the methodology. The overall view is that the study is designed to be an exploratory qualitative study relying on primary textual data gathered from a global search of keywords and from a systematic review of peer-reviewed literature acquired from Google Scholar and One Search. The data is then analyzed using content analysis following key themes and concepts under big data, processing the data machine learning, and artificial intelligence applications in real estate.

CHAPTER FOUR

DATA ANALYSIS AND FINDING

Introduction

The aim of the study is to explore the applications of big data in real estate. The study was guided by four research questions focusing on the benefits, challenges, trends, and processes of applying big data in real estate. The qualitative study collected textual data that was then analyzed using textual analysis. The following g section of the report presents the results and findings from the study.

Data Analysis

The study gathered data from 15 reports, inclusive of industry reports and empirical research data from peer reviewed resources. The reports have all been published in the last five years. Table 2 below summarizes the number of research reports that were explored in the completion of this study.

Table 3. Summarizes the Number of Research Reports

Source	Date	Where it Comes from
Big data and real estate: A review of literature	2019	Journal of Physics: Conference Series
Digital Transformation in Real Estate Marketing: A Review	2022	Big Data: A Road Map for Successful Digital Marketing

Big data visualization and analytics: Future research challenges and emerging applications	2020	BigVis 2020-3rd International Workshop on Big Data Visual Exploration and Analytics.
Scenarios for applying Big Data in boosting construction: A review	2018	Proceedings of the 21st International Symposium on Advancement of Construction Management and Real Estate
Big Data in Construction: Current Applications and Future Opportunities	2022	Big Data and Cognitive Computing
Machine Learning Applications in Real Estate: Critical Review of Recent Development	2022	IFIP International Conference on Artificial Intelligence Applications and Innovations
Real estate data marketplace	2021	AI and Ethics
Barriers to the digitalisation and innovation of Australian Smart Real Estate: A	2021	Environmental Technology & Innovation

managerial perspective on the technology non-adoption		
How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries–A review and research agenda	2022	International Journal of Information Management Data Insights
A Study On Big Data–Its Opportunities, Challenges and Risks	2020	International Journal of Engineering, Science
Data intelligence: trends and challenges	2020	Systems Engineering Theory & Practice
The research development of hedonic price model-based real estate appraisal in the era of big data	2022	Land
Towards a critical housing studies research agenda on platform real estate	2021	Housing, theory and society

On big data, artificial intelligence and smart cities	2019	Cities
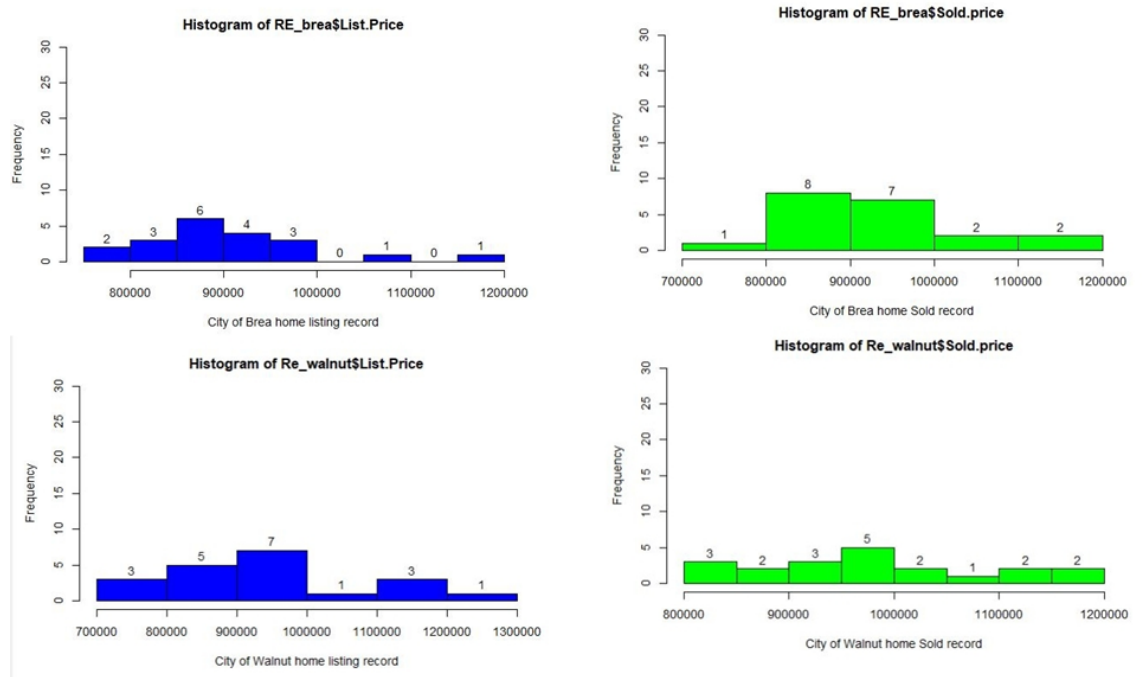
Figure 2. The Histogram Listing Price and Sold Price in Both Cities

Using a small size of big data will be easier to extract the values from the

data. Figure 4.1 is the Histogram of Brea and Walnut. This histogram shows the

list price and sold price. As we can see, different cities have different price

trends. In the city of Brea, there were only two houses sold at prices over 1

million. In Walnut, there were four houses above 1 million, and there were more

aggressive buyers in Walnut. Wei et al. (2022) identified that school districts and

business districts are a great way to improve real estate appraisal. We should understand the location before doing any data exploration.

Some of the useful variables in the data set will be easy to explore, such as sold prices and listing prices. This information is very useful either in the financial industry, agents, or bankers. Table 4.2 shows the percentage of each house transaction between the listing price and the selling price. People are willing to pay more to make buyers' offers more competitive.

Table 4. The Table Contains the Results of Sold Price

```
[1]   0.010632698  0.082051282 -0.016666667  0.130001130  0.000000000  0.000000000  0.000000000
[8]   0.062857143 -0.011098901  0.047058824 -0.062500000 -0.004651163  0.067853170  0.147058824
[15]  0.073414905  0.069782330 -0.098818475 -0.124999125  0.029423876  0.141714286


[1]   0.053571429  0.025000000 -0.136219856  0.089385475  0.023444545 -0.013026052  0.063732156
[8]   0.063474388  0.072727273  0.035415931  0.025882353  0.058201058  0.100131752  0.025862069
[15] -0.018789144  0.084656085  0.064102564  0.064705882  0.000000000 -0.009483667
```

The top side is the percentage in Brea, one buyer willing to pay more than 14% of the listing price to own the house. In Walnut, the highest rate of buyers paid 10% above to own the property.

Benefits of Big Data in Real Estate

At conception, this study hypothesized that big data applications would benefit the valuation and pricing of real estate property and improve the decision-making process in the buying and selling process. Analysis of literature reveals that the valuation, selling, and buying decisions present simplistic decision-making scenarios in the real estate industry (Guo, 2020). As such, the scenarios do not require the application of big data. This is the reason why buyers, sellers,

valuation experts, real estate agents, and even mortgage companies may not find much value in the use of big data in the industry (Kumar, 2020). Big data finds a lot of utility in evaluation of risk in the real estate industry.

Risk management in the real estate industry plays a crucial role in the sustainability of the industry and the stability of the financial system. At the level of the industry, risk management is important in informing the interest rates charged on borrowers. Some of the challenges that have been recorded in the industry include lending to subprime borrowers whose inability to pay back their mortgages resulted in the global financial crisis of 2008/09 (Hall, Yoder & Karandikar, 2021). The most crucial level of risk management is in the development of derivatives in the industry. Derivatives are conceptually high-risk financial products whose risk profile is influenced by the risk of the underlying asset, but the risk profile may not be correlated. The product involves bundling together multiple mortgages of different risk profiles. Sometimes, there are also derivatives of derivatives, meaning that the products can be so complicated as to reveal the true level of the risk underlying the assets.

The investment industry has been using computation methods in assessing the risk of investments for years. In the past, the computations were based on structured data from the markets and from the commercial banks (Oluwunmi, Role, Akinwale, Oladayo & Afolabi, 2019). While such analyses may include fairly large datasets, they were not applying the many forms of unstructured data that is now made available through the Internet. The

implication is that the industry could still benefit further by using the large

datasets that are created using technologies by consumers of financial products,

specifically mortgages, that are the basic structure in the building of derivatives

and compound derivatives.

Risk in the real estate industry is influenced by multiple factors. Economic

factors such as the rate of inflation, rate of interest, rate of unemployment, and

foreign exchange rates. However, such data is only macro-level data for which

the industry is aware. The more critical factors affect the way the loans are

offered to the borrowers, credit worthiness of the borrowers, and default rates.

Multiple factors affect the ability of the mortgage consumers to pay back the

loans. Similarly, these issues influence the stability of the industry and with-it

sustainability of the investment products created at the national levels. It is such

risks that create the need for the industry to consider big data solutions.

Data Applications in Real Estate

Current practice indicates that industry relies on housing price indices.

The indices track the prices of land and buildings in different locations (Yang, Yin

& Li, 2021). The indices are updated regularly monthly, quarterly, semiannually,

and annually. Primarily, the reports are provided by the real estate industry

based on data of recent sales of land, housing, and commercial properties. Some

of the indices track rental and leasing rates of properties. For the most part, the

indices are only used in indicating prices of property, thereby influencing the

prices (Yang, Yin & Li, 2021). However, there are weaknesses in the use of price indices as the primary or only source of real estate data.

The housing price indices are not primarily designed to inform decision-making in the industry. Rather, the housing price indices are designed to be indicators of health and stability in the real estate sector. The implication is that some of the players in the industry lack the necessary tools to inform decision-making (Singh, Chen, Singhania, Nanavati & Gupta, 2022). It can be considered that the structure of available data in the market favors the real estate agents and not the owners, sellers, and buyers of the real estate property. It is such challenges that create a need develop more comprehensive data for decision-making in the industry.

The alternative is the development of a big data solution for decision-making in the real estate industry. The concept of big data seeks to include more information in the computation of property values. The analysis does to by focusing on the information that users would consider in making the decision to sell or buy real estate. The concept of big data considers the real estate buying process as a journey that leaves and utilizes a lot of data. The application of big data seeks to incorporate such data in the decisions, not just by focusing on the information that users are consciously aware of but also the data that the users may not have interacted with. The concept achieves that by applying supervised and unsupervised machine learning models in the applications of big data in the

47

real estate industry (Singh, Chen, Singhania, Nanavati & Gupta, 2022). The concept is as presented in the decision model shown below.

Trends in Adoption of Big Data in Real Estate

A search of information on the application of big data in the real estate industry indicates that multiple suppositions on how big data can be employed in the industry. However, there is no evidence that the industry has already adopted big data solutions (Obinna & Udo, 2022). If there are real estate entities that have already started using big data, then such information may not be publicly available. The greatest likelihood is that the real estate is yet to adopt big data in their operations.

While it is not clear how the real estate industry is applying big data, there are various commentaries and analyses of potential that the industry can unlock through big data. The industry could unlock improved property evaluations, better property marketing, and improved pricing of property insurance (Cheryshenko & Pomernyuk, 2021). Information also indicates that big data could lead to better real estate products, product offering with personalized solutions, and quality management of real estate property. Additionally, analysis has indicated that the industry could unlock improvements in occupancy rates and rents collectible from the property. However, such evidence is only presented in suppositions of what could be if the industry employed big data to inform decisions.

The complexity of implementing and using big data solutions is captured accurately. Building algorithms takes time and cost (Obinna & Udo, 2022). Data mining is costly. Businesses in the real estate industry already use analytics to the extent of projecting demand and prices for real estate property. Therefore, for many of the businesses, the current data already serves them well and may not see the need for further investments in big data solutions. The implication of such findings is that the big data discussion in the real estate industry remains in infancy. Thus, analysis of data indicates there is still an opportunity for the industry to consider big data.

Challenges in the Application of Big Data

Complexity in the implementation of big data solutions in the real estate sector was considered the major problem or challenge. Big data applications require advanced technological skills to develop the models (Chauhan, 2020). Whilst complexity is still a challenge, evidence indicates that the utility of the big data models could be the key factor influencing the adoption and non-adoption of big data in the industry. The real estate industry involves many players. The players include the customers or buyers at the bottom of the pyramid, the real estate agents and owners or properties, the financiers or mortgage lending companies, and special investment vehicles that include investment banks (Guo, 2020). All the players require large volumes of real estate data to inform their decisions.

49
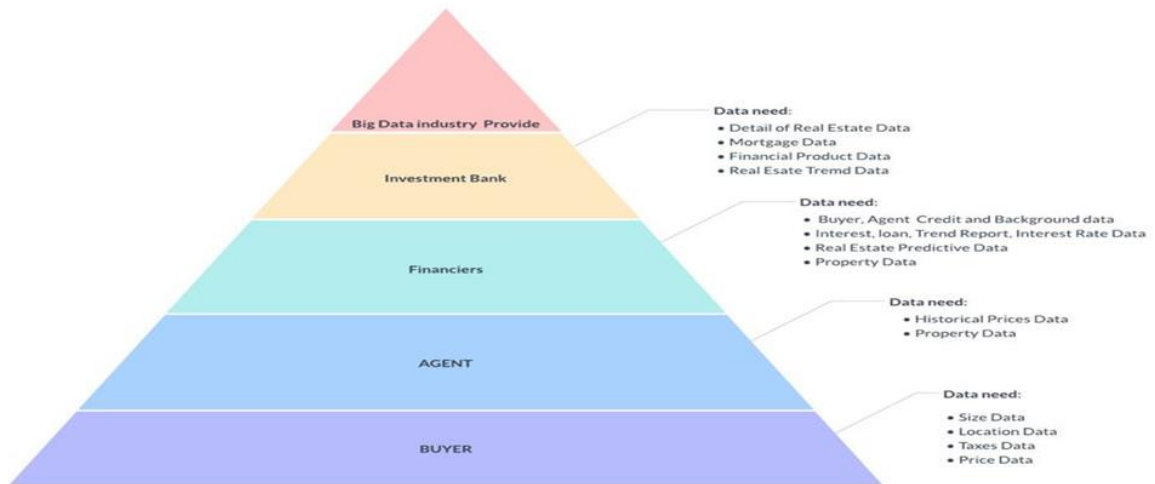
Figure 3. The Data Speared by Level

The majority of stakeholders in the real estate industry do not consider big data useful in the real estate industry. A buyer who is interested in one or two homes is simply interested in the prices of houses in the specific region of interest. For most of the stakeholders at the bottom of the pyramid, the key decision variable is information about the particular type of property that they require (Chauhan, 2020). The size of property, location, price, taxes, and amenities close to the property are the key factors. For these stakeholders, there is limited to no utility for big data. A property index combined with information about property taxes and renovations is adequate to inform the purchase decisions.

The next level is that of real estate agents. The agents could be viewed as the market makers in the real estate industry. Their role is mainly linking buyers

with sellers and advising the interested parties on the prices of real estate. The services of the agents may be limited to localities so that a particular real estate agent specializes in buying and selling property in a particular area (Babii, Chen & Ghysels, 2019). The information that that agents require include historical prices of property in the particular location, the size of property, the age of the property, real estate laws and local laws affecting zoning of properties in the region, taxes, and other such information. The real estate agents use this information mainly to inform the price and valuation advisories. Notably, the real estate agents are also the major players in the valuation of the properties.

The amount of data that the real estate agents require is more than that required by the buyers and sellers. These agents also have access to more information. However, that information does not qualify as big data, which is especially true because real estate agents specialize in particular locations and because the real estate industry is also highly segmented. The utility of big data to the real estate agents is of limited nature. As a result, the agents may also not be interested in expending huge resources in development of big data solutions considering that such developments require huge capital outlays. For man agents, such spending may be unjustified.

The financiers are the third level of players in the real estate industry. They require more information than the buyers or sellers and the real estate agents. Their services extend beyond a single or particular location and are not just interested in the prices of the properties but also the health of the real estate

sector. They take interest in the repayments of mortgage loans, the uptake of mortgage loans, trends in interest rates, credit worthiness of the borrowers, and other such information that affects the macro-level decision-making in the real estate industry (Babii, Chen & Ghysels, 2019). Their use of data in the real estate industry creates a case for big data applications. More importantly, the industry needs to apply predictive analytics in determining future trends in the performance of the real estate industry. The main reason why they require that much information is that they are exposed more to the performance of the real estate industry. Precisely, many of the real estate properties are purchased on mortgage loans and remain on the balance sheets of financial institutions for years at a time (Babii, Chen & Ghysels, 2019). Consequently, financial institutions do not only track the data to determine how the industry would perform at a single point in time but through years, sometimes through decades (Breeden & Leonova, 2019). For financial institutions, the utility of big data creates incentive for the development of big data solutions.

The last major level is the investment banking level at which the focus of the company is in managing risk (. Investment banks play the role of structuring deals in the real estate industry by the sale of bonds and other large scale financial products for the purposes of sale to the market. For the investment banks, the role of bundling together mortgages from banks and selling them as derivatives means that they require much more data and information to price the derivatives (Cao, Chen, Hull & Poulos, 2021). Such derivatives include

mortgage-backed securities, collateralized debt obligations, credit default swaps, and other similar products (Babii, Chen & Ghysels, 2019). It is such products that were involved in the global financial crisis of 2008-09 when default rates on subprime mortgages resulted in the collapse of the financial markets (Hall, Yoder & Karandikar, 2021). The cases and the scale of operations of investment banks means that they require big data. Besides, the investment banks lack the same grassroots networks that would allow them access and apply the data. The only solution that can provide lasting solutions to the institutions is the application of big data.

The overall view from this analysis is that the major concern with the application of big data is with the ability of the industry to find utility in big data. Apart from the investment banks and mortgage institutions that would consider potential applications of big data in real estate industry, real estate agents may not find much utility in the solutions because the property indices serve them well (Cao, Chen, Hull & Poulos, 2021). Hence, the analysis concludes that big data may not be considered for application by entities that only operate in a small scale within the industry.

Methods and Processes of Applying Big Data

The application of big data in real estate valuations and appraisals requires the development of algorithms. A simple way to understand algorithms is comparing them to the valuation models that have been in use in the real

estate industry (Pérez-Rave, Correa-Morales & González-Echavarría, 2019).

However, instead of having a model that the business has to develop on its own,

maintain, and update with each new use case, the application of big data

requires the use of machine learning and artificial intelligence so that the

algorithms go through a process of learning, whether supervised or unsupervised

(Singh, Chen, Singhania, Nanavati & Gupta, 2022). Once the model is fully

developed, it can be used in predicting the performance of the real estate

industry and also automate the valuations of properties and projects.

The main differentiator of algorithms in big data from the models currently

in use by the real estate industry is the amount of data and information that he

algorithms can process. While the regular models are primed to apply

quantitative and mainly financial data with a limited number of variables, there is

no limit as to the amount of information that the algorithm in big data can process

(Singh, Chen, Singhania, Nanavati & Gupta, 2022). Further, algorithms in big

data use large data samples to make predictions. The implication is that the

process results in highly accurate predictions. Another possibility is for the users

to see trends in data that would not be visible when the business is only using

the limited quantitative data that is commonly used in the valuations.

The key to unlocking the potential of big data in real estate is in being able

translate the data into business decision for the real estate industry. One of the

key challenges that continues to face the adoption of big data in many industries

is the shortage of candidates who can translate the data into business goals so

that the models or algorithms that are developed communicate to strategic

business objective (Tabesh, Mousavidin & Hasani, 2019. A focus on how big

data translates to business development for the real estate industry is therefore a

major consideration on how the industry can be developed to adopt big data.

Summary

In summary, analysis of information from both grey and white literature

shows a low adoption of big data in the real estate industry. Though complexity

of applying big data is recorded as a problem faced by the industry, the major

issue is the inability of the industry to translate big data into business goals. Big

data finds applications more readily in mortgage companies and in investment

banks as key players in the real estate industry (Cao, Chen, Hull & Poulos,

2021). While real estate valuation and appraisals was considered a major area of

application of big data, the risk management application would benefit the most

from big data. The next chapter presents the conclusion and recommendations of

the study.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

This is the last chapter in the study. It builds on Chapter Four findings by presenting summary of those findings, conclusions on the findings, and recommendations that include direction of future research.

Summary

The study focused on the application of big data for comprehensive analysis of the real estate industry. It was guided by four research objectives and questions encompassing the determination of benefits and advantages of utilizing big data in the real estate market. It also studied the trends in the application of big data in the real estate market. Thirdly, the research explored the challenges in applying big data in the real estate market. Lastly, the study looked into the methods and processes of applying big data in appraisal of assets in the real estate market.

The design of the study was qualitative and exploratory. Qualitative textual data was collected through systematic search of white and grey literature both on the worldwide web and on Google Scholar. Keen interest was directed towards working papers, conference papers, industry reports, and peer reviewed research reports published by scholarly journals. In gathering the information, attention was directed to research work published within the last five years, focusing directly on big data applications in real estate, and authored in the

English language. Upon acquisition of data, thematic content analysis was conducted, guided by the research objectives outlined above.

The study summative finds that a lot of the published literature only indicates how the industry could benefit from the adoption of big data and not on experiences of the industry in applying big data. In other words, the real estate industry is yet to adopt big data solutions in their business models. The main challenge is that the industry struggles to find real use cases of big data. The current model of using housing price indices to evaluate real estate property appears to serve the industry just fine (Yang, Yin & Li, 2021). However, real use cases might emerge among mortgage and commercial banking institutions as well as investment banking institutions which experience significant exposure to the real estate industry and therefore require a better understanding of the risk and financial health of the industry. The large financial and investment institutions also have the incentive to invest the huge capital outlays that are required to implement big data solutions in the real estate industry.

Conclusions

The research concludes that the real estate industry has a low adoption of big data solutions because many of the players in the industry have not yet learned how to translate the big data to business objectives. The low adoption of big data is further complicated by the complexity of the algorithms required to make good use of them. The study has demonstrated that adoption of big data

requires the businesses to invest in skilled manpower for the development and training of the models for use in big data. The industry also requires capital investment in big data solutions, and it also requires investment in time for data scrubbing or mining. The analysis finds that such investment is unjustified if the business does not find real links to business outcomes.

While there are many potential benefits in applying big data in the real estate industry, the main benefits would be in risk management in the industry. Therefore, the industry players who would benefit the most from the adoption of big data are the players with the greatest risk exposure. They include the mortgage companies and commercial banks which carry the loans in their books for extended periods of time. They also include the investment banks and other financial industry players that bundle the real estate products into derivatives. The development of mortgage-backed securities and collateralized debt obligations results in an amplification of risk. Big data applications would be useful in better understanding of the risk exposure for such institutions.

<center>Recommendations</center>

This research recommends further studies on the development of models and algorithms for use by the real estate industry. At the time of this research, the industry lacked evidence of practical applications of big data. The study considered that if the industry was provided with fully developed big data solutions, then adoption of the solutions would be much easier.

<center>58</center>

Industry-level collaboration in the development of big data solutions would help speed up the adoption of the solutions. While larger players such as investment banks, commercial banks, and mortgage companies have the capacity to develop the big data solutions, they may need to rely heavily on data from real estate agent who interact directly with buyers and sellers of real estate property. Collaboration would result in the fast unlocking of the opportunities that the industry provides.

Limitations of the Study

The study was not able to identify businesses in the real estate industry that may have tried to adopt and implement big data solutions. Consequently, the study misses out on the voice of industry players. The limitation could be overcome by conducting surveys with the companies in the real estate industry to determine their experiences with big data solutions in the industry.

APPENDIX A

ORIGINAL DATA COLLECTED FROM RSTUDIO

A TABLE CONTAINING NINE VARIABLES WITH FORTY ENTRIES

	Month	List.Price	Sold.price	Per.Sqf	Sqft	Sqft.lot	Bedroom	Bath	Address
1	9-Jul	949900	960000	359	2676	0.32acres	5	4.0	1440 Bexley Ln, Brea, CA 92821
2	8-Jul	975000	1055000	418	2526	5115	4	2.5	2101 Windbreak Cir, Brea, CA 92821
3	7-Jul	900000	885000	407	2176	5984	3	2.5	1116 Niguel Canyon Way, Brea, CA 92821
4	7-Jul	999999	1130000	523	2160	5700	3	2.5	859 Zion Canyon Way, Brea, CA 92821
5	6-Jul	920000	920000	368	2500	5936	5	4.0	300 Saint Crispen Ave, Brea, CA 92821
6	2-Jul	1080000	1080000	491	2198	5034	4	2.5	3061 E Coalinga Dr, Brea, CA 92821
7	1-Jul	1200000	1200000	650	1723	6970	2	2.5	448 S Madrona Ave, Brea, CA 92821
8	1-Jul	875000	930000	421	2207	9594	4	3.0	1327 LA Canada Dr, Brea, CA 92821
9	30-Jun	910000	899900	498	1827	2769	3	2.5	3016 E Walking Beam Pl, Brea, CA 92821
10	30-Jun	850000	890000	491	1814	2336	3	2.5	882 Skyler Way, Brea, CA 92821
11	30-Jun	800000	750000	499	1603	7015	3	2.0	905 McCart Cir, Brea, CA 92821
12	29-Jun	860000	856000	412	2077	5000	4	2.5	1713 Harvest Ln, Brea, CA 92821
13	29-Jun	899000	960000	507	1893	5474	4	3.0	2228 Arts Ave, Brea, CA 92821
14	25-Jun	850000	975000	458	2130	5000	4	2.5	2265 Timbercreek Cir, Brea, CA 92821
15	25-Jun	899000	965000	653	1478	8960	3	2.0	453 Castlegate Ln, Brea, CA 92821
16	24-Jun	781000	835500	543	1540	2227	3	2.5	357 Aqua Way, Brea, CA 92821
17	21-Jun	931000	839000	613	1519	6710	3	2.0	408 Juniper St, Brea, CA 92821
18	21-Jun	999999	875000	504	1983	7597	4	3.0	1791 Ravencrest Dr, Brea, CA 92821
19	18-Jun	849990	875000	556	1575	7150	4	2.0	1406 Baywood Dr, Brea, CA 92821
20	18-Jun	875000	999000	469	2130	5100	4	2.5	2297 Shadetree Cir, Brea, CA 92821

Showing 1 to 20 of 20 entries, 9 total columns

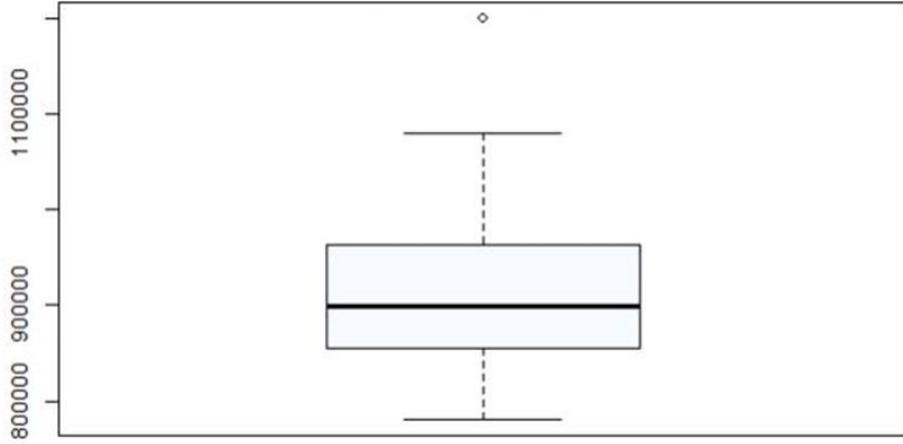|	Month	List.Price	Sold.price	Per.Sqf	Sqft	Sqft.lot	Bedroom	Bath	Address
1	2-Jul	1120000	1180000	539	2188	8857	4	3.0	20512 Julliard Dr, Walnut, CA 91789
2	1-Jul	800000	820000	490	1679	5002	3	2.0	21533 Magnolia St, Walnut, CA 91789
3	29-Jun	1239900	1071001	306	3500	1.87acres	4	5.0	19830 E Saddle Ridge Ln, Walnut, CA 91789
4	30-Jun	895000	975000	531	1836	9401	4	2.0	20309 Shadow Mountain Rd, Walnut, CA 91789
5	28-Jun	1109000	1135000	461	2460	0.25acres	4	3.0	20534 Camelback Dr, Walnut, CA 91789
6	25-Jun	998000	985000	484	2060	0.23acres	3	3.0	625 Whistling Wind Cir, Walnut, CA 91789
7	25-Jun	958888	1020000	376	2712	8846	5	3.0	20545 Missionary Ridge St, Walnut, CA 91789
8	24-Jun	898000	955000	455	2100	7465	4	3.0	316 LA Tortola Dr, Walnut, CA 91789
9	23-Jun	1100000	1180000	598	1974	0.26acres	4	3.0	632 Bronco Way, Walnut, CA 91789
10	22-Jun	849900	880000	450	1954	6463	4	3.5	19702 Heathridge Cir, Walnut, CA 91789
11	17-Jun	850000	872000	388	2250	7293	5	3.0	1524 Leanne Ter, Walnut, CA 91789
12	16-Jun	945000	1000000	403	2480	7667	4	4.0	315 Eola Dr, Walnut, CA 91789
13	16-Jun	759000	835000	584	1429	7774	3	2.0	248 Centinary Dr, Walnut, CA 91789
14	14-Jun	928000	952000	413	2303	0.3acres	4	3.0	20148 Evening Breeze Dr, Walnut, CA 91789
15	11-Jun	958000	940000	382	2459	8462	4	3.0	1141 Hare Ave, Diamond Bar, CA 91789
16	10-Jun	945000	1025000	525	1951	8354	4	3.0	20405 Shadow Mountain Rd, Walnut, CA 91789
17	9-Jun	780000	830000	546	1520	7826	3	2.0	422 Radcliff Dr, Walnut, CA 91789
18	9-Jun	850000	905000	571	1584	6620	4	2.0	1526 Hallgreen Dr, Walnut, CA 91789
19	8-Jun	1150000	1150000	375	3063	0.28acres	5	5.0	20454 Lake Canyon Dr, Walnut, CA 91789
20	4-Jun	949000	940000	387	2432	7992	4	3.0	1336 Garzon Pl, Walnut, CA 91789

Showing 1 to 20 of 20 entries, 9 total columns
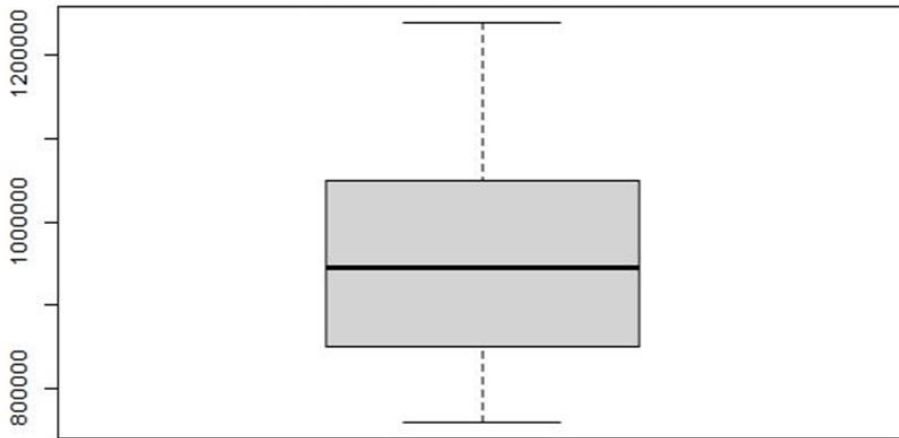
61

APPENDIX B

BOXPLOT TO CHECK OUTLINER

FIGURE CONTAINING PRICE RANGE IN TWO CITIES

Brea House Listing record

Walnut House Listing record

APPENDIX C

STRUCTURE TABLE IN RSTUDIO

STRUCTURE TABLE SUMMARISE NINE VARIABLES

```
      Month            List.Price        Sold.price         Per.Sqf           Sqft         Sqft.lot
 Length:20          Min.   : 759000   Min.   : 820000   Min.   :306.0   Min.   :1429   Length:20
 Class :character   1st Qu.: 850000   1st Qu.: 898750   1st Qu.:387.8   1st Qu.:1922   Class :character
 Mode  :character   Median : 945000   Median : 965000   Median :458.0   Median :2144   Mode  :character
                    Mean   : 954134   Mean   : 982500   Mean   :463.2   Mean   :2197
                    3rd Qu.:1023500   3rd Qu.:1036500   3rd Qu.:533.0   3rd Qu.:2459
                    Max.   :1239900   Max.   :1180000   Max.   :598.0   Max.   :3500
    Bedroom            Bath           Address
 Min.   :3.00   Min.   :2.000   Length:20
 1st Qu.:4.00   1st Qu.:2.750   Class :character
 Median :4.00   Median :3.000   Mode  :character
 Mean   :3.95   Mean   :3.025
 3rd Qu.:4.00   3rd Qu.:3.000
 Max.   :5.00   Max.   :5.000
```

```
      Month            List.Price        Sold.price         Per.Sqf           Sqft         Sqft.lot
 Length:20          Min.   : 781000   Min.   : 750000   Min.   :359.0   Min.   :1478   Length:20
 Class :character   1st Qu.: 857500   1st Qu.: 875000   1st Qu.:420.2   1st Qu.:1693   Class :character
 Mode  :character   Median : 899500   Median : 925000   Median :494.5   Median :2030   Mode  :character
                    Mean   : 920244   Mean   : 943970   Mean   :492.0   Mean   :1987
                    3rd Qu.: 956175   3rd Qu.: 981000   3rd Qu.:528.0   3rd Qu.:2182
                    Max.   :1200000   Max.   :1200000   Max.   :653.0   Max.   :2676
    Bedroom            Bath           Address
 Min.   :2.0    Min.   :2.000   Length:20
 1st Qu.:3.0    1st Qu.:2.500   Class :character
 Median :4.0    Median :2.500   Mode  :character
 Mean   :3.6    Mean   :2.625
 3rd Qu.:4.0    3rd Qu.:2.625
 Max.   :5.0    Max.   :4.000
```

REFERENCES

Abraham, C. M., Elayidom, M. S., & Santhanakrishnan, T. (2020). Big Data
Analysis for Trend Recognition Using Machine Learning Techniques.
International Journal of Sensors Wireless Communications and Control,
10(4), 540-550.

Ali, B., & Siniak, N. (2020). Integrating big data into decision-making in real
estate industry. *Management (16487974)*, *36*(2).

Allam, Z., & Dhunny, Z. A. (2019). On big data, artificial intelligence and smart
cities. *Cities*, *89*, 80-91.

Al-Qawasmi, J. (2022). Machine Learning Applications in Real Estate: Critical
review of Recent Development. In *IFIP International Conference on
Artificial Intelligence Applications and Innovations* (pp. 231-249). Springer,
Cham.

Andrienko, G., Andrienko, N., Drucker, S., Fekete, J. D., Fisher, D., Idreos, S., ...
& Sharaf, M. (2020, March). Big data visualization and analytics: Future
research challenges and emerging applications. In *BigVis 2020-3rd
International Workshop on Big Data Visual Exploration and Analytics*.

Assarroudi, A., Heshmati Nabavi, F., Armat, M. R., Ebadi, A., & Vaismoradi, M.
(2018). Directed qualitative content analysis: The description and
elaboration of its underpinning methods and data analysis
process. *Journal of Research in Nursing*, *23*(1), 42-55.

Babii, A., Chen, X., & Ghysels, E. (2019). Commercial and residential mortgage

defaults: Spatial dependence with frailty. *Journal of econometrics*, *212*(1),

47-77.

Barkham, R., Bokhari, S., & Saiz, A. (2022). Urban big data: City management

and real estate markets. In *Artificial Intelligence, Machine Learning, and

Optimization Tools for Smart Cities* (pp. 177-209). Springer, Cham.

Battisti, E., Shams, S. R., Sakka, G., & Miglietta, N. (2019). Big data and risk

management in business processes: Implications for corporate real estate.

Business Process Management Journal.

Bazeley, P. (2006). The contribution of computer software to integrating

qualitative and quantitative data and analyses. *Research in the

Schools*, *13*(1), 64-74.

Breeden, J. L., & Leonova, E. (2019, April). When Big Data Isn't Enough: Solving

the long-range forecasting problem in supervised learning. In 2019

*International Conference on Modeling, Simulation, Optimization and

Numerical Techniques* (SMONT 2019) (pp. 229-232). Atlantis Press.

Cao, J., Chen, J., Hull, J., & Poulos, Z. (2021). Deep hedging of derivatives using

reinforcement learning. *The Journal of Financial Data Science*, *3*(1), 10-

27.

Chauhan, V. (2020). A Study On Big Data–Its Opportunities, Challenges and

Risks. *International Journal of Engineering, Science*, *1*(1).

Chen, X., & Lu, W. S. (2018). Scenarios for applying Big Data in boosting construction: A review. In *Proceedings of the 21st International Symposium on Advancement of Construction Management and Real Estate* (pp. 1299-1306). Springer, Singapore.

Cheryshenko, M. S., & Pomernyuk, Y. Y. (2021, April). Integration of big data in the decision-making process in the real estate sector. In *IOP Conference Series: Earth and Environmental Science* (Vol. 751, No. 1, p. 012096). IOP Publishing.

Coleman, W., Johann, B., Pasternak, N., Vellayan, J., Foutz, N., & Shakeri, H. (2022). Using machine learning to evaluate real estate prices using location big data. *arXiv preprint arXiv:2205.01180*.

Conway, J. J. E. (2018). Artificial intelligence and machine learning: Current applications in real estate.

Du, D., Li, A., & Zhang, L. (2014). Survey on the applications of big data in Chinese real estate enterprise. *Procedia Computer Science*, *30*, 24-33.

Fields, D., & Rogers, D. (2021). Towards a critical housing studies research agenda on platform real estate. *Housing, theory and society*, *38*(1), 72-94.

Grybauskas, A., Pilinkienė, V., & Stundžienė, A. (2021). Predictive analytics using Big Data for the real estate market during the COVID-19 pandemic. *Journal of Big Data*, *8*(1), 1-20.

Guo, Y. (2020). Credit risk assessment of P2P lending platform towards big data

    based on BP neural network. *Journal of Visual Communication and Image*

    Representation, 71, 102730.

Hall, A. B., Yoder, J., & Karandikar, N. (2021). Economic distress and voting:

    Evidence from the subprime mortgage crisis. *Political Science Research*

    and Methods, 9(2), 327-344.

Kang, Y., Zhang, F., Peng, W., Gao, S., Rao, J., Duarte, F., & Ratti, C. (2021).

    Understanding house price appreciation using multi-source big geo-data

    and machine learning. *Land Use Policy*, *111*, 104919.

Kleinheksel, A. J., Rockich-Winston, N., Tawfik, H., & Wyatt, T. R. (2020).

    Demystifying content analysis. *American journal of pharmaceutical*

    education, *84*(1).

Kok, N., Koponen, E. L., & Martínez-Barbosa, C. A. (2017). Big data in real

    estate? From manual appraisal to automated valuation. *The Journal of*

    Portfolio Management, *43*(6), 202-211.

Kuckartz, U. (2019). Qualitative content analysis: From Kracauer's beginnings to

    today's challenges. In *Forum Qualitative Sozialforschung/Forum:*

    Qualitative Social Research (Vol. 20, No. 3, p. 20). DEU.

Kumar, D. (2020). A review: Big data and machine learning.

    https://www.ijsdr.org/papers/IJSDR2008057.pdf

Locurcio, M., Tajani, F., & Morano, P. (2020). Computational methods applied to

    data analysis for modeling complex real estate

    systems. *Complexity*, *2020*.

Li, B. (2021). Research on Real Estate Information System of the Real Estate

    Market Based on Big Data Technology. In *E3S Web of Conferences* (Vol.

    257, p. 02037). EDP Sciences.

Meier, Z., Pan, W. T., Zhuohong, S., Yingying, Z., & Zuchang, Z. (2019, August).

    Application of data mining technology in evaluating real estate investment

    plan based on GRA-AHP. In *Journal of Physics: Conference Series* (Vol.

    1284, No. 1, p. 012037). IOP Publishing.

Munawar, H. S., Ullah, F., Qayyum, S., & Shahzad, D. (2022). Big Data in

    Construction: Current applications and Future Opportunities. *Big Data and*

    Cognitive Computing, *6*(1), 18.

Munawar, H. S., Qayyum, S., Ullah, F., & Sepasgozar, S. (2020). Big data and its

    applications in smart real estate and the disaster management life cycle:

    A systematic analysis. *Big Data and Cognitive Computing*, *4*(2), 4.

Obinna, W. K., & Udo, M. J. (2022). Improving online Real Estate Management

    System using data analytics. *Journal of Emerging Technologies*, *2*(2), 66-

    75.

Oluwunmi, A. O., Role, B. A., Akinwale, O. M., Oladayo, O. P., & Afolabi, T. O.
(2019, December). Big data and real estate: a review of literature.
In *Journal of Physics: Conference Series* (Vol. 1378, No. 3, p. 032015).
IOP Publishing.

Pai, P. F., & Wang, W. C. (2020). Using machine learning models and actual
transaction data for predicting real estate prices. *Applied
Sciences*, *10*(17), 5832.

Pérez-Rave, J. I., Correa-Morales, J. C., & González-Echavarría, F. (2019). A
machine learning approach to big data regression analysis of real estate
prices for inferential and predictive purposes. *Journal of Property
Research*, *36*(1), 59-96.

Rabby, F., Chimhundu, R., & Hassan, R. (2022). Digital Transformation in Real
Estate Marketing: A review. *Big Data: A Road Map for Successful Digital
Marketing*, 39.

Rossini, P. (2000, January). Using expert systems and artificial intelligence for
real estate forecasting. In *Sixth Annual Pacific-Rim Real Estate Society
Conference, Sydney, Australia* (pp. 24-27).

Selvi, A. F. (2019). Qualitative content analysis. In *The Routledge handbook of
research methods in applied linguistics* (pp. 440-452). Routledge.

Singh, V., Chen, S. S., Singhania, M., Nanavati, B., & Gupta, A. (2022). How are
reinforcement learning and deep learning algorithms used for big data
based decision making in financial industries–A review and research

agenda. *International Journal of Information Management Data Insights*, *2*(2), 100094.

Tabesh, P., Mousavidin, E., & Hasani, S. (2019). Implementing big data strategies: A managerial perspective. *Business Horizons*, *62*(3), 347-358.

Treleaven, P., Barnett, J., Knight, A., & Serrano, W. (2021). Real estate data marketplace. *AI and Ethics*, *1*(4), 445-462.

Ullah, F., Sepasgozar, S. M., Thaheem, M. J., & Al-Turjman, F. (2021). Barriers to the digitalisation and innovation of Australian Smart Real Estate: A managerial perspective on the technology non-adoption. *Environmental Technology & Innovation*, *22*, 101527.

U.S. House Prices Rise 17.7 Percent over the Last Year; Up 4.0 Percent from the First Quarter | Federal Housing Finance Agency. (n.d.). Retrieved October 30, 2022, from https://www.fhfa.gov/Media/PublicAffairs/Pages/US-House-Prices-Rise-17pt7-Percent-over-the-Last-Year.aspx

Wei, C., Fu, M., Wang, L., Yang, H., Tang, F., & Xiong, Y. (2022). The research development of hedonic price model-based real estate appraisal in the era of big data. *Land*, *11*(3), 334.

Winson-Geideman, K., & Krause, A. (2016, January). Transformations in real estate research: The big data revolution. In *Proceedings of the 22nd*

Annual Pacific-Rim Real Estate Society Conference, Queensland,

Australia (pp. 17-20).

Wu, J., Liu, G., Wang, J., Zuo, Y., Bu, H., & Lin, H. (2020). Data intelligence:

Trends and challenges. *Syst. Eng.-Theory Pract*, *40*, 2116-2149.

Wei, L., & Zhang, W.-X. (2013). RMB/USD exchange rate prediction based on

support vector machine and principal component analysis. *Information

Technology Journal*, *12*(16), 3660–3664.

https://doi.org/10.3923/itj.2013.3660.3664

Yang, X., Yin, Z., & Li, J. (2021, December). Housing price mathematical

prediction method through big data analysis and improved linear

regression model. In 2021 *IEEE Conference on Telecommunications,

Optics and Computer Science (TOCS)* (pp. 751-754). IEEE.

Yousuf, H., & Zainal, A. Y. (2020). Quantitative approach in enhancing decision

making through big data as an advanced technology. *Adv. Sci. Technol.

Eng. Syst. J*, *5*(5), 109-116.