© July 2026 | IJIRT | Volume 13 Issue 2 | ISSN: 2349-6002

AI-Powered Buyer Segmentation for Real Estate Using Machine Learning

Ishita Das
Department of Computer Science and Engineering, Kalinga Institute of Industrial Technology (KIIT), Bhubaneswar, India

***Abstract*—The rapid digital transformation of the real estate industry has resulted in the generation of large volumes of customer and transaction data, creating opportunities for data-driven decision-making while simultaneously introducing challenges in identifying meaningful buyer patterns. Traditional customer segmentation techniques often rely on manual classification or demographic information, which fail to capture complex behavioural and investment characteristics of buyers. Consequently, organizations face difficulties in developing targeted marketing strategies, optimizing customer engagement, and improving investment planning.**

This paper presents an AI-powered buyer segmentation framework that leverages machine learning techniques to identify distinct groups of real estate buyers based on demographic, financial, and behavioural attributes. The proposed framework utilizes customer and property transaction datasets containing buyer demographics, investment behaviour, property ownership details, financing preferences, and satisfaction metrics. After performing comprehensive data preprocessing, feature engineering, and data transformation, the K-Means clustering algorithm was employed to discover hidden customer segments without requiring predefined class labels. The optimal number of clusters was determined using the Elbow Method and Silhouette Score to ensure meaningful segmentation.

To enhance the practical applicability of the proposed framework, an interactive business intelligence dashboard was developed using Streamlit and Plotly. The dashboard enables stakeholders to explore buyer behaviour through dynamic filters, key performance indicators, and interactive visualizations, supporting real-time business analysis and decision-making. The identified buyer segments provide valuable insights into customer investment capacity, purchasing behaviour, financing patterns, and regional distribution, enabling organizations to implement personalized marketing strategies and improve customer relationship management.

The experimental results demonstrate that the proposed framework effectively transforms raw real estate data into actionable business intelligence. By integrating machine learning with interactive visualization, the developed system provides a scalable and practical solution for customer segmentation, strategic planning, and data-driven decision-making within the real estate industry.

***Index Terms*—Buyer Segmentation, Machine Learning, K-Means Clustering, Real Estate Analytics, Business Intelligence, Streamlit, Customer Segmentation.**

I. INTRODUCTION

The real estate industry has undergone a significant transformation with the rapid advancement of digital technologies and the increasing availability of data-driven solutions. Every customer interaction, property transaction, financing activity, and investment decision generates valuable information that can be utilized to understand buyer behaviour and support strategic business planning. However, the growing volume, variety, and complexity of real estate data have made traditional analytical approaches increasingly inadequate for extracting meaningful insights. Organizations often possess large datasets but struggle to convert them into actionable knowledge that can improve customer engagement, optimize marketing strategies, and enhance investment decisions.

Customer segmentation has emerged as one of the most effective approaches for understanding diverse buyer groups within the real estate market. Rather than treating all customers as a homogeneous population, segmentation enables organizations to classify buyers based on common demographic, behavioural, and financial characteristics. Such categorization helps businesses design personalized marketing campaigns, recommend suitable properties, improve customer relationship management, and allocate organizational resources more efficiently. Despite these advantages, conventional segmentation techniques primarily rely on manual analysis or simple demographic classifications, which frequently overlook hidden behavioural patterns and complex relationships among customers.

Recent developments in Artificial Intelligence (AI) and Machine Learning (ML) have significantly enhanced the capability of organizations to analyze large-scale datasets and automatically discover meaningful patterns. Among various unsupervised learning techniques, K-Means clustering has gained widespread acceptance due to its simplicity, computational efficiency, and ability to partition data into distinct groups based on similarity. Unlike supervised learning methods, clustering algorithms do not require predefined class labels, making them particularly suitable for customer segmentation problems where the underlying buyer categories are unknown. By identifying naturally occurring customer groups, machine learning enables businesses to develop more accurate and data-driven strategies for customer acquisition, retention, and investment planning.

In addition to machine learning models, modern Business Intelligence (BI) platforms have become essential for translating analytical results into practical decision-support systems. Interactive dashboards allow business stakeholders to explore complex datasets through intuitive visualizations, performance indicators, and dynamic filtering mechanisms without requiring extensive technical expertise. The integration of machine learning with business intelligence therefore provides a comprehensive solution that bridges the gap between advanced analytics and real-world business decision-making.

This paper presents an AI-powered buyer segmentation framework developed for the real estate sector using machine learning techniques. The proposed approach utilizes customer demographic information, property ownership records, financing preferences, investment behaviour, and customer satisfaction data to identify meaningful buyer segments. Comprehensive data preprocessing, feature engineering, and transformation techniques are applied prior to implementing the K-Means clustering algorithm. The resulting customer segments are subsequently integrated into an interactive Streamlit-based dashboard that enables real-time visualization and exploration of buyer characteristics, investment trends, and geographical distributions.

The proposed framework demonstrates how artificial intelligence and business intelligence can be combined to transform raw real estate data into actionable insights. The identified buyer segments assist organizations in designing personalized marketing campaigns, improving customer engagement, optimizing financing strategies, and supporting strategic business planning. Furthermore, the developed dashboard provides decision-makers with an accessible platform for analyzing customer behaviour and deriving data-driven recommendations that enhance organizational performance.

The major contributions of this research are summarized as follows:

- Development of an AI-powered framework for buyer segmentation using machine learning techniques.
- Integration of customer demographic, financial, and property transaction data through comprehensive preprocessing and feature engineering.
- Implementation of the K-Means clustering algorithm to identify meaningful buyer segments based on behavioural and investment characteristics.
- Development of an interactive Streamlit dashboard for real-time visualization and business analysis of customer segments.
- Generation of actionable business insights that support personalized marketing strategies, customer relationship management, and informed decision-making in the real estate industry.

The remainder of this paper is organized as follows. Section II reviews the existing literature related to customer segmentation and machine learning applications in the real estate domain. Section III presents the proposed methodology, including dataset preparation, preprocessing, feature engineering, clustering, and dashboard development. Section IV discusses the experimental results and business interpretation of the identified buyer segments, while Section V concludes the paper and outlines potential directions for future research.

II. LITERATURE REVIEW

The increasing adoption of Artificial Intelligence (AI) and Machine Learning (ML) has significantly transformed customer analytics across various industries, including retail, banking, healthcare, and real estate. Customer segmentation has emerged as one of the most important applications of machine learning, enabling organizations to identify groups of customers with similar behavioural and financial characteristics. By understanding these customer groups, businesses can improve marketing effectiveness, optimize resource allocation, enhance customer satisfaction, and support strategic decision-making.

Traditional customer segmentation methods primarily rely on demographic information, geographic location, or manually defined business rules. Although these approaches provide basic customer categorization, they often fail to capture hidden behavioural patterns present within large and complex datasets. As organizations continue to generate vast amounts of customer information, there is an increasing need for intelligent analytical techniques capable of automatically discovering meaningful customer segments from multidimensional data.

Han et al. emphasized the importance of data mining techniques for extracting useful knowledge from large datasets and highlighted clustering as one of the fundamental methods for discovering hidden structures within unlabeled data. Their work demonstrated that clustering algorithms provide an effective foundation for customer profiling and market segmentation by grouping similar observations based on feature similarity. Similarly, Aggarwal discussed various clustering techniques for customer analytics and concluded that unsupervised learning methods are particularly effective when predefined customer categories are unavailable.
Among the available clustering algorithms, K-Means has become one of the most widely adopted approaches because of its simplicity, computational efficiency, and scalability. The algorithm partitions observations into distinct clusters by minimizing the distance between data points and their corresponding cluster centroids. Numerous studies have successfully applied K-Means clustering in domains such as banking, e-commerce, healthcare, and customer relationship management to identify customer groups and improve business performance. Its ability to process large datasets while maintaining relatively low computational complexity makes it especially suitable for real-world business applications.

Recent studies in the real estate sector have increasingly focused on the application of machine learning for property valuation, customer behaviour analysis, demand forecasting, and investment planning. Researchers have utilized demographic information, purchasing history, property characteristics, and financial indicators to better understand buyer preferences and market trends. These studies demonstrate that machine learning can reveal valuable insights that are often overlooked using conventional analytical methods, thereby supporting more informed business decisions and improving customer engagement.

The emergence of Business Intelligence (BI) tools has further enhanced the practical implementation of machine learning models. Interactive dashboards enable organizations to convert analytical results into intuitive visualizations that can be interpreted by business stakeholders without requiring technical expertise. Modern visualization platforms provide dynamic filtering, key performance indicators, comparative analysis, and real-time reporting capabilities, allowing organizations to make faster and more informed decisions based on analytical outcomes. The integration of machine learning with interactive dashboards therefore represents an important advancement in data-driven decision support systems.

Despite these advancements, several limitations remain in existing research. Many published studies primarily focus on clustering techniques without providing comprehensive business interpretation of the identified customer segments. Some research emphasizes predictive modeling while neglecting interactive visualization and decision-support mechanisms. Additionally, relatively few studies integrate customer demographics, investment behaviour, financing preferences, property ownership characteristics, and business intelligence within a single analytical framework specifically designed for the real estate industry.

To address these limitations, the present study proposes an integrated AI-powered buyer segmentation framework that combines comprehensive data preprocessing, feature engineering, K-Means clustering, and an interactive Streamlit-based business intelligence dashboard. Unlike traditional approaches that present clustering results only through statistical analysis, the proposed framework enables real-time exploration of buyer segments using dynamic visualizations and filtering capabilities. This integration improves the interpretability of machine learning results while providing actionable insights that support personalized marketing, customer relationship management, investment planning, and strategic decision-making within the real estate sector.

The proposed framework demonstrates how machine learning and business intelligence can be effectively combined to transform complex real estate datasets into meaningful customer insights. By providing both analytical accuracy and interactive visualization, the developed system offers a practical solution for organizations seeking to implement AI-driven customer segmentation and data-informed business strategies.

III. PROPOSED METHODOLOGY

This study proposes an AI-powered buyer segmentation framework that integrates data preprocessing, feature engineering, machine learning, and business intelligence to identify meaningful customer segments within the real estate industry. The proposed methodology transforms raw customer and property transaction data into actionable business insights through a systematic analytical workflow. The framework consists of seven sequential stages, including data acquisition, preprocessing, feature engineering, clustering, buyer segmentation, business insight generation, and interactive dashboard development.

The workflow begins with collecting customer demographic information and property transaction records from the Clients and Properties datasets. Since real-world datasets frequently contain inconsistencies, missing values, and heterogeneous data formats, extensive preprocessing is performed before analysis. Data cleaning, missing value handling, categorical encoding, and feature standardization ensure that the processed dataset is reliable and suitable for machine learning.

Following preprocessing, meaningful features such as customer age, total investment, property ownership count, average property price, and average floor area are derived to better represent buyer behaviour. The processed dataset is then standardized and analyzed using the K-Means clustering algorithm. The optimal number of clusters is determined through the Elbow Method and Silhouette Score analysis. Finally, the identified buyer segments are integrated into an interactive Business Intelligence dashboard developed using Streamlit and Plotly, enabling stakeholders to explore customer behaviour through dynamic visualizations and real-time analytics.

The proposed methodology follows a structured pipeline in which each stage progressively refines the data and contributes to the overall analytical framework. Raw customer and property datasets are transformed into meaningful buyer segments through systematic preprocessing, feature engineering, clustering, and business interpretation. The final dashboard bridges the gap between machine learning and business intelligence by presenting analytical insights in an interactive format that supports data-driven decision-making.

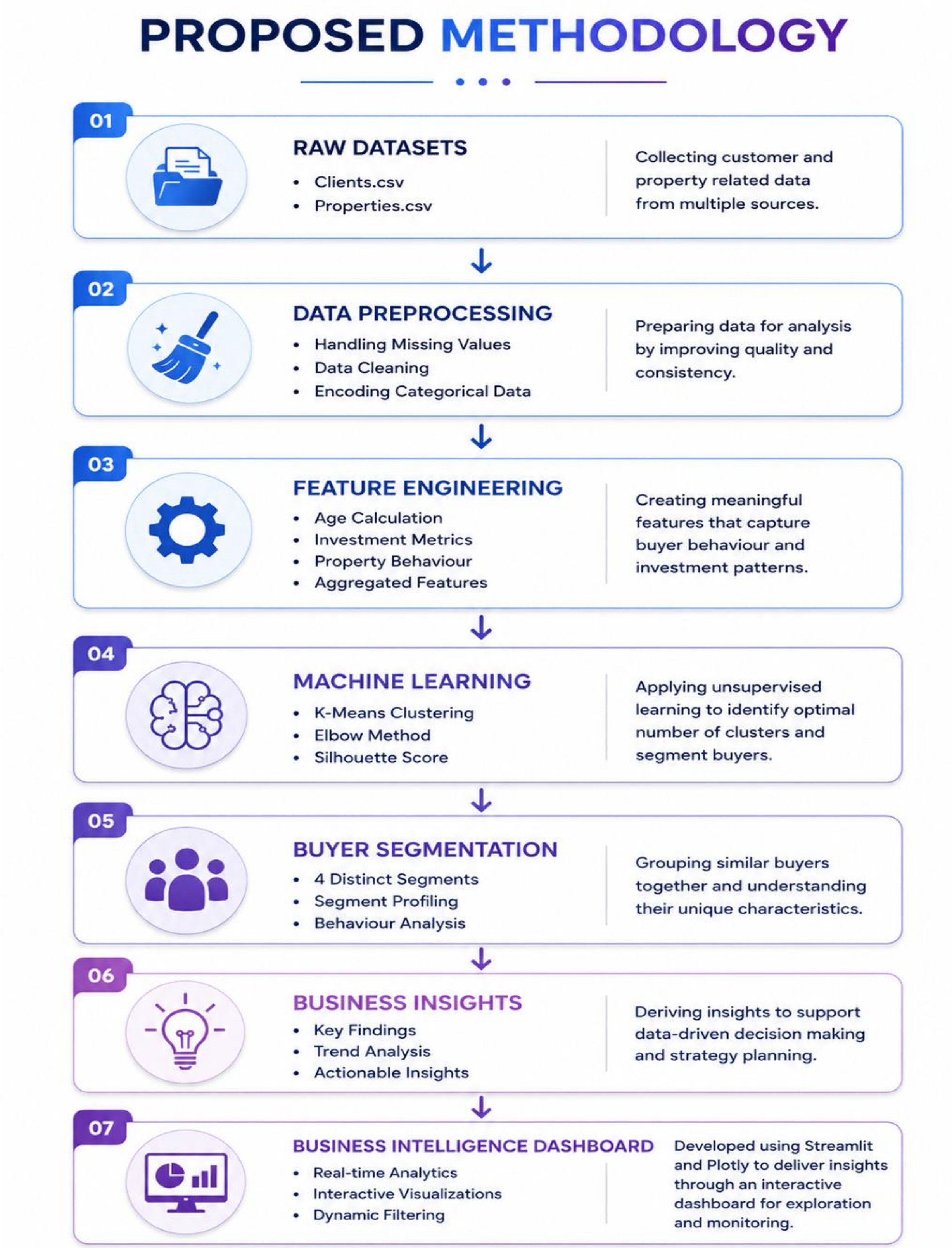


Figure 1. Overall Workflow of the Proposed AI-Powered Buyer Segmentation Framework

3.1. Dataset Description

The proposed buyer segmentation framework utilizes two structured datasets obtained during the internship at Parcl Co. Limited (Unified Mentor). These datasets collectively capture customer demographics, property transaction details, investment behaviour, and financing information required for identifying meaningful buyer segments through machine learning. The integration of customer-specific and property-related information enables a comprehensive analysis of purchasing patterns and investment characteristics within the real estate domain.

The first dataset, namely the Clients dataset, contains demographic and customer-related information such as client type, age, gender, country, region, acquisition purpose, satisfaction score, loan application status, and referral channel. These attributes provide valuable insights into customer profiles and behavioural characteristics that influence purchasing decisions.

The second dataset, referred to as the Properties dataset, contains transaction-level information associated with purchased properties. It includes attributes such as listing identification, transaction date, unit category, floor area, sale price, listing status, and client reference. These variables describe property ownership, investment value, and purchasing behaviour, thereby complementing the customer information available in the Clients dataset.

To facilitate buyer segmentation, both datasets were integrated using the client reference identifier to create a unified analytical dataset. This merged dataset combines demographic, financial, and property ownership information for each customer, providing a comprehensive representation of buyer behaviour suitable for machine learning analysis.

Table 1 summarizes the characteristics of the datasets used in this study.

Table 1. Dataset Summary

Dataset	Description	Records	Features
Clients	Customer demographic and behavioural information	2,000	12
Properties	Property transaction details	7,305	9

3.2. Data Preprocessing

Data preprocessing is a fundamental stage in the proposed buyer segmentation framework, as the quality of the input data directly influences the accuracy and reliability of the machine learning model. The raw datasets obtained from the Clients and Properties tables contained information collected from different operational processes and therefore required systematic preprocessing before analytical modeling. The objective of this phase was to improve data quality, eliminate inconsistencies, and prepare the datasets for clustering.

Initially, both datasets were examined to understand their structure, attribute types, and completeness. Descriptive statistics and data inspection techniques were employed to identify missing values, duplicate records, inconsistent formats, and invalid entries. This preliminary analysis provided a clear understanding of the overall data quality and highlighted the preprocessing tasks required before further analysis.

Missing value analysis revealed that the Clients dataset contained complete customer information, whereas the Properties dataset included missing values in the client reference field. These missing references corresponded to unsold property listings and therefore did not contribute to buyer behaviour analysis. Instead of imputing artificial values, only property records associated with valid customer references were retained for subsequent analysis, ensuring that the final dataset represented genuine customer-property relationships.

To maintain data consistency, duplicate records were examined using unique identifiers and row-level comparisons. No significant duplicate entries were identified, indicating that both datasets maintained data integrity and required minimal structural modification. Data cleaning procedures were further applied to standardize categorical variables, correct inconsistent labels, and convert numerical attributes into appropriate formats for analysis. For example, customer categories representing corporate organizations were standardized under a common label, while monetary values were converted into numerical data types after removing formatting symbols.

Date-related attributes were also transformed into standardized datetime formats to facilitate feature extraction. Customer age was calculated from the Date of Birth attribute and incorporated into the analytical dataset as a derived feature. Since buyer age plays an important role in understanding purchasing behaviour and investment preferences, accurate date conversion was essential for subsequent analysis.

After cleaning and transformation, categorical variables were encoded into numerical representations to enable their utilization within the machine learning model. Numerical features exhibiting different scales were standardized using the StandardScaler technique. Feature scaling ensured that variables such as total investment, property value, and floor area did not dominate the distance calculations performed by the K-Means clustering algorithm, thereby improving clustering performance and segmentation quality.

The preprocessing stage produced a clean, consistent, and integrated analytical dataset that served as the foundation for feature engineering and machine learning. By systematically addressing data quality issues and standardizing the input variables, the proposed framework ensured that subsequent analytical results were reliable, interpretable, and suitable for identifying meaningful buyer segments within the real estate market.

3.3. Feature Engineering

Feature engineering plays a crucial role in enhancing the effectiveness of machine learning models by transforming raw data into meaningful attributes that better represent customer behaviour. While the original datasets contained valuable demographic and transaction-level information, several additional features were derived to capture long-term investment patterns and purchasing characteristics more effectively. These engineered features provide a comprehensive representation of buyer profiles and improve the quality of customer segmentation.

Customer age was calculated from the Date of Birth attribute after converting it into a standardized datetime format. Age serves as an important demographic indicator that influences purchasing preferences, investment decisions, and financing behaviour. Rather than using raw date values, the derived age attribute provides a more meaningful numerical representation for clustering.

To analyze investment behaviour, the total number of properties owned by each customer was calculated and recorded as the Property Count feature. This attribute reflects purchasing frequency and distinguishes occasional buyers from customers with multiple property investments. Buyers with higher property counts generally exhibit different purchasing behaviour compared to first-time or single-property buyers.

Another important engineered feature is Total Investment, which represents the cumulative monetary value of all properties purchased by an individual customer. This variable provides a direct measure of a buyer's investment capacity and enables the identification of high-value customers who contribute significantly to overall business revenue.

In addition to total investment, the Average Property Price was computed by dividing the total investment by the number of properties owned. This feature represents the average value of properties purchased by each customer and helps differentiate buyers who consistently invest in premium properties from those focusing on affordable housing options.

The Average Floor Area feature was also derived to represent the typical size of properties purchased by individual buyers. Property size often reflects customer preferences, purchasing power, and intended usage, making it a valuable attribute for identifying distinct buyer segments within the dataset.

After generating these derived attributes, they were combined with the original demographic and financial variables to create a unified analytical dataset. The inclusion of engineered features significantly enriched the representation of customer behaviour by incorporating both demographic characteristics and long-term investment patterns. Consequently, the clustering algorithm was able to identify buyer groups based on comprehensive behavioural profiles rather than relying solely on the original transactional attributes.

Table 2 presents the engineered features incorporated into the proposed buyer segmentation framework.

Table 2. Engineered Features

Engineered Feature	Description
Age	Calculated from the customer's Date of Birth to represent demographic characteristics.
Property Count	Total number of properties owned by each customer.
Total Investment	Cumulative value of all properties purchased by a customer.
Average Property Price	Mean purchase price of properties owned by a customer.
Average Floor Area	Average floor area of properties purchased by a customer.

The engineered features collectively provide a richer representation of buyer behaviour, investment capacity, and property ownership characteristics. These variables form the primary input for the K-Means clustering algorithm and contribute significantly to the identification of meaningful customer segments that support strategic business decision-making.

3.4. K-Means Clustering

Customer segmentation was performed using the K-Means clustering algorithm, an unsupervised machine learning technique widely employed for identifying hidden patterns within unlabeled datasets. Unlike supervised learning methods, K-Means does not require predefined class labels. Instead, it partitions observations into homogeneous groups based on feature similarity, enabling the automatic discovery of distinct buyer segments from customer demographic, financial, and behavioural attributes.

Prior to clustering, all numerical features were standardized using the **StandardScaler** technique to eliminate differences in measurement scales. Since attributes such as total investment, average property price, and floor area possess significantly larger numerical values than demographic variables such as age and satisfaction score, standardization was essential to ensure that no single feature disproportionately influenced the Euclidean distance calculations used by the clustering algorithm.

The K-Means algorithm operates by randomly initializing a predefined number of cluster centroids and iteratively assigning each observation to its nearest centroid based on Euclidean distance. After assigning all observations, the centroid positions are recalculated using the mean of the observations within each cluster. This iterative optimization process continues until the centroid locations stabilize or the maximum number of iterations is reached, resulting in compact and well-defined customer groups.

Selecting an appropriate number of clusters is a critical step in clustering analysis. To determine the optimal value of **K**, two complementary evaluation techniques were employed: the Elbow Method and the Silhouette Score. The Elbow Method examines the relationship between the number of clusters and the Within-Cluster Sum of Squares (WCSS). The optimal cluster count is identified at the point where increasing the number of clusters results in only marginal improvements in reducing WCSS. While the Elbow Method provides a visual indication of the optimal value of K, the Silhouette Score quantitatively evaluates clustering quality by measuring both intra-cluster cohesion and inter-cluster separation. Higher silhouette values indicate that observations are well matched to their assigned clusters while remaining distinct from neighbouring clusters.

Based on the combined evaluation of the Elbow Method and Silhouette Score, the optimal clustering configuration was selected for the buyer segmentation framework. Each customer was subsequently assigned to a cluster according to similarities in demographic characteristics, investment behaviour, financing preferences, and property ownership patterns. Rather than interpreting clusters solely as numerical labels, each segment was analyzed using descriptive statistics to identify its distinguishing characteristics and business significance.

The resulting buyer segments represent groups of customers with similar purchasing behaviour and investment profiles. These segments provide valuable business insights that enable organizations to design targeted marketing campaigns, personalize customer engagement strategies, optimize financing recommendations, and improve overall decision-making within the real estate sector. The clustered dataset also serves as the primary input for the interactive Business Intelligence dashboard, where users can dynamically explore customer segments through visual analytics.

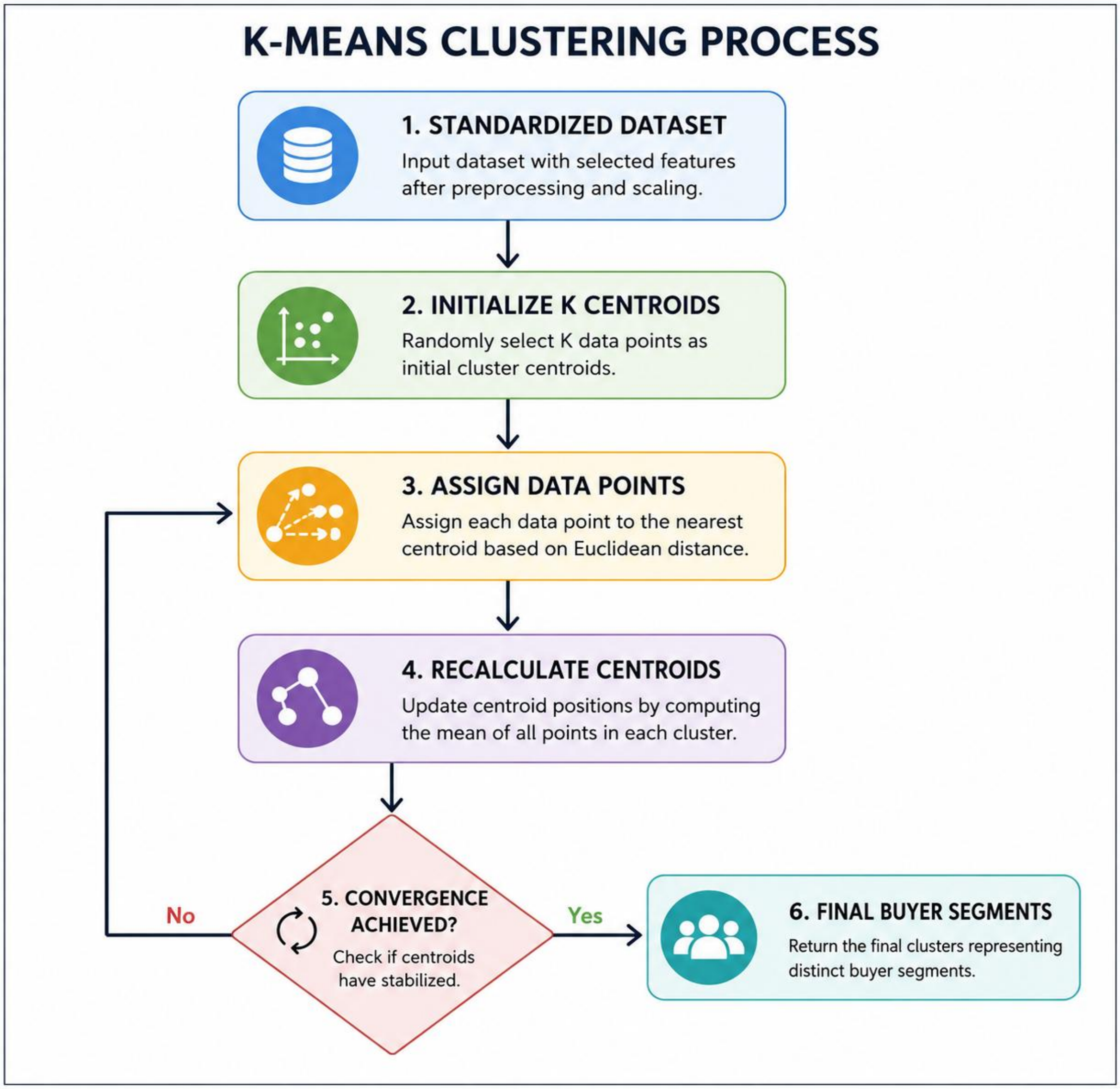


Figure 2. K-Means Clustering Process

3.5. Business Intelligence Dashboard

To enhance the practical applicability of the proposed buyer segmentation framework, an interactive Business Intelligence (BI) dashboard was developed using Streamlit and Plotly. While the K-Means algorithm successfully identifies distinct buyer segments, the generated clusters alone are insufficient for business stakeholders who require intuitive and actionable insights. Therefore, the dashboard serves as a decision-support system that transforms analytical results into interactive visualizations, enabling users to explore customer behaviour without requiring technical expertise.

The dashboard was designed to provide a comprehensive overview of buyer segmentation through a user-friendly interface. It integrates the clustered dataset with multiple visualization components, including key performance indicators (KPIs), distribution charts, comparative graphs, geographic analysis, and investment behaviour summaries. These visual elements enable stakeholders to analyze customer demographics, purchasing patterns, financing preferences, and property ownership characteristics from multiple perspectives.

To improve usability, interactive filtering capabilities were incorporated into the dashboard. Users can dynamically filter customer segments based on attributes such as country, region, gender, client type, and acquisition purpose. All visualizations and performance metrics update automatically according to the selected filters, allowing users to perform customized analyses and compare different buyer groups in real time.

The dashboard also presents several business-oriented performance indicators, including the total number of buyers, average investment value, average property price, average property ownership count, average floor area, customer age, and satisfaction score. These metrics provide a concise summary of buyer characteristics and enable organizations to quickly identify high-value customer segments and emerging investment trends.

In addition to numerical indicators, interactive charts developed using Plotly facilitate deeper exploration of customer behaviour. Visualizations such as buyer segment distribution, investment comparison, property ownership analysis, financing patterns, and geographical insights enable stakeholders to recognize hidden trends and evaluate differences among customer groups. The interactive nature of these visualizations allows users to zoom, filter, and examine individual segments without modifying the underlying analytical model.

The integration of machine learning with business intelligence significantly improves the practical value of the proposed framework. Rather than presenting clustering results as static outputs, the dashboard converts them into an accessible analytical platform that supports evidence-based decision-making. Consequently, real estate organizations can utilize the dashboard to design targeted marketing campaigns, optimize financing strategies, personalize customer engagement, and monitor buyer behaviour through an intuitive visual interface.

Figure 3 illustrates the Business Intelligence dashboard developed as part of the proposed buyer segmentation framework.

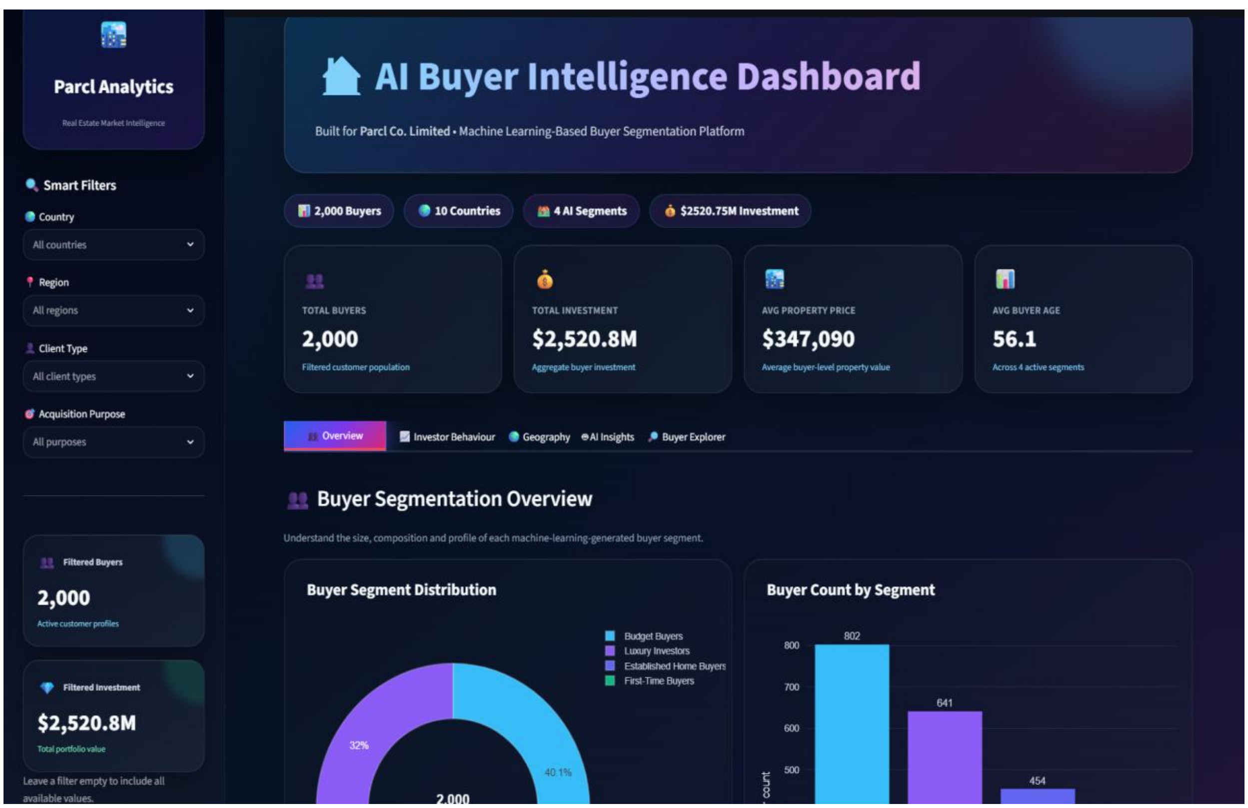


Figure 3. Complete Business Intelligence Dashboard

IV. RESULTS AND DISCUSSION

The proposed AI-powered buyer segmentation framework was implemented to identify meaningful customer groups within the real estate dataset using unsupervised machine learning techniques. The analytical pipeline included data preprocessing, feature engineering, K-Means clustering, and interactive visualization through a Business Intelligence dashboard. This section presents the experimental results and discusses the effectiveness of the proposed framework in identifying distinct buyer segments.

4.1. Determination of Optimal Number of Clusters

Selecting an appropriate number of clusters is a critical step in K-Means clustering, as it directly influences the quality and interpretability of the generated customer segments. To determine the optimal value of K, the proposed framework employed two complementary evaluation techniques: the Elbow Method and the Silhouette Score.

The Elbow Method evaluates the Within-Cluster Sum of Squares (WCSS) for different values of **K**. As the number of clusters increases, WCSS decreases because observations become more closely grouped around their respective centroids. However, beyond a certain point, the rate of improvement diminishes significantly, forming an "elbow" in the graph. This point represents a suitable balance between cluster compactness and model complexity.

To further validate the clustering performance, Silhouette Score analysis was conducted. The Silhouette Score measures both intra-cluster cohesion and inter-cluster separation, providing an overall assessment of clustering quality. Higher silhouette values indicate that observations are well grouped within their assigned clusters while remaining distinct from neighbouring clusters.

The Silhouette Score attained its highest value of 0.201 at K = 3, indicating strong statistical separation for a three-cluster configuration. However, the Elbow Method showed a continued meaningful reduction in within-cluster variation around the three-to-four cluster range. A comparative examination of the resulting cluster profiles revealed that the four-cluster configuration provided greater business interpretability by distinguishing Budget Buyers, Luxury Investors, Established Home Buyers, and First-Time Buyers as separate customer groups. Therefore, K = 4 was selected for the final buyer segmentation model, despite its slightly lower silhouette score, because it offered a more granular and actionable representation of buyer behaviour.

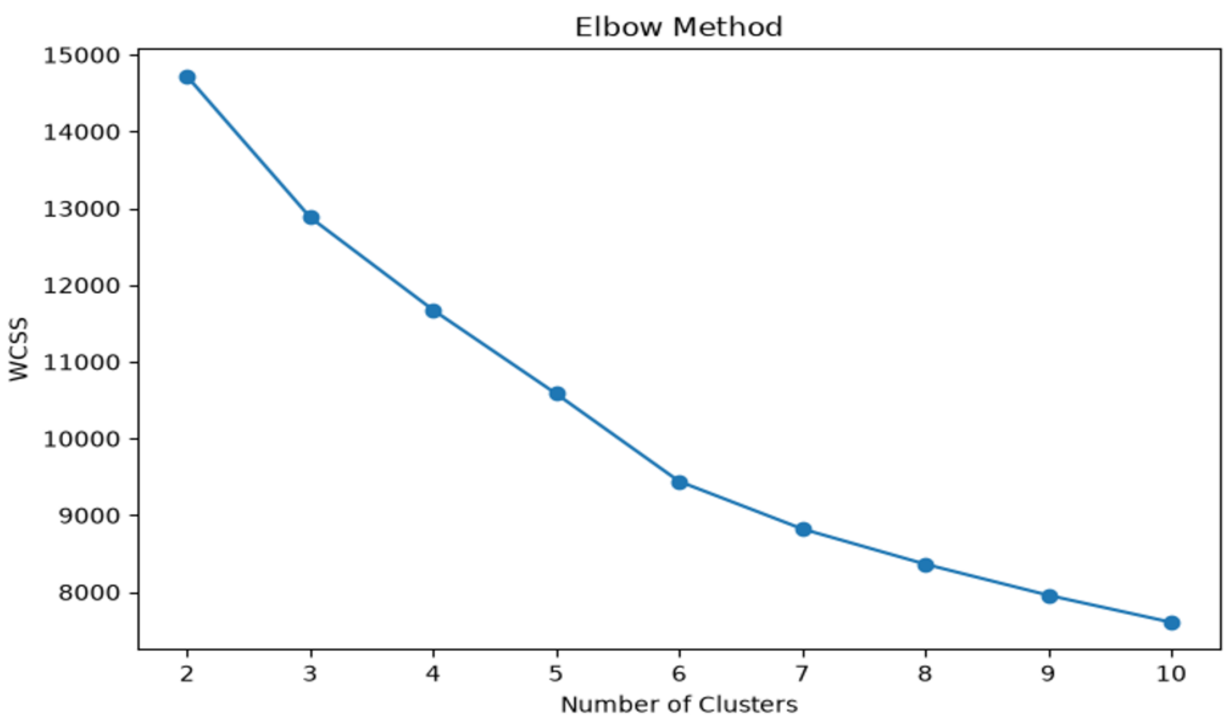


Figure 4. Elbow Method used to determine the optimal number of clusters.

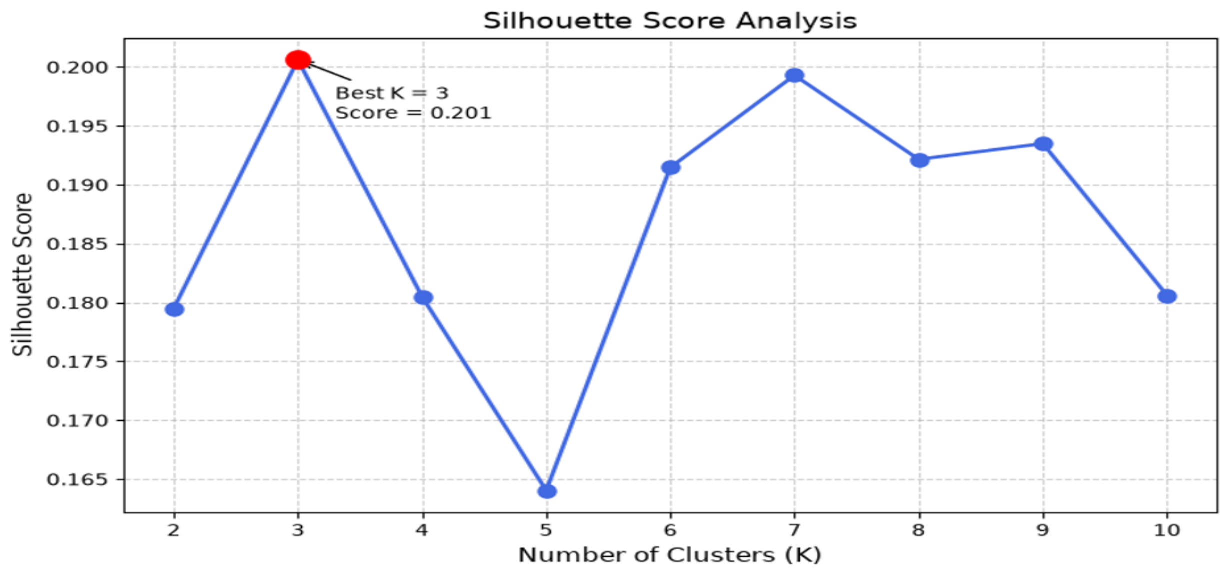


Figure 5. illustrates the corresponding Silhouette Score analysis.

Figure 5 indicates that K = 3 achieved the highest Silhouette Score. Nevertheless, K = 4 was adopted for the final model because the additional cluster produced a more meaningful separation of buyer profiles and supported more targeted business recommendations.

4.2. Buyer Segment Characteristics

After determining the optimal number of clusters, the K-Means algorithm grouped customers based on demographic characteristics, investment behaviour, financing preferences, and property ownership patterns. Each cluster exhibited distinct behavioural traits, enabling the identification of meaningful buyer segments with different investment capacities and purchasing preferences.

The resulting customer groups demonstrated noticeable variations in attributes such as total investment value, average property price, property ownership count, floor area, customer age, and satisfaction score. These differences indicate that buyers within the same cluster share similar purchasing behaviour while remaining significantly different from customers belonging to other clusters.

Rather than interpreting clusters solely through numerical labels, each cluster was analyzed using descriptive statistics to derive business-oriented insights. This interpretation enables organizations to distinguish between premium investors, moderate buyers, and value-oriented customers, thereby supporting targeted marketing campaigns, personalized financing solutions, and customer engagement strategies.

Table 3 summarizes the major characteristics of the identified buyer segments.

Table 3. Buyer Segment Characteristics

Buyer Segment	Key Characteristics	Business Opportunity
Budget Buyers	Moderate investment, value-conscious purchasing behaviour	Affordable housing campaigns and financing offers
Luxury Investors	High investment capacity and premium property ownership	Premium listings and exclusive investment opportunities
Established Home Buyers	Stable purchasing behaviour with long-term ownership	Customer retention and loyalty programs
First-Time Buyers	Younger customers entering the property market	Financial guidance and first-home assistance programs

4.3. Dashboard Analysis

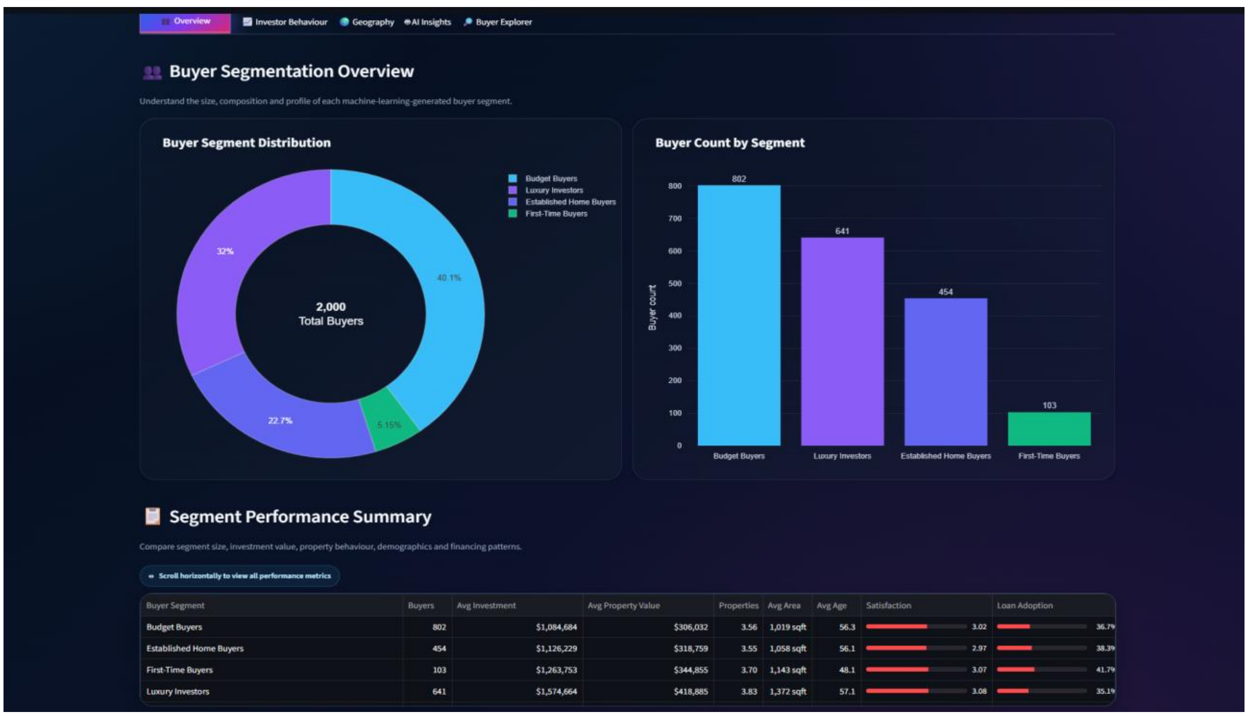
The clustered dataset was integrated into an interactive Business Intelligence dashboard developed using Streamlit and Plotly to facilitate real-time exploration of buyer behaviour. The dashboard provides an intuitive interface that enables users to analyze customer segments through interactive charts, dynamic filters, and key performance indicators.

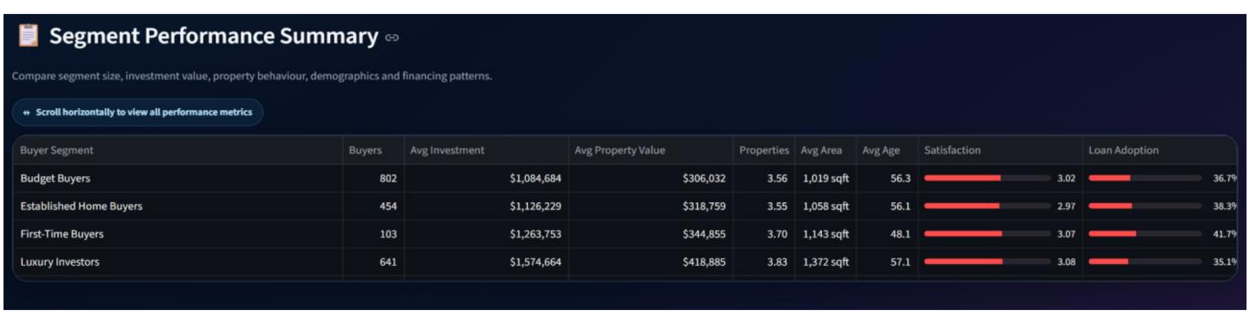
Users can filter the dataset based on demographic and behavioural attributes, including country, region, gender, client type, and acquisition purpose. All visualizations update dynamically according to the selected filters, enabling customized exploration of customer behaviour across different buyer categories.

The dashboard presents multiple analytical views, including buyer segment distribution, investment comparison, financing patterns, geographical insights, and customer demographics. These visualizations enable business stakeholders to identify high-value customer groups, monitor investment trends, evaluate regional purchasing behaviour, and compare buyer characteristics without requiring advanced analytical expertise.

The integration of machine learning with interactive visualization significantly improves the interpretability of clustering results. Instead of presenting static analytical outputs, the dashboard transforms complex customer data into actionable business intelligence that supports strategic decision-making and customer-centric marketing initiatives.

Figure 6 illustrates the developed Business Intelligence dashboard.





Buyer Segment	Buyers	Avg Investment	Avg Property Value	Properties	Avg Area	Avg Age	Satisfaction	Loan Adoption
Budget Buyers	802	$1,084,684	$306,032	3.56	1,019 sqft	56.3	3.02	36.7%
Established Home Buyers	454	$1,126,229	$318,759	3.55	1,058 sqft	56.1	2.97	38.3%
First-Time Buyers	103	$1,263,753	$344,855	3.70	1,143 sqft	48.1	3.07	41.7%
Luxury Investors	641	$1,574,664	$418,885	3.83	1,372 sqft	57.1	3.08	35.1%

Figure 6. Overview Dashboard (Results)

V. CONCLUSION AND FUTURE WORK

This study presented an AI-powered buyer segmentation framework for the real estate sector by integrating data preprocessing, feature engineering, K-Means clustering, and Business Intelligence visualization into a unified analytical workflow. The proposed framework successfully transformed raw customer and property transaction data into meaningful buyer segments that support data-driven decision-making. Through systematic preprocessing and feature engineering, customer demographic, financial, and behavioural information was converted into a structured analytical dataset suitable for machine learning.

The K-Means clustering algorithm effectively identified distinct buyer groups based on similarities in investment behaviour, property ownership patterns, and demographic characteristics. The optimal number of clusters was determined using the combined evaluation of the Elbow Method and Silhouette Score, ensuring that the generated segments were both statistically reliable and practically interpretable. These buyer segments provide valuable insights that enable real estate organizations to implement personalized marketing strategies, optimize financing solutions, improve customer engagement, and identify high-value investment opportunities.

To bridge the gap between analytical models and business applications, the clustered dataset was integrated into an interactive Business Intelligence dashboard developed using Streamlit and Plotly. The dashboard enables stakeholders to explore buyer segments through dynamic filters, key performance indicators, and interactive visualizations, thereby transforming complex analytical outputs into an accessible decision-support system.

Although the proposed framework demonstrates promising results, the study has certain limitations. The analysis was performed using a single real estate dataset, and the clustering model was limited to the K-Means algorithm. Furthermore, the framework primarily relies on structured demographic and transactional data, without incorporating external market factors such as economic indicators, property appreciation trends, or customer sentiment.

Future research can enhance the proposed framework by exploring advanced clustering techniques such as DBSCAN, Gaussian Mixture Models (GMM), or Hierarchical Clustering to compare segmentation performance. The integration of additional data sources, including real-time market trends, geographic information systems (GIS), social media sentiment, and customer feedback, may further improve segmentation accuracy. Moreover, incorporating predictive analytics and recommendation systems could enable real estate organizations to anticipate buyer preferences and deliver more personalized investment recommendations. These enhancements have the potential to transform the proposed framework into a comprehensive intelligent decision-support system for the modern real estate industry.

REFERENCES

[1] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.

[2] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.

[3] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY, USA: Springer, 2009.

[4] S. Lloyd, "Least Squares Quantization in PCM," IEEE Transactions on Information Theory, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[5] D. Arthur and S. Vassilvitskii, "K-Means++: The Advantages of Careful Seeding," in Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), New Orleans, LA, USA, 2007, pp. 1027–1035.

[6] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[7] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning with Applications in Python. Cham, Switzerland: Springer, 2023.

[8] Plotly Technologies Inc., "Plotly Python Graphing Library Documentation." Available: https://plotly.com/python/ (Accessed: Jul. 2026).

[9] Streamlit Inc., "Streamlit Documentation." Available: https://docs.streamlit.io/ (Accessed: Jul. 2026).

[10] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in

Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 1967, pp. 281–297.