Nepal Real Estate Market: Data Collection and Exploratory Analysis

Prafful Raj Thapa, Rojan Gautam

***Abstract*—This paper presents an Data Collection & Exploratory Data Analysis (EDA) on the Nepalese real estate market using property listing data collected from various online platforms. The dataset consists of 18 key features describing property characteristics such as location, price, area, type, number of rooms, age, and more. Through a series of descriptive statistics and visualizations, the analysis highlights important trends and relationships within the market, including the correlation between property price and area, the distribution of property types across different cities, and the impact of factors like property age and number of floors on pricing. The findings offer initial insights into Nepal's real estate market dynamics and serve as a foundation for more advanced market modeling and forecasting in the future.**

***Impact Statement*—This analysis contributes to the growing field of real estate data analytics by offering a detailed examination of property listings in Nepal. The findings can assist buyers, sellers, investors, and policymakers in making informed, data-driven decisions. Additionally, this project provides a foundation for predictive pricing models and spatial market analysis, highlighting how statistical insights can enhance the understanding of market trends, property valuations, and regional pricing dynamics. Ultimately, this work pushes the boundary of how data is used in real estate decision-making within emerging markets like Nepal.**

***Index Terms*—Exploratory Data Analysis (EDA)**

I. Introduction

THE real estate market plays a significant role in the economic development of any country, including Nepal. However, reliable and structured data on Nepal's real estate sector remains limited and fragmented. The lack of comprehensive datasets makes it difficult for stakeholders—such as investors, policymakers, realtors, and researchers—to make informed decisions or conduct meaningful analysis.

This project aims to address this gap by collecting and organizing real estate data specific to the Nepalese market. The dataset includes a variety of property attributes such as location, property type, size, price, number of rooms, and other relevant features. Data sources include online real estate listings and publicly available property databases.

Once the data collection phase was complete, the project carried out a simple Exploratory Data Analysis (EDA). This analysis serves as a preliminary step to identify trends, patterns, and correlations within the dataset. The EDA focuses on understanding property price distributions, size variations, location-based trends, and other key factors influencing the real estate market in Nepal.

By undertaking this project, we not only build a foundational dataset for future research but also provide basic analytical insights that can assist various stakeholders in understanding the current state of Nepal's real estate market.

II. Dataset Description

The dataset used in this project encompasses detailed statistics of real estate properties across Nepal. It comprises approximately 1,763 rows and 18 columns making it a valuable resource for analyzing the characteristics, trends, and pricing dynamics of the Nepalese real estate market. Each row in the dataset represents a single property listing, while the columns capture a wide range of features, including property type, location, area (in square feet), number of bedrooms and bathrooms, total floors, property age, price (in Nepalese Rupees), and other relevant attributes. The features includes:

1) **Property ID:** A unique identifier assigned to each property listing for reference and tracking.
2) **Property Name:** The title or name given to the property in the listing (e.g., "3BHK House in Baneshwor").
3) **Property Type:** Indicates the type of property, such as Residential, Semi-Residential, or Commercial.
4) **Bedrooms:** The total number of bedrooms available in the property.
5) **Living Rooms:** The number of living rooms in the property.
6) **Kitchen:** Total number of kitchens in the property.
7) **Bathrooms:** The number of bathrooms available in the property.
8) **Total Floors:** The total number of floors the building has.
9) **City:** The name of the city where the property is located (e.g., Kathmandu, Lalitpur, Pokhara).
10) **Area:** The sub-location or neighborhood within the city (e.g., Baneshwor, Gwarko, Jwalakhel).
11) **Property Face:** The direction the property faces, such as East, West, North, or South, often influenced by traditional and cultural preferences.
12) **Year Built:** The year when the property was constructed.
13) **Property Age:** The calculated age of the property in years, derived from the current year minus the year built.
14) **Property Area(sqft):** The total land or built-up area of the property in square feet.
15) **Price (Rs.):** The listed price of the property in Nepalese Rupees (NPR).
16) **Negotiable:** Indicates whether the listed price is negotiable or fixed, as mentioned in the original listing.
17) **Timestamp:** The date and time when the data was scraped or collected, used for data freshness tracking.

18) **Source System:** The website or online platform from which the property listing was scraped, used for reference and data source validation.

III. Methodology

The data utilized in this analysis was fully scraped from a popular Nepalese real estate trading platform. A structured dataset collection and Exploratory Data Analysis (EDA) pipeline was followed to ensure both data quality and analytical value.

1) **Platform Identification** The first step involved identifying popular online real estate trading platforms in Nepal. Various websites were explored to ensure they offered sufficient and relevant property listing data, such as property details, location, price, and other features. The selection was based on data availability, website accessibility, and relevance to the Nepalese property market. This step ensured that the final dataset would represent a broad and diverse range of property types and locations.
2) **Website Structure Analysis** After selecting the target platform, the next step was to analyze its website structure and underlying HTML layout. This involved inspecting page elements, class names, and data containers to understand how property information was organized on the site. This analysis helped in formulating an efficient scraping strategy, minimizing errors during data extraction. Understanding pagination, dynamic content loading, and other technical aspects was crucial for successful scraping.
3) **Selenium Experimentation** Before moving to large-scale scraping, small experimental runs were conducted using Selenium WebDriver. This phase involved writing sample Python scripts to automate browser actions like navigating through pages, clicking buttons, and extracting targeted data points. These tests helped identify potential challenges such as JavaScript-rendered content, dynamic loading, or anti-scraping measures. The experimentation allowed refinement of scraping techniques and ensured the robustness of the final scraper.
4) **Full-Scale Data Scraping** With a clear understanding of the site's structure and a working Selenium script, the project moved to full-scale data collection. The scraper was deployed to extract thousands of property listings while handling pagination and dynamically loaded content. The scraping process was monitored to avoid data duplication, minimize missing entries, and reduce server request overload to prevent being blocked. This phase resulted in a large, raw dataset ready for cleaning and analysis.
5) **Data Storage** The collected data was then stored in a CSV (Comma-Separated Values) file format, which is widely supported for data analysis tasks. This format allowed easy import into analytical tools like Python (Pandas), Excel, or R. Storing in CSV also made the data portable and easily shareable, facilitating quick exploratory analysis and visualization in later stages. Proper file encoding and field separation were maintained to avoid data corruption.
6) **EDA Feature Identification** Before conducting any analysis, the dataset was carefully reviewed to identify key features and variables that could provide meaningful insights. Features such as property type, location, area (in square feet), number of rooms, and price were marked as primary variables for analysis. Irrelevant columns, if any, were dropped to maintain dataset focus and reduce noise. This step was crucial to ensure that subsequent analysis was targeted and meaningful.
7) **Handling Missing Data** During initial inspection, the dataset contained missing or incomplete values for certain properties. Different strategies were employed to address this issue, including data imputation, removal of rows with excessive missingness, or using default values where appropriate. The goal was to preserve as much data as possible without compromising analytical accuracy. This step ensured the dataset was clean and reliable for analysis.
8) **Correlation Analysis** A correlation analysis was conducted among numeric variables in the dataset to explore relationships between different features. For example, the project analyzed how property area affects price, or how the number of floors or bedrooms correlates with property value. A correlation heatmap was generated to visually represent these relationships, helping to spot any strong positive or negative correlations. This step offered a quantitative understanding of inter-feature dependencies.
9) **Feature Summarization** After cleaning, summary statistics were generated for each key feature in the dataset. This included measures like mean, median, mode, minimum, maximum, and standard deviation for numerical columns such as area and price. Categorical variables like property type and city were also summarized through frequency counts and distributions. These summaries provided a quick overview of data ranges and distributions, setting the foundation for deeper analysis.
10) **Visualization and Pattern Recognition** Finally, various visualizations were created to identify trends, patterns, and outliers within the dataset. Tools like Matplotlib and Seaborn were used to generate scatter plots, bar charts, and correlation heatmaps. Visualizing the data allowed quick interpretation of property price trends across locations, property area distribution, and relationships between multiple variables. This step was essential for translating raw data into understandable insights and making the findings visually accessible.

This systematic approach ensured a thorough and insightful EDA, setting the stage for deeper statistical modeling and interpretation.

IV. Analysis & Result

The Exploratory Data Analysis (EDA) conducted on the collected Nepal real estate dataset revealed several key insights into property characteristics, pricing trends, and general distribution of properties.

A. Houses Distribution

An initial analysis of the dataset focused on the geographical distribution of property listings across different cities in Nepal. The dataset revealed a significant concentration of properties in a few major urban centers.

Approximately 80% of the total property listings were found to be from Kathmandu, Bhaktapur, Lalitpur, and Pokhara. Among these, Kathmandu accounted for the largest share, followed by Lalitpur and Bhaktapur, reflecting the higher level of real estate activity in the Kathmandu Valley.

Pokhara, being a popular tourist and residential destination, also had a noticeable number of listings.

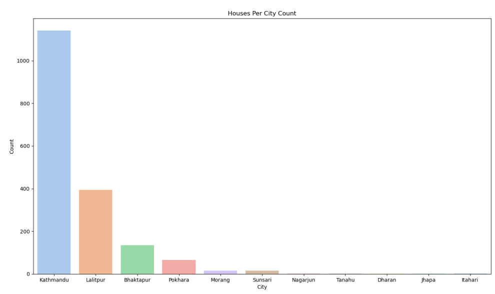

Fig. 1. Houses Distribution By City

The analysis of house distribution across different sub-locations (areas within cities) shows that a majority of property listings are concentrated in the well-known and commercially active neighborhoods of major cities like Kathmandu, Lalitpur, Bhaktapur, and Pokhara.

Areas such as Budanilkhanta, Kapan Area, Imadol, Kalanki, and Balkhu in Kathmandu, along with Tikathali, Jorpati and Satdobato in Lalitpur, and Thimi, Surbinayak and Balkot in Bhaktapur, appear frequently in the dataset. This indicates that real estate trading and new property developments are heavily focused within these popular urban and semi-urban localities.

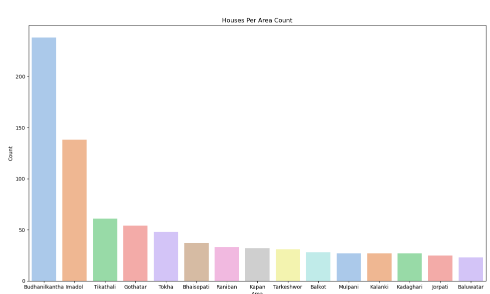

Fig. 2. Top 15 Houses Distribution By Area

B. Numeric Correlation Heatmap Analysis

Before diving into more specific visualizations, a numeric correlation heatmap was generated to gain an initial understanding of how different numeric features in the dataset relate to each other. This analysis helps identify relationships and patterns between variables, guiding further targeted visual exploration.

Key observations from the heatmap include:

1) A moderate positive correlation (0.62) between the number of kitchens and living rooms, suggesting that larger houses with more living space tend to include more kitchens—likely indicating multi-family or rental properties.
2) A correlation of 0.54 between bedrooms and bathrooms, which is expected, as houses with more bedrooms generally come with more bathrooms.
3) A weaker but noticeable correlation (0.40) between kitchens and bedrooms, again reflecting that bigger houses generally feature more rooms and facilities.
4) Interestingly, price and total floors showed only a low correlation (0.16), indicating that the number of floors alone is not a strong determinant of property price. This suggests that other factors, such as total property area, location, and overall amenities, play a more significant role in pricing.

Fig. 3. Numerical Correlation Heatmap

It is also important to note that certain expected correlations, like Total Floors vs. number of bedrooms, bathrooms, and living rooms, did not appear strongly in the heatmap. This is likely due to inconsistent or incomplete data entries, where many users input incorrect or missing values for floor counts and room counts.

C. Numerical Data Statistics Summary

To better understand the range, central tendency, and variability of key numerical features, a descriptive statistics table was generated for the dataset. Table I summarizes important statistical metrics such as count, mean, standard deviation, minimum, maximum, and quartile values for each numeric column.

TABLE I
NUMERICAL DATA STATISTICS

statistic	Bedrooms	Living Rooms	Kitchen	Bathrooms
Count	1,718	1,718	1,718	1,718
Mean	6.413853	1.877183	1.878347	4.182771
STD	2.961287	0.749718	0.813963	1.622607
Min	1	1	1	1
25%	5.0	1.0	1.0	3.0
50%	6.0	2.0	2.0	4.0
75%	7.0	2.0	2.0	5.0
Max	23.0	8.0	8.0	20.0
statistic	Total Floors	Property Area(sqft)	Price (Rs.)	Property Age
Count	1,718	1,718	1,718	1,718
Mean	2.821769	1587.450175	4.0203e7	6.82014
STD	0.677763	832.174771	3.0946e7	5.149564
Min	1.0	2.0	130,000	0.0
25%	2.5	1,026	2.4e7	2. 0
50%	2.5	1,369	3.3e7	6.0
75%	3.0	1,950	4.75e7	10.0
Max	5.0	5,000	3.2e8	45.0

Key Takeaways:

1) The average house has 6.41 bedrooms, with most properties ranging between 5 to 7 bedrooms. Outliers exist with some houses having up to 23 bedrooms, which likely represent commercial or multi-unit buildings.
2) Both average just under 2 per property, suggesting standard family-sized homes, though outliers with up to 8 kitchens and living rooms suggest possible apartment-style or multi-family dwellings.
3) Most properties have 2 to 3 floors, with the maximum recorded at 5 floors. The mean of 2.82 floors aligns with typical Nepalese urban housing structures.
4) The average property area is around 1,587 sqft, with 75% of properties below 1,950 sqft, indicating a dominance of moderately sized plots typical for city housing. The largest recorded property is 5,000 sqft.
5) The average property price stands at around 4 Crore NPR (40 million), with a wide range—from as low as Rs. 130,000 to as high as Rs. 32 Crore. This suggests a diverse market including both small rural listings and high-end city properties.
6) The average property is around 7 years old, with many properties between 2 to 10 years. The oldest property in the dataset is listed as 45 years old, indicating the presence of both new and older properties.

D. Distribution of Property Type & Property Face

An analysis of the property type distribution reveals that the Nepalese real estate market is heavily dominated by residential properties. Approximately 80% of the listings fall under the residential category, reflecting the high demand for family homes and private housing units.

The remaining 20% consists of semi-residential and commercial properties, such as mixed-use buildings, shops with attached living spaces, and purely commercial plots. This imbalance towards residential listings highlights that most property trading activity in Nepal is centered around private housing needs, rather than commercial development.

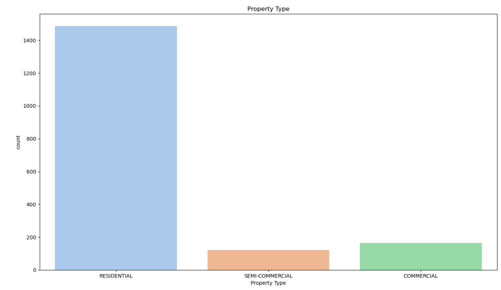

Fig. 4. Property Type Distribution

Another interesting aspect observed in the dataset is the preference for property orientation or ”face direction”. The analysis shows that a majority of the properties face east, which aligns with traditional Nepali beliefs that consider east-facing houses auspicious and favorable according to Vastu and Feng Shui principles.

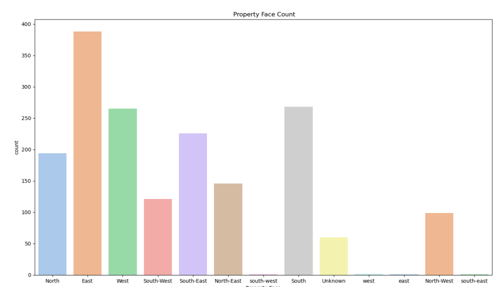

Fig. 5. Property Face Distribution

E. Distribution of Average Price By City & Area

1) Average Price by City: An analysis of the average property price distribution across different cities highlights a clear concentration of high-value real estate in a few urban centers. The cities with the highest average property prices include:

- Kathmandu
- Lalitpur
- Nagarjun
- Pokhara
- Bhaktapur

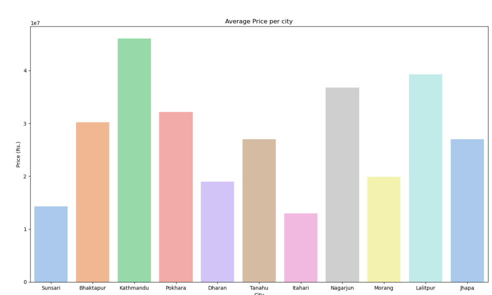

Fig. 6. Average Price By City

2) Average Price by Area (Sub-location): At a more granular level, the average property price by area (neighborhood/locality) shows that specific high-demand neighborhoods within these cities command significantly higher prices. Some of the areas with the highest average property prices include:

- Mahalaxmisthan
- Bansbari
- Dhobighat
- Jawalakhel
- Maharajgunj
- New Road

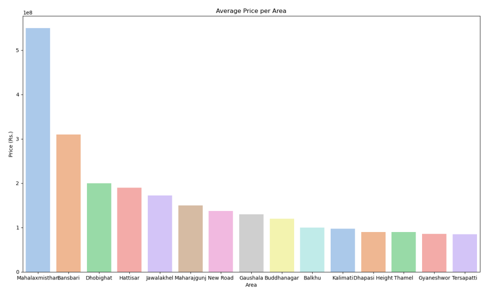

Fig. 7. Average Price By Area

These areas are known for their central location, better infrastructure, access to amenities, and overall desirability among buyers. The higher prices in these neighborhoods reflect both demand-driven market forces and their socio-economic value within the city.

F. Distribution By Property Age

A histogram analysis of property age reveals that most properties in the dataset fall within the 0 to 20 years age range. This indicates that Nepal's real estate market primarily consists of relatively new to moderately aged buildings, likely driven by ongoing urban development and new construction projects.

There is a sharp decline in the number of properties older than 20 years, and very few properties exceed 40 years of age. This scarcity of older properties may reflect market preferences for newer buildings, as well as redevelopment trends in urban areas where older structures are often replaced.

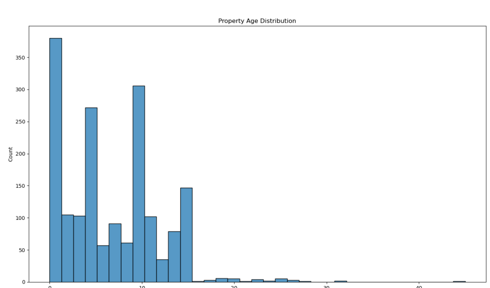

Fig. 8. Distribution Over Property Age

Further analysis of property price distribution across different property ages shows that price remains relatively stable regardless of the building's age. This supports the widely accepted market belief that land value plays a more dominant role in determining property prices than the age or condition of the built structure itself. In other words, location and land area continue to be the primary drivers of real estate value in Nepal, with building age having comparatively less impact.

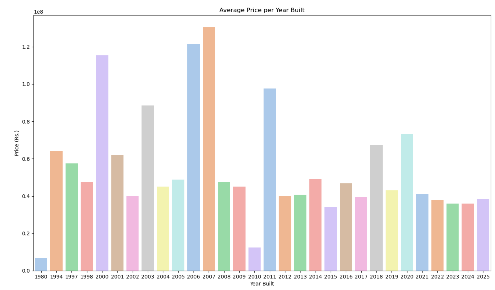

Fig. 9. Average Price By Property Age

G. Pair Plots For Each Property Type

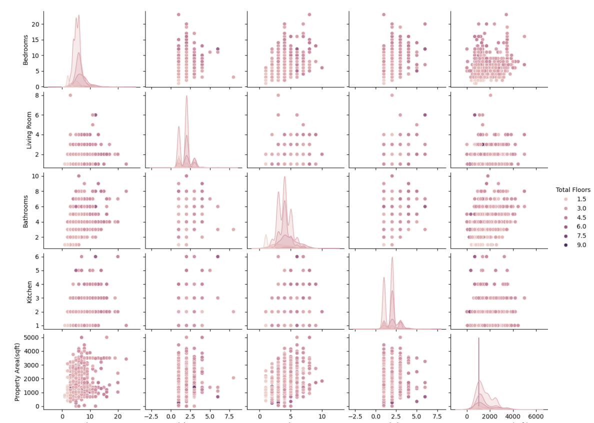

Fig. 10. Residential Property Pair Plot

1) For Residential Property:

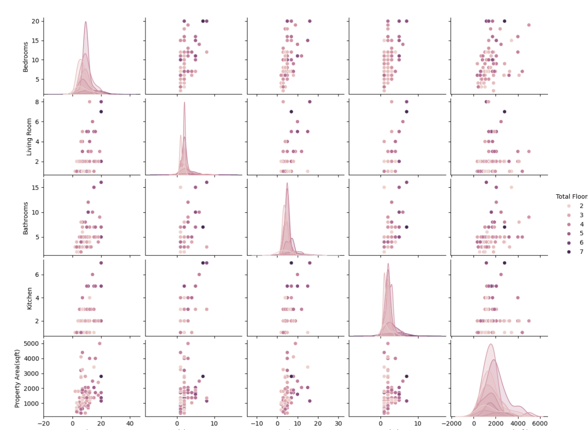

Fig. 11. Semi Commercial Property Pair Plot

2) For Semi Commercial Property:

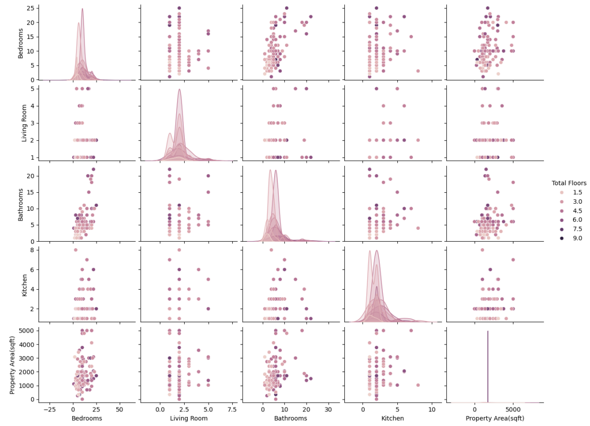

Fig. 12. Commercial Property Pair Plot

3) *For Commercial Property:*

V. Discussion

The exploratory analysis of the Nepal real estate dataset has provided valuable insights into the structure, trends, and characteristics of the current property market in Nepal. Several key patterns emerged from the data that reflect both market realities and cultural preferences in property buying and selling behavior.

Firstly, the geographical distribution of properties highlights a strong urban concentration, with Kathmandu, Lalitpur, Bhaktapur, and Pokhara accounting for the majority of listings. This reflects Nepal's urban-centric real estate development, where economic opportunities, better infrastructure, and population density drive both supply and demand.

The area-wise (sub-location) distribution further reinforced this trend, showing that well-known neighborhoods within these cities like Mahalaxmisthan, Bansbari, Dhobighat, and Jwalakhel command higher average property prices. This indicates that location continues to be a major factor in real estate valuation, as buyers prefer areas with good connectivity, facilities, and social status appeal.

The correlation analysis between numerical features revealed meaningful relationships between house features like kitchens, living rooms, bedrooms, and bathrooms, reflecting how property size scales with room count. However, total floors showed surprisingly low correlation with price, largely due to data inconsistencies and missing values in floor-related fields. This points to data quality issues often encountered when scraping from user-generated listings, where sellers might leave out or misreport property details.

The age-wise distribution of properties showed that most buildings are under 20 years old, with very few properties exceeding 40 years. Interestingly, the property price remained relatively consistent across different property ages, suggesting that buyers in Nepal prioritize land value over the age or condition of the existing building structure. This insight aligns with local market behavior where land is often seen as the primary long-term investment asset.

Finally, the overall numerical statistics and visualization patterns reinforce that the Nepalese real estate market is diverse but highly skewed towards urban residential properties, with significant price concentration in high-demand neighborhoods and cultural preferences (e.g., east-facing houses) also influencing property design and marketability.

Overall, this project not only sheds light on current real estate market dynamics in Nepal but also highlights the importance of structured data collection and cleaning when dealing with web-scraped datasets. The findings offer a solid foundation for future in-depth analyses, such as predictive modeling or more targeted market segmentation studies.

VI. Conclusion

This project successfully achieved its primary goal of collecting and organizing a structured real estate dataset focused on Nepal's property market. Through careful web scraping and exploratory data analysis, we were able to uncover key trends and patterns, such as the urban concentration of property listings, price variation across cities and areas, and the dominant role of land value over building age in determining property prices.

The analysis also highlighted data quality challenges common in web-scraped datasets, such as missing or inconsistent values, especially in fields like total floors and room counts. Despite these limitations, the project provides a valuable starting point for further research, market studies, and predictive modeling in Nepal's real estate sector.

Code: Link