Texila International Journal of Management
ISSN: 2520-310X
DOI: 10.21522/TIJMG.2015.11.02.Art011

Can Artificial Intelligence (AI) use Public Data to Predict Real Estate Price Patterns?

Umesh J Jadunandan
School of Business and Management, Texila American University, Guyana, South America

Abstract

Artificial Intelligence (AI) is evolving rapidly; however, the Real Estate Industry is slow to adopt AI Technologies. From the Real Estate point of view, Artificial Intelligence may be defined as utilization of computing methodologies upon real estate related data, providing insight into patters or trends of real estate sales prices beyond human cognitive capabilities. In this publication, the basics of Real Estate types, their corresponding recording processes, and data cleaning in preparation for AI modelling are discussed. Data from low-cost reliable source will enable results that will build confidence in utilization of emerging AI. The Dataset selected for this analysis is obtained from the public records of the Palm Beach County Property Appraiser's website. Palm Beach County is in the State of Florida, USA. Here, all real estate transactions are recorded with the Public Property Appraiser. This data is available to the public, thus providing a reliable low, or no cost dataset. A method to gather, clean, and process this data for analysis for an AI model is discuss.

Keywords: *AI Modelling, Data Acquisition, Property Data, Property Appraiser Data, Real Estate Prices Trends.*

Introduction

The typical Players participating in the transactions of real estate are Real Estate Agents/Brokers who facilitate transactions between buyers and sellers. Investors are Individuals or groups who buy properties to generate income or for appreciation. Developers are typically companies or can be individuals who build new properties or renovate existing ones. Finally, Property Managers handle the day-to-day operations of rental properties [22].

In all real estate transitions, the Market Dynamics, such as supply and demand and valuation play an important part. Market conditions fluctuate based on buyer/seller activity, economic conditions, and interest rates. Property values are also influenced by location, size, condition, and other market trends. Participating is real estate in not without its challenges. Some of the challenges include market volatility, economic downturns, and changes in regulations can pose risks. Investing in real estate can be a lucrative opportunity, but it also comes with risks and challenges. As AI technology continues to evolve, its role in real estate acquisition will expand further. AI-powered property models, with the aid of chatbots, will increasingly handle inquires, property searches, all with a personalized buyers' preferences [21]. This study examines the availability of public data and if that data can be obtained consistently and reliable for AI. A successful AI Model that can utilize this data will mitigate risks and challenges with acquisition Real Properties [9].

Methodology (Data Availability)

Palm Beach County, Florida USA is selected to obtain the real estate sales data, since this data is readily downloadable from the County's website. This is the largest County in the State of Florida. This County consists of over half million unique real estate parcels. The County is about 2.39 square miles with a

Received: 05.05.2025 ***Accepted: 26.05.2025*** ***Published on: 29.08.2025***
Corresponding Author: ujadu@yahoo.com

population of about 1.52 million people, calculated with a density of 773 people per square mile [17]. The County's Property Appraiser as of 2025 is Dorothy Jacks. Dorthy Jacks "is committed to having the best staff and systems in place to ensure a fair and equitable determination of property values for all Palm Beach County property Owners." [1]. From the Real Property Search tab, the owner's name, address, or a unique parcel control number (PCN) [1]. In addition, there is an Advance Sale Search option where one can search for residential, commercial, or vacant parcel types. In the methodology, only residential research is conducted. The countywide option is selected to have a large sample of data, since the more data is available for AI, the better the model prediction [5]. According to a report by VentureBeat, 80% of data scientist's time is spent collecting, cleaning, and organizing data instead of building models. This emphasis the importance the data phase is [13]. The sale type selected is to include all residential sales above $1000.00 USD. The important feature of the dataset is that each parcel is assigned a property or parcel control number. This unique number does not change and is the key identifier that will allow the model to track the historical sales of each parcel over time.

The research methodology followed the steps outlined below.

1. Obtain all PCN property identifier for the targeted geographic area
2. Obtain detail sales data per PCN identifier
3. Store data in a database
4. Build a script to process/clean data
5. Clean data, removing outliers
6. Prepare data for AI model training

Sales of real estate parcels are recorded and published continuously. In this study, the data is scheduled for download once per week. The data is processed to remove outliers, and each parcel is check for change in sale price with its corresponding its sale date. The percentage change in price is calculated over time and exported to a separate table for analysis. Since each parcel also showed the exact geographical location, the area with prices changes can be determined. Further development can integrate Graphical Information.

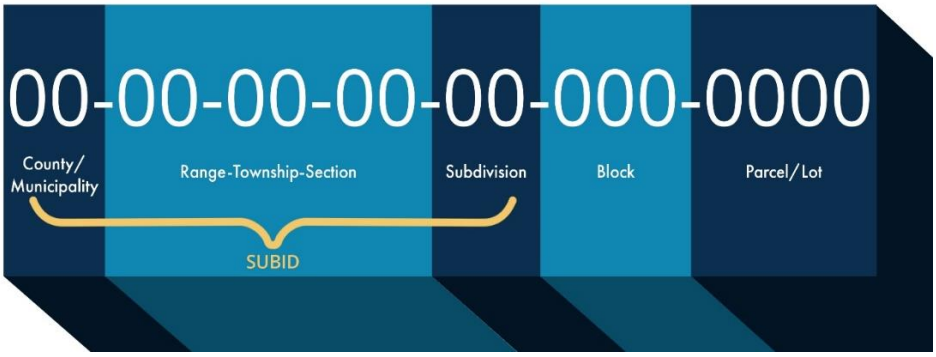

Figure 1. Parcel Control Number (PCN). Palm Beach County Property Appraiser [3]

Systems (GIS) to depict trends locally or Countywide. Figure 1 depict a pictorial representation of this 17-digit identifier. Each parcel of land and each living unit within a condominium or cooperative in Palm Beach County, Florida, USA is assigned a property or parcel control number, called a PCN. It is a unique 17-digit identifier that is used to locate and identify the parcel or unit. In Palm Beach County, Florida, USA are at least 600,000 unique PCN [1]. For this article, a sample set is used for demonstration.

The first two digits identify the county or municipality. If the parcel is in unincorporated Palm Beach County, these numbers will be 00, and if it if is in one of the municipalities, the number will reflect the municipality code (PDF). These two digits are also tied to the taxing district or millage code as verification that the ad valorem tax calculation is in sync.

The next six digits are the range, township, and section. The three groups of two digits reflect components of the Public Land Survey System and places the parcel within a specific square mile. The initial point in Florida is in the capitol, Tallahassee. The relationship between Tallahassee and Palm Beach Count is to the south and east, allowing us to drop the direction indicators on range (42E) and township (44S).

The following two digits denote the subdivision. The numbers are assigned based on the development of a subdivision or a condominium. The assigned number is tallied in chronological order. If a parcel is not in a subdivision, but remains un-platted acreage, the digits remain 00.

These first ten digits of the PCN are termed the SUBID and allow the selection of all the parcels in a development or un-platted parcels in a section.

The next three digits are loosely defined as the block. Definitions can vary. If blocks are identified in a subdivision and they will be part of the legal conveyance, they are noted in this field. In a multi-story condominium with one building, this field may reflect the floor. If a condominium has more than one building, the building number may be placed in this field. If blocks are not required, the number will remain 000.

The final four digits show the parcel or lot. The fourth digit is an assumed point number that is used to capture the genealogy of a parcel when it is split. The last digit or point number is reserved when a parcel is a portion of a full lot [1].

Data available from the Property Appraiser Public Access (PAPA) included the following [1]:

1. Property Characteristics
2. Location Information
3. Historical Sales Data
4. Taxable Value

Property Characteristics

The features included square footage, number of bedrooms, number of bathrooms, lot size, year built. It also contains zoning information to identify the property allowed use type. This zoning information will help separate the different classifications of property uses, such as commercial, industrial or residential. The Data available also allows detailing of each property by type, of construction materials, the cooling/heating mechanical types [18].

Location Information

This data shows the exact address of each parcel and includes geographical details such as parcel outlines, road names, public school zones, the property use and an aerial view with boundary lines. However, these data characteristics are not used to determine changes in sale prices.

Data Cleaning and Preprocessing

AI model requires clean and consistent dataset to identify patterns. Hence an important step in the process is gathering and cleaning the data. Therefore, the data is clean and preprocess to ensure it is accurate, consistent, and ready for analysis. This involves handling missing values, removing outliers, and transforming data into a suitable format for machine learning algorithms [4].

Feature Selection identify the features (variables) that are most relevant to the analysis. This involve selecting key factors such as property size, location, amenities, economic indicators, and others that are likely to impact real estate trends [8].

Real Properties are frequently transferred as the following:

1. Gifts.
2. Donation.
3. Family transfers.
4. Business exchanges.

These transactions are often recorded with lower than market value of the real property. To

this end, it is important to remove these outliers from the dataset.

Results

Data was collected using the advance option from the County's website. Information include the PCN, last sale price, sale date, owner 's name and property address. For this research, only the PCN, last sale price and date are the key selection most relevant. The data is collected and store in a local storage. Table 1 and Table 2 below depicted the raw data and cleaned data respectively.

Downloaded Sample

The data sample illustrated the PCN identifier, last sale price, the sale date, location and property size. However, some sales were recorded as $10.00 whilst others were zero dollars. These data need to be removed to ensure that the data represent the true market value of the real property sale. Any sale price below $1000.00 were removed.

Table 1. Raw Data with Outliers. Palm Beach County Property Appraiser [1]

Parcel Number	Sale Price	Sale Date	Sq. Ft
22-42-43-26-28-000-0260	0	6/18/2007	1534
22-42-43-26-24-000-0070	10	5/12/2015	1566
22-42-43-35-05-025-0030	825000	9/1/2022	1442
22-42-43-26-06-000-0090	10	3/9/2017	1953
22-42-43-35-15-000-0350	250000	11/3/2023	1740
22-42-43-36-21-000-0080	1	6/19/2015	2374
22-42-43-36-21-000-0080	0	6/21/2018	2374
22-42-43-25-01-000-0081	100	6/11/2024	1292
22-42-43-25-01-000-0162	10	3/7/2022	726
22-42-43-36-02-000-0141	10	5/23/2023	1986
22-42-43-25-00-000-7410	100	8/10/2019	1732
22-42-43-35-05-025-0080	84900	12/4/2015	1598
22-42-43-35-15-000-0060	10	11/20/2007	2306
22-42-43-35-22-000-0380	13100	9/5/2014	1337
22-42-43-36-00-000-3060	10	3/7/2022	2581
22-42-43-36-21-000-0060	176000	1/28/2016	2240
22-42-43-36-00-000-3130	10	8/14/2024	3743
22-42-43-36-21-000-0130	145900	8/13/2013	2224
22-42-43-35-01-008-0030	100	1/1/1980	1217
22-42-43-35-01-008-0030	10	1/24/2024	1217
22-42-43-35-22-000-0470	360000	9/23/2021	1292
22-42-43-35-24-001-0120	320000	2/17/2006	1952

Table 2. Example of Cleaned Dataset Palm Beach County Property Appraiser [1]

Parcel Number	Sale Price	Sale Date	Sq. Ft
22-42-43-36-21-000-0060	176000.00	1/28/2016	2240
22-42-43-36-21-000-0130	145900.00	8/13/2013	2224
22-42-43-35-22-000-0470	360000.00	9/23/2021	1292
22-42-43-35-24-001-0120	320000.00	2/17/2006	1952
22-42-43-35-22-000-0190	349900.00	9/2/2021	1349
22-42-43-25-00-000-7150	15200.00	1/1/1979	3005
22-42-43-36-21-000-0130	230000.00	4/7/2014	2224

22-42-43-26-28-000-0140	211000.00	6/23/2016	1534
22-42-43-36-27-000-0130	325000.00	8/5/2021	1789
22-42-43-36-27-000-0220	310000.00	6/9/2016	1789
22-42-43-26-06-000-0061	86500.00	8/1/1995	1238
22-42-43-36-28-000-0070	380000.00	4/10/2024	2526
22-42-43-35-19-004-0050	122000.00	12/1/1995	1836
22-42-43-35-15-000-0170	28300.00	2/10/2012	1056
22-42-43-25-00-000-7430	325000.00	8/2/2018	1277
22-42-43-26-24-000-0070	126000.00	4/15/2000	1566

Sampling Technique

The sampling technique is simple and straightforward, since all real estate sales prices below 1000.00 are removed. The downloaded data is copied into a local MYSQL database [3]. A table is created called Data_table and the column checked is called value. In any Machine Learning process, Data Preprocessing is the primary step wherein the raw/unclean data are transformed into cleaned data, so that in the later stage, machine learning algorithms can be applied. This python package makes the data preprocessing very easy in just 2 lines of code [2]. All you must do is input a raw data (CSV file), this library will clean your data and return you the cleaned data frame on which further you can apply feature engineering, feature selection and modelling.

A SQL scrip

DELETE FROM data_table

Where value < 1000;

The data is stored to local depository at a fix frequency. There, the SQL scripts are executed removing any outliers. Each PCNs is stored in a table, where each sale amount and date of sale is recorded. Each time a new dataset is uploaded, each PCN is checked for sale price change and its corresponding sale date is added. The percentage increase over time is calculated and outputted to a file for future AI Analysis.

Discussion

Real estate data collection involves gathering and organizing information about properties, markets, and transactions for various purposes like property valuation, market analysis, investment strategies, and tenant onboarding. This data can be collected through various sources, including public records, surveys, and property inspections [19]. The collected data is then used to analyse trends, assess property values, and make informed decisions within the real estate industry. Of these methods, the public records are readily available at most County's website. For this reason, the Palm Beach County, Florida USA is the largest County in the State of Florida was chosen as the preferred location to obtain data.

Other sources may involve additional steps to retrieved data are the Census Bureau, Local property listing sites, such as the Multiple Listing Service (MLS). Data as a cost and more fragmented are from surveyors of recent home sales, property characteristics. Buyer/seller interviews can provide valuable insights into market conditions. In addition, data collectors, often employed by appraisal management companies (AMCs), as they physically inspect properties, taking photos, measurements, and videos [20].

Some of the challenges in Data Collection can be attributed to incomplete or inaccurate public records, since public records may not always be complete or accurate, requiring data cleaning and validation [12]. Data privacy is also a concern. Ensuring data privacy and compliance with regulations is crucial when collecting and managing real estate data. Not to mentioned, technological limitations and lack of standardization. Technological advancements can impact data collection methods, requiring adaptation and innovation and standardizing data collection processes can help improve accuracy and consistency [16]. Perhaps, the most widely known AI-driven real estate tool is Zillow Zestimate, where the company

continuously refines its algorithm to predict a quick estimate on a property worth [14].

One of the benefits of using Palm Beach County Property Appraiser Website is that the data is not fragmented as typical of the nature of data. Sometimes data may be stored in various formats and locations, requiring integration and analysis.

Conclusion

The process demonstrated that low-cost reliable dataset is achievable for AI Modelling and analysis. Dataset is readily available with detailed real property sales information. The prediction of sale price trends in real estate using AI patterns recognition is an achievable objective [10]. One of the remaining challenges is automating the data collection procedure from the public source.

Key Takeaways:

1. Real estate data collection is a crucial process for informed decision-making in the industry.
2. Data collection can involve various methods, including public records, surveys, and property inspections.
3. Data collection is essential for property valuation, market analysis, and investment strategies.
4. Challenges in data collection include data accuracy, privacy, and technological limitations.
5. Data analytics can help leverage real estate data for more precise valuations, better investment decisions, and improved tenant onboarding.

With AI, a vast amount of data can be analysed continually, thereby providing insights not currently available.

Conflict of Interest

Conflict of Interest I, Umesh Jadunandan, declare that there is no conflict of interest in the data collection, literature review, writing of the script, and its publication by the Journal. However, in this discussion a brief review of where conflict of interest may arise with integration of AI into real estate is discussed below.

The integration of Artificial Intelligence (AI) into real estate presents exciting opportunities but also raises significant concerns about potential conflicts of interest. These conflicts can arise in various ways, potentially impacting buyers, sellers, and other stakeholders in the real estate market [6].

Some key areas where conflicts of interest may emerge in AI-driven real estate are with Automated Valuation Models (AVMs) and Appraisal Bias which use algorithms to estimate property values. This may be susceptible to biases present in the data they are trained on, leading to inaccurate or unfair valuations. For example, If an AVM is trained on historical data that reflects past discriminatory practices in lending or housing, it might perpetuate those biases in its valuations, unfairly impacting certain neighbourhoods or demographics [11]. Conflict of Interest may exist in Regulation also. To address this, regulatory bodies must work on implementing safeguards for AVMs to ensure accuracy, avoid conflicts of interest, and prevent discriminatory outcomes. Conflict of Interest may also arise when AI algorithms might prioritize properties that benefit the platform or its affiliated agents, rather than the properties that best match the buyer's needs and preferences. An AI platform might steer buyers towards properties that are listed by its affiliated agents or those that generate higher commissions, even if those properties are not the most suitable for the buyer [15].

A big concern also is the lack of transparency in AI algorithms. The "black box" nature of some AI algorithms can make it difficult to understand how they arrive at their recommendations, hindering accountability and potentially concealing biases or conflicts of interest. For example, if an AI-powered investment tool recommends certain properties, it may be unclear why those properties were selected and whether the selection process was

influenced by factors that benefit the platform or its developers, rather than the investor. AI systems often rely on vast amounts of data, raising concerns about data privacy and the potential for misuse of sensitive information about individuals and properties. If an AI tool collects and analyses confidential data about property owners, there's a risk that this information could be used for purposes that are not aligned with their interests, potentially leading to conflicts of interest or even discriminatory practices [7].

To mitigate these risks, it is crucial to:

1. Implement robust quality control standards for AVMs. This includes ensuring data accuracy, security, and preventing discriminatory impacts.
2. Promote transparency in AI algorithms. Make efforts to ensure that AI models are understandable and that their decision-making processes are auditable.
3. Develop ethical guidelines for AI in real estate. These guidelines should prioritize fairness, transparency, and accountability in the use of AI technologies.
4. Ensure data privacy and security. Establish strong data governance frameworks to protect sensitive information and prevent misuse.
5. Minimizing potential conflicts of interest and ensuring a fair and equitable market for all stakeholders is possible by addressing these concerns proactively.

Acknowledgements

The advancement of this study is a result of the encouragement, support and generosity of individuals and organizations to whom I must show unreserved appreciation. With all humility, I am thankful to Texila American University for providing me with a platform to study and conduct this research. Moreover, I am grateful to the employees at the County's Property Appraiser Office and my guide, who helped me throughout this work.

References

[1]. Jacks, D., 2023, "Get Sales Search." [Online]. Available: https://pbcpao.gov/AdvSearch/GetSalesSearch

[2]. Murali, K. M., 2025, "Machine Learning with Python." *GitHub*, [Online]. Available: https://github.com/muralimopidevi/Machine_Learning_With_Python

[3]. Abid, A. A., 2025, "AI-Powered Features for Developers and Data Pros." [Online]. Available: https://www.datacamp.com/tutorial/warp-terminal-tutorial

[4]. Alastal, A. I., & Shaqfa A. H., 2022, "*Journal of Data Analysis and Information Processing*." [Online]. Available: https://scirp.org/journal/paperinformation/paperid=116308

[5]. Mobile Application Development, 2024, FullEstop, "Top 10 AI Solutions for Real Estate Industry." [Online]. Available: https://www.fullestop.com/blog/top-10-ai-solutions-for-real-estate-industry

[6]. Gabriella, T., 2020, "Impacts and Application of AI in Real Estate." [Online]. Available: https://www.cbre.com/insights/articles/the-rise-of-the-machine-impacts-and-applications-of-ai-in-real-estate

[7]. Thomas, M., "The Future of AI: How Artificial Intelligence will change the world." [Online]. Available: https://builtin.com/artificial-intelligence/artificial-intelligence-future

[8]. Rai, A., 2023, Coherent Market Insights, "Geospatial analysis market analysis." [Online]. Available: https://www.coherentmarketinsights.com/market-insight/geospatial-analytics-market-5874

[9]. Greiner, B., 2023, "11 Ways Artificial Intelligence Will Disrupt Real Estate Too." [Online]. Available: https://openaire.co/11-ways-artificial-intelligence-will-disrupt-real-estate-too/

[10]. Greengard, S., Week, 2019, "Best predictive analytics solutions 2019." [Online]. Available:

https://www.eweek.com/big-data-and-analytics/top-predictive-analytics-vendors/
[11]. Thompson, H., 2021, Forbes, "The Rise of Location in Advanced Analytics." [Online]. Available: https://www.forbes.com/sites/esri/2021/01/22/the-new-analyst-the-rise-of-location-in-advanced-analytics/
[12]. Belyh, A., 2025, "14 Best Project Management Software for Real Estate in 2025." [Online]. Available: https://www.founderjar.com/real-estate-project-management/
[13]. Bhagyashree, 2025, "How to Source, Prepare, and Optimize Data for AI Models." [Online]. Available: https://www.promptcloud.com/blog/ai-training-data/
[14]. McFarland, A., 2025, "10 Best AI Real Estate Tools." [Online]. Available: https://www.unite.ai/best-ai-real-estate-tools/
[15]. Hobart, L. A., 2021, IAAO. "Glossary for Property Appraisal and Assessment, 3rd edition." [Online]. Available: https://www.iaao.org/wp-content/uploads/IAAO-Glossary_3rd-Ed_final.pdf
[16]. Asaftei, G., & Doshi, S., & Means, J., & Sanghvi., 2016, "How big data is transforming real estate." McKinsey & Company, [Online]. Available: https://www.mckinsey.com/industries/real-estate/our-insights/getting-ahead-of-the-market-how-big-data-is-transforming-real-estate
[17]. Wikipedia, "Palm Beach County, Florida", 2025, [Online]. Available: https://en.wikipedia.org/wiki/Palm_Beach_County,_Florida
[18]. Jacks D., 2023, Palm Beach County Property Appraiser, [Online]. Available: https://pbcpao.gov/index.htm
[19]. Weir, D., 2023, "Property Data Collection: Everything You Need to Know." [Online]. Available: https://www.mckissock.com/blog/real-estate/property-data-collection-everything-you-need-to-know/
[20]. Joshi, J., 2025, "6 Effective Data Collection Tips for Better Real Estate Marketplaces." [Online]. Available: https://www.habiledata.com/blog/data-collection-tips-for-real-estate-marketplaces/
[21]. Castle, 111, G. H., & Hemmady, S., 2024, "Artificial Intelligence, Applications Throughout the Real Estate Industry." *Consolidated Predictions*, Pages 354-355, [Online]. Available: http://www.realestatebook.ai
[22]. Florida Statutes, Department of Business & Professional Regulation, (DBPR), 2022, "*Florida Real Estate Law Book,"* [Online]. Available: https://www2.myfloridalicense.com/servop/testing/documents/FREC_printable_LawBook.pdf