Interactive Real Estate Profiler

CSE 6242 Spring 2023 **Team 32:** Bishal Regmi, Eduardo Quiros Campos, Sushma Ghimire, Sulekha Singh, Robert Testerman, Kyle Kemper

March 31, 2023

Introduction/Motivation

Our tool is designed to address the needs of novice investors who are interested in entering the real estate market but may lack the expertise or resources to evaluate the investment potential of different properties. By providing accurate and reliable estimates of expected price growth, rental profits, and return on investment, our solution empowers users to make informed decisions and navigate the complex real estate market with confidence.

Problem Definition

Real estate prices and investment opportunities have been a popular topic among investors, particularly during the COVID-19 pandemic [2] [16]. With the market experiencing significant growth, many people have adopted the "invest over rent" strategy to take advantage of the "meaty" returns offered by real estate investment [12]. However, investing in real estate can be complex and challenging, with several important factors to consider, such as location, property condition, financing options, and potential returns on investment. We have developed a tool that enables users to conduct a preliminary market analysis and visualize the estimated return on investment for potential properties.

Literature Survey

Real estate pricing and investment are highly complex and dynamic processes that are influenced by a wide range of factors, such as market trends, local amenities, property characteristics, economic conditions, and regulatory changes [6][1]. Several techniques have been proposed in the literature to model this complex system to predict property prices, including multiple linear regression [19], machine learning [11][5], spatiotemporal models [17], and others[9]. A recent paper by Wei. et al. emphasizes the importance of utilizing a hedonic pricing model for real estate appraisal and building a multi-model to avoid shortcomings of a single model [18]. However, the complexity of the real estate market and the vast amount of data available can make it challenging to extract meaningful insights[3][10]. To address this, interactive visualization techniques have been used to represent real estate data and provide a more intuitive and accessible way of exploring patterns and trends in the market [1][13].

Our tool serves as a pre-screening mechanism to assist users in navigating the market and filtering through available properties. By providing users with relevant market data and estimated financial metrics, we aim to help individuals make more informed decisions about potential property investments. Ultimately, our tool will help users avoid future frustrations, such as failed deals or unaffordable properties, by providing a realistic overview of the investment potential of their desired property. Our model can help investors with longer holding periods make informed decisions by providing accurate predictions of the sale vs. rent value of a property based on various attributes. The impact can be measured through user studies and experiments, and by comparing the model's predictions with ground truth data.

The pre-screening tool for novice real estate investors offers potential benefits in the form of providing valuable insights for making informed decisions, but there are inherent risks associated with investing in real estate, including market volatility[15], changes in interest rates, and unexpected expenses such as repairs or maintenance[8]. Investment in real estate inherently comes with risk to investors[7][4]. To mitigate these risks, the tool must provide accurate and easy-to-understand information[6], while also emphasizing the importance of conducting due diligence. Ultimately, the tool has the potential to empower novice investors to enter the real estate market with confidence.

Proposed Method

Why should it be better than the state of the art? / Innovations

Despite the availability of numerous tools for calculating returns, rental costs, and market predictions in real estate investment, there is currently a lack of an integrated tool that can synthesize these metrics and visualize them for the average consumer, particularly in the US consumer market. Although there may be proprietary tools for professional real estate investors, there is still a need for a tool that can cater to the needs of the average individual who is interested in purchasing a property for potential rental income.

We have leveraged cutting-edge algorithms from the literature to accurately predict the average price of a desired house type, rental profits, and expected price growth. In addition, we aim to enable users to easily compare the predictions with average stock market returns, and present all this information in a user-friendly visualization, providing users with unique and comprehensive housing market data and estimates[6][13]. We deploy this tool on the cloud to ensure it is readily available and free to use for anyone interested in investing in real estate. With an interactive dashboard with essential market information, our tool aims to help simplify the complex process of real estate investment for novice investors. Our commitment to maximizing user experience and accurate results make us confident in the success of our tool in empowering users to confidently enter the housing market.

We have leveraged open-source languages and modules, such as Python 3, to eliminate licensing costs for our tool. Our solution will use official Docker containers and popular Python-based web frameworks like FastAPI or Flask, all of which are also free. Scaling costs would only be incurred based on the number of users and can be minimized by utilizing cloud services like Amazon Web Services (AWS) Elastic Container Service (ECS) or EC2 with EMR Serverless. Our solution is designed to be scalable and cost-effective, using industry-standard practices to ensure optimal performance.

Detailed description of your approaches: algorithms, user interfaces, etc.

Data Collection: We are using Apache NiFi to support the data collection process. NiFi helps us automate the ingestion, processing and landing of the datasets for model training and UI development in AWS S3 buckets. Our training datasets are being pulled from Zillow, Redfin and the US Census Bureau. Zillow offers public access datasets with housing data such as the ZHVI (Zillow Home Value Index) which measures typical home values and market changes broken down by ZIP code. Zillow also provides for-sale home inventory data collected and segregated at the ZIP level as well. Redfin, a national real estate brokerage firm, provides housing market data which is available for download at the Redfin web site. We are also pulling data related to demographic variables and additional housing information for the ML models from the US Census Bureau.

API: We will be utilizing a docker micro-service architecture for our project. Currently, our front-end application is designed in one docker container and we will connect to a second container that houses our modeling solution. These two containers will communicate by utilizing a REST API created from FastAPI and use Uvicorn as the the ASGI (Asynchronous Server Gateway Interface). A POST command will be sent from the front-end to the model container with the associated information from the user and the return object will be the associated data. Additional calls or data modeling needed will be completed in a similar fashion.

Model Development: The dataset contains historical data on house prices, as well as property features such as the number of bedrooms, bathrooms, presence of a basement, interior area, lot size, and more. Similar data is available for rental properties. The general steps involved in building the ML model include data pre-processing, feature extraction, and feature selection.

We will experiment with random forest regression and autoregressive integrated moving average (ARIMA) to build the models. We will evaluate the model's performance through historical testing and validation, using metrics such as mean squared error and R-squared. We will follow a similar approach to build a separate model for rental property price forecasting. As new data becomes available, the model can be updated and retrained to improve accuracy and reliability.

Front-End Interface: We are developing a web-based front-end using React.JS and D3.JS. The web page will consist of two sections/tabs. The first section will feature an interactive map of the tri-state area color-coded by median home prices, displaying essential housing statistics when hovering over a zip code. The second section will be dedicated to housing predictions. Users will be able to select their desired features for a potential property, and the platform will make an API call to provide predictions for both housing and rental values. Additionally, the platform will display two line graphs, one showing the housing value appreciation compared to S&P 500 and national housing averages for 5, 10, and 30 years, and the other graph depicting a line graph of return on investment from renting, compared to S&P 500 returns.

Experiments/Evaluation

A number of experiments and evaluations can be performed to determine if the tool and model are working properly and delivery accurate results. The tool will be populated with data from New York City as the test case. A choropleth map will be created in the user interface. Summary data can be evaluated to ensure it is populating accurately. A sensitivity analysis will be conducted on different user inputs to make sure calculations and models are working as expected (e.g. if interest rate assumption is decreased by a certain percentage, does the financial analysis have the expected response). In this manner all user inputs will have a sample calculation performed by hand to validate accuracy. The tool, when fully implemented, should allow the user to answer the following questions:

- What is my expected return on investment in this sample area?
- How do different attributes of the property impact this expected return?
- How does this expected return compare to alternative investment options?
- What assumption(s) present the most risk to the investment (evaluate by user manipulating assumed values)?

In addition to guiding the user through these questions, more qualitative analysis can be performed on the tool. This will include time based analysis to ensure the tool is responsive enough when adjusting parameters or scaling. For example, how long does it take to update expected return figures when a user adjusts a property parameter? An important part of our tool is the machine learning model discussed above. It will be essential to validate the accuracy of this model. We can compare the machine learning real estate appreciation model to other real estate models available in the literature and on the web. We will perform a brief analysis to benchmark our model versus these others.

Finally, a usability evaluation will be performed by conducting a user study. We will select several users (coworkers, classmates, friends, family) and have them interact with our tool. Our tool is designed to provide insights to novices and experts alike, so this user group will contain a mix of people both knowledgeable and new to real estate investment. We will note their feedback on the user interface, and test their ability to utilize our tools to accurately answer the questions laid out above.

Plan of Activities The Gantt chart below shows the roles, responsibilities, and key checkpoints for the project. The primary start and end dates along with roles for each task is shown. This is the same as the initial plan of activities proposed, with some modifications to dates based on updated estimates from work performed to date. Most of the date ranges have been expanded, as the process has been iterative and developing the tool has caused the need to return to earlier steps to update our approach.

As of today, we have completed the framework for the project, and we are now focusing on developing the first page. The first page is about showing graphical representation in the map form with tooltips information about mean housing price, average rental value, neighborhood walk score, transit score, nearby school information. Users will also have the option to set the priority of various factors to highlight the areas favoring the user inputs. This feature is designed to help users easily identify neighborhoods that match their preferences. Next, the user will have various factors like sqft, no of rooms, type of residence, built year range, neighborhood score range to select. Based on those selections, users will be directed to the second page, where a comparison graph of predicted house price, rental price, and SP 500 investment will be shown for the next 10 years. We are still on track to finish the tool on schedule and team members are delivering their original scope.

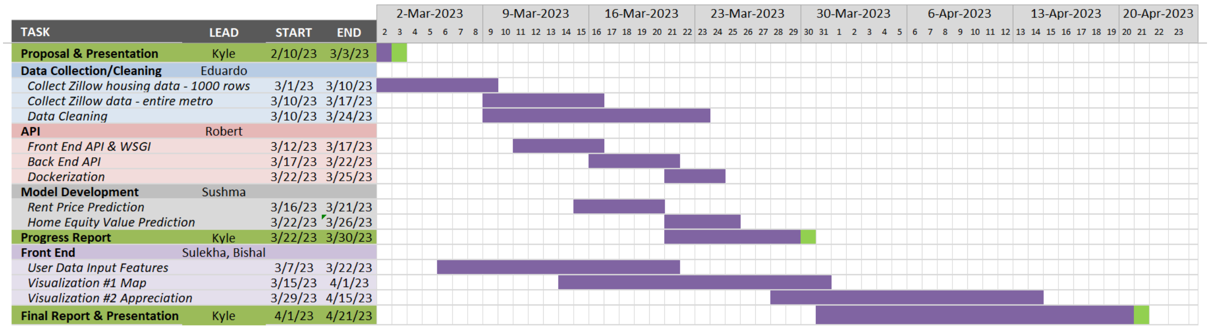

Figure 1: Original Gantt chart demonstrating checkpoints and plan of activities with roles

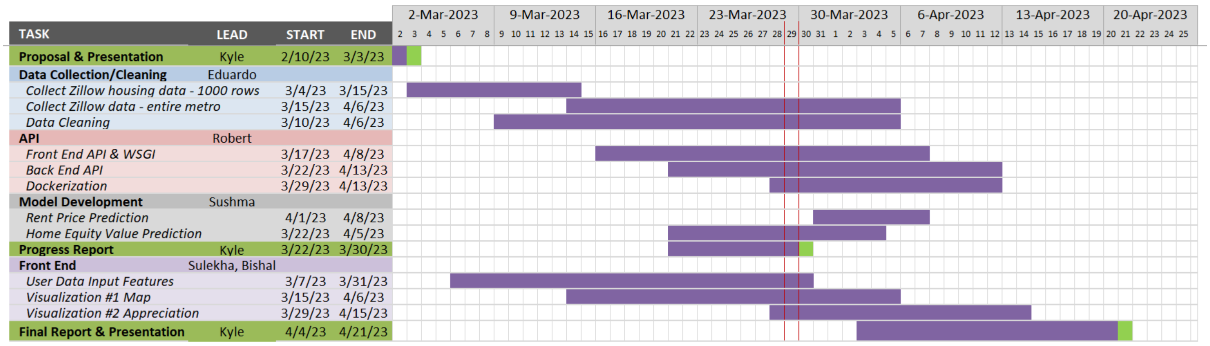

Figure 2: Revised Gantt chart demonstrating checkpoints and current plan of activities with roles

Distribution of team effort: All members have contributed a similar amount of effort

Conclusion and Discussion (Future work.)

We expect to be able to conclude meaningful insights about the expected return on real-estate investment for the region(s) analyzed in our demonstration. These results will be indexed versus alternative investment options. This tool uses advanced data analytics on a large real-world data set and presents results in a simple, interactive user interface to allow potential investors to make informed decisions. The final conclusion will be a recommendation to the user based on current market conditions.

References

[1] S. Agarwal, Y. Fan, D. P. McMillen, and T. F. Sing, "Tracking the pulse of a city—3d real estate price heat maps," *Journal of Regional Science*, vol. 61, no. 3, pp. 543–569, 2021.

[2] N. Balemi, R. Füss, and A. Weigand, "Covid-19's impact on real estate markets: review and outlook," *Journal of Property Investment & Finance*, vol. 35, no. 4, pp. 495–513, 2021. [Online]. Available: https://doi.org/10.1007/s11408-021-00384-6

[3] P. Bayer, C. Geissler, K. Mangum, and J. W. Roberts, "Speculators and Middlemen: The Strategy and Performance of Investors in the Housing Market," *The Review of Financial Studies*, vol. 33, no. 11, pp. 5212–5247, 04 2020. [Online]. Available: https://doi.org/10.1093/rfs/hhaa042

[4] E. Beracha, Z. Feng, and W. Hardin III, "Reit operational efficiency: Performance, risk, and return," *The Journal of Real Estate Finance and Economics*, vol. 58, pp. 1–30, 04 2019. [Online]. Available: 10.1007/s11146-018-9655-2

[5] F. D. Calainho, A. M. van de Minne, and M. K. Francke, "A Machine Learning Approach to Price Indices: Applications in Commercial Real Estate," *The Journal of Real Estate Finance and Economics*, 03 2022. [Online]. Available: https://doi.org/10.1007/s11146-022-09893-1

[6] R. Cellmer and R. Trojanek, "Towards increasing residential market transparency: Mapping local housing prices and dynamics," *SPRS international journal of geo-information*, vol. 9, p. 2, 2019. [Online]. Available: https://doi.org/10.3390/ijgi9010002

[7] D. Chambers, C. Spaenjers, and E. Steiner, "The Rate of Return on Real Estate: Long-Run Micro-Level Evidence," *The Review of Financial Studies*, vol. 34, no. 8, pp. 3572–3607, 03 2021. [Online]. Available: https://doi.org/10.1093/rfs/hhab028

[8] J. Clayton, D. C. Ling, and A. Naranjo, "Commercial real estate valuation: Fundamentals versus investor sentiment," *The Journal of Real Estate Finance and Economics*, vol. 38, pp. 5–37, 2009. [Online]. Available: https://doi.org/10.1007/s11146-008-9130-6

[9] J. Contat and W. D. Larson, "A Flexible Method of House Price Index Construction using Repeat-Sales Aggregates," *FHFA Staff Working Paper Series*, vol. 21-01, 2022. [Online]. Available: https://www.fhfa.gov/PolicyProgramsResearch/Research/PaperDocuments/wp2101.pdf

[10] M. Davis, A. Lehnert, and R. Martin, "The rent-price ratio for the aggregate stock of owner-occupied housing," *Review of Income and Wealth*, vol. 54, pp. 279–284, 02 2008. [Online]. Available: 10.1111/j.1475-4991.2008.00274.x

[11] Y. Kang, F. Zhang, W. Peng, S. Gao, J. Rao, F. Duarte, and C. Ratti, "Understanding house price appreciation using multi-source big geo-data and machine learning," *Land Use Policy*, vol. 111, p. 104919, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0264837719316746

[12] S. Lee and S. Stevenson, "Real estate in the mixed-asset portfolio: the question of consistency," *Journal of Property Investment & Finance*, vol. 24, no. 2, pp. 123–125, 2006. [Online]. Available: https://doi.org/10.1108/14635780610655085

[13] M. Li, Z. Bao, T. Sellis, S. Yan, and R. Zhang, "Homeseeker: A visual analytics system of real estate data," *Journal of Visual Languages & Computing*, vol. 45, pp. 1–16, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1045926X17301246

[14] J. S. Sagi, “Asset-Level Risk and Return in Real Estate Investments,” *The Review of Financial Studies*, vol. 34, no. 8, pp. 3647–3694, 10 2020. [Online]. Available: https://doi.org/10.1093/rfs/hhaa122

[15] R. J. Shiller, “Understanding recent trends in house prices and home ownership,” *Cowles Foundation Discussion Paper No. 1630, Yale Economics Department Working Paper No. 28*, 2007. [Online]. Available: https://ssrn.com/abstract=1017546

[16] C. W. Starr, J. Saginor, and E. Worzala, “The rise of proptech: emerging industrial technologies and their impact on real estate,” *Journal of Property Investment & Finance*, vol. 39, no. 2, pp. 157–169, 2021. [Online]. Available: https://doi.org/10.1108/JPIF-08-2020-0090

[17] L. Wang, G. Wang, H. Yu, and F. Wang, “Prediction and analysis of residential house price using a flexible spatiotemporal model,” *Journal of Applied Economics*, vol. 25, no. 1, pp. 503–522, 2022. [Online]. Available: https://doi.org/10.1080/15140326.2022.2045466

[18] C. Wei, M. Fu, L. Wang, H. Yang, F. Tang, and Y. Xiong, “The research development of hedonic price model-based real estate appraisal in the era of big data,” *Land*, vol. 11, no. 3, p. 334, 2022.

[19] Q. Zhang, “Housing Price Prediction Based on Multiple Linear Regression,” *Scientific Programming*, 2021. [Online]. Available: https://doi.org/10.1155/2021/7678931