LOCAL HOUSE PRICE PREDICTION AND VISUALIZATION

CSE6242 Final Report - Team No. 168

Wentao DUAN, Jingying GUAN, Li LI, Liuhui ZHAO, and Zehua ZHENG

1 Introduction

Housings are essential parts of human society: they serve basic needs such as shelter, and are significant assets for ownership and investment[1]. With more than 5 million houses sold in the US every year [2], an accurate prediction of house prices is of enormous importance and interests to almost everyone: everyday home buyers and sellers, real estate investors and agents, as well as policymakers, etc.

House prices are governed by numerous factors and sensitive to spatial variations and temporal disturbances. It is reasonable to expect that the house pricing attributes exhibit spatiotemporal heterogeneity within large housing markets due to localized supply and demand imbalances. Especially within a metropolitan area such as Denver, U.S., the supply of specific housing characteristics often exhibits strong patterns due to various needs of potential buyers, and people may hold different evaluations on house features.

Existing works on house price predictions generally fail to combine both spatial and temporal factors. They also lack visualization efforts on these factors, and hence make it hard for the general public to access their results. To bridge these gaps, a comprehensive and interactive tool that combines house price prediction and visualization would be a novel and effective approach to assist decision making in the housing market.

2 Problem Definition

In this project, we studied single family house market in the city of Denver, and explored spatial variation and temporal evolution of housing market characteristics. To tackle the problem, we divide it into the following three parts:

1. Develop visualization tools to demonstrate spatial and temporal distributions of key factors associated with house prices and housing market.
2. Build and compare machine learning models on house prices predictions and incorporate spatial and temporal factors.
3. Combine visualization and the chosen prediction model to develop the comprehensive and interactive tool mentioned above in the introduction.

Our study can be easily extended to other cities once historical data for local housing prices become available.

3 Survey

Conventional local house price prediction models are usually hedonic based, taking into account characteristics of the property and its surrounding environment [3]. Such hedonic approaches typically evaluate contributions of each individual feature to the total value of the property [4], but fail to incorporate spatial dependency in pricing of housing attributes. Spatial dependency/heterogeneity can arise from localized supply and demand imbalance, and geostatistical approaches (e.g., Kriging and co-Kriging, Geographically Weighted Regression (GWR)), or more often, a combination of hedonic and geostatistical models have included them in studying local house prices [1, 5–8]. Rather than a single global model on the entire housing market, geostatistical approaches fit separate local models for each sale point and weight observations by their distance to this point, thus allowing unique marginal-price estimates at each location [5].

Acknowledging that effects of house price predictors may change significantly across locations, and additionally, over time, several studies examined corresponding house price variations [9–13]. Inspired by Dynamic Model Averaging (DMA) and Dynamic Model Selection (DMS) models [14, 15], Bork and Møller investigated state-level house price forecastability across the US using these techniques, which allow for model adaptions over time, as well as the shift of parameters over space in each model. Forecast performance of the model is improved substantially, especially for states with highly volatile housing markets. Similarly, another study[12] analyzed and confirmed spatial and temporal dependence of house prices in the Dutch housing market, through error analysis in hedonic models using a weight matrix. Interestingly, spatial dependence was found to be dominant over temporal one.

Beyond these traditional regression-based methods, new machine learning algorithms have emerged in the field of house price estimation [16–24]. To forecast the average selling price in Chongqing, China, Wang et al. tuned the Support Vector Machine (SVM) model with particle swarm optimization method using historical house price data between 2000 and 2006, a higher forecasting accuracy than Back Propagation (BP) neural network algorithm [25]. Incorporating geographical factors into the house price prediction model, Limsombunc et al. applied Artificial Neural Network (ANN) model and compared its performance with a baseline hedonic model. The results showed that ANN provided better prediction performance for both in- and out-of samples [16]. However, in their work, some essential factors, e.g., economic factors and time dependent attributes, were not included in the model development. In a similar study conducted by Khamis and Kamarudin in New York City, the applied ANN model was found to yield a better performance, boosting r-square value by 26% compared with the applied Multiple Linear Regression (MLR) model [18]. To address model and parameter uncertainties that may not be captured by standard ANN or MLR models, Amri and Tularam applied a fuzzy-logic embedded neural network framework, i.e., Adaptive Neuro-Fuzzy System (ANFIS), to model house prices, where spatial attributes were represented in qualitative terms to describe their locational accessibility [26]. Authors concluded that although ANN and ANFIS do provide better performances in some cases, MLR is still considerably powerful when performing predictions.

Most often in existing work, spatial and temporal dependence of house prices have been treated as nuisance features and error terms, whereas their interaction with other features is neglected. Although more advanced techniques such as ANN or SVM were applied in previous studies, most of them only considered spatial heterogeneity without taking into account temporal impact on house prices. Moreover, the lack of visualization integrating spatial and temporal elements leads to lower accessibility of developed models to users in the market, which hinders their accessibility to the general public and applicability to specific local regions.

4 Proposed Method

4.1 Intuition

Our overall process and website structure are summarized in Fig.1. Our methodology starts with data processing (in Section 4.1.1), followed by exploratory data analysis (in Section 4.1.2), modeling (in Section 4.1.3), interactive platform building and publishing on Heroku (in Section 4.1.4), ended with user survey as evaluations (in Section 4.1.5). Initial data processing was done in QGIS, and then data analytics and visualization are coded in Python and incorporated with necessary modules.

Our key innovations and contributions are listed as follows:

1. To analyze contributing factors of single family house prices, we have collected related datasets (e.g., social-, economic-, geographical- data in the city and county of Denver) and conducted spatial joins to integrate and clean the data.
2. We applied the XGBoost model and identified key attributes for the single family prediction. Unlike previous studies, we combined spatial and temporal factors in the regression model, including neighborhood specific attributes (e.g., tree coverage, distance to food stores, number of traffic accidents), property specific attributes (e.g., area, lot size, bedrooms), and temporal specific attributes (e.g., consumer price index).
3. We developed a decision support platform for house hunting, which not only hosts historical single family house record and prediction, but also integrates city level and neighborhood level analytics of housing market - a key feature that is missing on most of the platforms.

In general, we offer a comprehensive and interactive platform that offers intuitive information and convenient access to the general public, bridging gaps that exist in the state of the art house price prediction studies and house hunting websites.

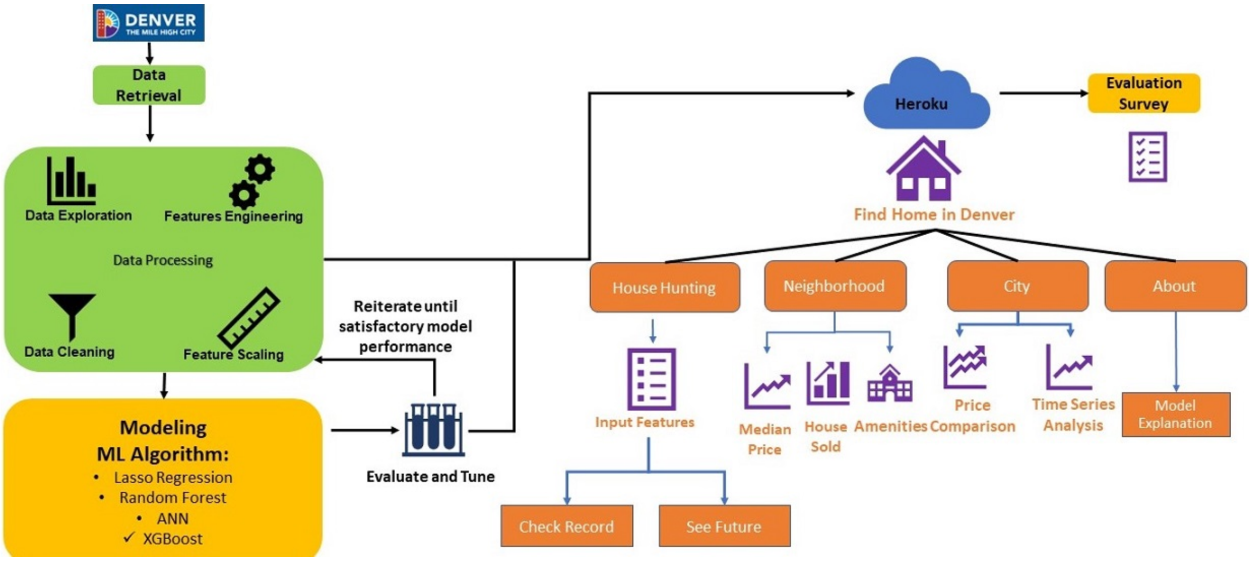


Figure 1: Methodology flowchart including website structure

4.1.1 Data Processing

We take the city and county of Denver as our case study in our project. The essential data set is the historical house selling price associated with the property details, which is collected from Denver open data website [27]. This data set contains over 200,000 house sales records dated from 1945 to 2020. Considering other related data sets, such as geographical data collected from Census Bureau [28] and housing market data from Redfin [29], we selected house selling records for single family type sold between 2000 and 2020. We also included in our datasets all the relevant neighborhood attributes, such as social-economic features, geographical features, traffic conditions, etc. Data processing steps are listed below.

1. *Collect data from Denver open-data website*: house sales record, neighborhood information, etc, and related data from Census Bureau and Redfin. Our raw data files are approximately 2 GB in size.
2. *Assign external keys for each data set for processing*: spatial join is applied based on spatial relationship among different datasets.
3. *Join data and remove records with null values*: assign neighborhood features into each house record.
4. *Remove irrelevant features*: any descriptive attributes such as house address, owner information, etc. are removed before modeling.
5. *Check data distribution and apply transform*: inflation adjustment is done with consumer price index; categorical data are one-hot encoded. Both predictors and response are transformed accordingly.
6. *Generate new features based on original ones.*
7. *Finish data processing for exploratory data analysis and modeling.*

4.1.2 Exploratory Data Analysis

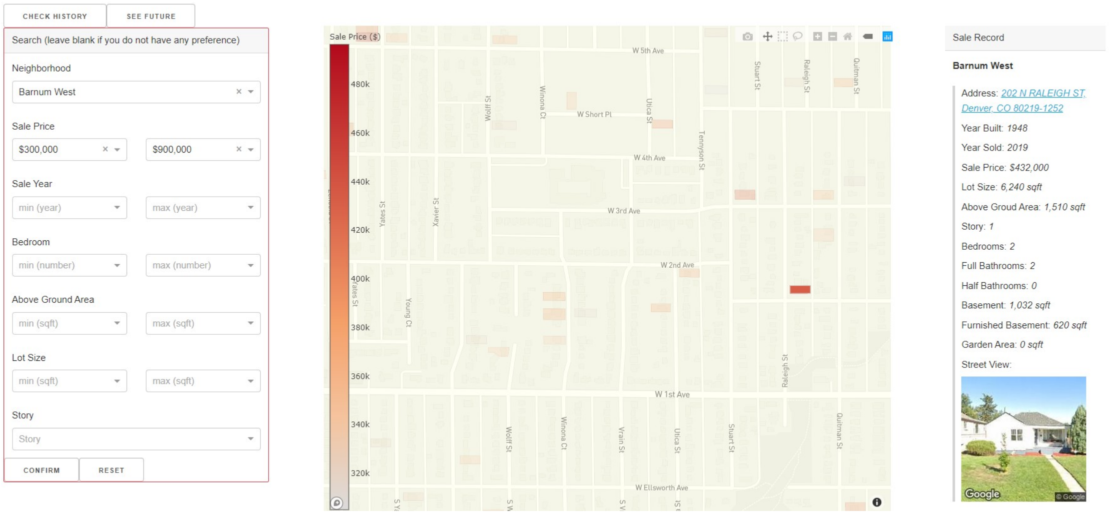
With processed data from Section 4.1.1, we moved on to verify spatiotemporal heterogeneity in Denver housing market through exploratory data analysis (EDA). With tools such as choropleth map, we explored how house prices and other housing market metrics (e.g. house sold) vary geographically and over time. Specifically, we compared these metrics over the years at national and state levels, among different cities similar to Denver, and across neighborhoods in the City and County of Denver. In addition, with exponential smoothing, we conducted time series analysis on the seasonality of median house price and house sold for comparison purpose. Besides housing market data, we also included amenities such as schools as parks in different neighborhoods of Denver to illustrate the influence of neighborhood quality on individual house price prediction. Details of our observations would be discussed later in section 5, and the EDA results confirmed our assumption on spatiotemporal heterogeneity and dependency.

4.1.3 Modeling

After confirming spatiotemporal effects, we incorporated related attributes in the modeling process along with other house and environmental features. To balance prediction accuracy and interpretability, we aim to select the best model based on prediction accuracy, and then extract important features from such a model to serve as input parameters for our website users. We tested four candidate machine learning models: linear regression with lasso regularization, artificial neural network (ANN) with different numbers of layers, and ensemble methods including random forest and XGBoost. Important features were also identified in different models, such as area above ground, and finished basement area. We split our data into training and test sets (75%-25%), and run grid search cross-validation to tune hyperparameters in each model. We then employed root mean square error (RMSE) and mean absolute error (MAE) as criteria to compare model performances, and have selected XGBoost as the best model for further employment. Details on model tuning and results will be discussed in detail in section 5.

4.1.4 Building Interactive Platform

We then combined our results in sections 4.1.2 and 4.1.3 into an interactive platform: EDA results serve as exploration tools, important features from the XGBoost model as user inputs, and the model itself for prediction. Our aim is to provide users with comprehensive knowledge and extensive information related to their house hunting in Denver by integrating historical transaction data, neighborhood characteristics and the proposed house price prediction model. Thus, we are creating an application mainly consisting of three levels of data presentation: 1) house price check and prediction; 2) neighborhood characteristics consisting of time series analysis and amenities and 3) price comparison with different cities. A snapshot of historical house record lookup is shown in Figure 2, and additional screenshots are included in Appendix A.



Search (leave blank if you do not have any preference)

Neighborhood: Barnum West

Sale Price: $300,000 – $900,000

Sale Year: min (year) – max (year)

Bedroom: min (number) – max (number)

Above Ground Area: min (sqft) – max (sqft)

Lot Size: min (sqft) – max (sqft)

Story: Story

Sale Record

Barnum West

Address: 202 N RALEIGH ST, Denver, CO 80219-1252
Year Built: *1948*
Year Sold: *2019*
Sale Price: *$432,000*
Lot Size: *6,240 sqft*
Above Groud Area: *1,510 sqft*
Story: *1*
Bedrooms: *2*
Full Bathrooms: *2*
Half Bathrooms: *0*
Basement: *1,032 sqft*
Furnished Basement: *620 sqft*
Garden Area: *0 sqft*
Street View:

Figure 2: Denver House Page (Check History)

With Plotly Dash in Python, we have created the web application including four tabs - *Denver House*, *Denver Neighborhood*, *Denver City*, and *About*.

1. *Denver House* page: this is where the users see historical house records and house price prediction by the choices of features, such as neighborhood, number of bedrooms and bathrooms, etc.
2. *Denver Neighborhood* page: this is where all neighborhood features and median sale prices can be queried as a reference for house hunting, which is different from other online real estate websites

3. *Denver City* page: here we provide a comprehensive comparison on key housing market metrics among the cities that are similar to Denver in population, including median house price and sale/list price ratios, etc. Seasonality analysis with exponential smoothing is also included for comparison, which is another innovation from other platforms.
4. *About* page: here we provide a brief explanation on the data and models we have used in this project.

4.1.5 User Survey

To test the effectiveness of our website, we designed an online survey, and distributed it to potential users. Our survey questions fall into the following two major categories: experience with the application in aspects of visualization, accessibility and functions, and potential problems and improvements. We have then improved our website according to some of these feedback.

5 Experiments and Evaluations

5.1 Description of Test Beds

Our experiments focus on the two following questions:

1. Which model gives the best prediction accuracy, and what are important features in each model?
2. How to select function and visualization combinations that provide the best user experience?

5.2 Experiments on Modeling

Spatial and temporal factors are of importance in our model development process. Detailed experiments are described below:

1. *Spatial factors*: To consider spatial dependence of individual house prices, we first identified two aggregation levels for data collection and manipulating: census tract level and neighborhood level. Census tracts are smaller than neighborhoods – $\sim$150 census tracts in the City and County of Denver, and $\sim$80 neighborhoods. In terms of data analysis and modeling, these two aggregation levels are similar. Thus, considering general knowledge of neighborhood, we chose to use neighborhood level information for modeling individual house prices.
2. *Temporal factors*: Understand that our data has a range of 20 years, to use such data, we need to consider inflation adjustment. There are two major temporal factors we have considered: house appreciation (i.e., house price index), and inflation (i.e., consumer price index). Although both indices are applicable to create a fair comparison of house prices over the years, it seems that house price index already include a lot of hidden features in house price prediction. Thus, we considered consumer price index to be a more fundamental factor and used it in our model.

In the modeling part, we tested four models, and for each model, hyperparameters were tuned through random search to find the model configuration of best performance. We have compared a range of lambda values for Lasso regression model, layer configurations and optimizers for ANN model, maximum depth and number of estimators random forest model, and similar hyperparameters for XGBoost model. We then compared their performance, which is summarized in Figure 3. We found XGBoost to be the best model among four candidates, with a mean prediction error of $44K, and a median prediction error of $13K.

5.3 Experiments on Website

We aim to provide comprehensive and accessible information to users, and a key issue in experimental design is to balance functionality and user friendliness. Our approach is first to guarantee key functionalities, and then provide the best combination of visualization elements. Below are the examples of some experiments we have conducted on our website:

Model	Configurations	Performance		
		R^2	RMSE	MAE
Lasso Regression	Alpha: 12.83	0.69	72,679	29,329
ANN	Layers:(60,30,60) Optimizer: AdaMax	0.69	73,169	28,937
	Layers:(60,60) Optimizer: AdaMax	0.69	72,272	28,583
Random Forest	Estimators: 1000	0.86	48,454	14,541
XGBoost	Gamma: 0.17 Learning rate: 0.18 Estimators: 134	0.89	44,112	13,332

Figure 3: Model Configuration and Performance

- *Denver House tab*: we experimented and compared two designs: one with both functions (check history and price prediction) provided upfront, and input options shared between them; and the other one with input options separated and hidden until button of a certain function is triggered. Although the two designs share the same functionality, our internal evaluation suggests that the second design is more user-friendly, and hence we have selected it for as our final design.
- *Denver Neighborhood tab*: Our experiments in this part focused on two questions: what is the most effective way to display a certain piece of information, and how to build interactions between these visualization elements. We have tested different combinations in this part: For example, to show the amenities in selected neighborhoods, we have tested two designs of this functionality: the first one only display amenities in a map of one single neighborhood, and allows users to switch among selected neighborhoods through a dropdown box; and the second design adds amenities of all selected neighborhoods on a single map, and allows users to zoom into a certain neighborhood. Our internal evaluation suggests that the second design provides a better overview and connection between neighborhoods, and hence is our final choice.

5.4 Website Evaluation through a User Survey

Beyond internal experiments and evaluations, we have also received external evaluations on our website, as discussed in Section 4.1.5. We received a total of 34 responses and the results of the evaluation are stored and accessed through this link. We have included screenshots for the feedback of the questions of our survey in Appendix B. Although there is a limitation in the survey audience, overall the end users are satisfied with our visualization efforts and website functionality. We have also received constructive feedback such as how to make the website easier to understand. We note that this is an ongoing effort and by the time of this report, we have resolved the clarity issues mentioned in the survey comments by adding more descriptions and removing unnecessary add-ons.

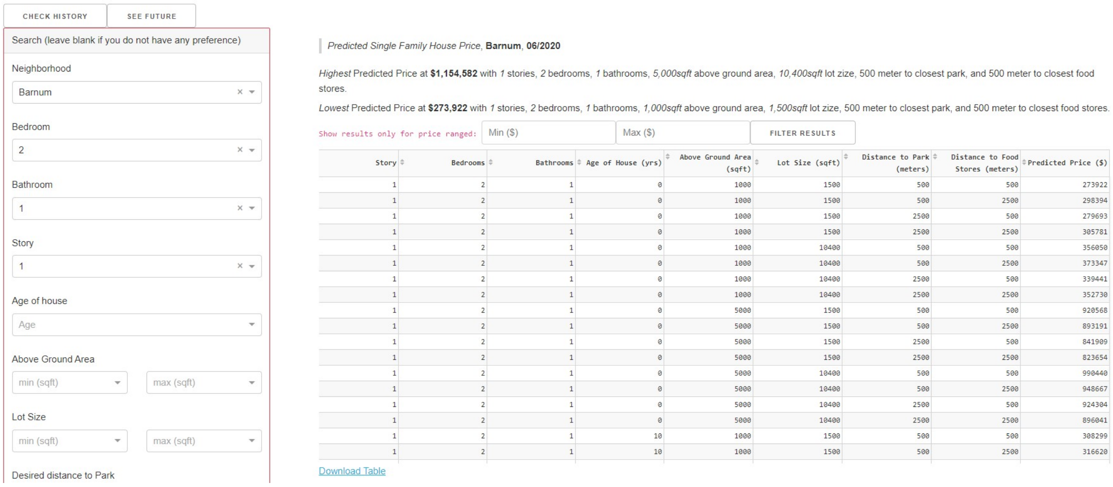
6 Conclusions and Discussion

In conclusion, we have developed a comprehensive website portal. Incorporating machine learning and visualization techniques, this platform offers interactive house price prediction, visual exploration and decision support that are easily accessible and of high interests to the general public. A key feature of our project is the exploration and verification of the impact of spatial and temporal factors on house sale prices.

7 Distribution of Team Member Effort

All team members have contributed similar amount of effort.

Appendix A Website Design



Search (leave blank if you do not have any preference)

Neighborhood: Barnum

Bedroom: 2

Bathroom: 1

Story: 1

Age of house: Age

Above Ground Area: min (sqft) max (sqft)

Lot Size: min (sqft) max (sqft)

Desired distance to Park

Predicted Single Family House Price, **Barnum**, **06/2020**

Highest Predicted Price at **$1,154,582** with *1* stories, *2* bedrooms, *1* bathrooms, *5,000sqft* above ground area, *10,400sqft* lot zize, 500 meter to closest park, and 500 meter to closest food stores.

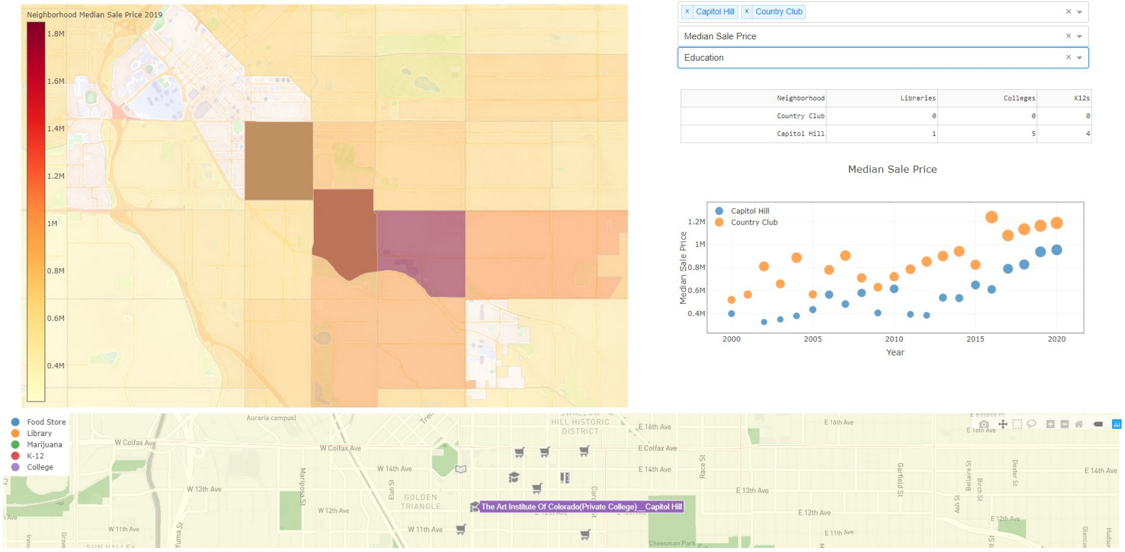
Lowest Predicted Price at **$273,922** with *1* stories, *2* bedrooms, *1* bathrooms, *1,000sqft* above ground area, *1,500sqft* lot zize, 500 meter to closest park, and 500 meter to closest food stores.

Show results only for price ranged: Min ($) Max ($) FILTER RESULTS

Story	Bedrooms	Bathrooms	Age of House (yrs)	Above Ground Area (sqft)	Lot Size (sqft)	Distance to Park (meters)	Distance to Food Stores (meters)	Predicted Price ($)
1	2	1	0	1000	1500	500	500	273922
1	2	1	0	1000	1500	500	2500	298394
1	2	1	0	1000	1500	2500	500	279693
1	2	1	0	1000	1500	2500	2500	305781
1	2	1	0	1000	10400	500	500	356050
1	2	1	0	1000	10400	500	2500	373347
1	2	1	0	1000	10400	2500	500	339441
1	2	1	0	1000	10400	2500	2500	352730
1	2	1	0	5000	1500	500	500	920568
1	2	1	0	5000	1500	500	2500	893191
1	2	1	0	5000	1500	2500	500	841909
1	2	1	0	5000	1500	2500	2500	823654
1	2	1	0	5000	10400	500	500	990440
1	2	1	0	5000	10400	500	2500	948667
1	2	1	0	5000	10400	2500	500	924304
1	2	1	0	5000	10400	2500	2500	896041
1	2	1	10	1000	1500	500	500	308299
1	2	1	10	1000	1500	500	2500	316620

Download Table

Figure 4: Denver House Page (See Future)



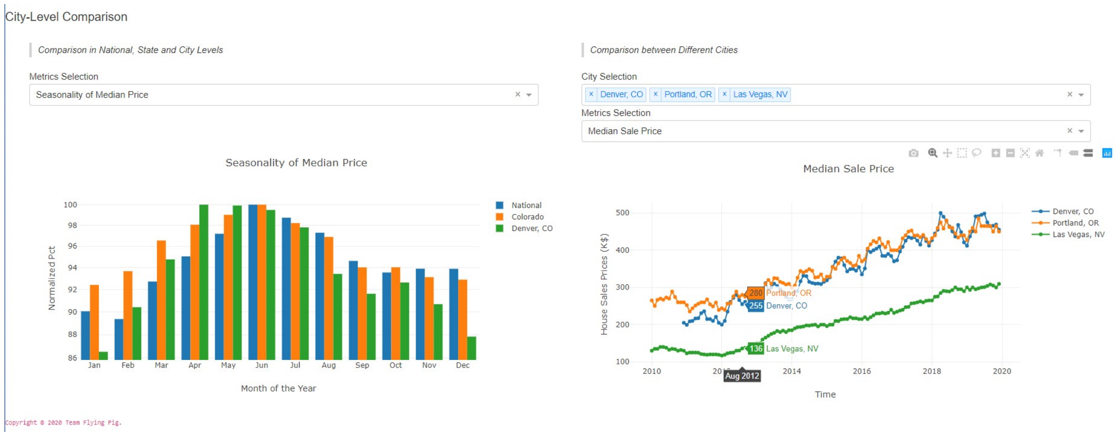
× Capitol Hill × Country Club

Median Sale Price

Education

Neighborhood	Libraries	Colleges	K12s
Country Club	0	0	0
Capitol Hill	1	5	4

Figure 5: Denver Neighborhood Page



City-Level Comparison

Comparison in National, State and City Levels

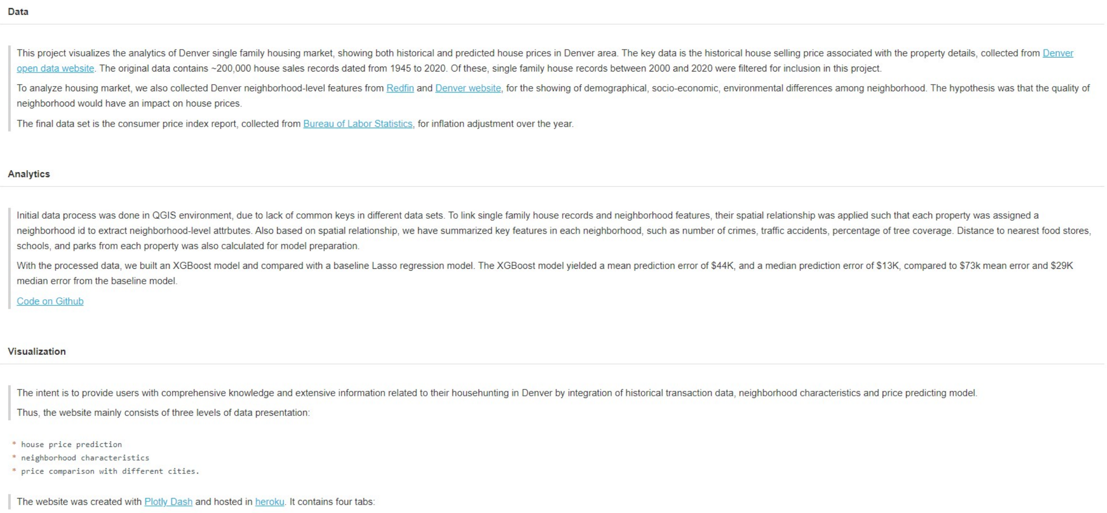
Metrics Selection: Seasonality of Median Price

Comparison between Different Cities

City Selection: Denver, CO; Portland, OR; Las Vegas, NV

Metrics Selection: Median Sale Price

Figure 6: Denver City Page



Data

This project visualizes the analytics of Denver single family housing market, showing both historical and predicted house prices in Denver area. The key data is the historical house selling price associated with the property details, collected from Denver open data website. The original data contains ~200,000 house sales records dated from 1945 to 2020. Of these, single family house records between 2000 and 2020 were filtered for inclusion in this project.

To analyze housing market, we also collected Denver neighborhood-level features from Redfin and Denver website, for the showing of demographical, socio-economic, environmental differences among neighborhood. The hypothesis was that the quality of neighborhood would have an impact on house prices.

The final data set is the consumer price index report, collected from Bureau of Labor Statistics, for inflation adjustment over the year.

Analytics

Initial data process was done in QGIS environment, due to lack of common keys in different data sets. To link single family house records and neighborhood features, their spatial relationship was applied such that each property was assigned a neighborhood id to extract neighborhood-level attrbutes. Also based on spatial relationship, we have summarized key features in each neighborhood, such as number of crimes, traffic accidents, percentage of tree coverage. Distance to nearest food stores, schools, and parks from each property was also calculated for model preparation.

With the processed data, we built an XGBoost model and compared with a baseline Lasso regression model. The XGBoost model yielded a mean prediction error of $44K, and a median prediction error of $13K, compared to $73k mean error and $29K median error from the baseline model.

Code on Github

Visualization

The intent is to provide users with comprehensive knowledge and extensive information related to their househunting in Denver by integration of historical transaction data, neighborhood characteristics and price predicting model.

Thus, the website mainly consists of three levels of data presentation:

```
* house price prediction
* neighborhood characteristics
* price comparison with different cities.
```

The website was created with Plotly Dash and hosted in heroku. It contains four tabs:

Figure 7: About Page

Appendix B Website Feedback

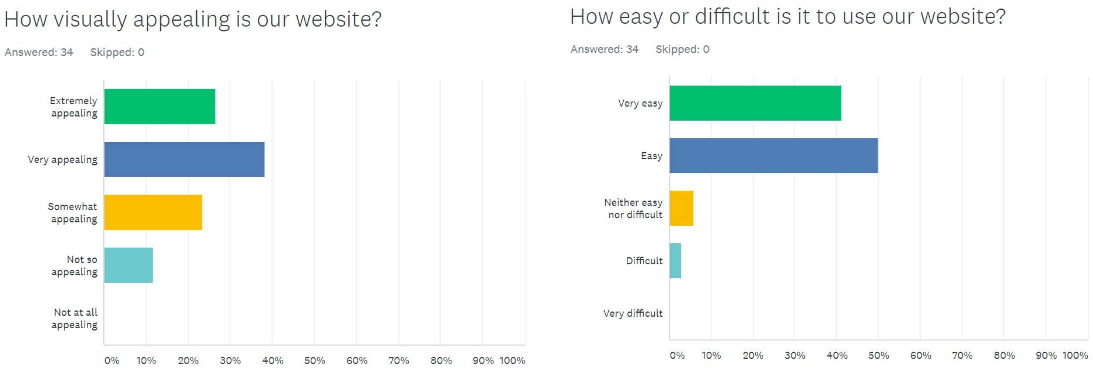


Figure 8: Survey Feedback (a)

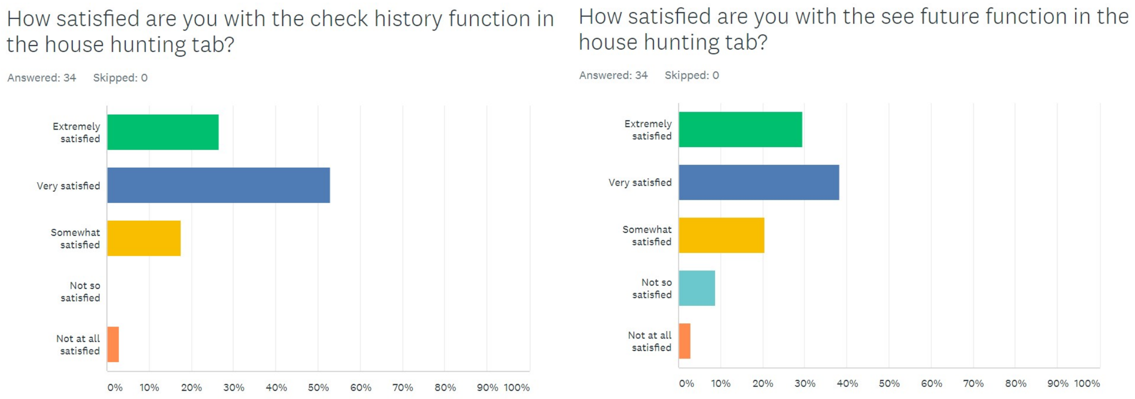


Figure 9: Survey Feedback (b)

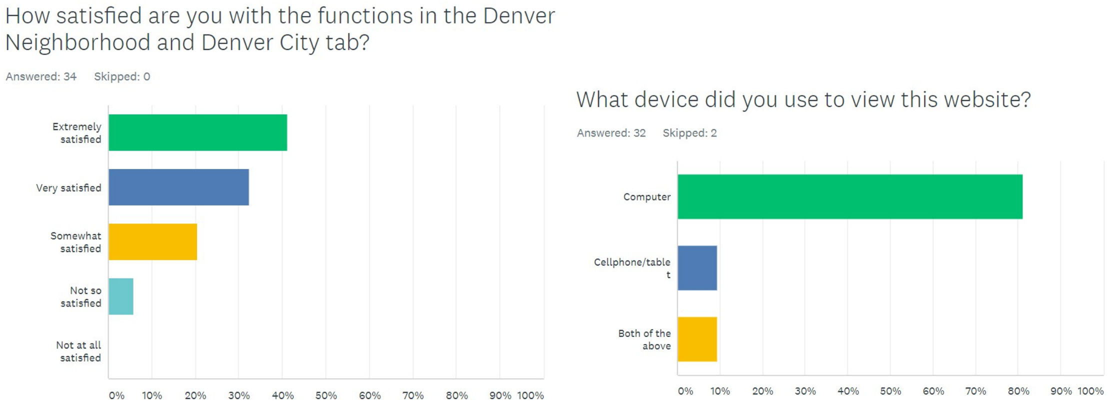


Figure 10: Survey Feedback (c)

References

[1] Radoslaw Cellmer and Radoslaw Trojanek. Towards increasing residential market transparency: Mapping local housing prices and dynamics. *ISPRS International Journal of Geo-Information*, 9(1), 2019. ISSN 2220-9964. doi: 10.3390/ijgi9010002. URL https://www.mdpi.com/2220-9964/9/1/2.

[2] Jennifer Rudden. *U.S. existing home sales 2005-2020*. URL https://www.statista.com/statistics/226144/us-existing-home-sales/.

[3] Stacy Sirmans, David Macpherson, and Emily Zietz. The composition of hedonic pricing models. *Journal of real estate literature*, 13(1):1–44, 2005.

[4] Stephen Malpezzi. *Hedonic Pricing Models: A Selective and Applied Review*, chapter 5, pages 67–89. John Wiley Sons, Ltd, 2008. ISBN 9780470690680. doi: 10.1002/9780470690680.ch5. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470690680.ch5.

[5] A Stewart Fotheringham, Chris Brunsdon, and Martin Charlton. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, 2003.

[6] Steven Bourassa, Eva Cantoni, and Martin Hoesli. Predicting house prices with spatial dependence: a comparison of alternative methods. *Journal of Real Estate Research*, 32(2):139–159, 2010.

[7] Alain Bonnafous, Marko Kryvobokov, and Pierre-Yves Peguy. Insight Into Apartment Attributes And Location With Factors And Principal Components Applying Oblique Rotation. ERES eres2010-153, European Real Estate Society (ERES), January 2010.

[8] Henry Crosby, Theo Damoulas, Alex Caton, Paul Davis, João Porto de Albuquerque, and Stephen A Jarvis. Road distance and travel time for an improved house price kriging predictor. *Geo-spatial Information Science*, 21(3):185–194, 2018.

[9] R Kelley Pace, Ronald Barry, John M Clapp, and Mauricio Rodriquez. Spatiotemporal autoregressive models of neighborhood effects. *The Journal of Real Estate Finance and Economics*, 17(1):15–33, 1998.

[10] Bradford Case, John Clapp, Robin Dubin, and Mauricio Rodriguez. Modeling spatial and temporal house price patterns: A comparison of four models. *Journal of Real Estate Finance and Economics*, 29 (2):167–191, 2004. ISSN 08955638. doi: 10.1023/B:REAL.0000035309.60607.53.

[11] Ingrid Nappi-Choulet and Tristan-Pierre Maury. A spatial and temporal autoregressive local estimation for the paris housing market. *Journal of Regional Science*, 51(4):732–750, 2011.

[12] Xiaolong Liu. Spatial and Temporal Dependence in House Price Prediction. *Journal of Real Estate Finance and Economics*, 47(2):341–369, 2013. ISSN 08955638. doi: 10.1007/s11146-011-9359-3.

[13] Lasse Bork and Stig V. Møller. Forecasting house prices in the 50 states using Dynamic Model Averaging and Dynamic Model Selection. *International Journal of Forecasting*, 31(1):63–78, 2015. ISSN 01692070. doi: 10.1016/j.ijforecast.2014.05.005. URL http://dx.doi.org/10.1016/j.ijforecast.2014.05.005.

[14] Gary Koop and Dimitris Korobilis. Forecasting inflation using dynamic model averaging. *International Economic Review*, 53(3):867–886, 2012. ISSN 00206598. doi: 10.1111/j.1468-2354.2012.00704.x.

[15] Adrian E. Raftery, Miroslav Kárný, and Pavel Ettler. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, 52(1):52–66, 2010. ISSN 00401706. doi: 10.1198/TECH.2009.08104.

[16] Visit Limsombunc, Christopher Gan, and Minsoo Lee. House Price Prediction: Hedonic Price Model vs. Artificial Neural Network. *American Journal of Applied Sciences*, 1(3):193–201, 2004. ISSN 15469239. doi: 10.3844/ajassp.2004.193.201.

[17] Hasan Selim. Determinants of house prices in turkey: Hedonic regression versus artificial neural network. *Expert Systems with Applications*, 36(2, Part 2):2843 – 2852, 2009. doi: https://doi.org/10.1016/j.eswa.2008.01.044.

[18] Azme Bin Khamis and Nur Khalidah Khalilah Binti Kamarudin. Comparative study on estimate house price using statistical and neural network model. *International Journal of Scientific & Technology Research*, 3(12):126–131, 2014.

[19] Xiaochen Chen, Lai Wei, and Jiaxin Xu. House Price Prediction Using LSTM. 2017. URL http://arxiv.org/abs/1709.08432.

[20] Quanzeng You, Ran Pang, Liangliang Cao, and Jiebo Luo. Image-Based Appraisal of Real Estate Properties. *IEEE Transactions on Multimedia*, 19(12):2751–2759, 2017. ISSN 15209210. doi: 10.1109/TMM.2017.2710804.

[21] Byeonghwa Park and Jae Kwon Bae. Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Systems with Applications*, 42(6):2928–2934, 2015. ISSN 09574174. doi: 10.1016/j.eswa.2014.11.040. URL http://dx.doi.org/10.1016/j.eswa.2014.11.040.

[22] Eman Ahmed and Mohamed Moustafa. House price estimation from visual and textual features. *arXiv preprint arXiv:1609.08399*, 2016.

[23] Omid Poursaeed, Tomáš Matera, and Serge Belongie. Vision-based real estate price estimation. *Machine Vision and Applications*, 29(4):667–676, 2018.

[24] Stephen Law, Brooks Paige, and Chris Russell. Take a look around: Using street view and satellite images to estimate house prices. *ACM Transactions on Intelligent Systems and Technology*, 10(5), 2019. ISSN 21576912. doi: 10.1145/3342240.

[25] Xibin Wang, Junhao Wen, Yihao Zhang, and Yubiao Wang. Real estate price forecasting based on svm optimized by pso. *Optik*, 125(3):1439–1443, 2014.

[26] Siti Amri and Gurudeo Anand Tularam. Performance of multiple linear regression and nonlinear neural networks and fuzzy logic techniques in modelling house prices. *Journal of Mathematics and Statistics*, 8(4):419–434, 2012.

[27] Denver open data, (accessed February 20, 2020). https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-parcels.

[28] U.S. Census Bureau. 2010 tiger/line shapefiles technical documentation. Technical report, 2012.

[29] Redfin data center, (accessed January 25, 2020). https://www.redfin.com/blog/data-center/.