Post-Pandemic Real Estate and Population Shift

Tripathi S, Ho M, Bijanapalli R, Mow M, Duong G, Eduardo E

1 INTRODUCTION

Our project aims to study and visualize the influence of the COVID-19 pandemic on migration patterns and real estate prices throughout the United States, namely due to the growing movement for employees to work from home in technology related fields. Our objectives[1] are to 1) identify, quantify, and interactively visualize post-pandemic migration patterns and corresponding changes in local real estate prices and 2) predict future patterns in local real estate prices. This project will support future migration and real estate research and offer individuals looking to buy a home in a new area or those interested in national real estate investment an interactive tool to analyze and predict home prices.

2 PROBLEM DEFINITION

Our work aims to answer the problem statement of how home buyers and investors can understand the impact of COVID-19 and its effects including people migration due to remote work on housing markets and prices. Home buyers and investors would like to understand the national housing price behavior in order to make better decisions regarding relocation, purchasing, and investments. As part of this project we aim to provide visualizations of current housing price behavior and predictions of future housing prices, with an intent for home buyers to leverage this information to make better decisions.

3 SURVEY

Much research has gone into modeling the relationship between migration patterns and housing prices in the United States[2]. [18] introduces the idea of migration networks and shows that migration drives up housing demand and prices especially in cities with strong migration links. [5] looks at the local housing markets in the San Francisco Bay Area in relation to neighboring tech corporate campuses. The paper found that housing prices increased by an “additional 7.1% in immediate vicinity of the tech campus 2 years after arrival. [8] also concludes with similar results. Also [14] shows that there is compelling evidence for shifting housing towards suburbs from the city center quoting jobs and amenities like restaurants as the reason for high housing prices in dense urban areas. [22] suggests that housing prices in cities heavily relies on the service industries. [21] presents explanation of the volatility in home ownership between 1994 and 2015 in relationship to migration. [11] highlights research into the effect of Tesla’s and Amazon’s relocation to Austin, TX and Seattle/Bellevue, WA, respectively, on housing prices in the neighboring areas using difference-in-difference (DID) method to evaluate changes in housing price index around said locations. [20] also examines housing prices in a different market using the DID model, focusing on the cities in China’s Yangtze River Delta. It shows that for cities experiencing a sharp drop in housing prices, we can predict how fast the prices will recover by observing the city’s urban resilience index, which is equal to the city’s marginal travel intensity.

Other research specifically focuses on the relationship of housing prices to remote work. [4] expands on this research to show that the movement to remote work drives housing prices down in areas of high productivity as workers relocate to cheaper metro areas, thus increasing demand and prices in these connected cities. [9] offers research of the migratory patterns of people from different income brackets before and after the pandemic, finding that high income households are moving out of high population areas at a relatively faster rate. [17] specifically shows that the migration of workers leads to hollowing out of the city center to surrounding suburbs dubbed the "donut effect". This work

Table 1: Literature Survey: References Assigned to each Author

Assignee	Alias	References
Tripathi S	stripathi49	[3],[6],[23]
Ho M	mho74	[11], [5], [21]
Bijanpalli R	rbijanapalli3	[14], [17], [19]
Mow M	mmow3	[4],[18],[15]
Duong G	giangdvt177	[10], [22], [8]
Eduardo E	eeduardo3	[9], [20], [7]

The answers to the 9 Heilmeier questions are referenced using superscripts$^{1-9}$.

also demonstrates a relatively smaller impact on intercity movement. [15] quantifies the increase through regression analysis stating that a 1% increase in remote work leads to a 0.98% increase in housing prices.

Some studies also specifically address the effect that the COVID-19 pandemic had on population migration and housing prices through remote work policies. [23] explains that the effect of migration leads to property construction to be evenly distributed in urban, suburban, and rural areas. It also provides information on the effect of different remote work policies during the COVID-19 pandemic that led to people migration and hence affected the housing market. [6] discusses the impact of shutdown and social distancing that lead to virtual property tours, hence leading to people opting to tour properties at different locations. [3] helped to expand the understanding of COVID-19 policies on the property rental market in addition to the home buying market. [7] experiments with several machine learning models to find out which factors are most important for predicting rent and sale prices for real estate, concluding that TOM (time-on-the-market, or vacancy rate) is consistently the largest influencer of changing real estate prices. [10] further examines residential instability—including moving, crowding, and financial health—in the Bay Area during the pandemic. Finally, the work in [19] defines a 2-phase methodology for analysis of the effects of anomalous events such as a pandemic on housing market demand and uses this methodology on 6 Italian cities.

4 PROPOSED METHOD

4.1 Expected innovation

Much of the existing research on migration and housing prices in our literature review excluded[2] the COVID-19 pandemic and post-pandemic data [14], [17], [21]. Though much of the literature offered visualization tools through bar charts, line charts, and choropleth maps, none offered an interactive capability to view trends and changes in these statistics nor allowed the reader to study further in depth a specific migration path or city of choice. Lastly, though some of the literature offered mathematical [17] or regression based models relating migration to housing prices, none offered the reader a machine learning based model to predict future housing prices given the data. Some of the studies [8], [10], [5] were also limited to a single or a few metropolitan areas.

Our work focused on innovating[3] in these specific areas namely 1) focusing on the impact of the COVID-19 pandemic by adding the post pandemic data publicly available through the Redfin [1], and Federal Housing Financing Agency (FHFA) [2] APIs on regions across the entire United States and storing and fetching this data using Google Big Query, 2) offering interactive visualization tools to better view trends and perform deeper research into a selected state/county in USA using Plotly for choropleth and line chart visualizations and Dash to host the web application, and 3) offering a model to predict housing prices in a given state and county using a time series[13] ML model such as SARIMA[16] from the statsmodels library.

Our visualization should help individuals understand the effect of the COVID-19 pandemic on real estate and people migration. This will help government[4] and city planners to plan for the areas of growth, new home buyers to look for areas where new communities are being formed, and investors[4] to invest in areas of future growth.

4.2 Data Setup

We first were able to utilize the Redfin API and Federal Housing Financing Agency (FHFA) to pull the data from these sources. Our setup consists of a Google BigQuery instance on Google Cloud Platform. The Redfin county and state data and FHFA HPI data were stored into our BigQuery tables. We used the SARIMA model from statsmodels library to predict future home prices across different states and counties. The forecast data was also added into our BigQuery tables.

4.3 Visualizing Data

The visualizations in this project were generated using the Plotly package. The color scheme for our visualizations of the Redfin data start from light green for low values up to dark, red colors for high values. This was designed to make high values more prominent. The color scheme for our visualizations of the Federal Housing Finance Agency (FHFA) data uses light green at low values to dark green at high values. The visualization components were hosted as webpages using Dash.

4.4 Modeling

We used both an ARIMA model and SARIMA model from the statsmodel library. We trained the models using two different time series datasets. Training dataset 1 included housing prices prior to Covid i.e. before 2020-01-31, and training dataset 2 included house prices with Covid pandemic period housing sales.

We ended up using the data below:

- Redfin [1] housing market data
- FHFA [2] house price index data

4.5 Workflow

1 Given below are different steps taken:

- `Storing`: Store the existing Redfin and FHFA data into the GCP BigQuery table
- `Predictions Without Covid data in training`: Run predictions using SARIMA with training data excluding the Covid time period
- `Predictions With Covid data in the training`: Run predictions using SARIMA with training data including the Covid time period
- `Comparison`: Find the percentage difference between both above predicted prices
- `Analysis`: Develop visualizations using Plotly and develop the interactive visualization using Dash

On the front end we have a web server running on DASH with an interactive web page powered by Plotly. When users generate an event such as selecting a metric or property type from the drop down,

- A query is sent back to the GCP Big Query Table
- BigQuery returns the housing and forecast data
- Plotly is used to display the data to the user in the form of choropleth maps and line charts

5 EXPERIMENTS/EVALUATION

5.1 Testbed

The visualizations and prediction models were designed and built to be able to answer the following questions:

- How do real estate prices compare between states and/or counties?
- How did COVID-19 affect real estate prices and migration patterns?
- What are the predictions for state and county real estate prices up to 2026?

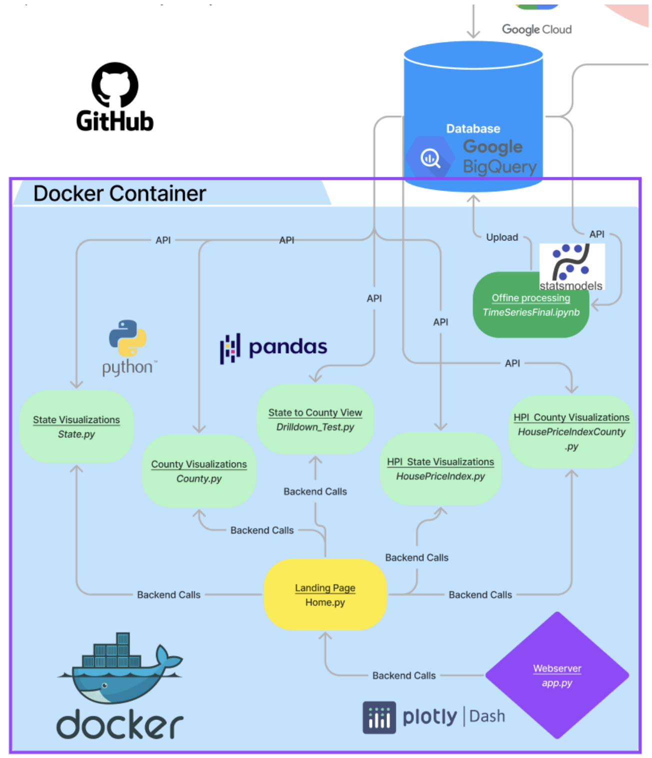

Figure 1: Data and compute Setup Along with the technologies used

Figure 2: Home Page of the webapp

5.2 State View

When a user opens our project, a web server python script, called app.py, will load all of our python scripts and the user will start at the home page, called Home.py. This page contains information on definitions of metrics used in our visualizations and the team members in our project.

Within the homepage are section headings that can be selected to navigate through each python script containing our visualizations. When "State" is selected, a choropleth of US states for various metrics and property types, as defined in the Home Page, will be accessible and a slider can be used to view the data for different time periods between 2012 and 2022:

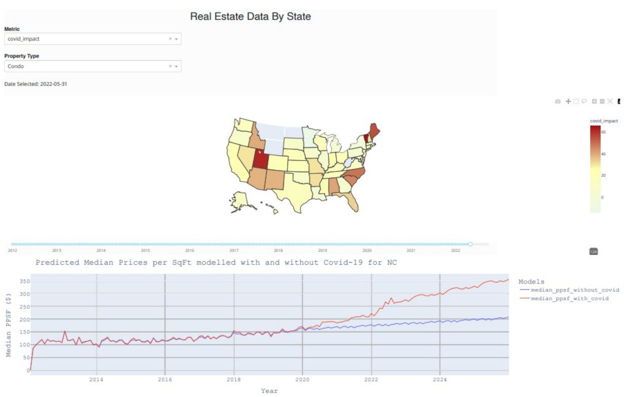

Figure 3: US State Choropleth with Line Chart

Additionally, hovering over states will populate a line chart using data from our trained SARIMA models, see section 5.3 Predictions, which predict future median price per square feet (ppsf) for a particular state.

5.3 County Drilldown View

Clicking the "Drilldowntocounty" heading displays the same choropleth of US states, but when a state is clicked, a zoomed in view of the counties for said state is displayed, which uses a Federal Information Processing Standard (FIPS) code specific to each county to populate the choropleth, Figure 3:

5.4 Housing Price Index View

As a separate visualization, using data from the Federal Housing Finance Agency (FHFA) we also visualized the metric of housing price index (HPI) for US states and counties as shown in Figures 4 and 5, respectively:

Alongside the choropleths of HPI data for US state and counties were line charts using data from our trained SARIMA models, see section 5.3 Predictions.

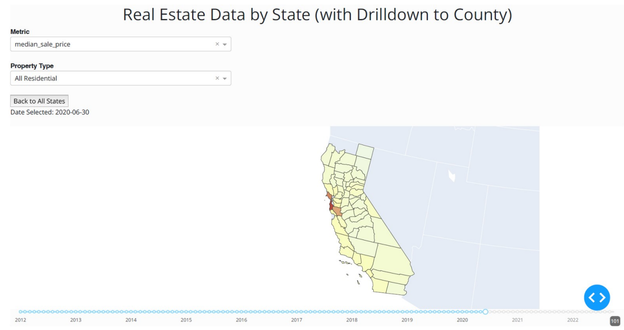

Figure 4: US State to County Drill Down

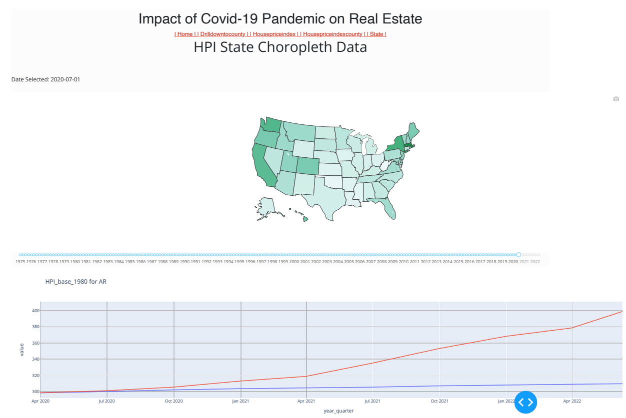

Figure 5: HPI State Choropleth with Line Chart

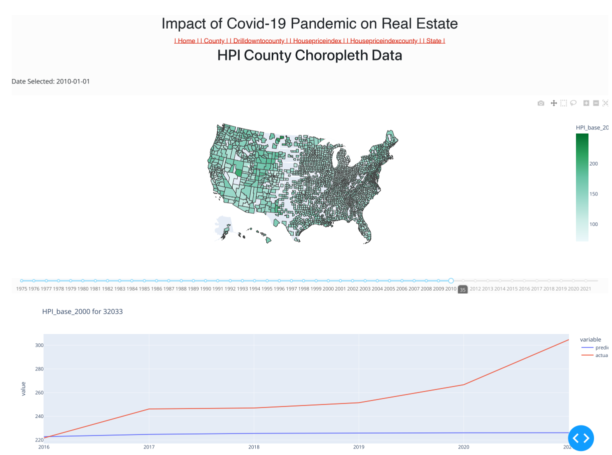

Figure 6: HPI County Choropleth with Line Chart

5.5 Predictions using SARIMA Model

We trained our models to make predictions, as mentioned in section 5.2 Modeling, for 2023, 2024, and 2025, on metrics such as median price per square foot for home sale prices. The target date range of data we used to train our models was between 2016-2022. Below are examples of the two ways we trained our models:

- Without Covid Data : train with data from 2016 to 2020 Jan, test with 2020 Feb to 2022 Aug, and predict on 2022 Sept to 2025 Dec
- With Covid Data : train with data from 2016 to 2022 Aug, test with 2020 Feb to 2022 Aug, and predict on 2022 Sept to 2025 Dec

5.6 Evaluation of Model

We ran two models (ARIMA and SARIMA) for forecasting future housing prices. We used Root Mean Square Error (RSME) evaluation technique to find the accuracy of the predicted data. RMSE was calculated between predicted and test data set. Higher RMSE means low prediction accuracy and vice versa. Based on this evaluation technique we finalized that SARIMA model was better for future house price predictions as its RSME was lower.

We also used heuristic approaches to visually analyze the choropleth and line charts to identify the states which our model predicted to have significant increases in price during and post covid as more people started migrating to these areas. Examples include Vermont, Idaho, and the Sun Belt region (Arizona). Remote work and the cost of living have incentivized people to relocate to these areas. Our model was able to confirm the increase in housing prices due to this population migration in these areas post Covid.

5.7 Results Summary

One major finding was the significant increase in real estate prices throughout the US during the COVID-19 pandemic, which had a large influence on the output of our time-series models when the model was trained with data prior to COVID-19 and tested on data during the pandemic. A closer look at the predictions of our models using models trained on pre-COVID data and post-COVID data showed that counties further away from cities had a larger difference between the predictions of these models, which may be explained by an increasing migration of people out of cities into suburbs.

More specifically, the significant effect COVID-19 had on increasing real estate median ppsf when compared to our SARIMA model trained on Pre-COVID 19 data, specified in section 4.4 Modeling, can be seen in Figure 6:

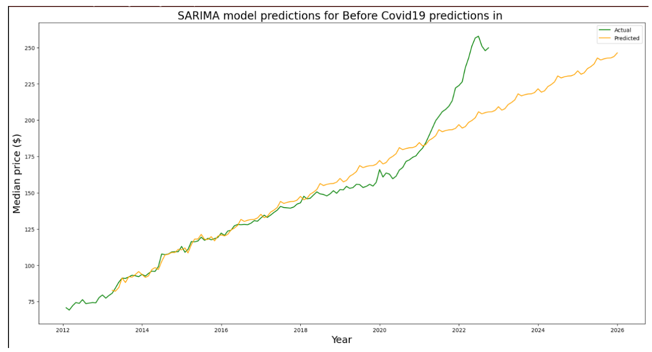

Figure 7: Sarima Model Forecasts Trained with Pre-COVID Data Compared with Actual Data

The orange line represents our SARIMA model and the green line represents the actual data of median ppsf. The deviation between the two, at around 2020, showed that COVID-19 had a significant impact on increasing the trajectory of real estate prices relative to what would have been expected with the trend of prices prior to COVID-19.

Addressing our testbed questions, first, we were able to successfully build a visualization tool that allowed for real estate prices to be compared across the state and county level. This included a drilldown feature that allows the user to dig deeper into counties in any given state for data ranging from 1975 to present across two views using the Redfin housing price data or the FHFA housing price index data. In addition, the choropleth maps and line charts successfully show that COVID-19 had a significant effect on housing prices. This effect was not equally distributed to all areas of the United States, but primarily affected low-cost areas likely due to migration of people to these areas as a result of remote work and other factors stemming from the COVID-19 pandemic. Lastly, we were able to develop a model to predict housing prices up to 2026 that have reasonably shown to identify low-cost areas such as Vermont, Idaho, and the Sunbelt to have significantly higher housing price increases post-covid as a result of migration.

6 CONCLUSION

In conclusion, we successfully 1) identified, quantified, and interactively visualized post-pandemic migration patterns and corresponding changes in local real estate prices and 2) predicted future patterns in local real estate prices. In addition, our team learned a lot about developing data visualization tools using large data sets and industry tools, and gained experience working together as a team on complex projects such as this.

6.1 Limitations and Future Extensions

One limitation to our project was not being able to process real estate prices by zip code, due to computational processing constraints of local and free online cloud computing resources. Due to the inability to efficiently process and visualize all of the zip code data, we instead used US counties and state. The size of the data from Redfin for zip code and county data were 2.29 GB and 301.32 MB, respectively. Additionally, there are 41,692 zip codes compared to 3,143 counties in the US. This change in scope effectively reduced the size of data by about a factor of 7.6x, which allowed us to generate a county choropleth visualizations from Section 5.1. However, even with county data, there was a computational limitation of adding a line chart, with data from our SARIMA models, onto the county choropleth. If this project were to be picked up again, we would add a line chart onto the county choropleth and add zip code data to provide a more localized view throughout the US.

Another limitation was not being able to host our work on a web server through Google Cloud web hosting, as was originally intended due to unexpected bugs with web hosting. In the future, adding web hosting functionality would improve the accessibility and the ability to share our work more easily with the public.

The datasets used in this project were downloaded. Another future extension would be to create a connection between our BigQuery database, using an API, to constantly query new data added from Redfin and the Federal Housing Finance Agency (FHFA) to update our visualizations and models. This extension would keep our visualizations and models up to date by providing live insights on the state of real estate prices.

As a future task we can also use deep learning model using LSTM [12] from PyTorch to forecast house prices in future. LSTM model can be a more sophisticated way of predicted future prices.

6.2 Team Effort

The following team members contributed a similar amount of effort. Michael Ho worked on the visualizations, Michael Mow worked on GCP and general set up and dash server, Ramakrishna worked on setting up the Docker and time-series models, Enrico worked on the HPI model analysis and visualizations, Shashank worked on timeseries models for county and data persistance.

REFERENCES

[1] 2021. Data center. https://www.redfin.com/news/data-center/

[2] 2022. House Price Index Datasets. https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx

[3] Mafalda Batalha, Duarte Gonçalves, Susana Peralta, and João Pereira dos Santos. 2022. The virus that devastated tourism: The impact of covid-19 on the housing market. *Regional Science and Urban Economics* 95 (2022), 103774.

[4] Jan Brueckner, Matthew E Kahn, and Gary C Lin. 2021. *A new spatial hedonic equilibrium in the emerging work-from-home economy?* Technical Report. National Bureau of Economic Research.

[5] Karen Chapple and Jae Sik Jeon. 2021. Big tech on the block: examining the impact of tech campuses on local housing markets in the San Francisco Bay area. *Economic Development Quarterly* 35, 4 (2021), 351–369.

[6] Walter D'Lima, Luis Arturo Lopez, and Archana Pradhan. 2022. COVID-19 and housing market effects: Evidence from US shutdown orders. *Real Estate Economics* 50, 2 (2022), 303–339.

[7] Andrius Grybauskas, Vaida Pilinkienė, and Alina Stundžienė. 2021. Predictive analytics using Big Data for the real estate market during the COVID-19 pandemic. *Journal of big data* 8, 1 (2021), 1–20.

[8] Louis Hansen. 2020. Coronavirus lockdowns no match for Bay Area Home Buyers. https://www.mercurynews.com/2020/06/04/coronavirus-lockdowns-no-match-for-bay-area-home-buyers/

[9] Peter H Haslag and Daniel Weagley. 2022. From LA to Boise: How migration has changed during the COVID-19 pandemic. *Available at SSRN 3808326* (2022).

[10] Jackelyn Hwang, Vasudha Kumar, Becky Liang, Jason Vargo, et al. 2022. Residential Instability in the Bay Area through the COVID-19 Pandemic. *Community Development Research Brief* 2022, 04 (2022), 1–37.

[11] Junfeng Jiao, Mira R Bhat, Amin Azimian, Akhil Mandalapu, and Arya Farahi. 2022. Housing market price movements under tech industry expansion during COVID-19. *International Journal of Housing Markets and Analysis* ahead-of-print (2022).

[12] Michael Keith. 2022. Exploring the LSTM neural network model for Time Series. https://towardsdatascience.com/exploring-the-lstm-neural-network-model-for-time-series-8b7685aa8cf

[13] Konstantin Kutzkov. 2022. Arima vs Prophet vs LSTM for time series prediction. https://neptune.ai/blog/arima-vs-prophet-vs-lstm

[14] Sitian Liu and Yichen Su. 2021. The impact of the COVID-19 pandemic on the demand for density: Evidence from the US housing market. *Economics letters* 207 (2021), 110010.

[15] John A Mondragon and Johannes Wieland. 2022. *Housing Demand and Remote Work.* Technical Report. National Bureau of Economic Research.

[16] Marco Peixeiro. 2022. Time Series Forecasting with Sarima in python. https://towardsdatascience.com/time-series-forecasting-with-sarima-in-python-cda5b793977b

[17] Arjun Ramani and Nicholas Bloom. 2021. *The Donut effect of COVID-19 on cities.* Technical Report. National Bureau of Economic Research.

[18] Gregor Schubert. 2021. House Price Contagion and US City Migration Networks. *Job Market Paper. Cambridge, Mass.: Joint Center for Housing Studies, Harvard University* (2021).

[19] Francesco Tajani, Felicia Di Liddo, Maria Rosaria Guarini, Rossana Ranieri, and Debora Anelli. 2021. An Assessment Methodology for the Evaluation of the Impacts of the COVID-19 Pandemic on the Italian Housing Market Demand. *Buildings* 11, 12 (2021), 592.

[20] Chuanhao Tian, Xintian Peng, and Xiang Zhang. 2021. COVID-19 pandemic, urban resilience and real estate prices: the experience of cities in the Yangtze River Delta in China. *Land* 10, 9 (2021), 960.

[21] Susan Wachter and Arthur Acolin. 2016. Owning or renting in the US: Shifting dynamics of the housing market. *Penn IUR Brief (Philadelphia, PA: Penn Institute for Urban Research* (2016).

[22] Bingbing Wang. 2021. How Does COVID-19 Affect House Prices? A Cross-City Analysis. *Journal of Risk and Financial Management* 14, 2 (Jan 2021), 47. https://doi.org/10.3390/jrfm14020047

[23] Yunhui Zhao. 2020. US housing market during COVID-19: aggregate and distributional evidence. *Available at SSRN 3744679* (2020).