Real Estate Analysis Project

Presented by: Makonnen Ramsey, Kinjal Sakhida and Kristin Bell

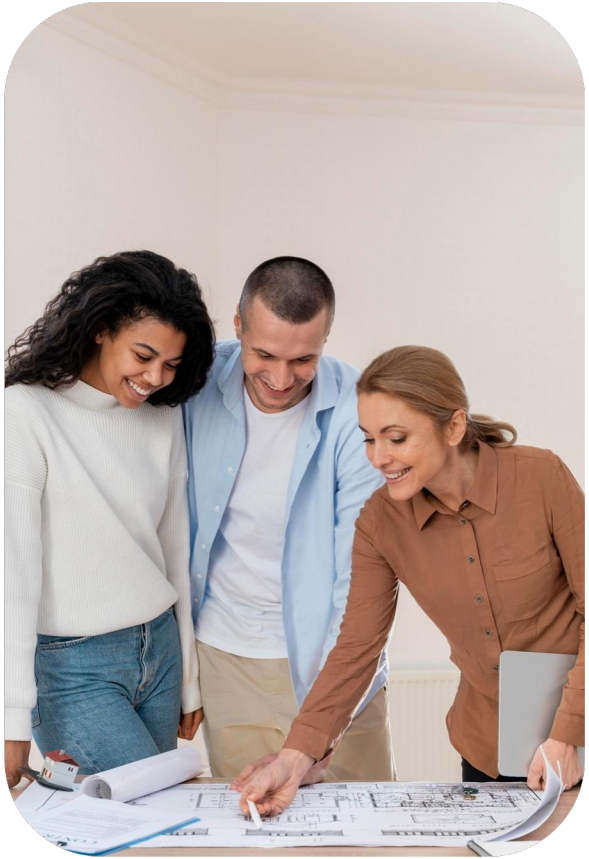
Table of contents

01 Introduction

02 Key Questions of Interest

03 Analysis Process

04 Analysis Conclusions

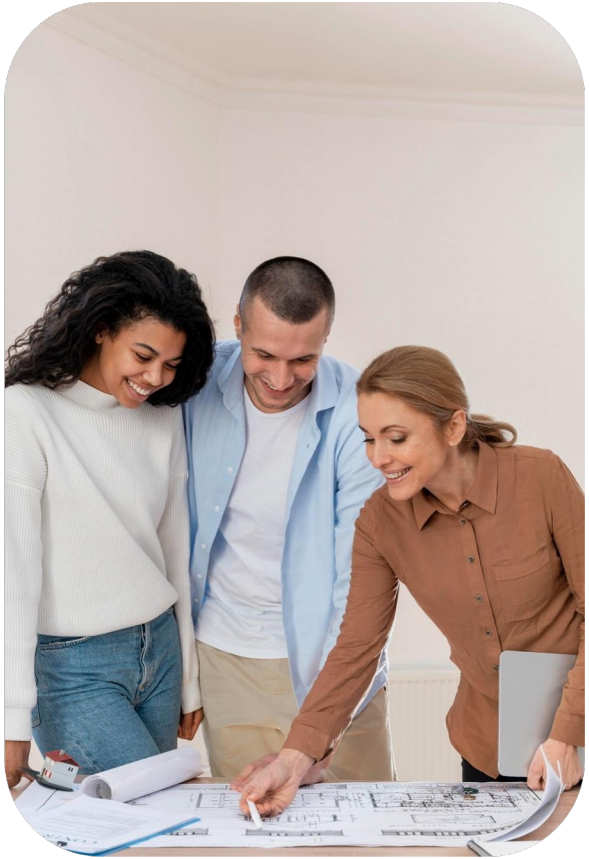


01 Introduction

01

In this project, we will be taking a closer look at identifying trends within the United States Real Estate industry utilizing a well-known Real Estate listings database. We chose to segment the data to represent certain criteria including:

- Three distinct states (Pennsylvania, Wisconsin & Washington) to represent three distinct regions of the United States (East, Midwest & West)
- 5 bedrooms or less
- Listings from 2015 onwards (~10 years)



02
Key Questions

02

What correlations exist for variables such as number of bedrooms, bathrooms and square footage, etc. and how much impact do they have on house price?

How do these correlations compare amongst the geographical areas we chose to analyze?

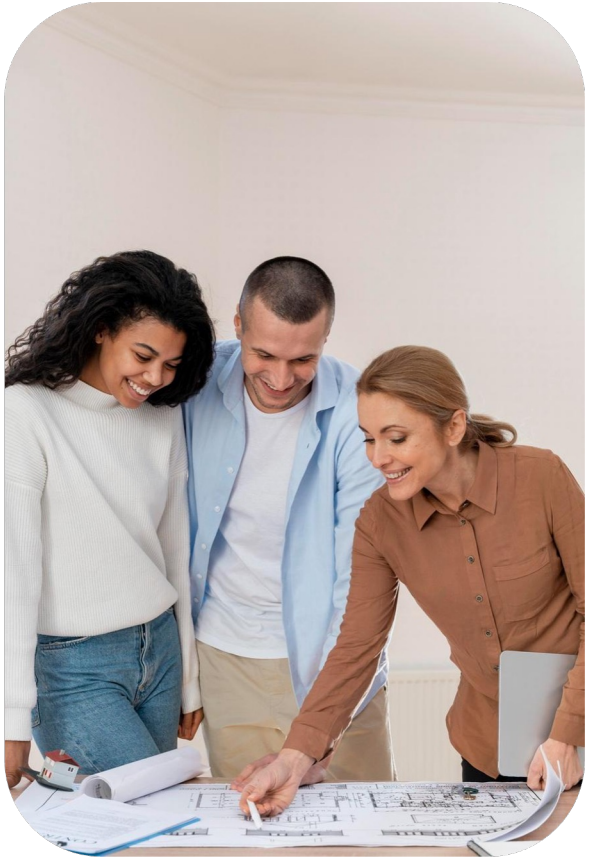
Which variable has the largest impact in each location?

What are the measures of central tendency for pricing in the areas we chose to examine?

How have housing prices changed over a time period of ~ 10 years in each location?

How has the volume of houses being sold changed over the years in each location?

What are some conclusions from our analysis that we could relay to a real estate market participant?



03 Analysis Process

Selecting and cleaning the data

The initial dataset had more than 2M datapoints and was significantly larger than GitHub allowed (170.57 MB). In order to work with the data, we first trimmed it down by removing Puerto Rico and Virgin Islands and removed unnecessary columns (”Brokered By” and “Street”)

We then filtered our dataset to represent listings with 5 bedrooms or less as we felt it was appropriate for the average real estate market participant (sellers, buyers, investors, developers and real estate professionals) and only included listings from the past decade by converting the raw data within the "previous sold date" column to "datetime" format.

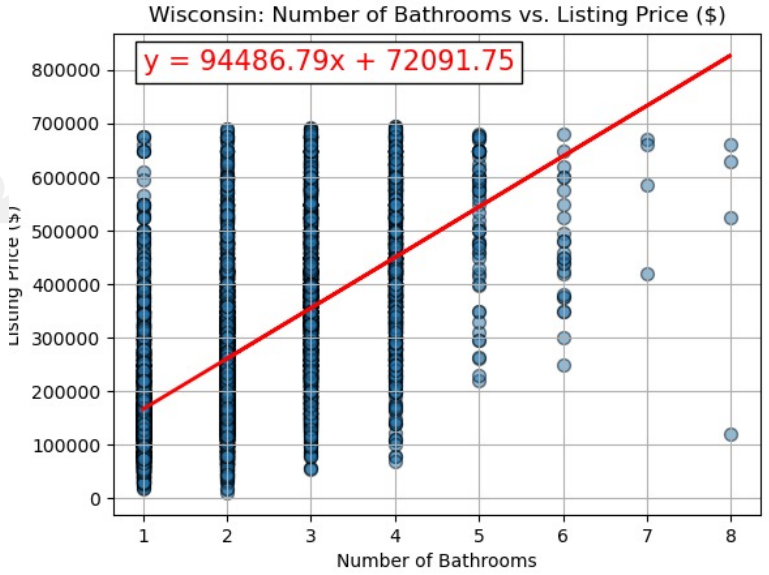
Next we chose Washington, Wisconsin and Pennsylvania as a cross section of the United States (West, Mid-West and East) to examine in depth.

Finally, we determined and eliminated outliers via interquartile range calculations. Also, we eliminated listings with a price of less than $5,000. This was executed to remove inaccurate listings due to data entry error (i.e. listings with a $1.00 value listings with information entered into the wrong field, etc.)

Regression Analysis: Wisconsin

We found that in Wisconsin, living space had the strongest correlation to house price with an "r value" of 0.64. The resulting linear equation was y=110.12x + 66893.54, indicating that each additional square foot of living space will equate to an increase of $110.12 in overall house price. In other words, the price per square foot for a house in Wisconsin is $110.12.

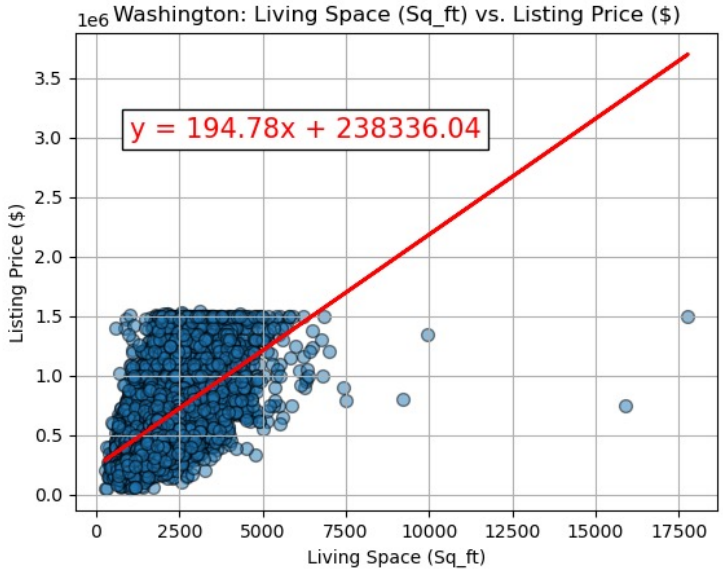
The second strongest correlation to house price in Wisconsin was number of bathrooms, with an "r value" of 0.63. The resulting linear equation was y=94486.79x + 72091.75, indicating that each additional bathroom will equate to an increase of $94,486.79 in overall house price. In other words, each additional bathroom will increase the house price by $94,486.79.

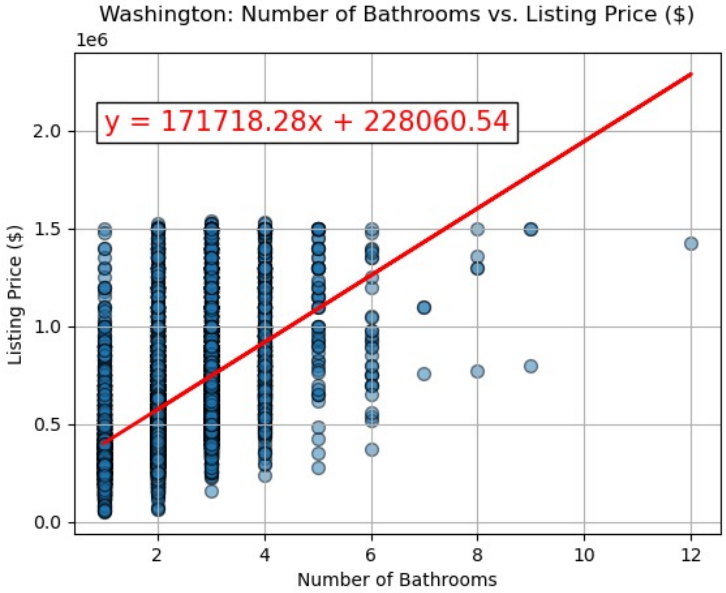


Regression Analysis: Washington

We found that in Washington, living space had the strongest correlation to house price with an "r value" of 0.57. The resulting linear equation was y=194.78x + 238336.04, indicating that each additional square foot of living space will equate to an increase of $194.78 in overall house price. In other words, the price per square foot for a house in Washington is $194.78.

The second strongest correlation to house price in Washington was number of bathrooms, with an “r value” of 0.49. The resulting linear equation was y=171718.28x + 228060.54, indicating that each additional bathroom will equate to an increase of $171,718.28 in overall house price. In other words, each additional bathroom will increase the house price by $171,718.28.

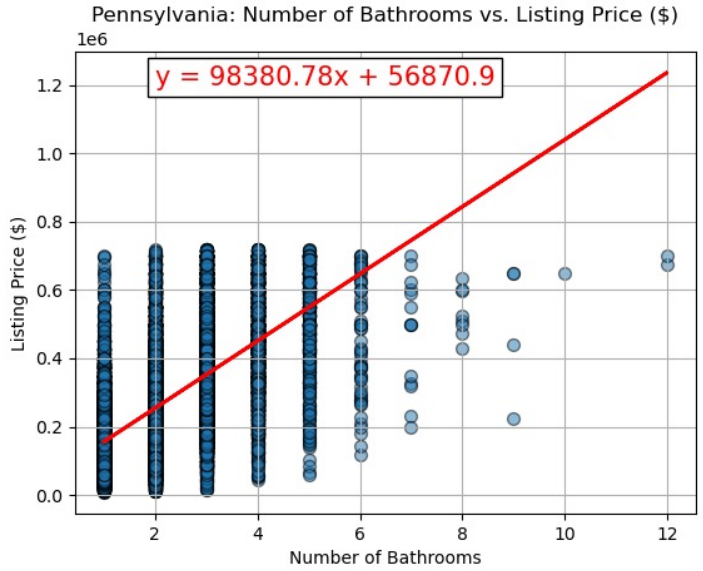


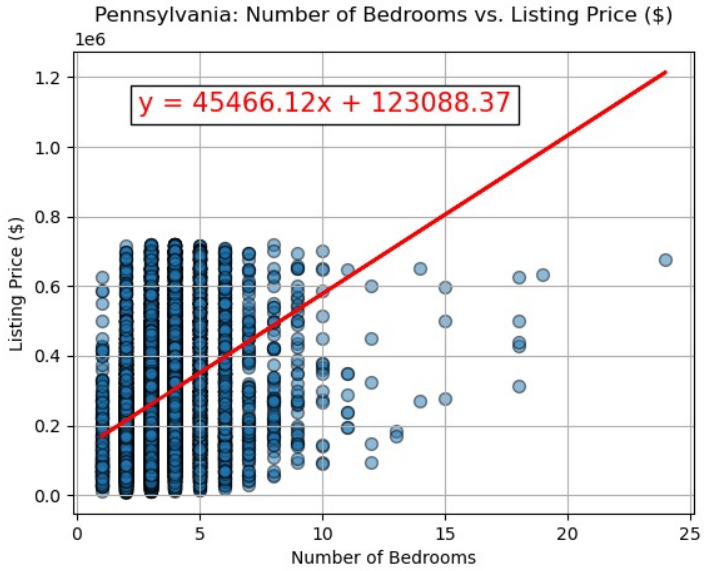


Regression Analysis: Pennsylvania

We found that in Pennsylvania, the number of bathrooms had the strongest correlation to house price with an "r value" of 0.62. The resulting linear equation was y=98380.78x + 56870.90, indicating that each additional bathroom will equate to an increase of $98,380.78 in overall house price. In other words, each additional bathroom will increase the house price by $98,380.78.

The second strongest correlation to house price in Pennsylvania was number of bedrooms, with an “r value” of 0.29. The resulting linear equation was y=45466.12x + 123088.37, indicating that each additional bedroom will equate to an increase of $45,466.12 in overall house price. In other words, each additional bedroom will increase the house price by $45,466.12.

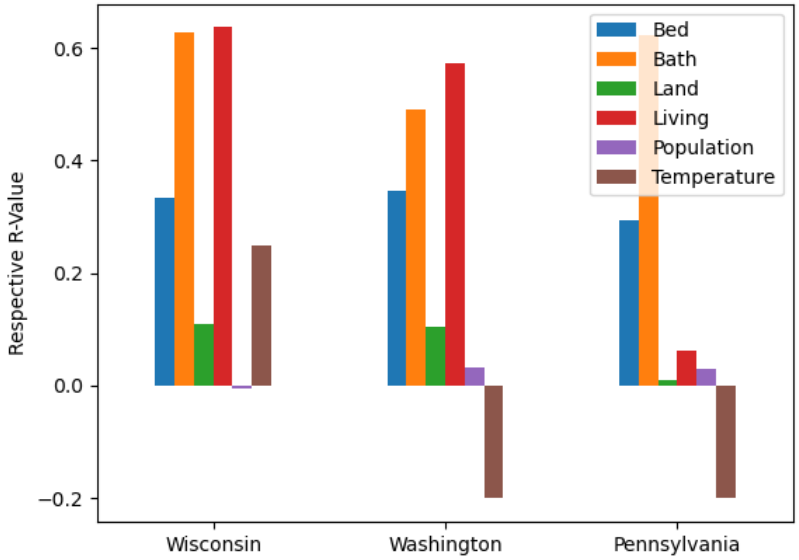


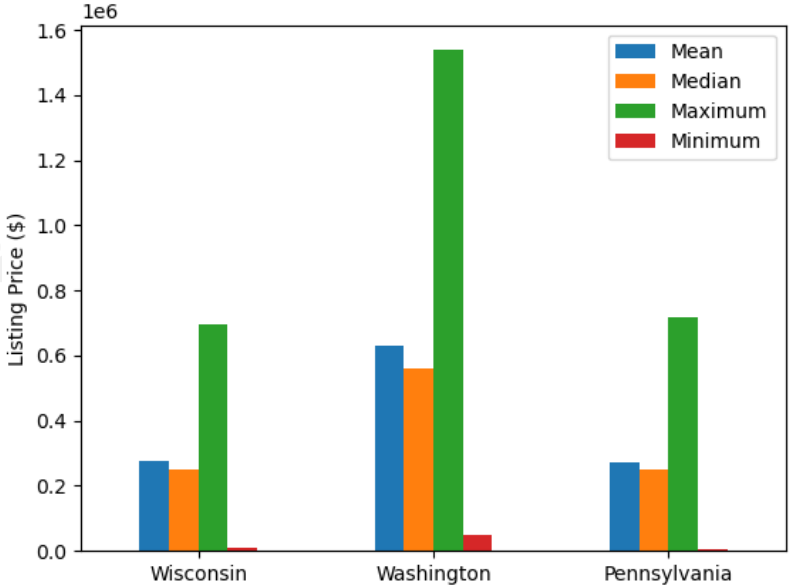


Regression Analysis: Collective

The graph on the left shows all respective ‘r values’ for each variables’ correlation to price.

The graphs on the right shows that Washington has the highest price measures of central tendency/max and min.

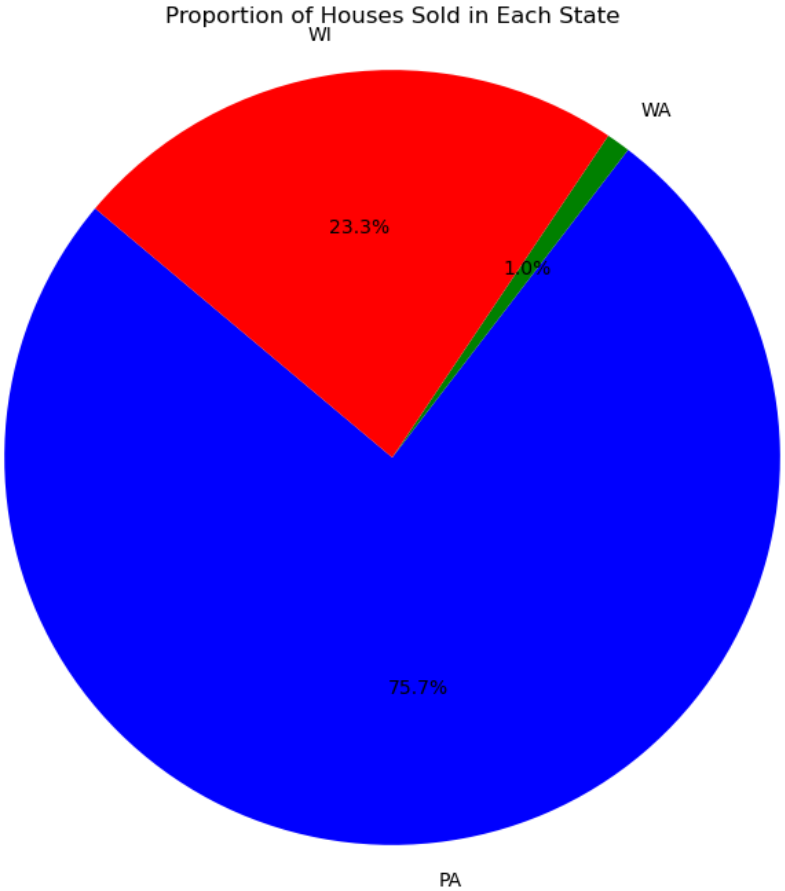




03

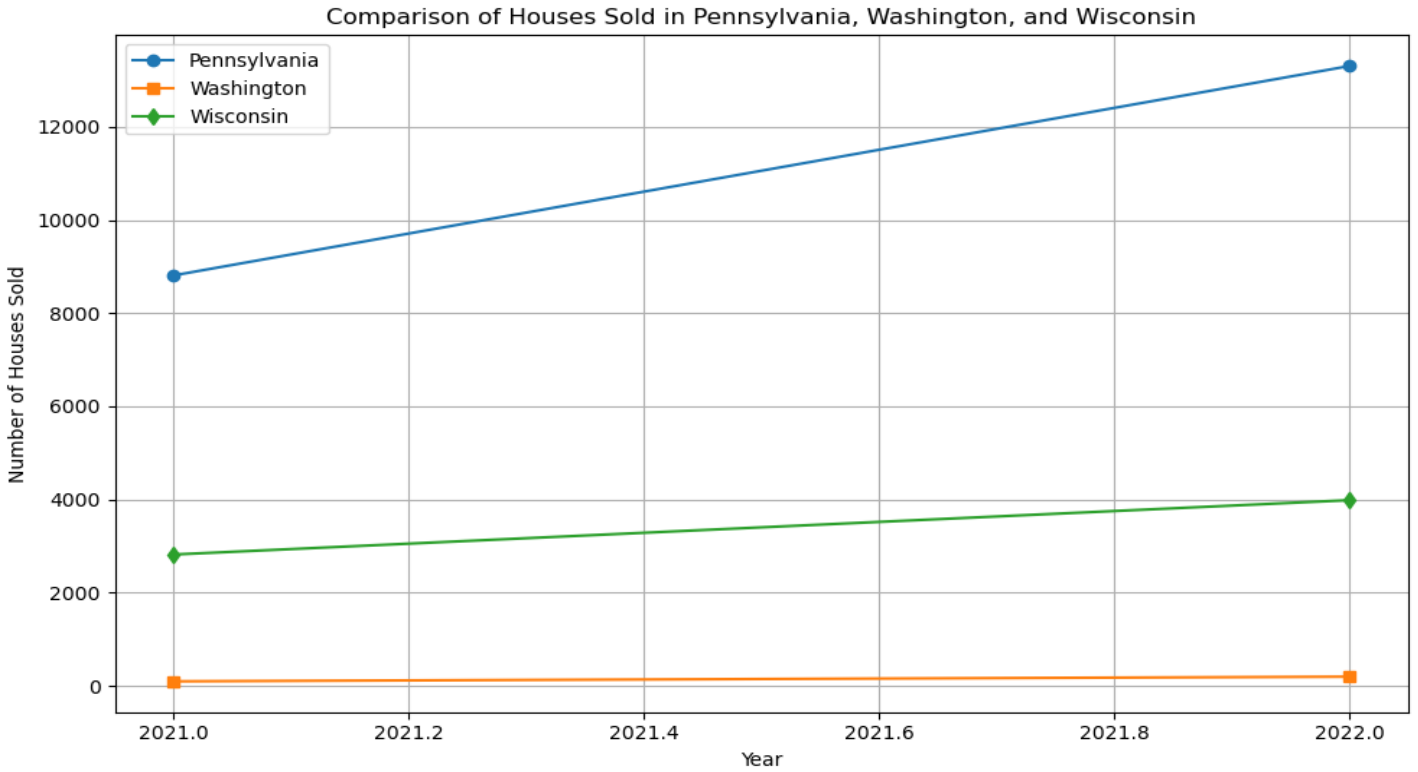
Time Analysis

Pennsylvania greatly outperforms its counterparts with 75.7% of all the “sold” homes within our dataset. Coming in second is Wisconsin with 23.3% followed by Washington at only 1%. This data clearly shows the strength of the Pennsylvania real estate market, especially as it relates to inventory velocity/turnover.



Time Analysis

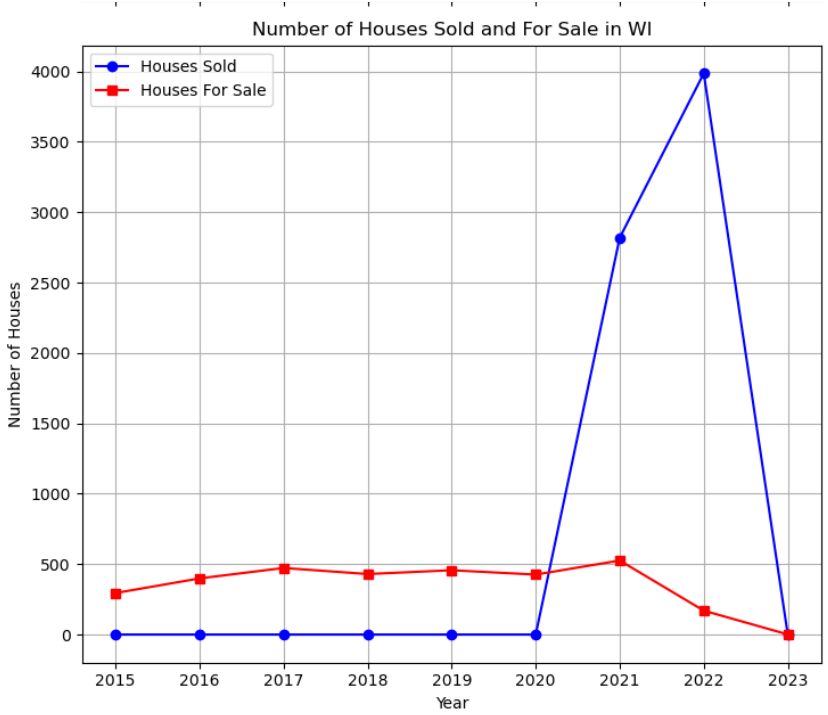
Adding to the validity of Pennsylvania's high degree of inventory velocity, we can see the rate of increase or slope for the state is significantly steeper than that of Wisconsin and Washington.

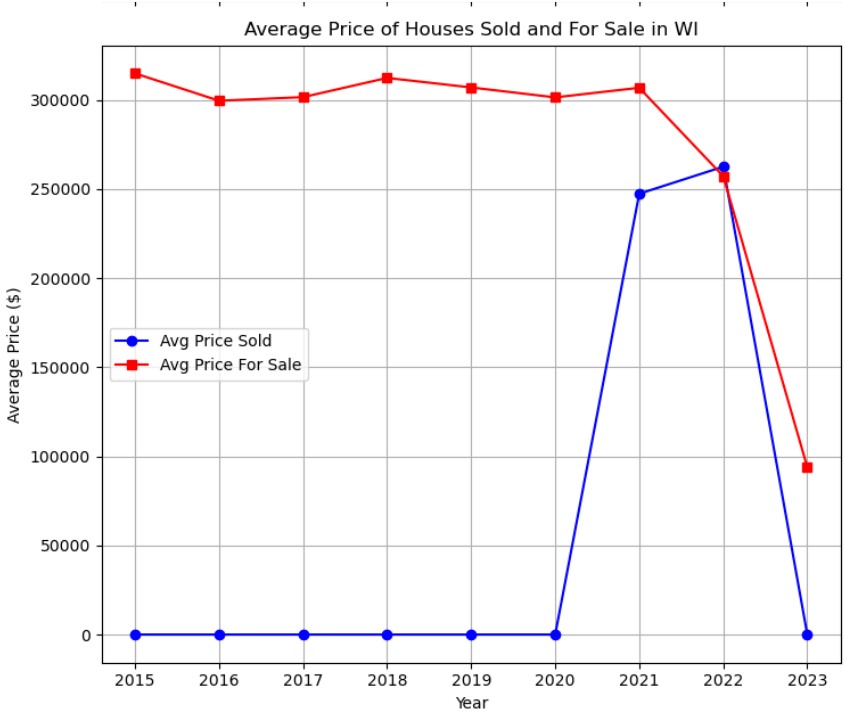


03

Time Analysis: number of houses/avg price (Wisconsin)

We can very clearly see that our hypothesis that the COVID-19 global pandemic had a significant impact on the U.S. real estate market is substantiated.

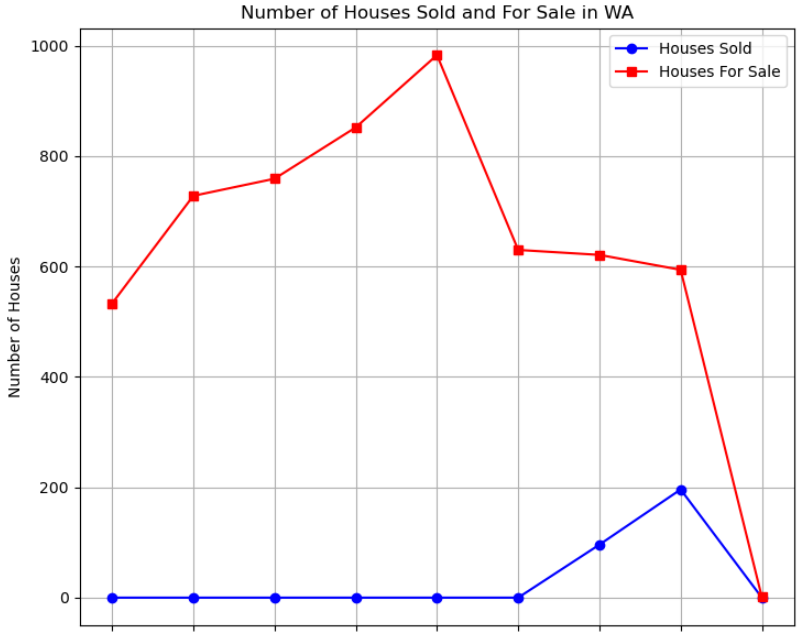


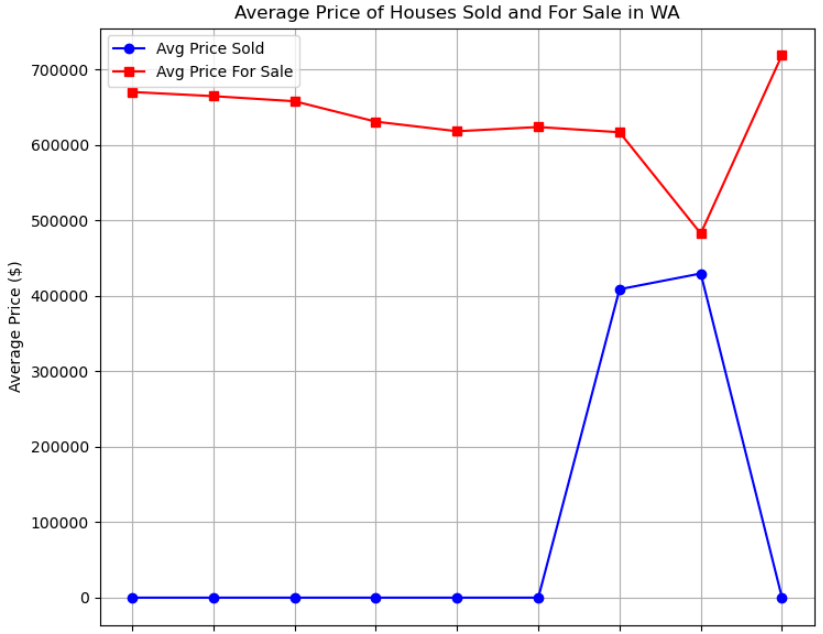


03

Time Analysis: number of houses/avg price (Washington)

According to the “Houses Sold” line depicted in blue for each graph, we see a remarkable increase in the total number of houses sold across all states in 2020 and thereafter. As an increase in houses sold represents an increase in market demand and subsequent decrease in available inventory, we can see the same distinct uptick visually in terms of average price for homes sold as well (right side graphs).

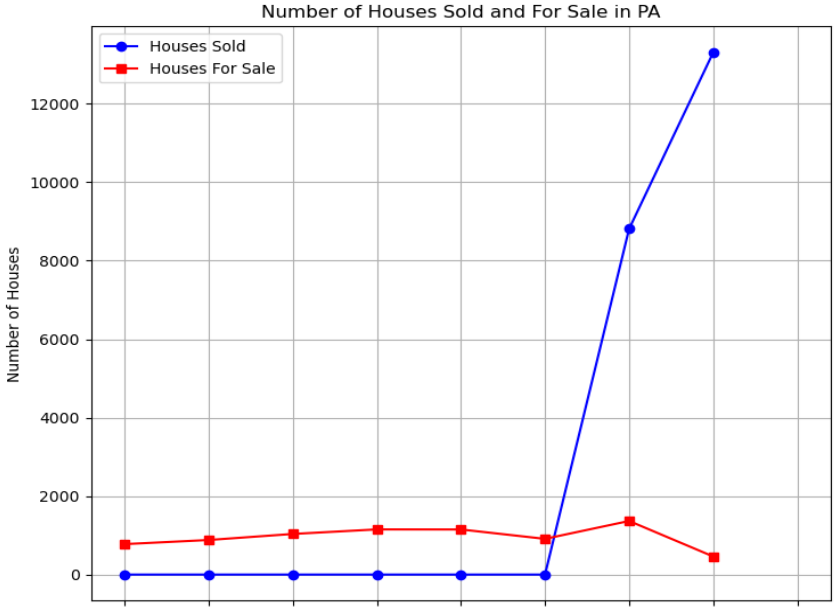


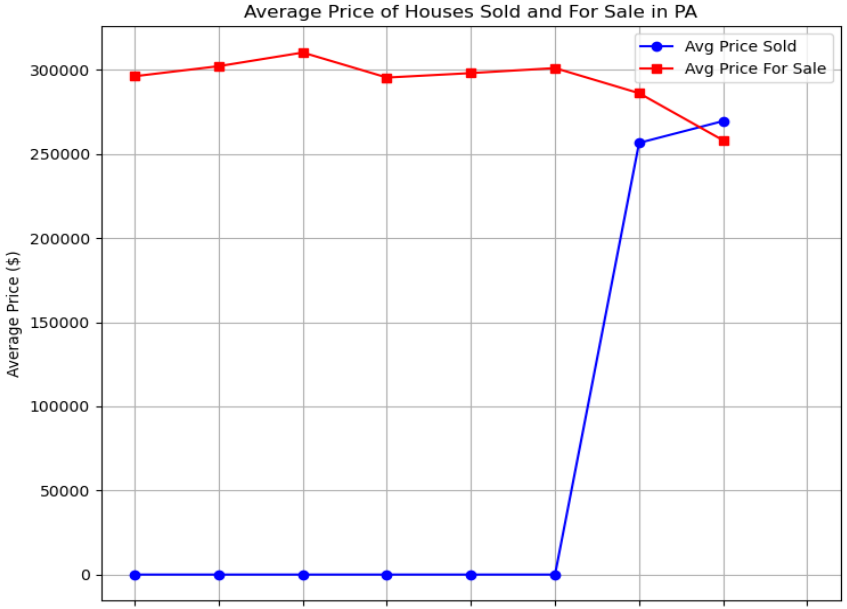


03

Time Analysis: number of houses/avg price (Pennsylvania)

For all three states we can see a fairly consistent delta between the number of homes “for sale” and homes “sold” and a subsequent shrinking occurring around 2020 (homes for sale out numbering homes actually sold). This shows that the pandemic was a key contributing factor to the housing market’s supply and demand change. In the cases of Wisconsin and Pennsylvania, this impact even results in the number of homes “sold” greatly outweighing the number of homes “for sale”.

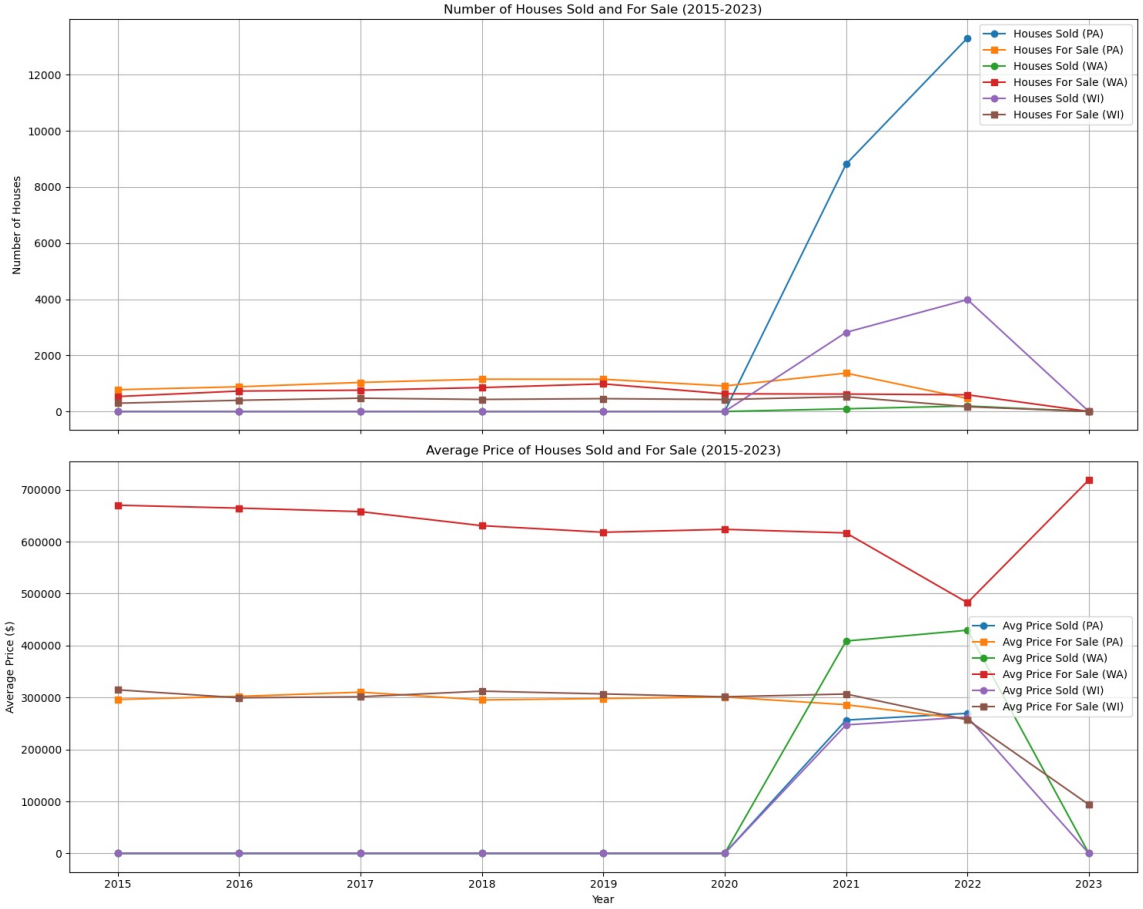




03

Time Analysis: comprehensive (all three states)

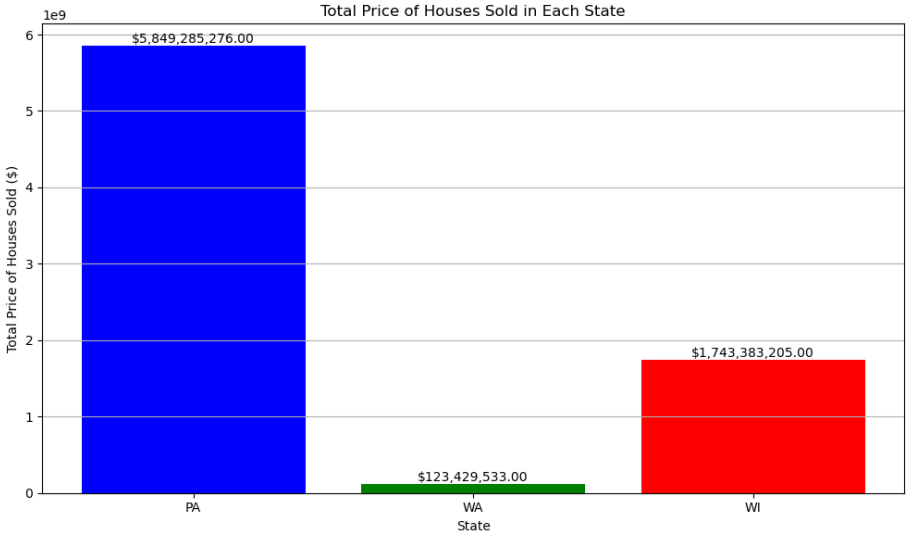
In order to show these relationships in a collective manner for our dataset, we lastly created a visualization showing the “for sale” and “sold” trendlines for each state on a single graph for number of houses sold and average price. This further demonstrates COVID-19’s impact on the real estate market and represents that impact for each state allowing comparison.

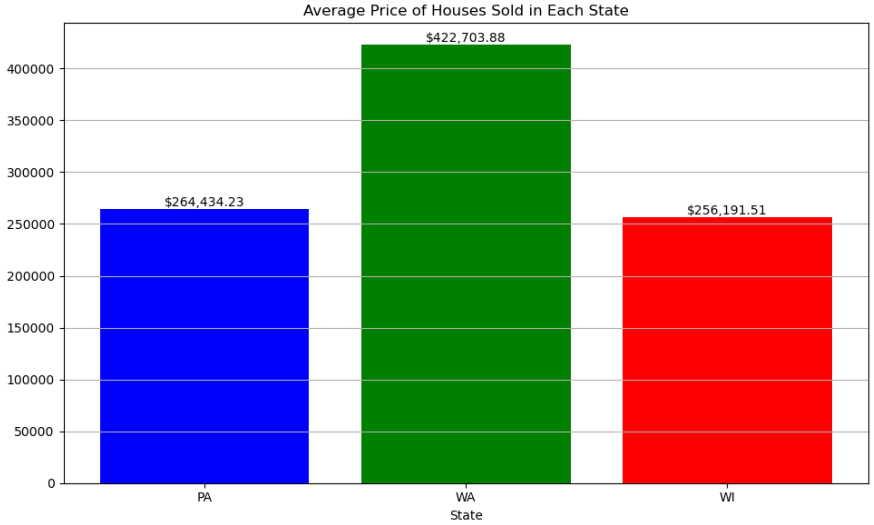


04

Collective Summary

Finally, pulling from a number of our conclusions we wanted to provide some implications which we could potentially provide to a market participant. In an effort to do so and in conjunction with our above analysis, we created two additional visualizations that display the total dollar amount sum of all houses sold as well as the average price for a home sold within each state. As you can see, the first graph shows the “best” performing state was Pennsylvania with $5,849,285,276, the “worst’ performing state was Washington with $123,429,533 and Wisconsin in-between the two with a total dollar amount sum of $1,743,383,205.When looking at the second graph, our calculations display Washington had the highest average price for a home sold at $422,703.88, followed by Pennsylvania at $264,434.23 which was very slightly ahead of Wisconsin with a value of $256,191.51.





Final Conclusion

Given the entirety of our analysis, we would suggest a few key points to consider for a potential real estate developer or investor looking to put money into the real estate market within one of our three states.

First, our data suggests it's a solid time to be investing in the real estate market as evidenced by a strong uptick in demand mirrored by an increase in average pricing across the board. Although, measures of central tendency, maximum/minimum and average price of a home sold results clearly are in favor of the Washington real estate market, the poor inventory velocity within the state could represent a "riskier" investment. While the return from a single house sold in Washington would indeed outweigh the return experienced in Wisconsin or Pennsylvania, the quantity of houses being sold and in-turn the total sum of dollars is maximized within Pennsylvania.

As such, if the investment strategy revolved around casting a "wide net" with multiple properties, we'd suggest Pennsylvania would represent the best option within our three states. Furthermore as concluded via our regression analysis, within Pennsylvania one should place more emphasis on a property's number of bathrooms first, followed by the number of bedrooms above other characteristics.

Lastly, outside of the real estate industry specifically, policy makers can utilize this analysis to inform decisions surrounding future growth, supply and demand, specific legislation/programs to adequately influence their local real estate market.

Thanks!

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**