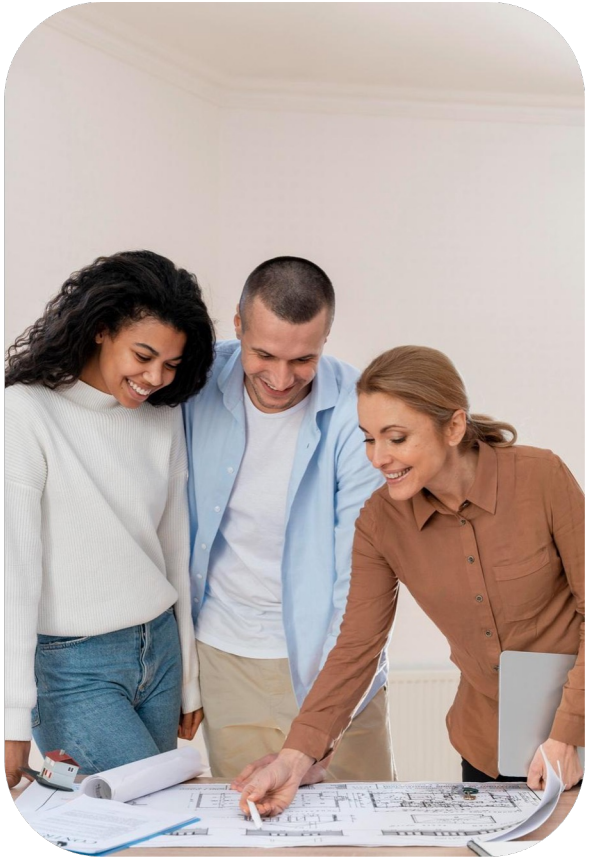
Real Estate Analysis Project

Presented by: Makonnen Ramsey, Kinjal Sakhida and Kristin Bell

Table of contents

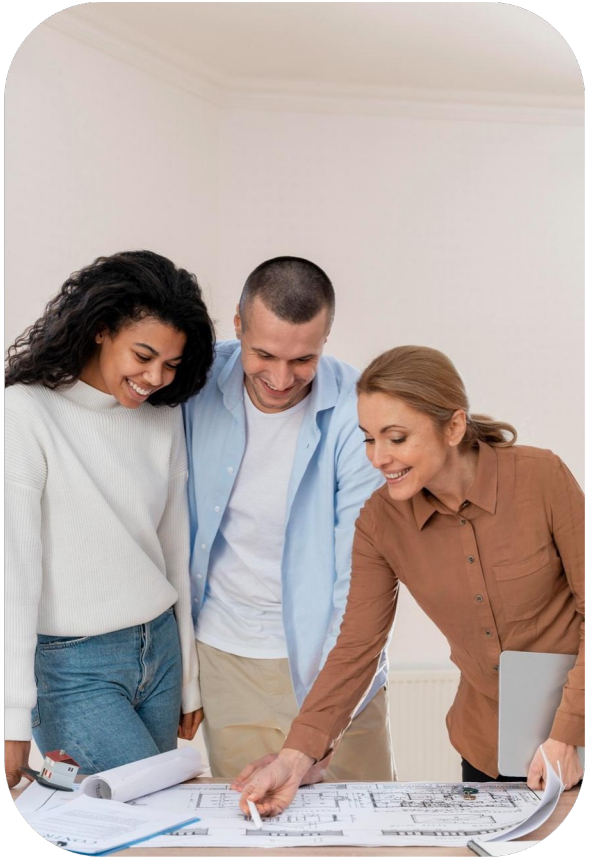


01 Introduction

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

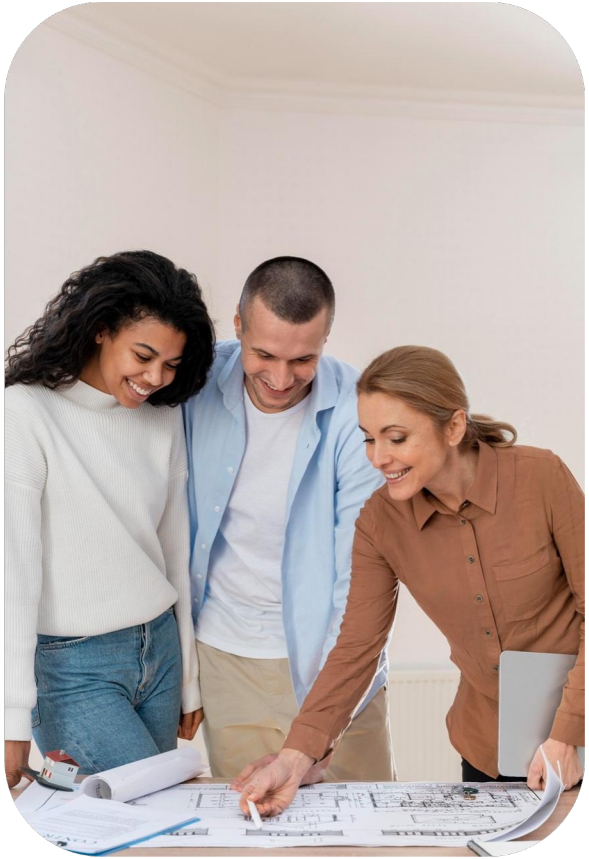
What correlations exist for variables such as number of bedrooms, bathrooms, square footage, etc. and how much impact do they have on house price for each location?

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

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

What represents a “typical” or “average” house for each state and how do these metrics compare?

What are some conclusions from our analysis that we could relay to a real estate market participant? Are there other professionals outside of the real estate industry that could gain insights from our analysis?



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.

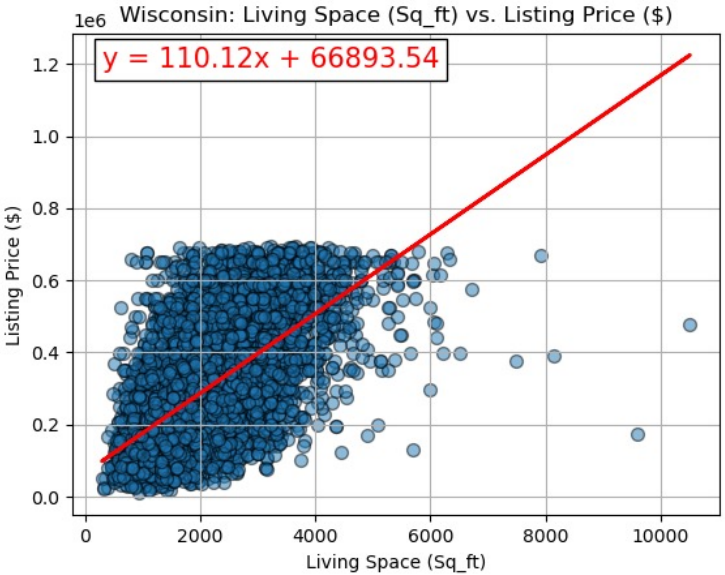
Next, we chose Washington, Wisconsin and Pennsylvania as a cross section of the United States (West, Mid-West and East) to examine in depth.

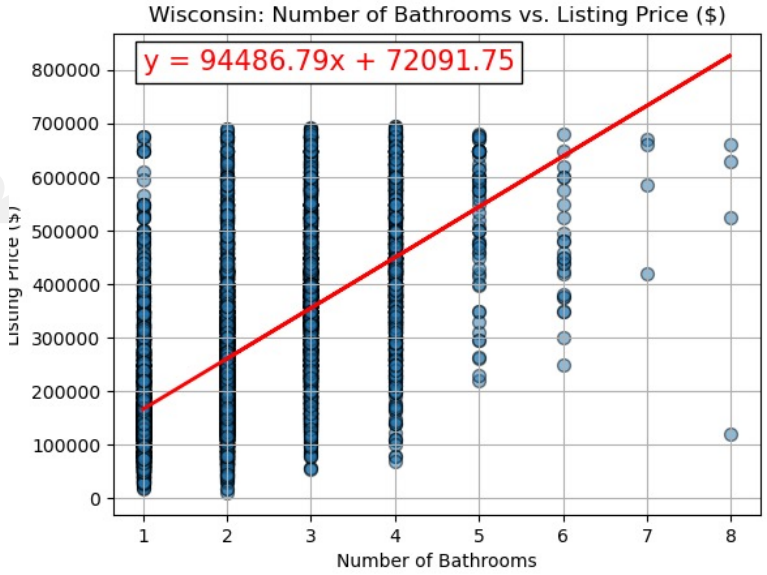
Finally, we determined and eliminated outliers via interquartile range calculations. Also, we eliminated listings with a price of less than $5,000. This was executed to remove inaccurate listings due to data entry error (i.e. listings with a $1.00 value listings with information entered into the wrong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

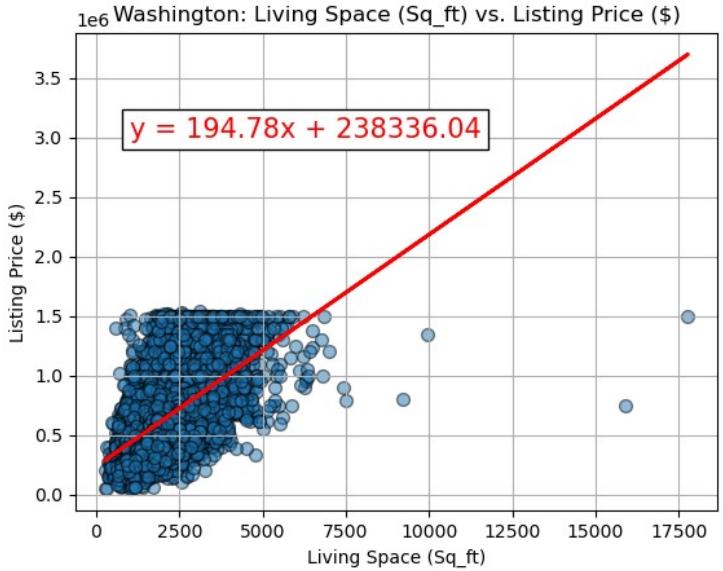


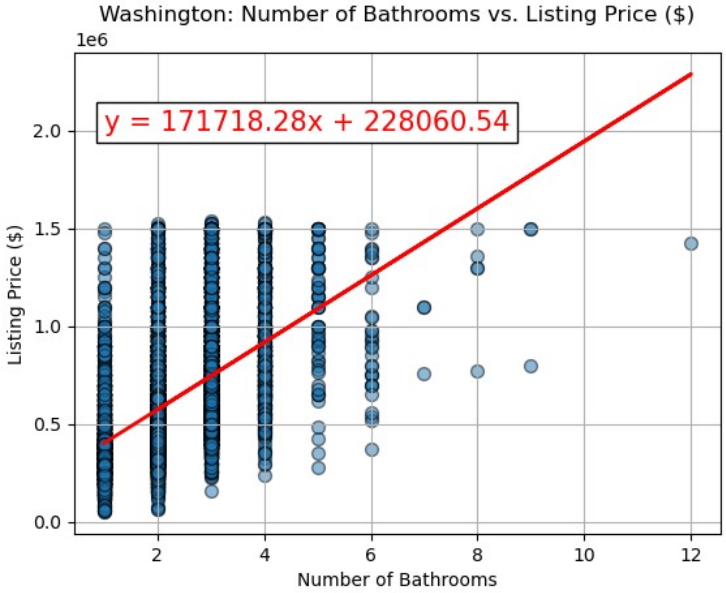


Regression Analysis: Washington

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.

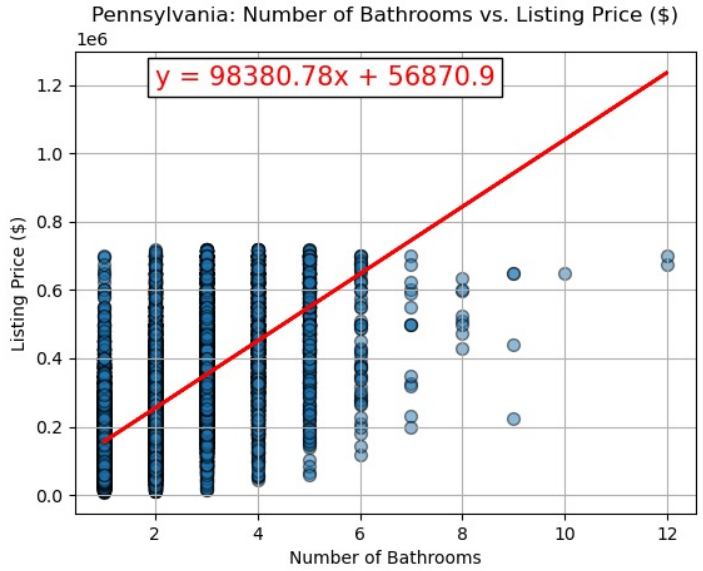


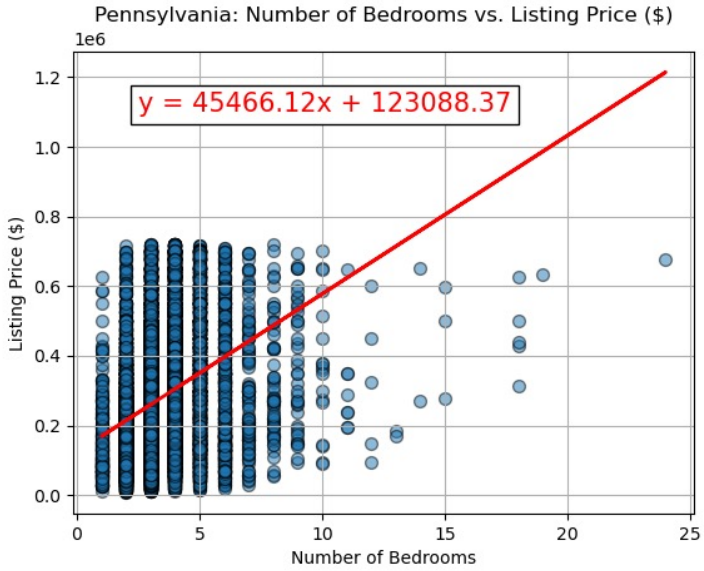


Regression Analysis: Pennsylvania

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.

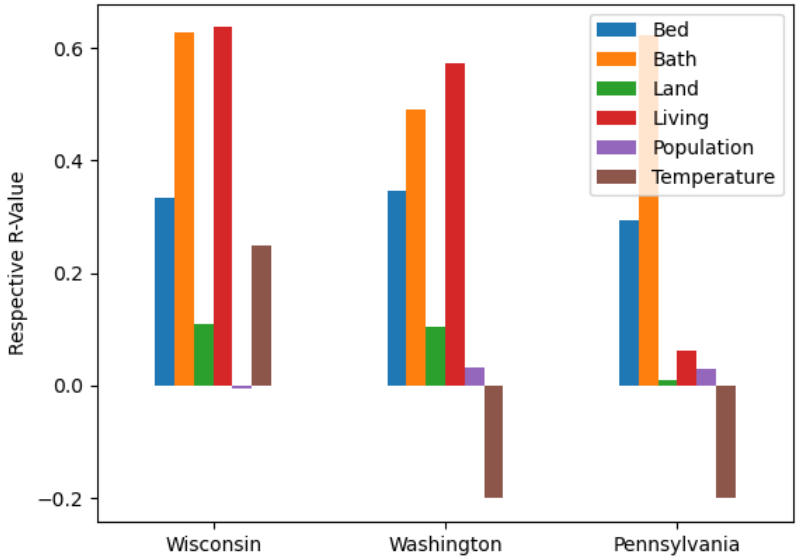


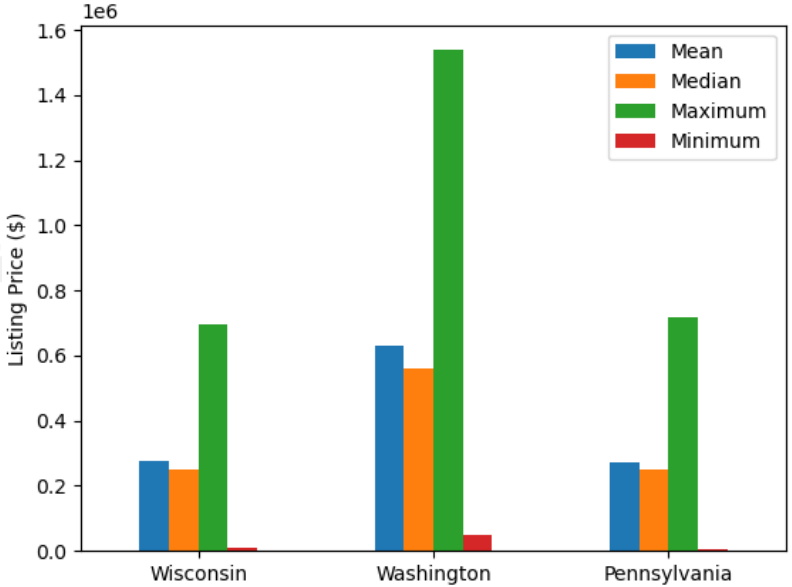


Regression Analysis: Collective

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.

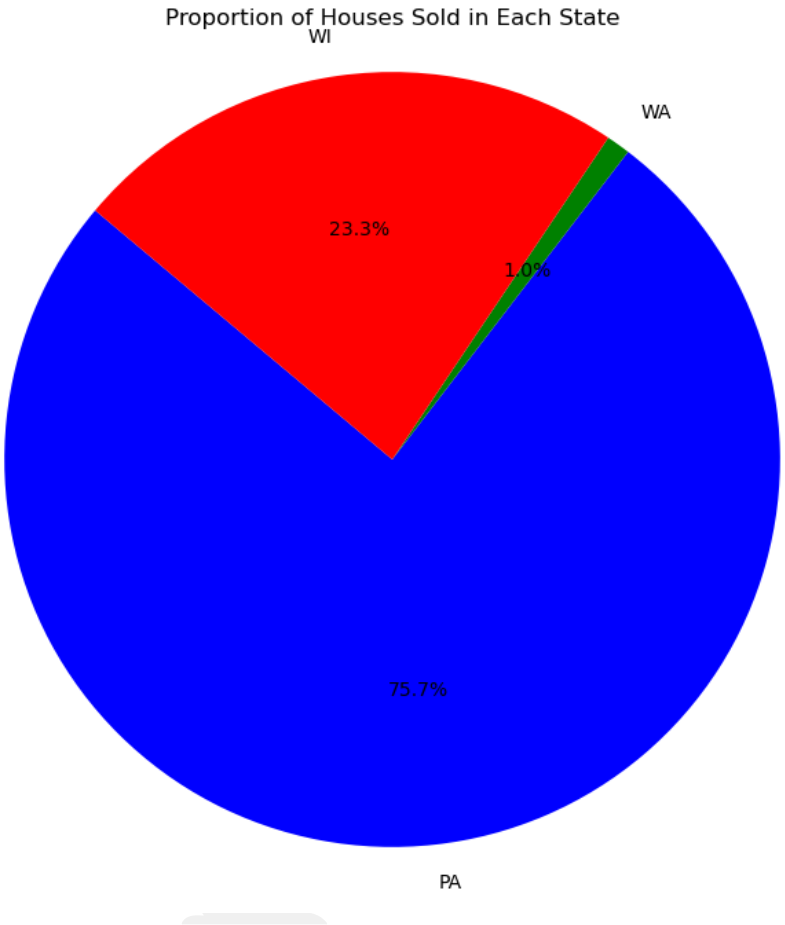




03

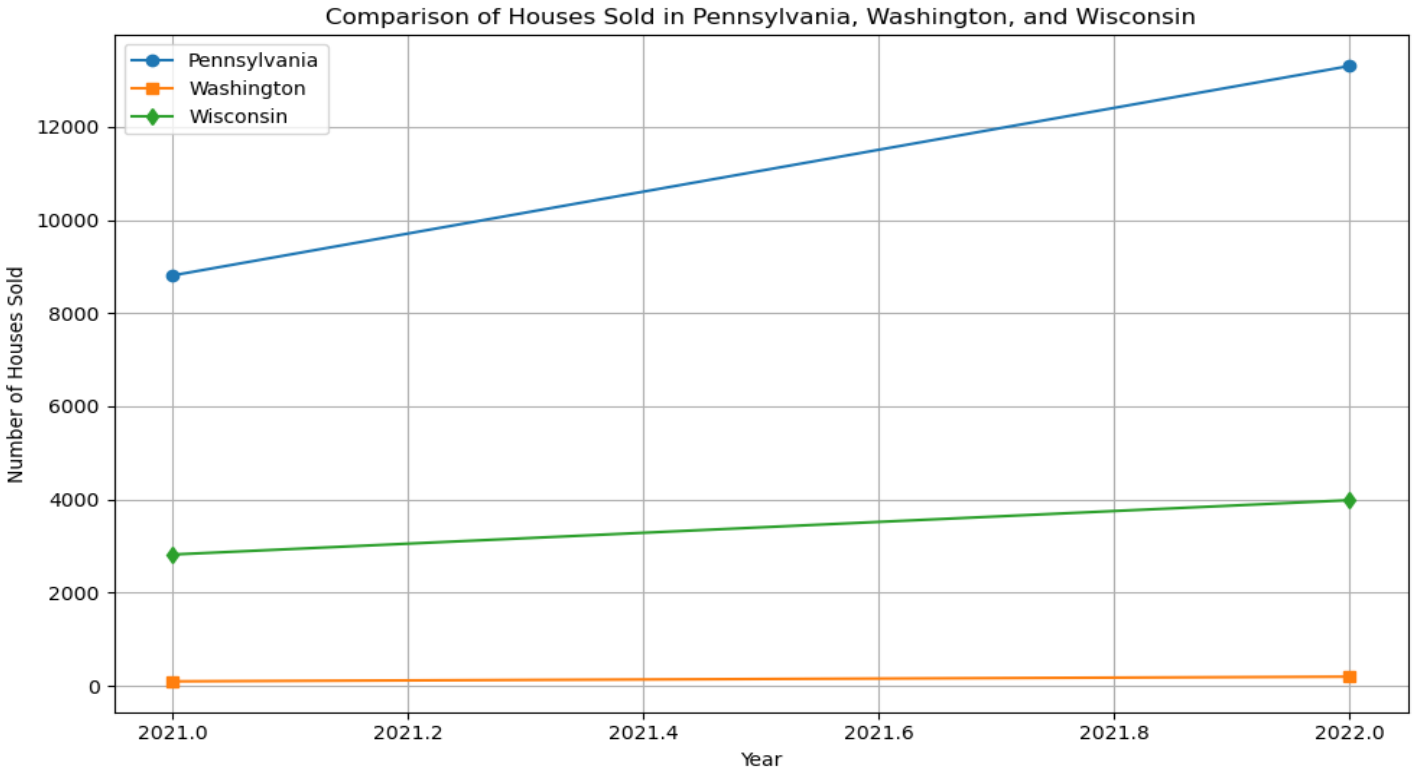
Investment Analysis

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.



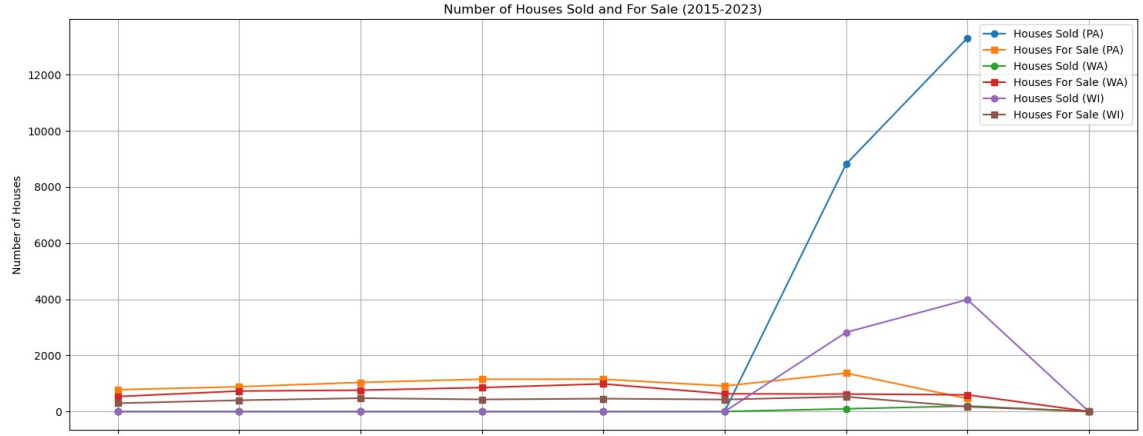
Investment Analysis

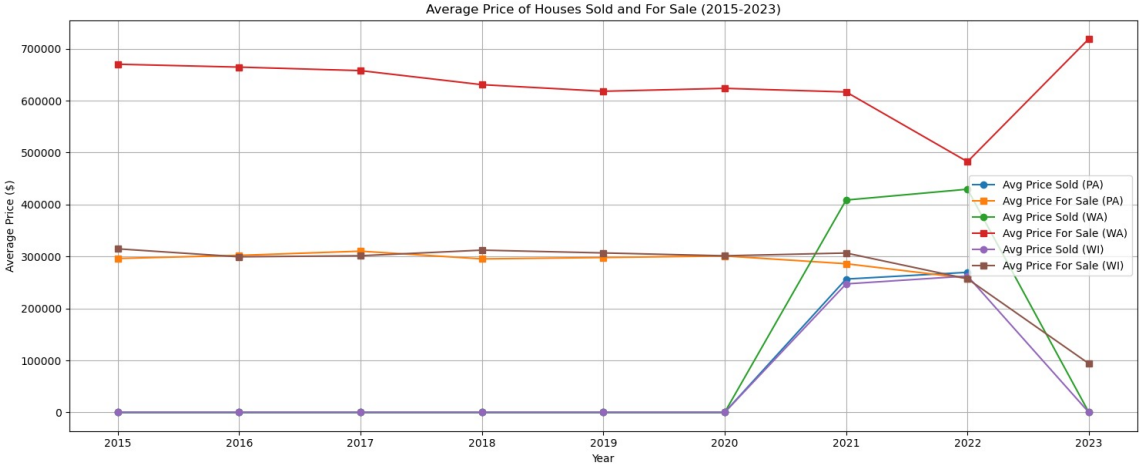
Adding to the validity of Pennsylvania's high degree of inventory velocity, we can see the rate of increase or slope for the state is significantly steeper than that of Wisconsin and Washington.



Investment Analysis - Time

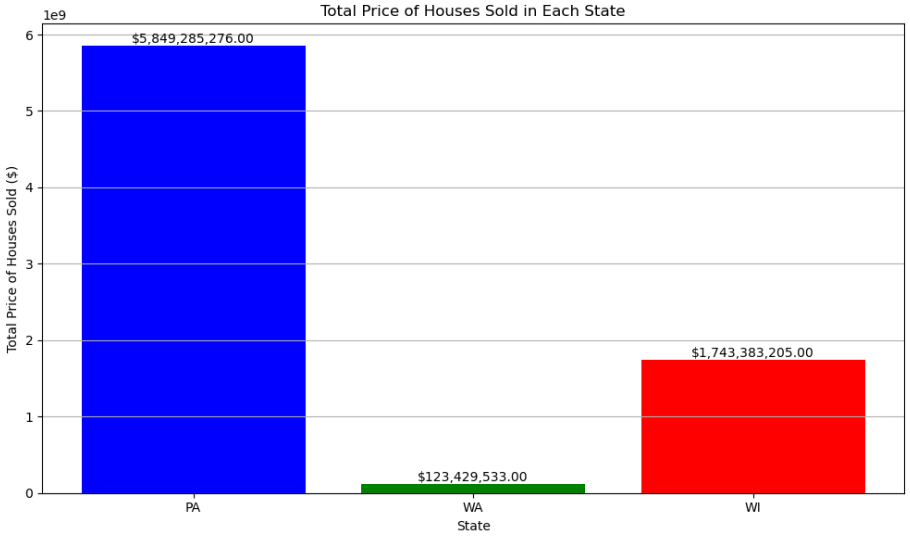
In order to show the impact of time in a collective manner for our dataset, we created a visualization showing the “for sale” and “sold” trendlines for each state on a single graph for number of houses sold and average price.

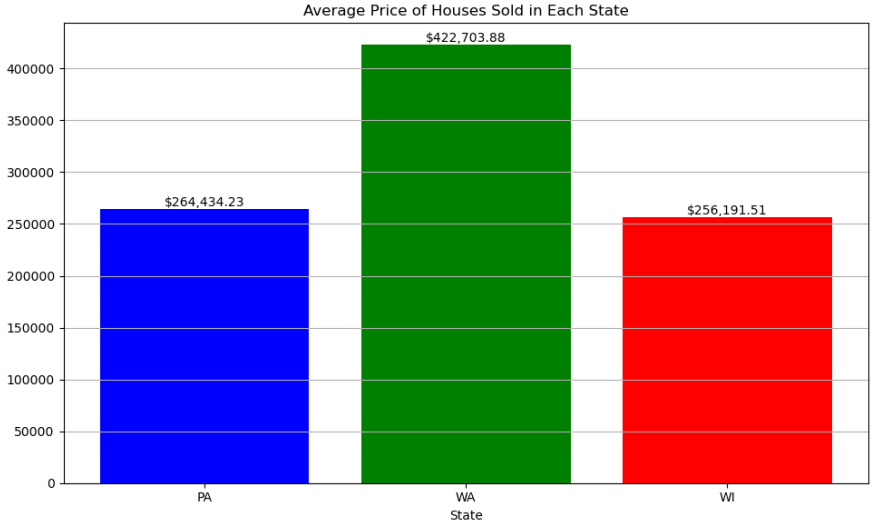




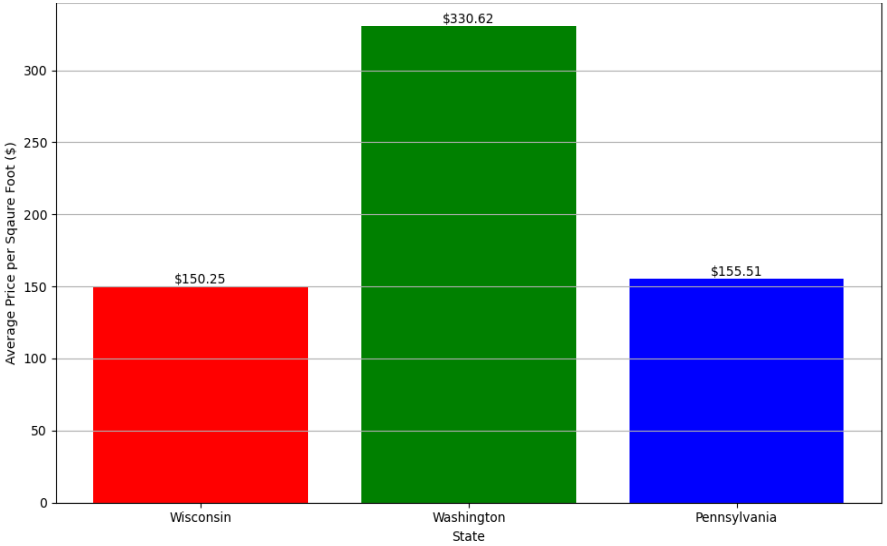
04 Investment Analysis & Collective Summary (1)

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 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 highest performing state was Pennsylvania with $5,849,285,276, the lowest 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.



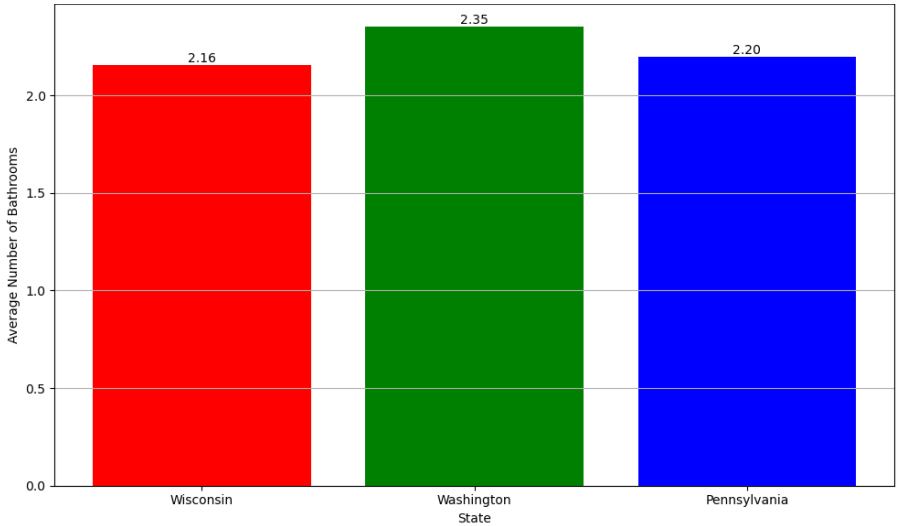


04 Investment Analysis & Collective Summary (2)



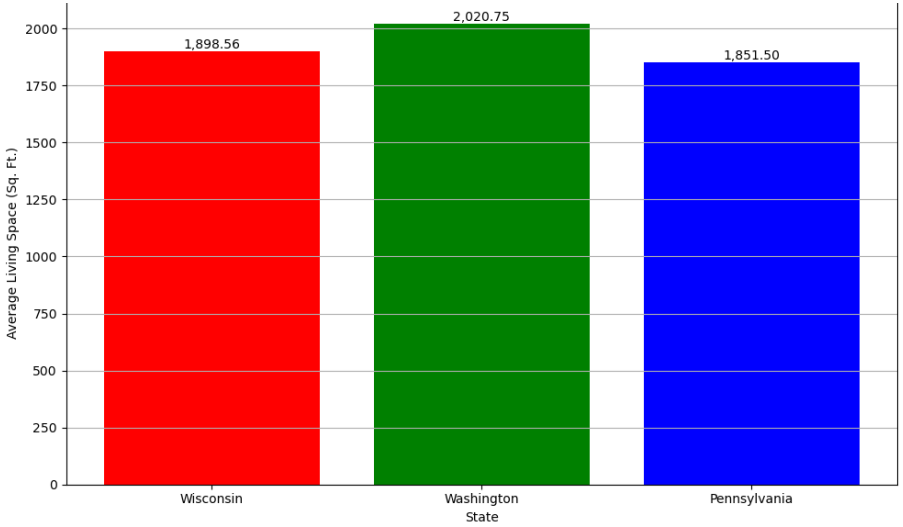
< Average Price / Square Foot

Average Number of Bathrooms >



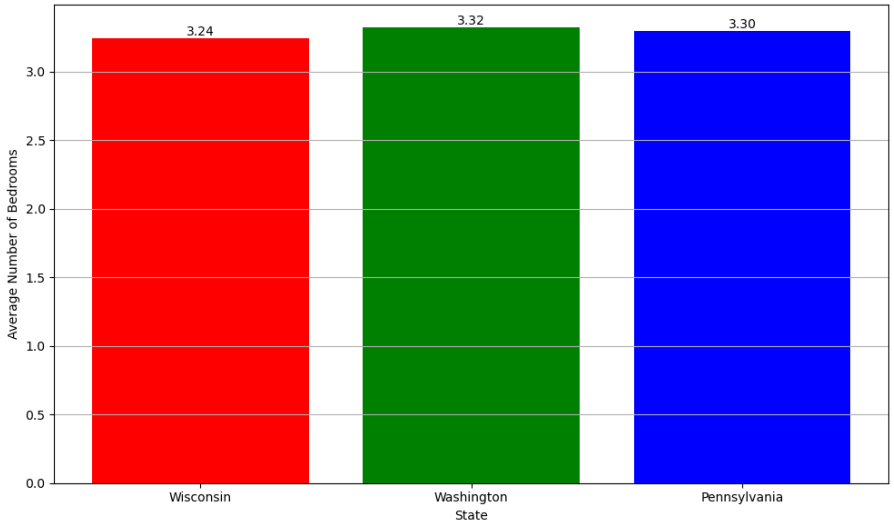
04

Investment Analysis & Collective Summary (3)



< Average Living Space (Sq. Ft.)

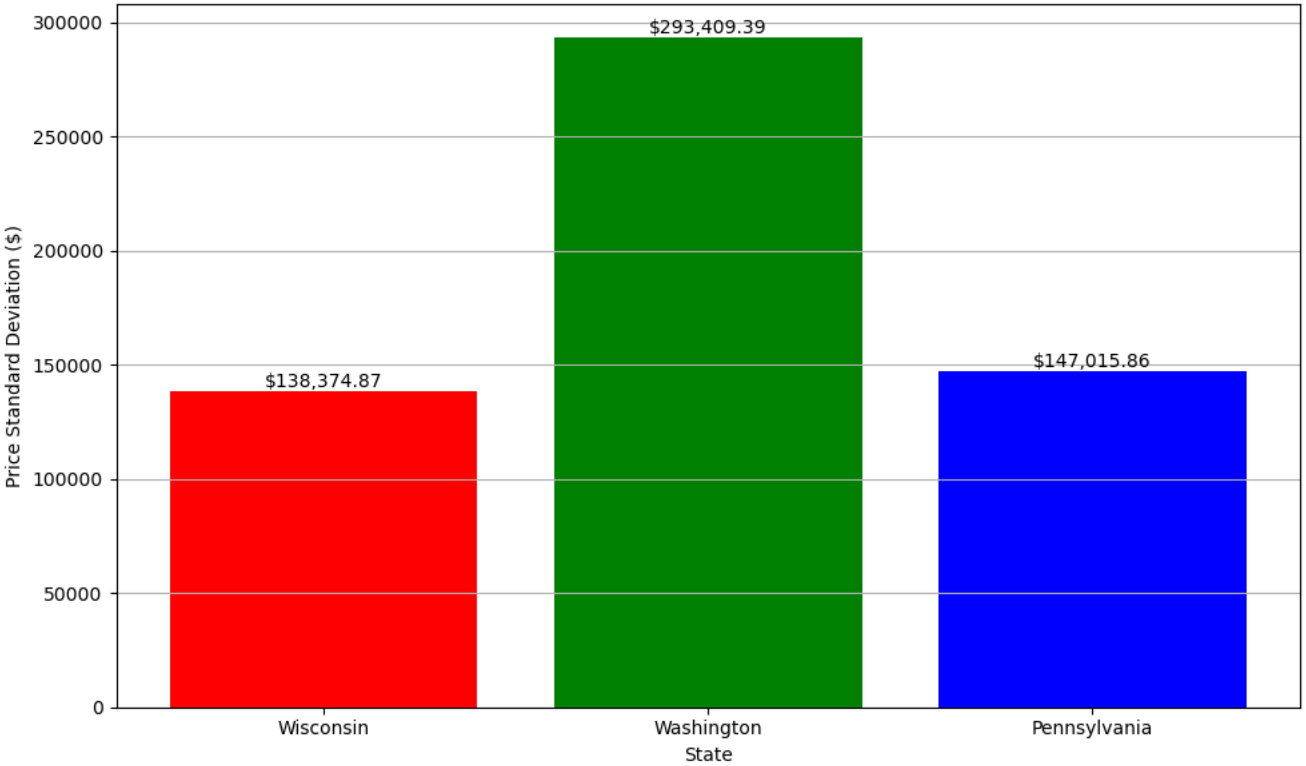
Average Number of Bedrooms >



04

Investment Analysis & Collective Summary (4)

We also included a bar graph to visualize and compare each state's standard deviation with regards to pricing to assess "riskiness". The listings within the state of Washington experienced the highest pricing standard deviation with a value of $293,409.39 followed by Pennsylvania at $147,015.86 very slightly ahead of Wisconsin with a value of $138,374.87.



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 every key metric (pricing measures of central tendency, maximum/minimum, average price of a home sold, average price per square foot, average living space, average number of bedrooms/bathrooms) is clearly in favor of the Washington real estate market, the poor inventory velocity within the state coupled with the highest standard deviation regarding price 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/highest rate of increase 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.

Outside of the real estate industry specifically, policy makers can also 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**