STAT 1361 Final Project Non-Technical Report by Clare Cruz

The purpose of this project is to increase PA-VA reality's pricing knowledge in the Pittsburgh and Richmond housing markets. This analysis involves pursuing two goals. The first one is investigating what influences the housing prices for the homes in Pittsburgh and Richmond. The second goal is to build a pricing model for the team of realtors to use for price comparison. To answer both of these questions, this analysis evaluated over a dozen models to see which one had the best accuracy rate. Then, based on the model with the best prediction accuracy, we calculated which variables had the greatest impact on housing prices. In the end, we found that the best model contains five variables that are the most influential on the housing prices in both markets. These variables are the property's age, square footage, the number of bathrooms, the lot area, and the zip code.

Looking at each variable individually, we see that they tell us a lot about the property pricing trends for these markets. To start, the property's age affects the pricing since the price generally increases as the age of the house increase. While this relationship is known, both Richmond and Pittsburgh are old cities. As a result, numerous older properties are still valuable and desired for their unique qualities. This attribute means that the newer properties are priced higher than older ones, but the difference in pricing between ages is not drastic. Next, we see that the number of bathrooms, square footage, and lot area are all important factors that pertain to the size of a property. These factors describe the limited capacity issue that is present in both markets. As a result, bigger properties are generally going to cost more. This relationship is especially true based on the location since large properties closer to downtown areas cost more than a property of the same size in the suburbs. This result leads to the biggest takeaway from this analysis, the impact of the zipcode on housing prices. In this analysis, zipcode is by far the most influential variable on the price. The reasoning behind zipcode's importance is the distinct and diverse neighborhoods present in both markets. In Pittsburgh, zip codes have a separate government, taxing structure, and landscape. Similarly, in Richmond, zip codes have definite characteristics, strict boundaries, and varying population demographics. So in both of these markets, location is key to pricing. In contrast, there is another takeaway about which factors do not contribute to the pricing. All of the cosmetic characteristics: exterior finish, roof type, the type of property, the number of stories, and whether or not the property has a basement, have little effect on the pricing.

Finally, we are confident that this best model will perform well and give our company an advantage with price comparison, given that the data is within the range given in the historical data set. The confidence is rooted in the model's predictive accuracy under testing conditions, which erred by a dollar on average. Given that the property prices ranged in the hundreds of thousands, we were satisfied with the model's average error. In the future, this model can be expanded to different pricing levels, property characteristics, and locations if the company wants to develop the model further.

In summary, this analysis provided PA-VA realty with a reliable model for price predictions and four critical takeaways about the pricing factors in the Richmond and Pittsburgh markets.

- The age of the property significantly affects the pricing. Older properties are generally priced less than newer properties, but some older properties can potentially have a higher value due to their uniqueness.
- The size of the property (number of bathrooms, square footage, and lot area) affects the pricing due to the size constraints within the two markets.
- The location of the property is critical to the pricing. The characteristics of the areas vary greatly between zip codes in both markets.
- Cosmetic characteristics (exterior finish, roof type, the type of property, the number of stories, and whether or not the property has a basement) do not matter for pricing in these markets.