EUROPEAN CENTRAL BANK

Market Research

"An Empirical Economic Investigation into the Corona Crisis' Impact on Fundamental Hedonic Features in Real Estate Valuation Models Using Machine Learning Methods."

Completed on 7th of March 2022 by:

Sawyer M. Benson
Sindlinger Str. 15
60326 Frankfurt am Main
E-mail: sawyer.benson.official@gmail.com

Acknowledgements

Throughout the process of writing this paper, I have received a great deal of support.

I would first and foremost like to acknowledge my supervisor, Yuki Sato, who provided me with constant direction and expert knowledge throughout the entirety of this paper. Your detailed feedback and suggestions pushed the quality of my research far beyond the point I could have accomplished without your support.

I would like to thank Real Estate Finance and Economics professors Prof. Dr. Daniel Ruf and Prof. Dr. Johannes Strobel and Machine Learning professor Dr. Jens Mehrhoff. Your courses directly provided the practical foundational basis and inspiration from which this paper was created.

I would also like to acknowledge the countless individuals who contributed brilliant advice on Stackoverflow.com, as well as the programmers behind the opensource packages for the programming languages R and Python. Your selfless work is pushing the many fields of scientific thinking forward and contributed greatly to this paper.

I would like to thank the Goethe Goes Global scholarship for funding my studies, allowing me to put my full energy into developing the skills needed to produce meaningful research in the field of Finance and Economics.

Lastly, I would like to thank my partner, Tonia Michaely, and my great friends, Vagan Esaian (Ваган Есаян) and Leo Bruckhuisen. Without your support, I would not have accomplished my academic and personal goals which underpinned this paper.

Abstract

The purpose of this paper is to understand the economic impacts of the Corona crisis on housing prices and how these impacts changed the relative demand of homeowners for certain hedonic features of a home. To determine which variables are most important, an eXtreme Gradient Boost (XGBoost) machine learning algorithm was trained on a large (i.e., 104×24412*) dataset of hedonic features from Louisiana, USA market data. An analysis of the XGBoost's variable-importance ranking shows that number of bedrooms, city centrality, total living area, age, and days on the market are key variables in determining a home's market price.*

A panel of controlled analyses shows that Corona, measured by the daily 3-month moving average of official infections at the time of sale, significantly increased average home prices by 8.97 USD per additional daily infection. This finding establishes the Corona-specific reaction of housing market prices to daily infections. When analyzing how this price shift is explained by changes in specific demand for certain property characteristics, I find: the premiums for each level of number of bedrooms is significantly increased in response to daily infections, with an average increase of 32, 37, 27, and 47 USD per additional daily infection for levels 2-through-5 bedrooms respectively; the premium for being central to a city increased by 5.17 USD per daily infection while properties located outside the city were not significantly impacted by daily infections; premiums for home size increased by 0.02 USD per square foot of living area per additional daily infection, with the premium for the smallest homes being significantly decreased by 3.15 USD per daily infection; premiums for property age, which are historically negative, decreased even further with the penalty for each additional year of age increasing by 0.06 USD per daily infections with the oldest homes experiencing the largest loss of 2.75 USD per daily infection per year of age and; an increase in the penalty for each additional day a home sits on the market of 0.04 USD per daily infection with homes sitting the longest period of time on the market experiencing the largest loss of 2.75 USD per daily infection per day.

These findings lead way to a greater understanding of how housing markets are impacted by wide-spread global pandemics such as the Corona crisis and contribute to the larger discussion around how homeowners, investors, and policy makers can act in the face of future crises.

Table of Contents

Index of Figures

Index of Tables

1. Introduction

In late 2019, a virus first detected in Wuhan, China would set in motion a global pandemic which, by the end of 2021, will have killed 4.3 million people, infected 238 million others, and disrupted the global economy across virtually every measurable dimension (WHO 2021). According to seven economic impact models constructed by McKibbin and Fernando (2020), estimates of the total global economic loss in terms of GDP are measured to be as large as 9.2 trillion USD. However, despite a generally positive correlation between GDP and real estate prices (see Figure 1), historically measured as high as 98%, real estate market prices in the US hit near-record heights following the outbreak of Corona (Anissanti 2021). According to a report released by Zillow Analytics (Manhertz 2021), U.S. real estate gained 2.5 trillion USD of value in 2020 alone, representing the largest single-year growth since 2005, despite an approximately 760 billion USD decrease in GDP in the same year (FRED 2021).

The focus of this paper is to investigate the economic impact the Corona crisis had on housing prices and how these impacts changed the relative demand of homeowners for certain hedonic features of a home.

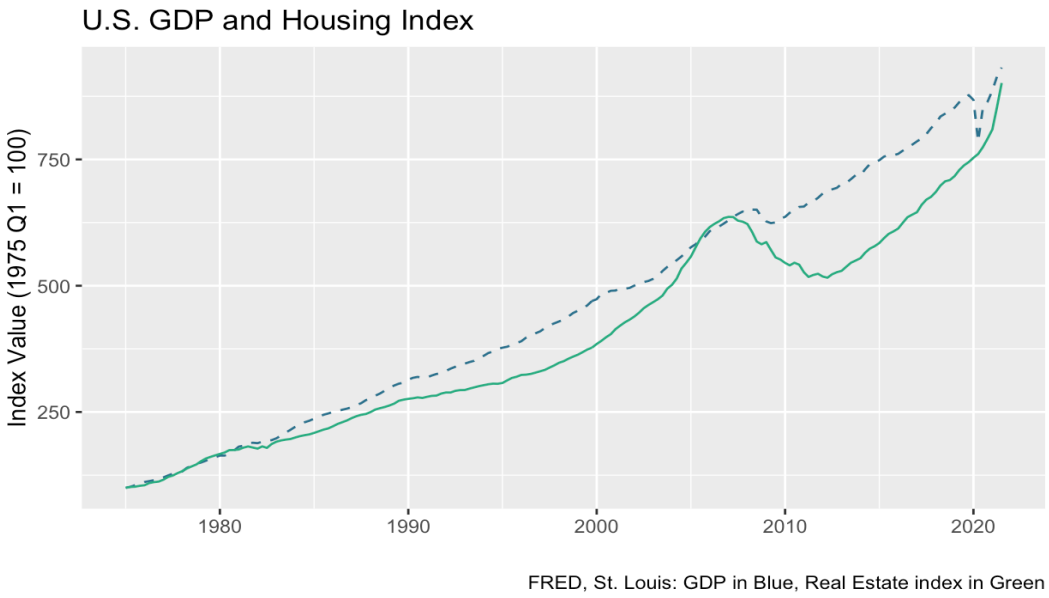

Figure 1 *U.S. GDP and Housing Index*

In the remainder of this paper, I will apply the Hedonic Pricing Method (HPM) to Louisiana housing market data in order to inferentially describe the economic impact of the global pandemic on residential housing market values. Furthermore, I will take advantage of the HPM's structural framework of using real estate properties' hedonic features (e.g., size, age, number of bedrooms, etc.) to test for changes in demand for specific property features pre versus post pandemic. The HPM will be econometrically modeled using an Ordinary Least Squares (OLS) regressions framework for specific variable analysis while several variations of machine learning (ML) prediction models will be estimated to test different independent variables' maximum explanatory power in predicting out-of-sample observations. The results of these models will shed light onto the real estate pricing dynamics within the Corona pandemic.

2. Literature Review

2.1 Background

The market value of a commodity is most often theoretically defined as the equilibrium price derived from the basic economic principal, or law, of supply and demand (Locke and Engels 1691; Epple 1987). However, the real estate market often violates this assumption due to its unique characteristics as an asset class (Wheaton 1999). For example, much of the underlying utility of a property is its use as a means of shelter by its owner (LING, OOI, and LE 2015). This rather unusual relationship to this asset introduces several behavioral biases which cause economic frictions not accounted for by traditional neoclassical economic theory (Nicolaides 1988). A notable example of behavioral bias impacting real estate price dynamics is the endowment effect. This behavioral finding was originally established by Kahneman, Knetsch, and Thaler (1990) in the late 20th century, and later applied to real estate markets by BAO and GONG (2016). The latter of the two stating that the predictably irrational behavior of market participants to overvalue their home due to sentimental attachment to the property forces market prices into sustained economic disequilibrium. Other highly cited unusual characteristics are that real estate assets are very infrequently traded due to high transaction costs (Collett, Lizieri, and Ward 2003; Guilkey, Miles, and Cole 1989), governments tend to interfere, both directly and indirectly with real estate markets through the creation of fiscal and monetary policies (Bingyang, Jie, and Yinhan 2013; Du, Ma, and An 2011), such as through creating renter-protections laws such as 'squatter's rights' laws which allow a renter to remain in a home for extended periods of time long after they have stopped paying rent (Hoy and Jimenez 1991; Gardiner 1997).

2.2 Real Estate Valuation Methods

The idiosyncratic asset features outlined in section 2.1 along with a high level of heterogeneity across many dimensions of the entire real estate asset class makes the creation of a generalized pricing model difficult and have led to a wide range of proposals and recommendations about what determines the market price of real estate assets and how to reliably models can model those pricing dynamics (Curcuru et al. 2010). In Pagourtzi et al. (2003), the authors outlines several of the currently accepted real estate valuation methods, ranging from what they categorize as the *traditional methods*, such as comparable-group, cost-basis, income-multiple, profit-multiple, and contractor's method, to the *advanced methods*, such as ANNs, spatial analysis methods, fuzzy logic, and the hedonic pricing method. According to a meta-analysis conducted by Sirmans et al. (2006), currently, the most widely used and accepted advanced methodological framework for real estate valuation modeling is the Hedonic Pricing Method.

2.2.1 Hedonic Pricing Model in Real Estate

First applied in 1939 on automobile data, according to Goodman (1978), the HPM is a model which estimates the value of distinct characteristics of a commodity which directly or indirectly contribute to its market value. Besides its implementation in real estate finance and economics, such as in this paper, this methodology has a wide range of applications such as its implementation in consumer and market research (Holbrook and Hirschman 1982; Arnold and Reynolds 2003), construction of consumer price indices (Moulton 1996; Schultze 2003), various tax assessments (Berry and Bednarz 1975; Bernasconi, Corazzini, and Seri 2014), automated automobile valuation (Cowling and

Cubbin 1972; Matas and Raymond 2009), and computer sales (Dulberger 1987; Wakefield and Whitten 2006).

Since its introduction, the HPM has gain significant popularity among housing market and commercial real estate researchers. The specific real estate-based topics include, but are not limited to, the construction of housing price indices (Gouriéroux and Laferrère 2009; Wallace and Meese 1997), the estimation and prediction of a property's market value in situations where market-transaction data is low-dimensional or non-existent (LeSage and Pace 2004) and, as in this paper, the specific analysis of changes in the demand for specific property characteristics across time, subgroups, or both (Clapp and Giaccotto 1998). As the broad search for a satisfactory modeling framework focuses in on the HPM, another debate arises regarding the best functional form of this method. Traditionally utilizing the standard OLS framework (Pace and Gilley 1998), researchers are increasingly utilizing a variety machine learning algorithms to accomplish an increasingly more refined set of findings.

2.2.2 OLS and the Hedonic Pricing Model in Real Estate

Unsurprisingly, regression analysis is the preferred estimation approach among real estate researchers when using HPM for price estimation. These multiple regression analysis methods are most often either an Ordinary Least Squares (OLS) regression or a Maximum Likelihood approximation of the log-likely equation derived directly from the hedonic function. Each of these estimation methods take a functionally similar path as they both estimate a vector of parameters (i.e., beta coefficients) that best fits the explanatory hedonic variables to the associated market price. They differ only by the loss function used in the identification of that best-fitted parameter vector.

The most used hedonic price regression equation, with respect to real estate markets, models the relationship between market rents or market property values to a list of hedonic characteristics. The classical construction of this model according to Herath, S. K. & Maier, G. (2010) is the following:

Equation 1

$$R = f(P, N, L, t)$$

where R is rent or price of the property; P is property related attributes; N is neighborhood characteristics; L is locational variables and t is an indicator of time.

2.2.3 Machine Learning and the Hedonic Pricing Model in Real Estate

Though first introduced by Turing (1950) under the broader umbrella term of *artificial intelligence*, the adoption of ML methods in real estate would take many years of software and hardware development, allowing or the subsequent collection of ever-growing data sets and central processing units (CPUs) capable of processing the often extraordinary number of calculation required to produce a solution for a given algorithm (Dutta 2018). The primary advantage of ML techniques is that ML algorithms learn and improve over time and across many iterations and variable combinations, while tradition statistical and econometric techniques produce static results across a single model (Anguita et al. 2010). The algorithm improves across those iterations as it seeks to minimize the model's error in predicting observations not previously seen, often called *out of sample* observations. By doing this, the algorithm seeks to find the strongest general relationship between the independent variables and the dependent variable instead of seeking to minimize the error within a given sample set. This in turn results in stronger, more generalized interpretation of the findings.

Mohd et al. (2020) provides a thorough overview of the various applications of ML to real estate valuation methods, including the Ridge and Lasso regression techniques, as well as artificial neural networks (ANN) and gradient boosting techniques used in this paper. The author concludes that ML models outperform other standard valuation methods such as multiples and OLS valuation methods in both accuracy and strength of interpretation. The best performing of these models was a gradient boosting model, which was able to capture the deep and often-hidden layers of interaction between hedonic features.

2.3 Corona Crisis' Impact on Real Estate Markets

In the wake of the Corona crisis, there were several papers and articles regarding the economic impact of the global pandemic on the housing market being expeditiously publishes in virtually every major journal. These papers investigate topics such as structural and temporal changes in the housing market using hedonic methods (Shimizu et al. 2010), changes in housing market demand for specific property types and features (Tajani et al. 2021), potential changes in housing preferences due to the Corona pandemic and highlight the challenges for policy making under such conditions (Nanda et al. 2021). Each of these papers result in findings which establish a statistically significant impact of Corona on relative demand for certain hedonic features in each respective real estate market. Later in this paper, I will establish a similar finding in the Louisiana housing market.

3. Data

3.1 Data Collection

The utilization of Big Data collected through a data-mining process called *web-scrapping* has increasingly become the method of choice for researchers across virtually all disciplines. The term web-scrapping simply refers to the process of collecting structured data from websites using algorithms to automate the collection process. Methods, such as the ones I have implemented in this paper, have been used by established authors such as Anguita et al. (2010)**.**

In this paper, I have used a mixture of the programming languages R and Python, supplemented by several packages created by Selenium, to write an algorithm that collects the required hedonic variables for this research from the Multiple Listing Services (MLS). **Table 1** is a summary of the original data set's key features.

Table 1 *Original Data Summary*

Variable List *Structure and short discription*	
Name	Information
Date Range	23.10.2010 - 12.12.2021
Location	Louisiana, USA
Number of Variables	49
Number of Observations	31,280
Pre-Corona Obs	6,256
Post-Corona Obs	25,024

Variable List *Structure and short discription*		
Variable Type	Variables	Observations
Continuous	11	31184
Factor	38	31280
Nominal Total	49	31280
Factor-Expanded Total	114	31280

3.2 Data Processing

Though the data-collecting algorithms return structured data, it is nevertheless far from being suitable for the rather picky models which will eventually analyze them. Therefore, the following processes were completed in order to render the raw data into a usable form:

1. **Missing values** (i.e., N/A values) were removed.

2. **Outliers** were identified and removed for all continuous variables. An 'outlier' is defined by being more than 1.5 standard deviations from the mean of the variable's own distribution.
3. **Multilevel factor** data was broken out by each level into binary representations through *Hot-One* coding. Some features which had many factor levels were simplified using Lasso regressive methods.
4. **High-leverage** point observations, according to diagnostic linear regression, were removed
5. **Duplicates**, defined by the MLS unique identification number, were removed
6. **Structural errors** (e.g., dates structured as string variable) were corrected
7. **Binary variables** were created for key variables (e.g., city limits) by the following standard method:

Equation 2

$$I(y) = \begin{cases} 1, & x \in A \\ 0, & x \notin A' \end{cases}$$

Where I is an indicator function with space A that composes dummy variable x into 1 if the condition is met and into 0 otherwise.

The results of the data-cleaning process can be seen in **Table 2**.

Table 2 *Cleaned Data Summary*

Variable List *Structure and short discription*		
Name	Raw	Clean
Date Range	23.10.2010 - 12.12.2021	23.10.2010 - 12.12.2021
Location	Louisiana, USA	Louisiana, USA
Number of Variables	49	49
Number of Observations	31,280	24,412
Pre-Corona Obs	6,256	4,882
Post-Corona Obs	25,024	19,529

Variable List *Structure and short discription*		
Variable Type	Variables	Observations
Continuous	11	24412
Factor	38	24412
Nominal Total	49	24412
Factor-Expanded Total	103	24412

3.3 Variable List

A list and short description of all variables used in this paper can be found in ***Table 3***

Table 3 *Used Variables and Short Description*

Variable List

Structure and short discription

Count	Name	Structure	Discription
1	list_price	Number	Original listing price
2	photo_count	Number	Number of photos on listing
3	area_living	Number	Total living area in sqft.
4	land_acres	Number	Size of land in acres
5	area_total	Number	Total area in sqft.
6	age	Number	Age of property
7	dom	Number	Days on the market
8	sold_price	Number	Actual sold price
9	infections_daily	Number	Daily public corona infections
10	infections_accum	Number	Accumulation of public corona infections
11	infections_3mma	Number	3-month moving average of daily public corona infections
12	sold_date	Date	Date on which the property was sold
13	beds_total	Factor	Total number of beds
14	bath_full	Factor	Total number of full bathrooms
15	bath_half	Factor	Total number of half bathrooms
16	property_type	Factor	Property type
17	property_condition	Factor	Property condition
18	property_style	Factor	Property Style
19	roof_type	Factor	Roof type
20	patio	Factor	Patio present
21	out_building	Factor	Detached building (e.g. shed) present
22	city_limits	Factor	Property within city limits
23	mls_number	Factor	MLS number (unique ID)
24	ac_type	Factor	Air conditioning type

Variable List

Structure and short discription

Count	Name	Structure	Discription
25	school_general	Factor	School in city limits
26	pool	Factor	Pool present
27	gas_type	Factor	Gas type
28	appliances	Factor	Appliances included
29	garage	Factor	Garage present
30	energy_efficient	Factor	Energy-efficient features present
31	exterior_type	Factor	Exterior type
32	exterior_features	Factor	Exterior features
33	fireplace	Factor	Fireplace present
34	foundation_type	Factor	Foundation type (e.g. slab)
35	sewer_type	Factor	Sewer type
36	subdivision	Factor	Property within subdivision
37	water_type	Factor	Water supply type
38	waterfront	Factor	Property has waterfront
39	corona_date_split	Factor	Date of first mandatory lockdowns in Louisiana (i.e. 23.03.2020)
40	top25_sold_price	Factor	Top 25th percentile of sold price
41	top50_sold_price	Factor	Top 50th percentile of sold price
42	bottom25_sold_price	Factor	Bottom 25th percentile of sold price
43	top25_area_living	Factor	Top 25th percentile of total living area
44	bottom25_area_living	Factor	Bottom 25th percentile of Living area
45	top25_age	Factor	Top 25th percentile of total age
46	bottom25_age	Factor	Bottom 25th percentile of age
47	top25_dom	Factor	Top 25th percentile of days on market
48	bottom25_dom	Factor	Bottom 25th percentile of Days on Market
49	infections_period	Factor	Period after accumulated infections > 1000 cases

3.3 Data Descriptive Statistics

3.3.1 Correlation

The correlation matrix (see ***Figure 2***) between all continuous variables shows that with exception to the variables which will have obvious correlations (e.g., sold_price and

list_price, area_total and area_living, and infections numbers), there are no other correlations which would cause concern.

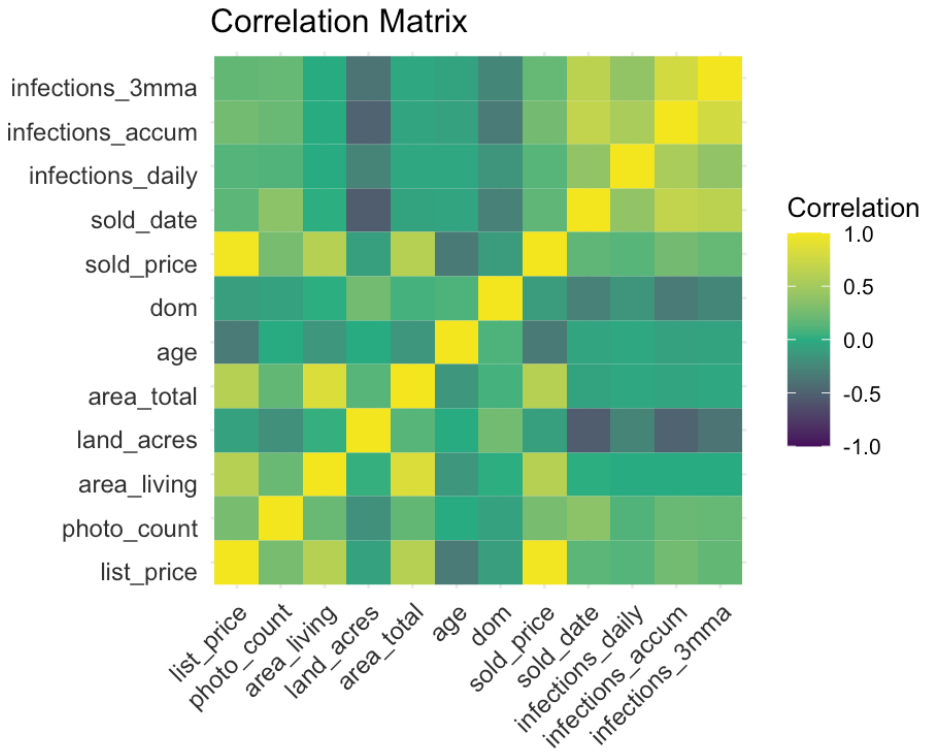

Figure 2 *Correlation Matrix*

3.3.2 Distributions of Select Variables

As the standard descriptive characteristics of a particular variable are considered (i.e., measures of frequency, central tendency, dispersion, and position), the matrix of density plots below (see ***Figure 3***) give us a good overview of the most relevant variables in this data set.

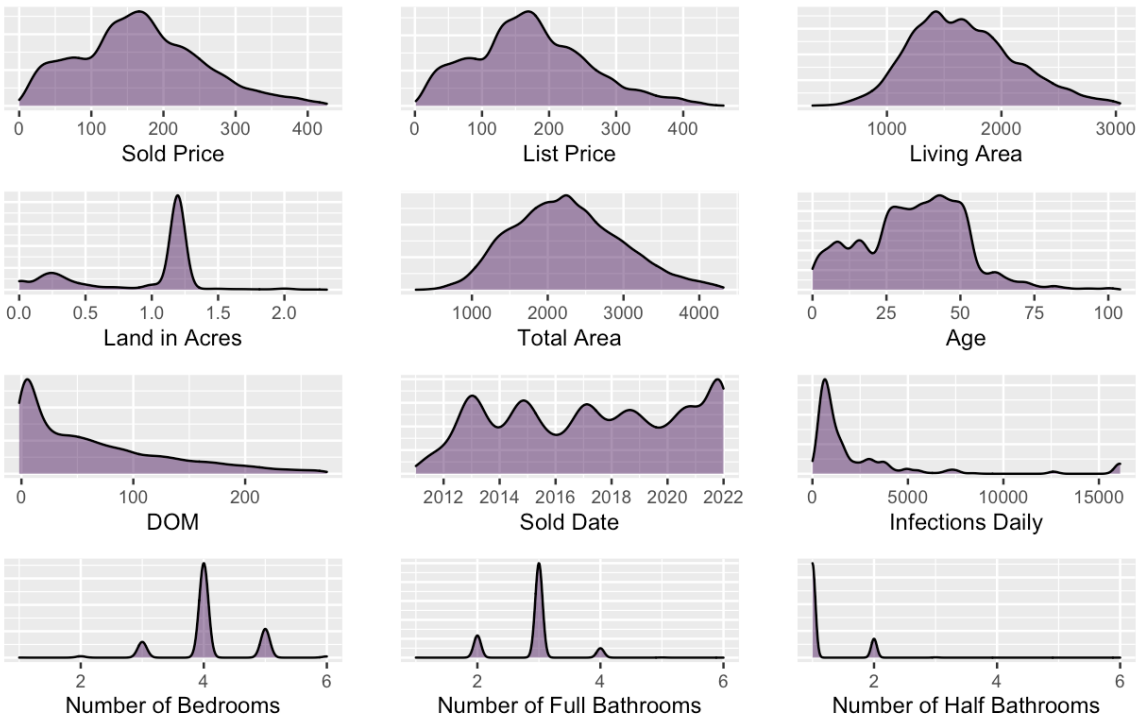

Figure 3 *Distributions of Continuous Variables*

3.3.3 Price Index

To understand the general shape and historical trend of the pricing market specifically within Louisiana, I have constructed a simple weighted-average price per sqft. index (see ***Figure 4***) using 1, 2, 3, and 4 month-moving averages of the entire data set. Furthermore, I show the St. Louis Federal Reserve's GDP (*green*) and price index (*purple*) for the Louisiana housing market (see ***Figure 5***). Though the methods used in both indices are not the same, which accounts for the general level differences between the two, one can see the shape of the data is fitting to the population.

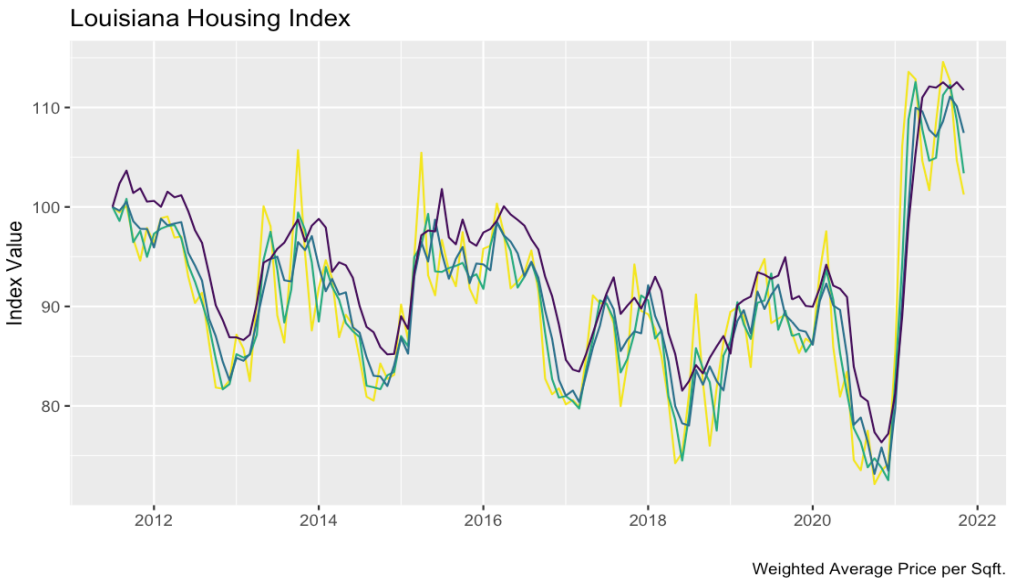

Figure 4 *Louisiana Housing Index: Data Set*

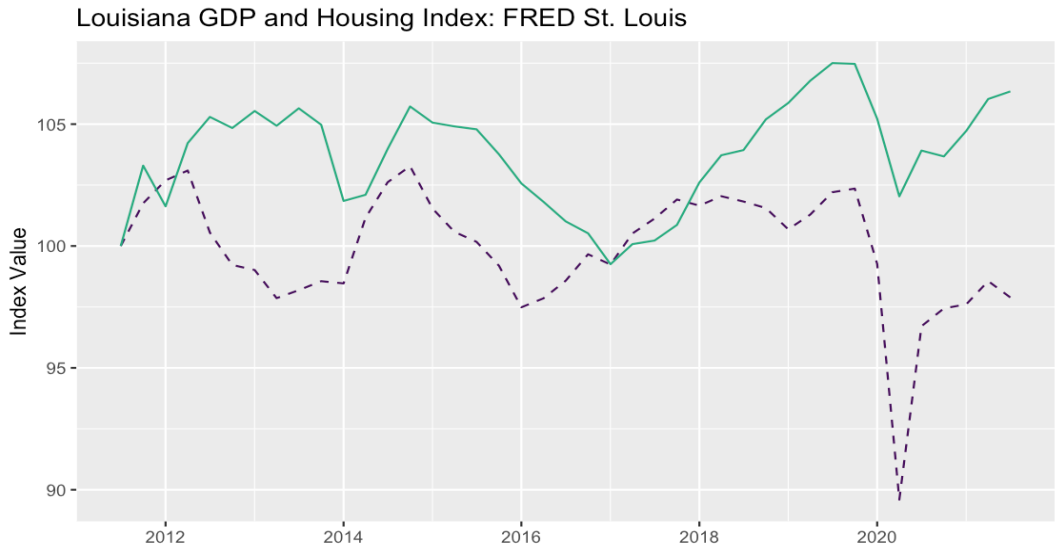

Figure 5 *Louisiana GDP and Housing Index: FRED*

4. Methodology

The overarching method used in this paper is the Hedonic Pricing Method (HPM), also often referred to as hedonic regression or hedonic demand theory. The fundamental theory behind the HPM is the following: commodities are distinguishable by their component parts, therefore, the market value of a given commodity can be calculated by summing the estimated values of its separate characteristics. For this theory to hold true, several critical requirements must be met. Primarily, that the commodity being valued can be reduced to its component parts and that the market is able to implicitly and independently value these characteristics. The fulfillment of these requirements is not obvious and will in some measure fall short of accounting for the complete nature of price dynamics in practically every asset class. However, this limitation offers an interesting problem to test. Namely, to find the limit of the accumulated power of these component parts to account for market values and their deviations across time and subgroups. These exact questions will be later examined by implementing a machine-learned predictive model to measure the theoretical maximum explanatory power of the included hedonic variables. In the following two subsections, we review the methods used in this paper to econometrically model the HPM on hedonic real estate data.

4.1 Multi-Variable Linear Regression

In this section, I will outline the construction of my base OLS model, termed the Alpha model, as well as the treatment process for heteroscedasticity, multicollinearity, non-linearity, and high-leverage points and outliers.

4.1.1 Basic Model Design

Following the OLS construction laid out by Herath and Maier (2010):

Equation 3

$$R = f(P, N, L, t)$$

where R is rent or price of the property; P is property related attributes; N is neighborhood characteristics; L is locational variables and t is an indicator of time.

This paper's base OLS model, named the Alpha model, is as follows:

Equation 4

$$P_{n\times 1} = A_{n\times 1} + B_{k\times 1} V_{n\times k} + \mathcal{E}_{n\times 1}$$

where P is a $n \times 1$ vector of sold prices; A is a $n \times 1$ vector of the model's intercepts; B is a $k \times 1$ vector of beta coefficients; V is a $n \times k$ matrix of all hedonic variables; $\mathcal{E}$ is a $n \times 1$ vector of the model's random error; subscript n is the number of observations and k is the length of the variable list.

4.1.2 Accounting for Heteroscedasticity

A Breusch-Pagan test was conducted on a standard linear regression model with sold price as the dependent variable and the rest of the dataset as regressors. The Breusch-Pagan (BP) test was established as a method in 1979 and follows the logic set by the Lagrange multiplier test principle (Breusch and Pagan 1979). This test tests the null hypothesis that the variance in the model's errors is independent from model's regressors (i.e., heteroscedasticity). The test's results in a rejection of the null hypothesis, thereby

finding the base model to be heteroscedastic. The results of this test are summarized in *table 4*.

Table 4 *Testing for Heteroscedasticity - Results*

Breusch Paga Test for Heteroskedasticity *Hypotheses*		
Hypotheses	Test Summary	-
Ho: the variance is constant	DF	1
Ha: the variance is not constant	Chi2	850.4231
-	Prob > Chi2	0.00

To resolve the heteroscedasticity found, I will produce heteroscedasticity-consistent (HC) standard errors, also known as heteroscedasticity-robust standard errors, through the refined method established by econometrician Halbert Lynn White (White 1980). This process is as follows:

If the model's errors u_i are independent but have distinct variances, σ_i^2 then $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ which can be estimated with $\hat{\sigma}_i^2 = \hat{u}_i^2$. This relationship produces the estimator found in White (1980):

Equation 5

$$v_{\mathrm{HCE}}[\hat{\beta}_{\mathrm{OLS}}] = \frac{1}{n}\left(\frac{1}{n}\sum_i X_i\, X_i{}'\right)^{-1}\left(\frac{1}{n}\sum_i X_i\, X_i{}'\hat{u}_i^2\right)\left(\frac{1}{n}\sum_i X_i\, X_i{}'\right)^{-1}$$

$$= (\mathbb{X}'\mathbb{X})^{-1}(\mathbb{X}'\mathrm{diag}(\hat{u}_1^2, \ldots, \hat{u}_n^2)\mathbb{X})(\mathbb{X}'\mathbb{X})^{-1},$$

where $\mathbb{X}$ denotes the matrix of stacked $X_i{}'$ values from the data. The estimator can be derived in terms of the generalized method of moments (GMM).

For the remainder of this paper, all results and figures regarding statistical significance will be referring to tests conducted with heteroscedasticity-consistent (HC) standard errors. As sample errors in my models will have equal variance and are uncorrelated, the least-squares estimates of each model's beta coefficients are regarded as Best Linear Unbiased Estimators (BLUEs).

4.1.3 Accounting for Multicollinearity

Multicollinearity is measured using Variance Inflation Factors (VIF). The VIF of a predictor measure how accurately that variable can be predicted using all other variables. For context, the square root of a VIF represents the increase in standard error of the estimated coefficient with respect to the case when that given variable is independent of all other variables. In line with current convention, all variables with a VIF larger than 5 are eliminated. A graphical representation of all variable multicollinearity, measured by VIF, is shown in figure 6.

This test resulted in the elimination of the variables ***list_price*** and ***area_total*** as there were highly multicollinear with ***sold_price*** and ***area_living*** respectively.

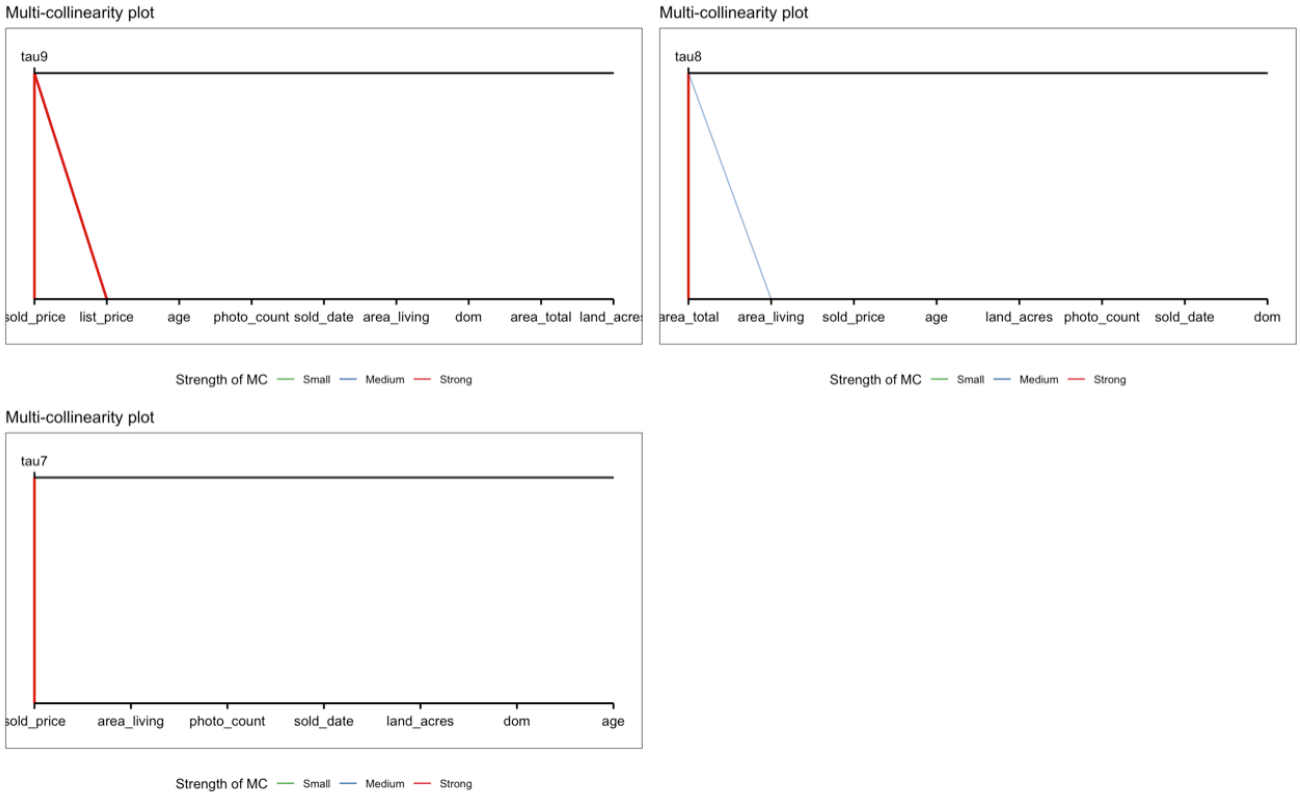

Figure 6 *VIF Testing of Variable Multicollinearity*

4.1.4 Accounting for Non-Linearities

Visual analysis was conducted on all continuous variables and non-linear variables transformation were added to ***age*** and ***area_living***

Age:

An analysis of age versus sold price shows a well-established u-shaped pattern. In order to allow the OLS model to better capture this relationship, a new variable age^2 is added to the model.

Figure 7 shows the Partial Dependency Plot (PDP) of age within the Alpha model. This plots the marginal prediction of the Alpha model across the full range of age. When the scale of the y-axis is reduced, we see the slight curvature in Alpha model's estimation of age effects. This addition improved R^2 by .08.

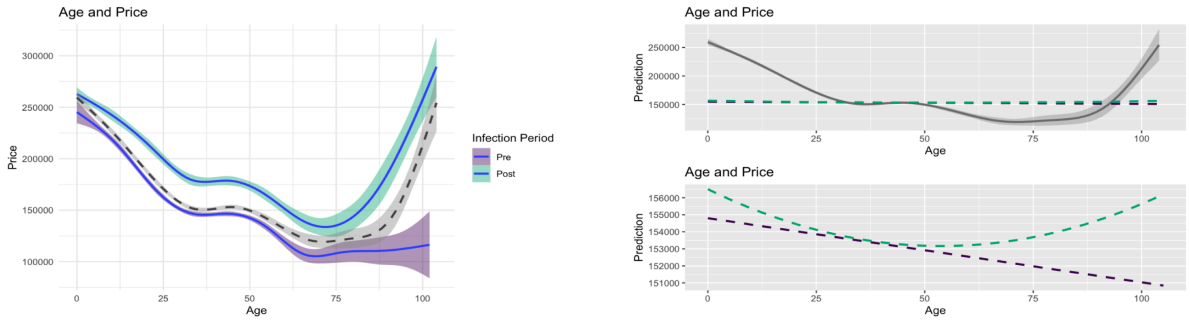

Figure 7 *PDP: Age within the Alpha Model*

Living Area:

An analysis of living-area versus sold price reveals a Sigmoid pattern between the two variables. In order to allow the OLS model to better capture this non-linear relationship, a new variable $living\ area^2$ is added to the model.

Figure 8 shows the Partial Dependency Plot (PDP) of $living\ area^2$ within the Alpha model. This addition improved R^2 by .06.

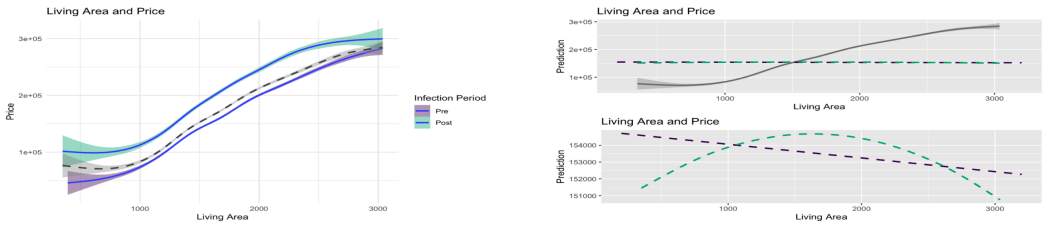

Figure 8 *PDP: Living Area within the Alpha model*

4.1.5 Accounting for High-Leverage Points and Outliers

After the adjustments of the previous sections have been made, a panel of visualizations are run on the Alpha model. The results show no extreme outliers and only one high-leverage point (obs. #23515), as shows by the residual versus leverage plot in quadrant two of figure 9. This observation is removed in the final Alpha model. These uniform results are mainly due to the previous removal of outliers and the overall quality of the data set.

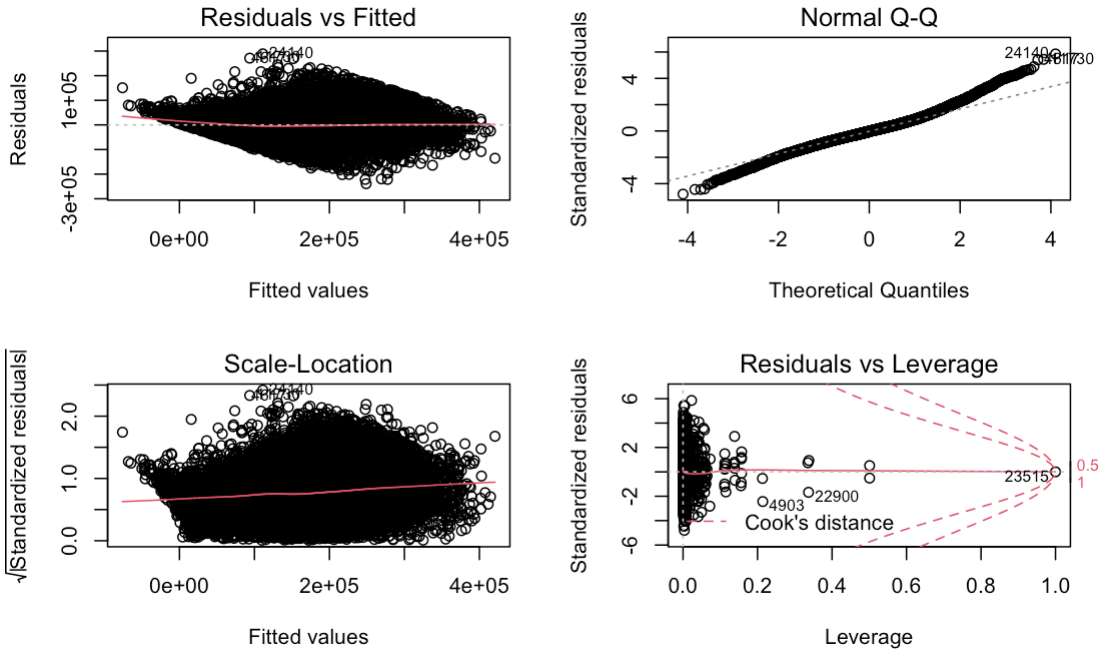

Figure 9 *Outlier Testing with Alpha Model*

4.1.6 Final Alpha Model

The Alpha model is the baseline OLS for this paper and is robust to heteroscedasticity, multicollinearity, non-linearities, high-leverage points and outliers. These meta-level adjustments increase our confidence in the statistical tests results which follow.

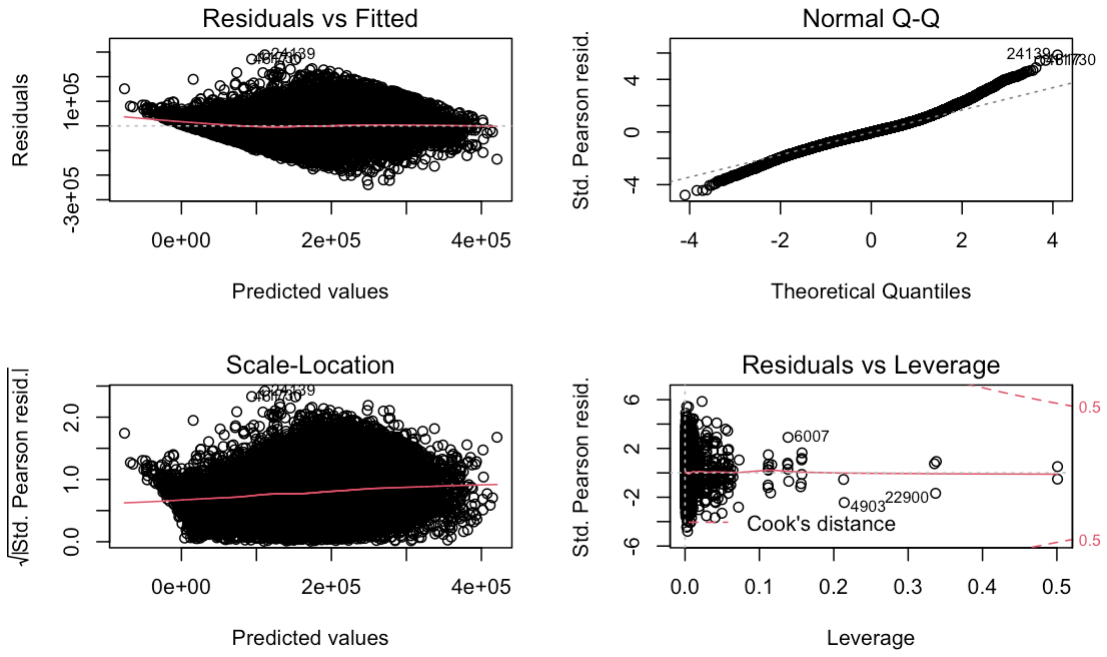

Figure 10 *Final Alpha Model Output Panel*

4.3 Modeling Changes in Demand for Hedonic Features

The focus of this paper is how the Covid crisis impacted housing prices and the relative levels of demand for specific hedonic features. In the case where the HPM is in the OLS functional form, the beta coefficients of this model represent relative demand for each associated hedonic feature (Shimizu et al. 2010). For example, $\beta_{pool=1} = 11{,}856$ USD is interpreted as the average consumer's willingness to pay for an average pool, ceteris

paribus. However, this paper wishes to measure the *changes* in the average consumer's willingness to pay for a given feature (e.g., pool) in relationship a measurement of Covid's economic impact.

4.3.1 Comparison Method

A method of statistically comparing changes in the beta coefficients of features of interest under multiple scenarios (e.g., post and pre-Corona period) is needed. To accomplish this, a method outlined by the UCLA Statistics department is implemented (Bruin 2011). The best way to understand this method is to see a simplest reproducible example.

Suppose we want to test the economic impact of Corona on the relative demand for swimming pools, $\beta_{pool=1}$ pre- versus post-Corona period. Using the UCLA method, we test the null hypothesis H_0: $\beta_{pool=1,\, corona=0} = \beta_{pool=1,\, corona=1}$ with the following OLS

Equation 6

$$sold\ price = \alpha\ + \beta_1 pool\ +\ \beta_2 corona\_period\ + \beta_3 (pool \times corona\_period)$$

Which results in the following:

Table 5 *Example Output: Presence of Pool and Infections*

		sold_price	
Predictors	*Estimates*	*CI*	*p*
(Intercept)	154123.97	152875.73 – 155372.21	**<0.001**
pool [1]	53118.17	48666.02 – 57570.33	**<0.001**
infections period [1]	41724.75	39257.17 – 44192.34	**<0.001**
pool [1] * infections period [1]	-7766.40	-16115.57 – 582.77	0.068
Observations	24412		
R^2 / R^2 adjusted	0.072 / 0.072		

Interpretation of this simplified model's estimates results:

- Intercept: Intercept for $pool = 0$
  - The omitted group
- pool: Slope for $pool = 1$
  - The included group
- infections period: (Intercept $infections\ period > 0$) − (intercept $infections\ period = 0$)
  - This can be seen by running individual regressions for each case
- (pool*infections period): Slope for $pool_{infections\ period=1} - pool_{infections\ period=0}$
  - This estimate tests the null hypothesis H_0: $\beta_{pool=1,\ corona=0} = \beta_{pool=1,\ corona=1}$

Therefore, we say:

The average premium for a property having a **swimming pool** fell by **7,767 USD** when compared to pre-Corona levels, ceteris paribus. However, this finding is only significant at the ***p < 0.10*** level.

4.3.2 Selecting a Corona Measurement

A good measurement variable for measuring the response of the market to the Corona crisis must be a Corona-related matric which is publicly available; common knowledge to the population and is a reasonable measurement of future economic shifts. For this reason, data collected from the Louisiana Department of Health (LaDH 2022) was used to calculate the 3-month moving average of Corona infections (infections_3mma).

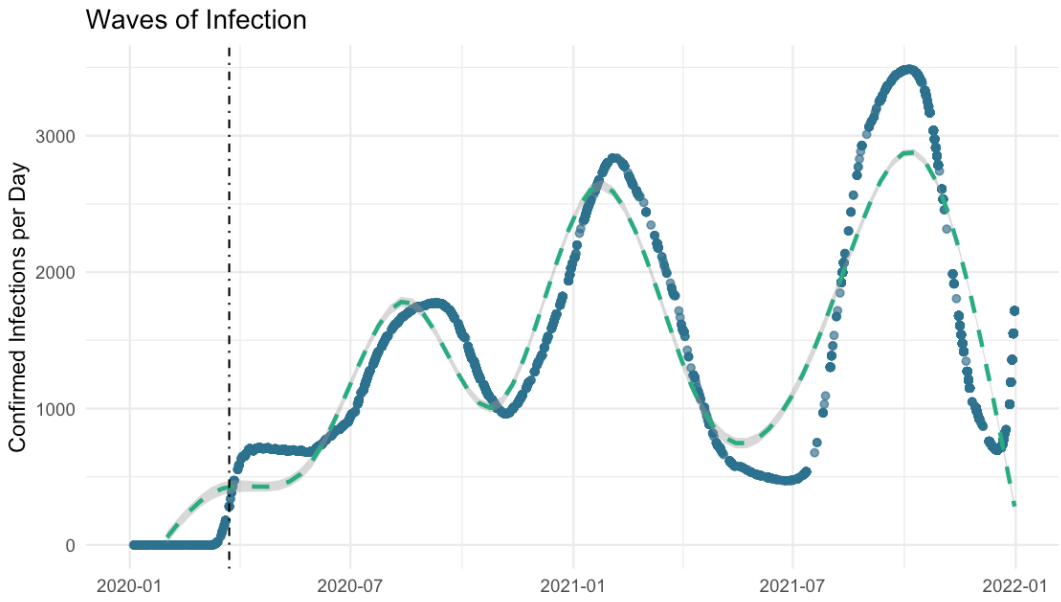

Figure 11 *Waves of Corona Infections*

This measurement fulfills the previously stated requirements of a good test measurement as it is purely related to the Corona crisis, is publicly available, is assumed to be publicly known as it is reported across all major news stations daily, and perhaps most importantly, is the primary metric used to decide when mandatory lockdowns are instituted. For this paper, I assume the market is responding to some lagged value of daily infections which are being used by consumers to estimate the likelihood of future lockdowns and the stringency, and duration of current lockdowns. With this rational, infections_3mma is selected as the primary measurement of Corona's impact.

4.4 Machine Learning

4.4.1 Machine Learning Methods

In the previous section, it was stated that the multivariable regression models estimate a vector of parameters (i.e., beta coefficients) that best fit the explanatory hedonic variables to the associated dependent variable. Intuitively, the resulting fitted coefficient vector is fitted to the entire data set, and therefore, the loss function minimizes the error in the model's ability to *explain* the very independent variable it was fitted to. Restated, these results are ultimately limited to their inferential value within the exact context of the data set the model it is trained on. If one is to establish a wider, more general relationship between dependent and independent variables that go beyond the context of the trained data set, supervised machine learning (ML) prediction models are an extremely powerful tool of accomplishing that goal. Though the models used in this paper differ across several key processes, they each generally follow a similar logic:

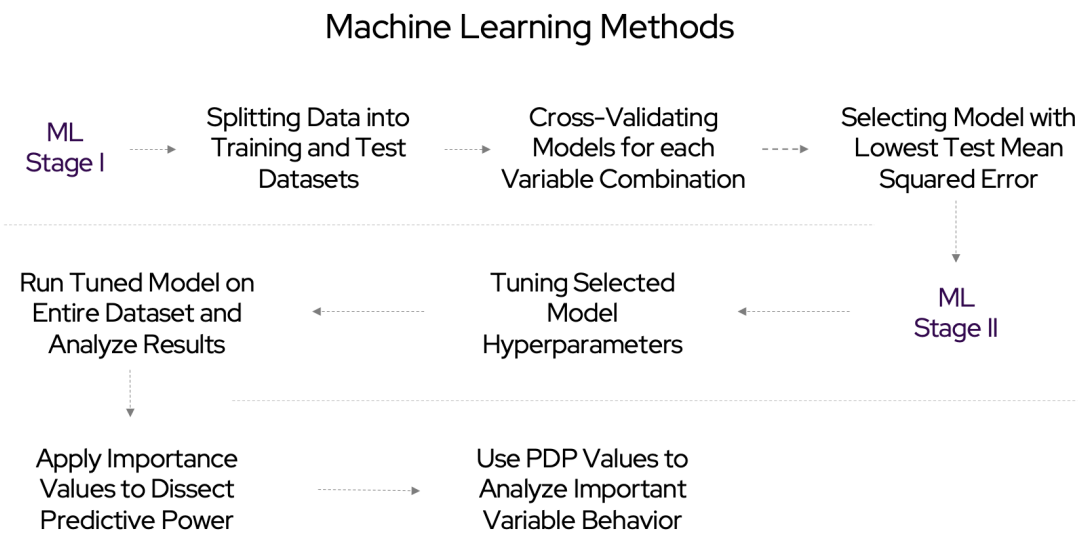

Figure 12 *Machine Learning Methods*

4.4.2 Model Evaluation with Cross-Validation

In order to rank order models, I perform a process called Cross-Validation (CV). First, the full data set must be split into 'test' and 'train' (i.e., validation) subsets. Each ML model will be fitted to the train data set and its performance will be evaluated based the model's ability to predicted out-of-sample observations in the test (validation) data set. In this way, these models are ranked based on their test mean squared errors (MSE). This process is often referred to as Cross Validation (CV). The two most used CV methods are the Validation Set Approach and K-Fold Cross Validation.

The Validation Set Approach (VSA) is the simplest case of cross validation data splitting as it randomly splits the entire data set into train and test subsets based on a certain percentage split. For example, the researcher can choose to split the data set with an 80%

training and 20% testing split. The estimation for the test MSE is simply the test error against the test data set.

Equation 7

$$CV_{vsa} = MSE_{vsa}$$

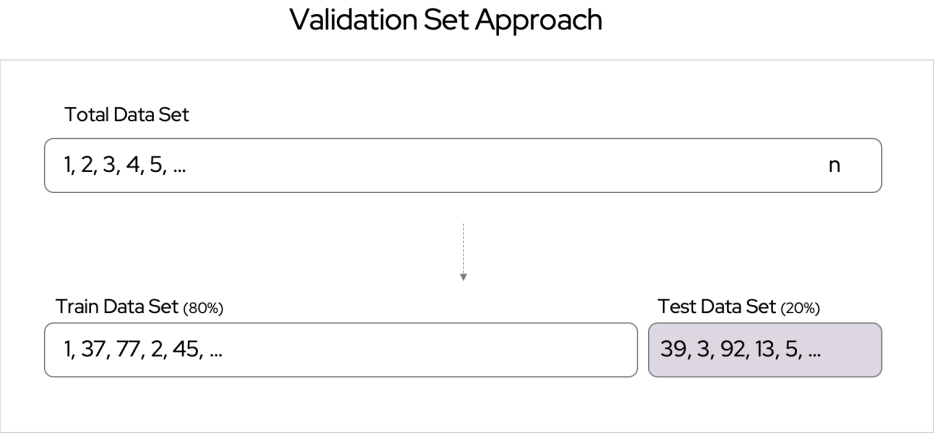

Figure 13 *Validation Set Approach*

The K-Fold CV method has increasingly been used by research as it offers a more comprehensive cross validation process when compared to other methods. K-Fold CV is the process of randomly splitting the entire data set into k groups, or folds, each with approximately an equal number of observations. The first fold is held out and the model is trained on the remaining k-1 folds. This process is repeated k times, each time holding out a different fold until every fold has been treated as the validation set. Finally, this will result in k estimations of the model's test error and the final estimation will be the average across all k model fits.

Equation 8

$$CV_{k-fold} = \frac{1}{k}\sum_{t=1}^{k} MSE_i$$

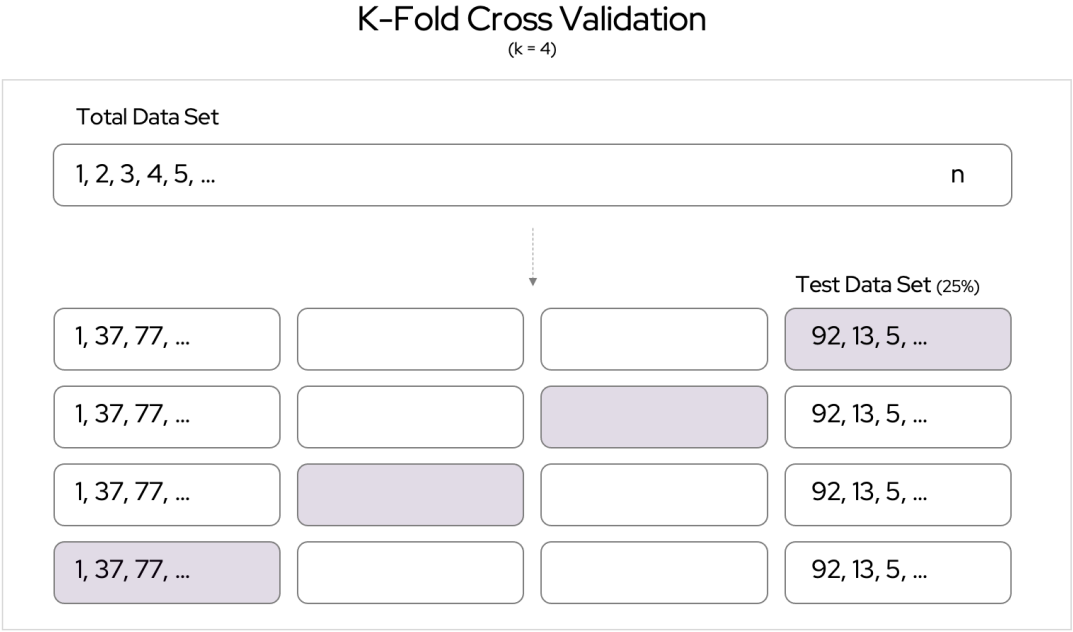

Figure 14 *K-Fold Cross Validation*

4.4.4 ML Model Selection

For this paper, ML models will be ranked based on two features: Accuracy, as measured by test MSE, and interpretability, qualitatively defined by the model's ability to provide insights into to which features are relevant to the ability to make correct predictions, and by how much are they relevant.

Accuracy

Since a method of model fitting and evaluation has been established in the previous section through the process of Cross Validation, we now have a way to rank different models to each other based on their ability to estimate test MSE. With this feature, it is possible to

compare several models to one another in terms of effectiveness in predictions. Five different ML models were built for this paper with the following results:

Table 6 *ML Models: Comparison of Results*

ML Models *Comparison of Results*		
Model	Train MSE	Test MSE
Ridge Regression	15.23%	22.97%
LASSO Regression	15.22%	22.98%
Artificial Neural Network	11.17%	20.06%
Gradient Boosting	10.76%	21.36%
eXtream Gradient Boosting	11.09%	19.57%

Interpretability

A major criticism of ML models is their lack of interpretability, explaining *how* the model makes such accurate predictions. A prime example of this critique can be seen in Artificial Neural Networks (ANN) machines, which often use many small nodes to sequentially generate a single prediction without any one of these nodes holding a clear interpretation as to which variables, or combination of variables, lead to a particular improvement in prediction. However, as ML methods become more commonly used, the demand for interpretation of these models has driven several useful interpretation methods across for a variety of models. These methods include interpretation for a single prediction, such as

the Local Interpretable Model-agnostic Explanations (LIME), as well as generalized methods of measuring relative feature importance, which uses various techniques to determine the average contribution of each variable to the model's ability to decrease its test MSE rate.

Of the models sampled, the eXtreme Gradient Boosting Machine, also referred to as XGBoost, outperforms the other models both in terms of having the lowest test MSE and having the best methods for detailed variable interpretation. For this reason, XGBoost is chosen as the primary ML model for this paper.

4.4.5 eXtreme Gradient Boosting Machine Algorithm

In order to understand the logic of XGBoost, one must first look at the compact, yet powerful algorithm behind its computation. In this section, I will lay out the mathematical formulation of the input, core algorithm, and final output of the XGBoost machine.

1. Algorithm input: We start with a training set $\{(x_i, y_i)\}_{i=1}^{N}$, a differentiable loss function $L\big(y, F(x)\big)$, several weak learners (shallow trees) M, and a learning rate α.
2. Algorithm:

2.1 Initialize model with a constant value:

Equation 9

$$\hat{f}_{(0)}(x) = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N} L\left(y_i, \theta\right)$$

2.2 For $m = 1$ to M:

2.2.1 Compute the 'gradients' and 'hessians

Equation 10

$$\hat{g}_m(x_i) = \left[\frac{\partial L\big(y_i, f(x_i)\big)}{\partial f(x_i)}\right]_{f(x)=\hat{f}_{(m-1)}(x)}$$

$$\hat{h}_m(x_i) = \left[\frac{\partial^2 L\big(y_i, f(x_i)\big)}{\partial f(x_i)^2}\right]_{f(x)=\hat{f}_{(m-1)}(x)}$$

2.2.2 Fit an original weak learner (e.g., shallow tree) using the training set $x_i, -\frac{\hat{g}_m(x_i)}{\hat{h}_m(x_i)}\}_{i=1}^{N}$ by solving the following optimization problem:

Equation 11

$$\hat{\phi}_m = \underset{\phi \in \mathbf{\Phi}}{\operatorname{argmin}} \sum_{i=1}^{N} \frac{1}{2} \hat{h}_m(x_i) \left[-\frac{\hat{g}_m(x_i)}{\hat{h}_m(x_i)} - \phi(x_i)\right]^2$$

$$\hat{f}_m(x) = \alpha \hat{\phi}_m(x).$$

2.3 Update the model:

Equation 12

$$\hat{f}_{(m)}(x) = \hat{f}_{(m-1)}(x) + \hat{f}_m(x).$$

3. Algorithm Output:

Equation 13

$$\hat{f}(x) = \hat{f}_{(M)}(x) = \sum_{m=0}^{M} \hat{f}_m(x)$$

4.4.6 XGBoost Model Fitting and Hyperparameter Tuning

In practice, when attempting to produce the most optimal results from an XGBoost machine, as is true with most other ML models, one must first transform the data set into an optimal form and then run set of hyperparameter tests to determine the appropriate level for each of the model's basic structural rules (i.e., hyperparameters).

Since the XGBoost machine requires only numerical data, factor data must be converted into numerical levels which can be the used in the construction of decision trees. To accomplish this with the added complexity of some factor variables having more than two levels, I have used a method called *One-Hot Encoding*, which encodes each individual factor variable level into a vector containing '1' if that factor and level are present, and '0' otherwise. In this way, the data frame is converted into a large matrix of continuous and binary columns.

Table 7 *One-Hot Encoding Example*

Unique Identifier	beds_total
#112345	beds_total_1
#22346	beds_total_2
#31452	beds_total_3
#49083	beds_total_4
#59867	beds_total_5

One-Hot Encoding

Unique Identifier	beds_total_1	beds_total_2	beds_total_3	beds_total_4	beds_total_5
#112345	1	0	0	0	0
#22346	0	1	0	0	0
#31452	0	0	1	0	0
#49083	0	0	0	1	0
#59867	0	0	0	0	1

Once the model is fitted to the data, the next procedure is to tune the hyperparameters which govern the algorithm's learning process and therefore determine the resulting values of estimated parameters. To do this, a large grid of hyperparameters is created; an individual model containing each unique combination of hyperparameter levels is generated and their resulting test MSE's are ranked and analyzed. The results from this large grid search determine which combination of hyperparameters are optimal, and those hyperparameters are used in the final model. The search grid from this research was so large, it took my computer *78 hours* to complete all the calculation necessary for the full tuning process.

4.4.7 XGBoost Partial Dependency Plots

Partial dependency plots (PDP) show a relationship, or dependence, between the model's response variable (i.e., *sold price*) and a chosen variable, or set of variables, of interests (VoI), resulting in the graphical representation of a variable's marginal contribution to the machine's prediction across the VoI's entire range. In order to analyze the results of the XGBoost machine at the variable-by-variable level, I have generated a panel of four graphical partial dependency plots for each variable of interest. paper will be the following: Basic PDP plot, Individual Conditional Expectation (ICE) plots, PDP heatmap with VoI against Corona infection, 3-dimensional PDP heatmap with VoI against Corona infection and sold price. An example of each of these graphs, using days on market as the variable of focus, can be seen in quadrants IV, I, III, and II respectively (see ***Figure 15***).

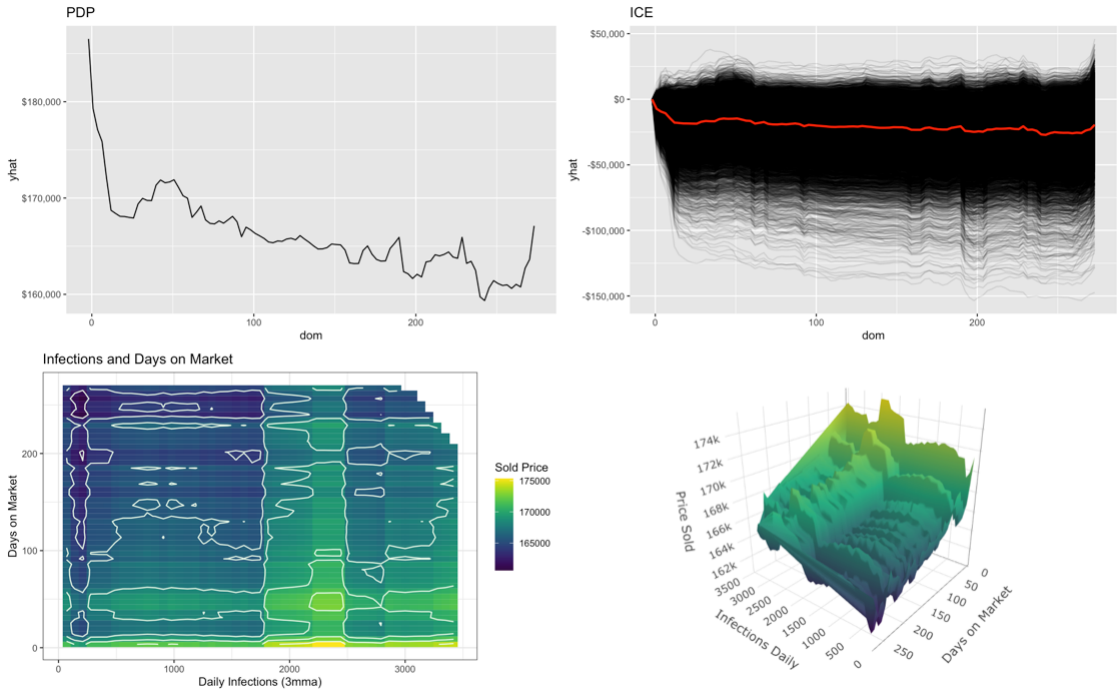

Figure 15 *PDP, ICE, Heatmap, and 3D Heatmap Examples*

5. Hypothesis Construction

5.1 Corona: General Case

Hypothesis I: The Corona crisis significantly increased housing prices

Reasoning: As many workers have permanently shifted to remote work, the total utility of residential housing is expected to have increased, thereby increasing the price households are willing to pay. This price shift can be explained through measuring the changes in relative demand for specific hedonic features, such as bedrooms, size, age, and others.

5.2 Corona: Premium for Bedrooms

Hypothesis II: The Corona crisis significantly increased demand-premiums for every level of number of bedrooms greater than 1

Reasoning: As many workers have permanently shifted to remote work, the premium for an additional bedroom is expected to increase across each level of total number of bedrooms as households need additional rooms for home offices and other activities. The premium for a single bedroom is expected to be insignificant since there are no recorded 0-bedroom houses to upgrade from.

5.3 Corona: Premium for City Centrality

Hypothesis III: The Corona crisis impacted properties within city limits more than those not within city limits

Reasoning: As workers who live in city limits are expected to hold jobs more likely to be compatible with remote work, such as jobs within the financial

industry, home prices within city limits are disproportionately impacted by the shift to remote work.

5.4 Corona: Premium for Size

Hypothesis IV: The Corona crisis increased the premium for property size

Reasoning: Same as for hypothesis I

Hypothesis V: The Corona crisis impacted properties in the top 25th percentile of property size more than the bottom 25th percentile

Reasoning: Households who live in the top 25th percentile of home sizes are more likely to hold white-collar jobs, which are more likely to be made remote, which increases the premium they are willing to pay for additional hedonic features.

5.5 Corona: Premium for Age

Hypothesis VI: The Corona crisis decreased the premium for property age

Reasoning: As the velocity of home sales increase, the general relationship between property age and price is expected to be exasperated as those upgrading houses will shift towards newer properties.

Hypothesis VII: The Corona crisis impacted properties in the top 25th percentile of age more than the bottom 25th percentile

Reasoning: The premium for the youngest (i.e., newest) properties is expected to increase while the premium for the oldest properties is expected to decrease. As the market shifts towards newer properties, they must necessarily shift away from older properties.

5.6 Corona: Change in Days on Market

Hypothesis VIII: The Corona crisis significantly decreased the premium for days on market

Reasoning: As the velocity of home sales increase, more homes are sold faster, and therefore the number of days on the market becomes less of a predictor of quality since even lower quality and over-priced homes are sold more quickly.

Hypothesis IX: The Corona crisis impacted properties in the top 25th percentile of days on market more than the bottom 25th percentile

Reasoning: The premium for a property being sold within the top 25th percentile of days on market (DOM), i.e., the properties which sit on the market the longest, is expected to be disproportionately affected when compared homes which sold the fastest, as households' elasticity of demand for specific characteristics decreases relative to price, it will take larger deviation in price and quality to make a home sell in the top 25th percentile of DOM

6. Results

This section will follow a consistent structure for each set of results. Each hypothesis will have: A *summary of the finding*; *feature review* of the characteristic being modeled; and *model results* from the OLS and ML models with standard hypothesis conclusion(s).

6.1 Corona: General Case

6.1.1 Summary of Findings

Preliminary analyses of housing prices and daily infections reveals a positive historical relationship between these two key variables. This relationship is strongly supported by the XGBoost ML algorithm, which shows that a maximal reduction in predictions error is achieved by increasing prediction price at all levels of daily infections greater than 0, holding all other factors constant. XGBoost also determined that from the 104 total variables used, daily infections are the 19th most important variable in reducing price prediction error.

In the fully controlled Alpha model, the beta coefficient for daily infections suggests that each additional infection is associated with an average home price *increase of 8.97 USD*, ceteris paribus. This finding is significant at the $p < 0.00$ level. Therefore, I reject H_0: $beta_{infections_3mma} = 0$ and conclude that the Corona crisis significantly increased housing prices (see ***Hypothesis I***).

6.1.2 Visual Review

We first look at the distribution of daily infections and the accumulation of infections across time. The variation in historical daily infections should provide a strong measurement for explaining variations in price related to this key variable.

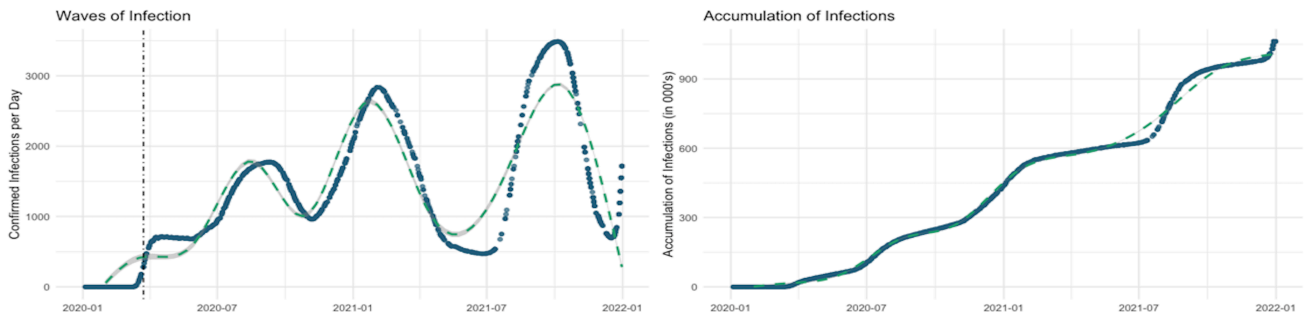

Figure 16 *Distribution of Daily Infections and Accumulation of Infections*

To investigate the raw historical relationship between infections rates and prices, we look at the trend line of price versus daily infections (rhs) and a comparison between the price distributions of pre- and post-infections period (lhs). This analysis suggests a historically positive relationship between infections and price. Though these graphs are promising, it is unclear if other factors could be influencing this relationship. To establish a controlled statistical relationship, we turn to the following OLS and ML models.

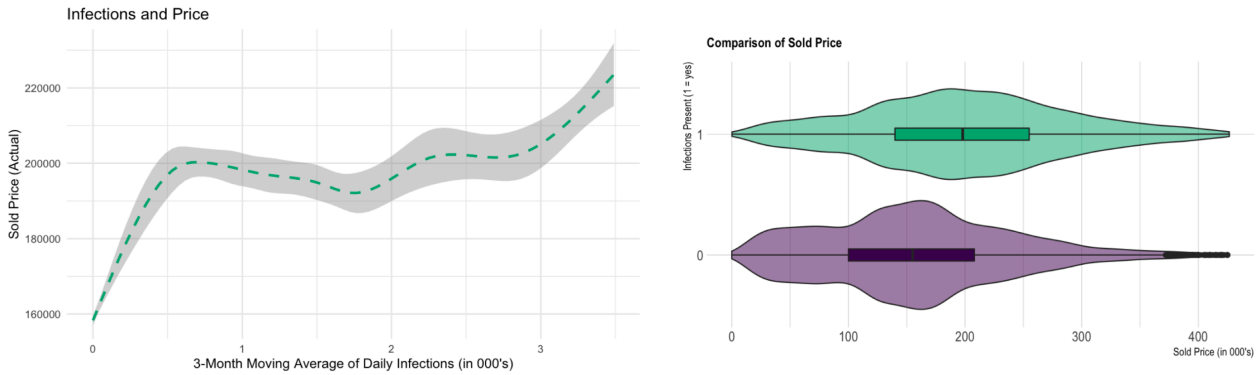

Figure 17 *Infections and Price: Historical*

6.1.3 OLS Modeling

Since displaying the full OLS output for every set of results is impractical, I will include a summarized version with only the key variables being analyzed in each respective section. Please note that if you wish to see each regression output table, code, and data associated with this paper, you can visit my public GitHub repository with the following link: https://github.com/Sawbenson15/HPM_Thesis.

> In the fully controlled Alpha model, the beta coefficient for daily infections suggest that each additional infection is associated with an average home price *increase of 8.97 USD*, ceteris paribus. This finding is significant at the $p < 0.00$ level. I therefore reject the H_0 of Hypothesis I, concluding that the Corona crisis significantly increased housing prices (see ***Table 8***).

This foundational finding lays the groundwork for the following results, which attempt to explain the significant relationship between average daily infections and price at the level of individual hedonic variables.

Table 8 *OLS Result: Infections and Price*

	sold_price	
Predictors	*Estimates*	*p*
(Intercept)	161669.85 (52648.79 – 270690.91)	**0.004**
infections 3mma	8.97 (7.92 – 10.02)	**<0.001**
Observations	24394	
R^2 / R^2 adjusted	0.663 / 0.662	

Figure 18 shows the historical relationship between daily infections and sold price with a 1SD error margin at all point (in grey), the best single-variable fit (in green), and the Alpha model's marginal fit (in blue). We see that the many controls of the Alpha model flatten the estimated marginal effect of each additional daily infection on price, however, this relationship remains positive and significant, as previously established.

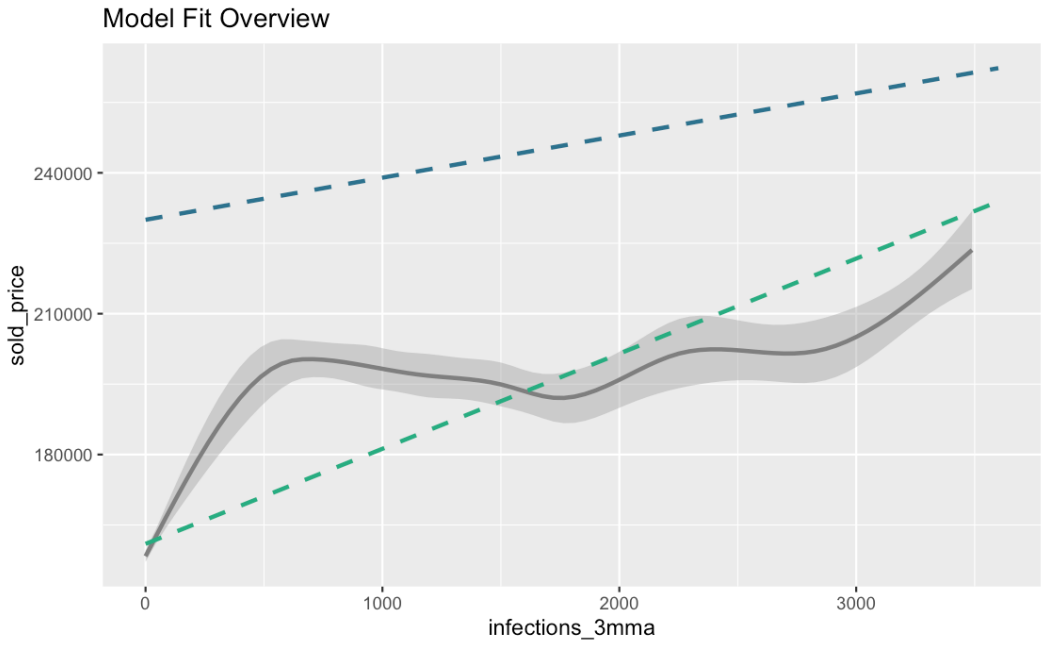

Figure 18 *Model Fit Overview: Infections and Price*

6.1.4 ML Modeling

To further test the relevance of Corona infections in determining prices, we look at this variable's relative ranking based on its ability to improve out-of-sample predictions within our XGBoost model. The measurement used is called Variable Importance and is defined by a combination of gain, cover, and frequency. Gain is how much prediction power is gained when the variable is added to the model; cover is how heavily the variable is weighted in each iteration, and frequency is how frequently the variable appears across all iterations.

We see that of the 104 unique variables offered to the XGBoost machine, the 3-month moving average of infection was the 19th most descriptive variable in predicting prices. I consider these results non-trivial and a very strong confirmation that the relationship found in the OLS model is correct.

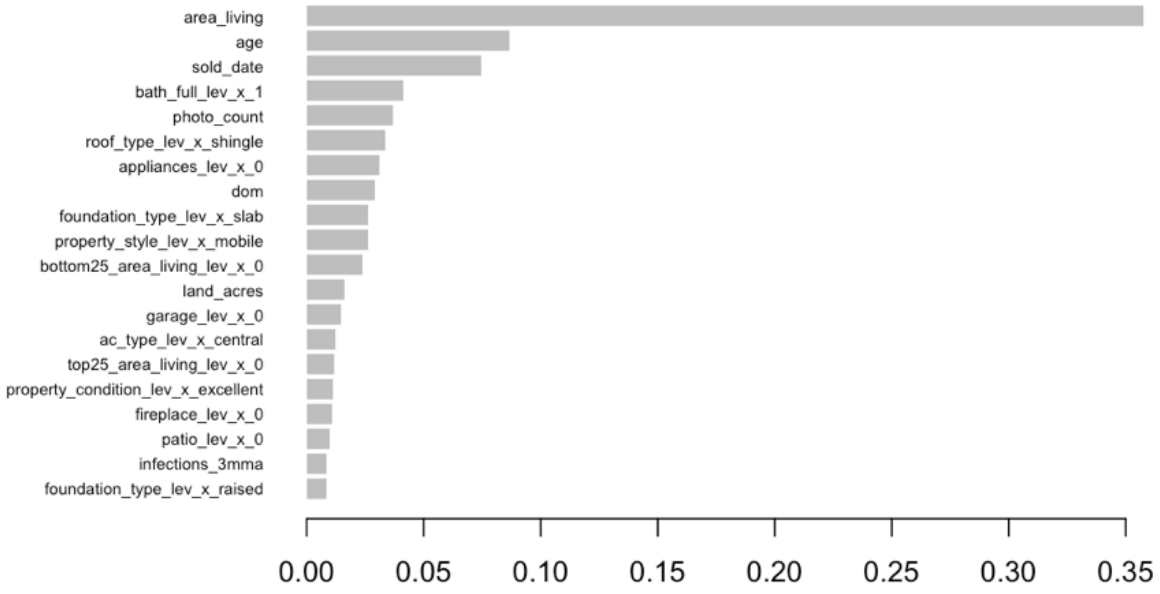

Figure 19 *XGBoost Variable Importance Ranking*

To get a detailed understanding of the marginal effect each additional daily infection has on price, we look at the PDP and the ICE of infections and price (see **Figure 20**). The PDP (top) shows the optimal price change with respect to daily infections which minimize the model's test MSE. The ICE (bottom) is the individual PDP for all 24,412 observations in the dataset, centered at infection = 0. The ICE allows us to understand the simple PDP, which is an average of the individual ICEs, in more detail and to pick up on any signs of heteroscedasticity among the individual sample predictions.

The PDP shows that on average, an increase in price at every level of daily infections > 0 reduces test MSE. Furthermore, this trend is generally upward trending, with a notable range of response between 1,800 and 2,500 infections, which I refer to as the *infection-price ridge.* Discussions regarding possible explanations for the general shape of this region will be explored in the final discussion.

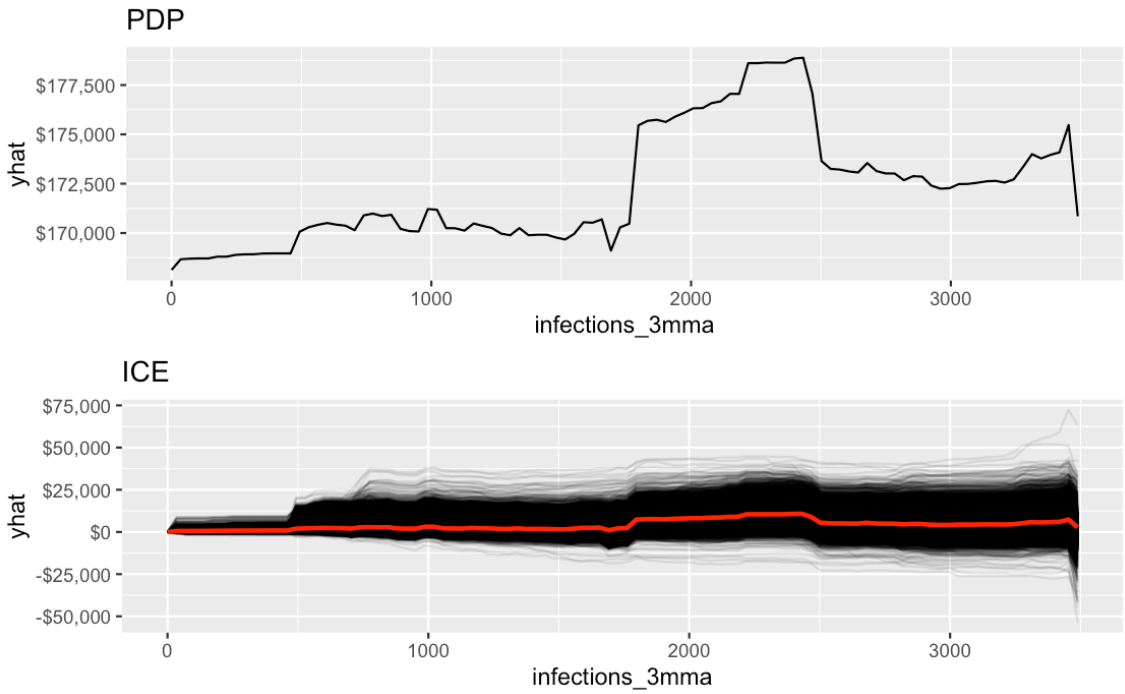

Figure 20 *PDP and ICE: Infections on Price*

6.2 Corona: Premium for Bedrooms

Foreword for this section and the following sections:

There are two contextual details which must be considered for one to appropriately interpret the following findings. First, the variables for total number bedrooms and the total living area of a home are highly correlated, and therefore when total living area is

included in the OLS model, the beta coefficients for each level of number of bedrooms represents the premium homeowners are willing to pay for the actual feature of an extra bedroom at a fixed living area, and not the extra living area itself. As this results in convoluted results, living area is excluded when testing the effect of bedrooms. This effectively measures the additional room and the average additional living area associated with that additional room at each level (e.g., from 1 bedroom to 2, 2 to 3, etc.).

Secondly, the OLS results listed below represent the *change in the premiums* (i.e., beta coefficients) from post versus pre-Corona and not the *absolute price premiums*, as in the case of measuring daily infection alone. This differentiation of *absolute price premiums* and *changes in premiums* will be represented in the following hedonic-feature analyses.

6.2.1 Summary of Findings

Preliminary findings show that the price distribution for every level of number of bedrooms increased after the beginning of the infection period (i.e., accumulation of infections >= 1000). The average number of total bedrooms between pre- and post-infections period deviate slightly, with post-infections period being on average 0.003 bedrooms higher. However, the key focus of this section is not absolute changes in the feature *number of bedrooms*, but rather in the change in premium for each level of number of bedrooms.

The fully controlled Alpha model results show that the average premium for each level of number of bedrooms significantly increased in response to increases in daily infections. Every increase found is at least significant at the $p < 0.05$ level except for single bedroom homes. Therefore, I reject the H_0: of hypothesis II and

conclude that the Corona crisis significantly increased demand-premiums for every level of number of bedrooms greater than 1. It should also be noted to the premium increase is increasingly larger at each level, suggesting that the premium for the 4th bedroom in a home increased more than the premium for the 3rd bedroom. Possible theories regarding this significant shift in premiums is explored in the final discussion.

6.2.2 Visual Review

Figure 21 shows the general distribution of number of bedrooms per residential property.

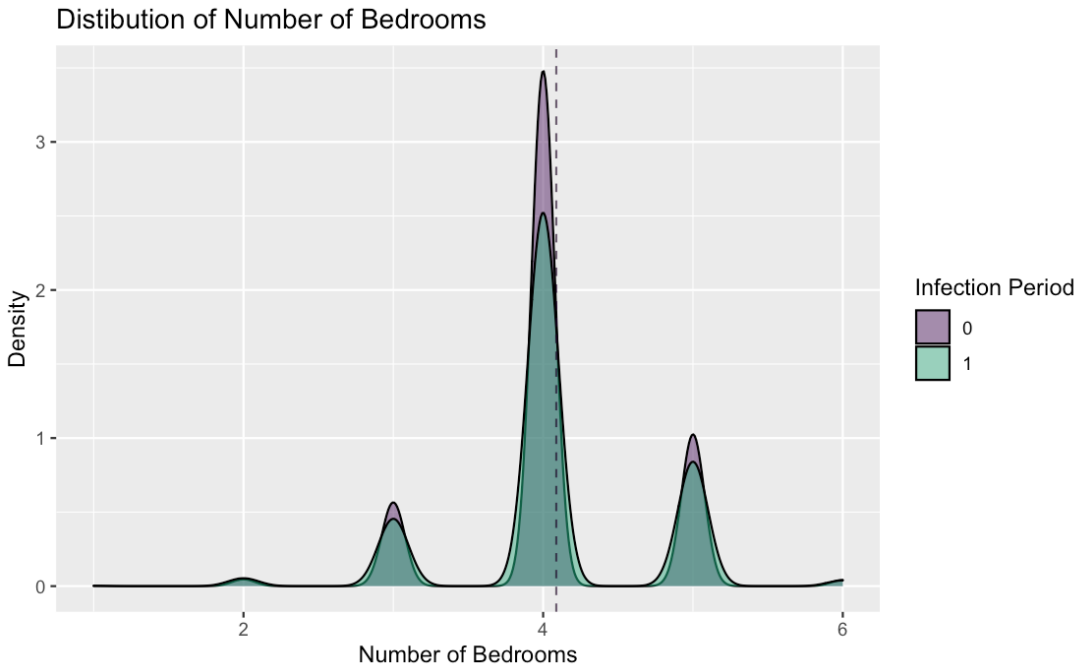

Figure 21 *Distribution of Number of Bedrooms*

Figure 22 shows the distributions of prices at each level of number of bedrooms (lhs) and the distributions of prices at each level of number of bedrooms split by before and after infection period (rhs).

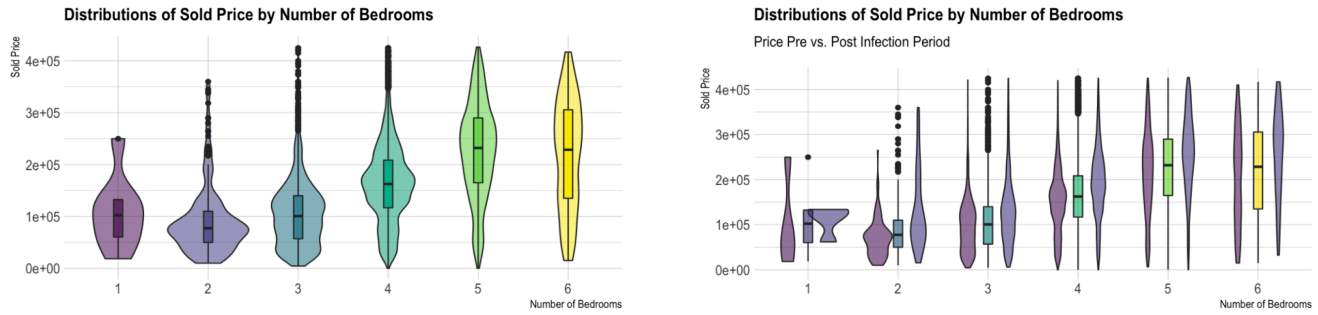

Figure 22 *Distribution of Sold Price and Number of Bedrooms*

It is also instructive to look at the same two plots above but normalized by square footage. We see a significant flattening between each level of number of bedrooms, however, a distinct change from pre versus post infection period is still visible (rhs). The next section will test if this shift from post to pre-Corona is significantly in general, and significantly related to daily infections.

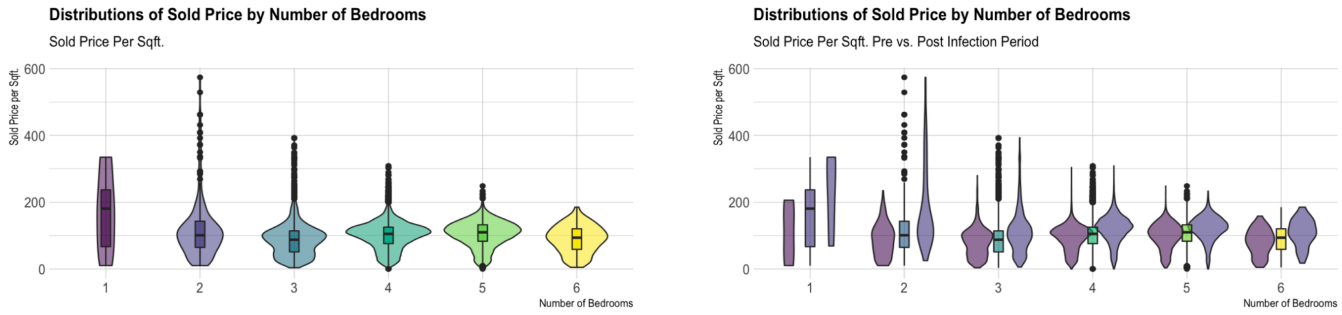

Figure 23 *Distribution of Sold Price and Number of Bedrooms psf.*

6.2.3 OLS Modeling

Under the fully controlled Alpha model, we see that Corona-driven premiums changes at every level of number of bedrooms is statistically significant except for single-bedroom properties. This finding suggests that for every additional 3-month-moving-average daily infection present at the time the property is sold is associated with an average increase of 25, 32, 37, 27, and 47 USD for levels 2-through-5 respectively (see ***Table 9***).

Table 9 *OLS Results: Number of Bedrooms, Infections, and Price*

	sold_price	
Predictors	*Estimates*	*p*
(Intercept)	234141.47 (112638.64 – 355644.31)	**<0.001**
beds total [1] * data factor$infections 3mma	25.21 (-7.69 – 58.12)	0.133
beds total [2] * data factor$infections 3mma	32.23 (-0.04 – 64.49)	0.050
beds total [3] * data factor$infections 3mma	36.77 (4.57 – 68.96)	**0.025**
beds total [4] * data factor$infections 3mma	37.09 (4.85 – 69.32)	**0.024**
beds total [5] * data factor$infections 3mma	47.40 (13.97 – 80.84)	**0.005**
Observations	24394	
R^2 / R^2 adjusted	0.568 / 0.567	

6.2.4 ML Modeling

Due to number of bedrooms being categorical, the results of the XGBoost model for these variables are not as easily visually represented as continuous variables, which have an

easier interpretation through PDPs. However, we can consider the results in written form and with adjusted graphs to represent the importance rankings.

The variable importance ranking of each factor level from 1-through-5 bedrooms is 84, 60, 47, 34, and 68 respectively from the total 104 variables used in the model. This relatively low relevance is expected, as much of the explanatory power of bedrooms on price is taken by total living area, ranked number 1 in the importance matrix. However, we see that test MSE is minimized by increasing price at each level of number of bedrooms (see ***Figure 24***).

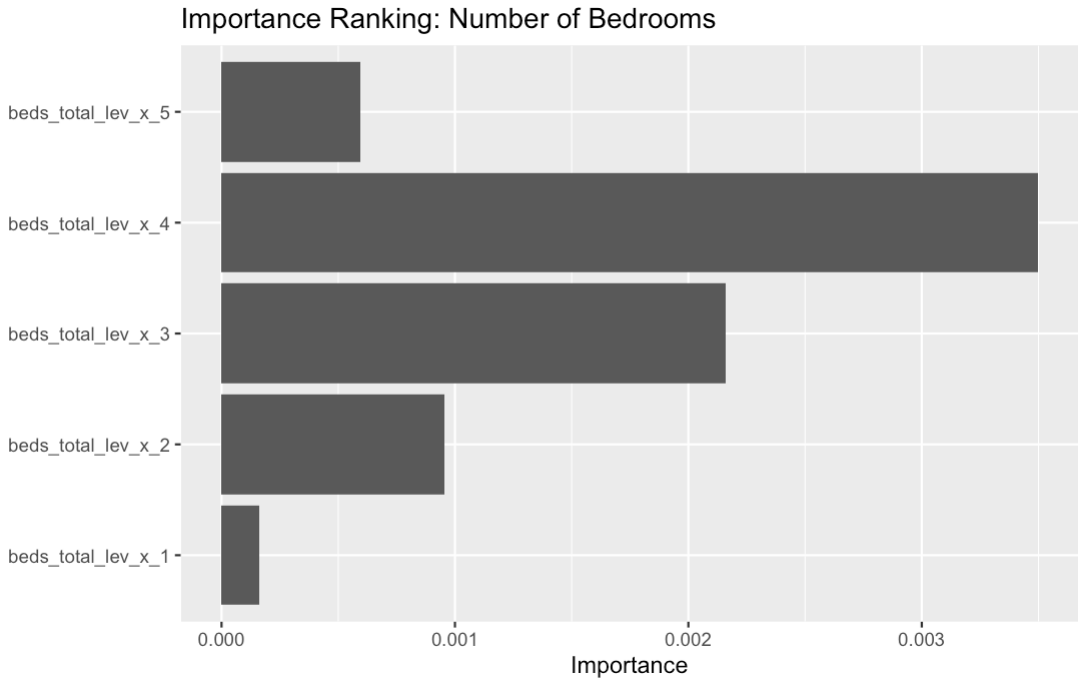

Figure 24 *Importance Ranking: Number of Bedrooms*

6.3 Corona: Premium for City Centrality

Foreword for this section:

In this section, it is important to note that a property being within city limits does not mean it should be viewed as being in the center of a major city. This categorization is better understood as a distinction from properties which are rural, such as properties on farmland and heavily wooded or otherwise remote areas. It should also be noted that results at several levels of analyses will be discussed, such as nominal differences in the populations, price reactions to daily infections for each sub-population, and the change in the premium for being within city limits.

6.3.1 Summary of Findings

Preliminary analysis of city centrality's impact on home prices shows that homes within city limits have a higher price on average when compared to rural properties. Further visual analysis suggests that prices grew in both sub-populations post versus pre-infection period, with city-central properties experiencing the larger nominal shift in average prices. The XGBoost model shows that an increase in prediction price when a property is within city limits maximally reduces the model's test MSE, suggesting this variable is a generally important feature, with a ranking of 42 in the importance matrix.

In the fully controlled Alpha model, the coefficient for daily infections and city limits suggests that each additional infection is associated with an average *increase of 5.17 USD* in the premium for being located within city limits compared to property not within city limits, ceteris paribus, while daily infections had no significant impact on rural properties (see **Table 11**). This finding is significant at

the $p < 0.00$ level. Therefore, I reject the H_0: of hypothesis III and conclude that the Corona crisis impacted properties within city limits more than those not within city limits.

6.3.2 Visual Review

To better understand the effect Corona had on properties specifically located within city limits, I split the variable city_limits' distribution by pre- versus post-infection period. This separation shows a clear increase in association with the infections period (see ***Figure 25***).

Figure 25 *Distributions of Price for City-Limits Properties*

6.3.3 OLS Modeling

First, we look at the main result which tests the change in premium for being within city limits, in contrast to rural properties, before versus after infection period. This finding shows each additional infection is associated with an average city_limits premium *increase of 5.17 USD,* compared to property not within city limits, ceteris paribus. This finding is significant at the $p < 0.00$ level.

Table 10 *OLS Results: City Limits and Infections on Price*

	sold_price	
Predictors	*Estimates*	*p*
(Intercept)	160389.19 (51378.41 – 269399.98)	**0.004**
data factor$infections 3mma	4.24 (1.25 – 7.22)	**0.005**
data factor$city limits [data factor$city limits1]	6380.32 (1990.11 – 10770.52)	**0.004**
data factor$infections 3mma * data factor$city limits [data factor$city limits1]	5.17 (2.14 – 8.19)	**0.001**
Observations	24394	
R^2 / R^2 adjusted	0.664 / 0.663	

To further highlight the differences in Corona's impact of properties within city limits versus outside of city limits, I ran a separate regression model for each sub-population with all the standard controls of the Alpha model to see the impact of daily infections on prices. This result shows that properties within city limits (lhs) have an average increase of 9.15 USD per additional daily infection which is significant at the $p < 0.00$ level, while the model for rural properties (rhs) shows that daily infections is not a significant variable in explaining variations in prices. It should also be noted how much smaller the number of observations is in the rural model.

Table 11 *OLS Results: City Limits versus Rural and Infections on Price*

	sold_price	
Predictors	*Estimates*	*p*
(Intercept)	164656.38 (54632.16 – 274680.60)	**0.003**
infections 3mma	9.15 (8.06 – 10.23)	**<0.001**
Observations	23381	
R^2 / R^2 adjusted	0.660 / 0.659	

	sold_price	
Predictors	*Estimates*	*p*
(Intercept)	-341427.54 (-492581.59 – -190273.49)	**<0.001**
infections 3mma	2.06 (-1.42 – 5.53)	0.246
Observations	1013	
R^2 / R^2 adjusted	0.799 / 0.787	

6.3.4 ML Modeling

The XGBoost PDP of the city_limits dummy variable shows that increasing property value when the property is within city limits maximally reduces test MSE.

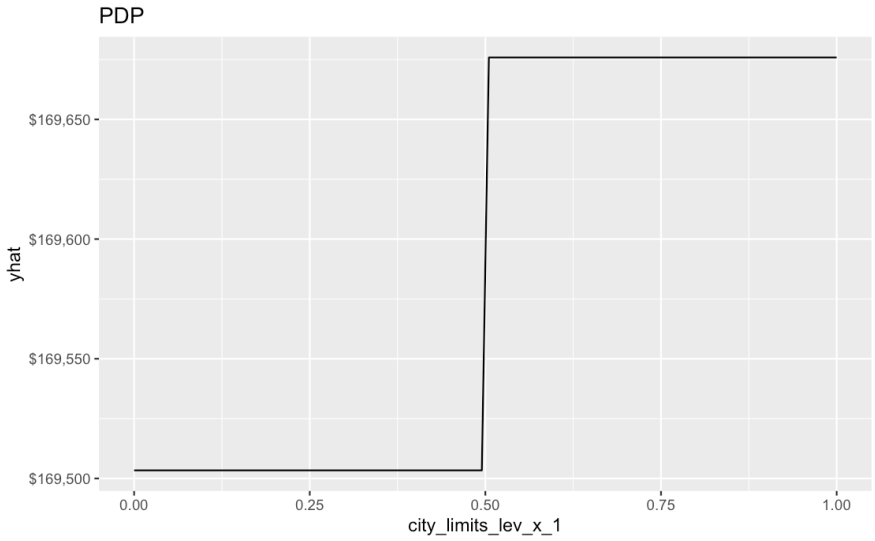

Figure 26 *PDP: City Limits*

6.4 Corona: Premium for Size

6.4.1 Summary of Findings

An analysis of size shows that this feature is smoothly distributed across the data set with the average total living area marginally higher in the post-infections period. The XGBoost model ranks this feature as the number one most important variable in predicting prices.

In the fully controlled Alpha model, the coefficient for daily infections and total living area suggests that each additional infection is associated with an average premium *increase of 0.02 USD* for total per square foot of living area, ceteris paribus. This finding

is significant at the $p < 0.00$ level. Therefore, I reject the H_0: of hypothesis IV and conclude that the Corona crisis increased the premium for property size.

Furthermore, I have run a separate Alpha model for the top and bottom 25th percentile of total living area. These results show that the premium for the smallest homes decreases per additional daily infection (-3.15 USD, p < 0.00), while homes in the top 25th percentile increase with respect to daily infections (0.97 USD, p < 0.34). This suggests a shift away from smaller homes more than a shift towards larger homes in response to the Corona crisis. Therefore, I fail to reject the H_0: of hypothesis V and conclude that the Corona crisis impacted properties in the top 25th percentile of property size more than the bottom 25th.

6.4.2 Visual Review

Total living area is relatively smoothly distributed across the data set with the average total living area marginally higher in the post-infections period. Since this feature is roughly stable in the short run, this result is expected (see ***Figure 27***).

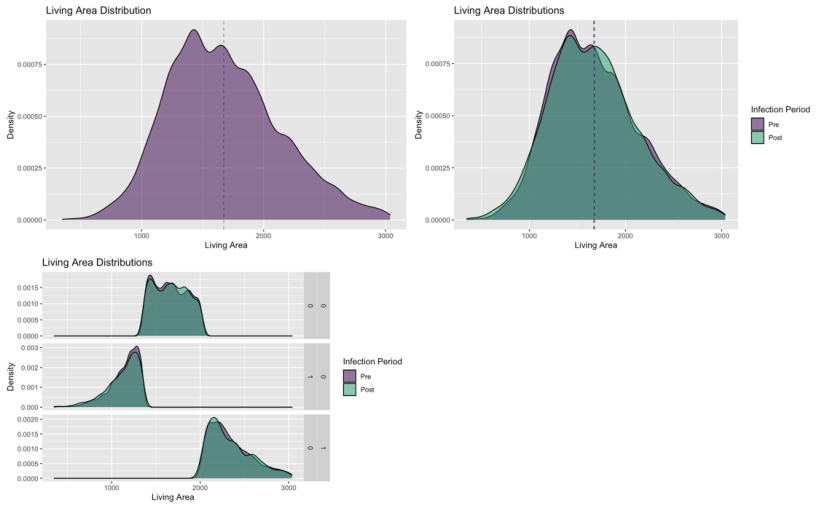

Figure 27 *Distribution of Living Area*

Price distribution, however, differ between pre- and post-infection period when normalized by per square foot (sqft.) (see ***Figure 28***).

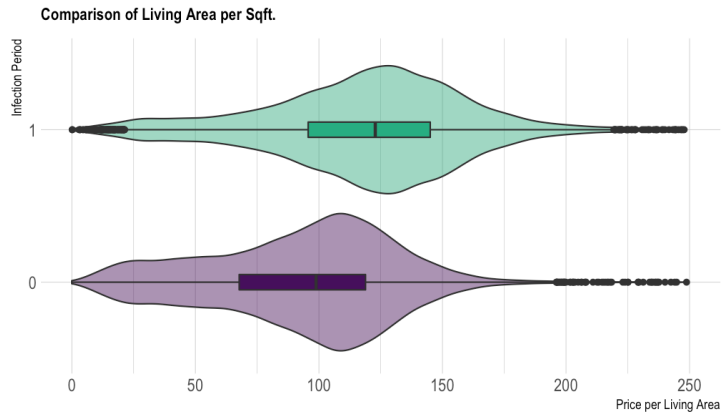

Figure 28 *Distribution of Living Area Before and After Infection Period*

6.4.3 OLS Modeling

The fully controlled Alpha model shows the coefficient for daily infections and total living area suggests that each additional infection is associated with an average premium *increase of 0.02 USD* for total living area, measured per square foot, ceteris paribus. This finding is significant at the $p < 0.00$ level.

Table 12 *OLS Results: Living Area and Infections*

	sold_price	
Predictors	*Estimates*	*p*
(Intercept)	213412.04 (98453.50 – 328370.57)	**<0.001**
area living * data factor$infections 3mma	0.02 (0.01 – 0.02)	**<0.001**
Observations	24394	
R^2 / R^2 adjusted	0.623 / 0.622	

To further test this relationship, I have run a separate Alpha model for the top and bottom 25th percentile of total living area. These results show that the premium for the smallest homes decreases per additional daily infection (-3.15 USD, $p < 0.00$), while homes in the top 25th percentile increase with respect to daily infections (0.97 USD, $p < 0.34$). This suggests a shift away from smaller homes and towards larger homes in response to the Corona crisis.

Table 13 *OLS Results: Top versus Bottom: Living Area and Infections*

	sold_price	
Predictors	*Estimates*	*p*
(Intercept)	204839.02 (91241.00 – 318437.04)	**<0.001**
data factor$infections 3mma * bottom25 area living	-3.15 (-5.00 – -1.29)	**0.001**
Observations	24394	
R^2 / R^2 adjusted	0.632 / 0.631	

	sold_price	
Predictors	*Estimates*	*p*
(Intercept)	217856.92 (105252.41 – 330461.42)	**<0.001**
data factor$infections 3mma * top25 area living	0.97 (-1.02 – 2.97)	0.340
Observations	24394	
R^2 / R^2 adjusted	0.638 / 0.638	

6.4.4 ML Modeling

The PDP and ICE plots show a stable and consistent increase in prices as total living area increases.

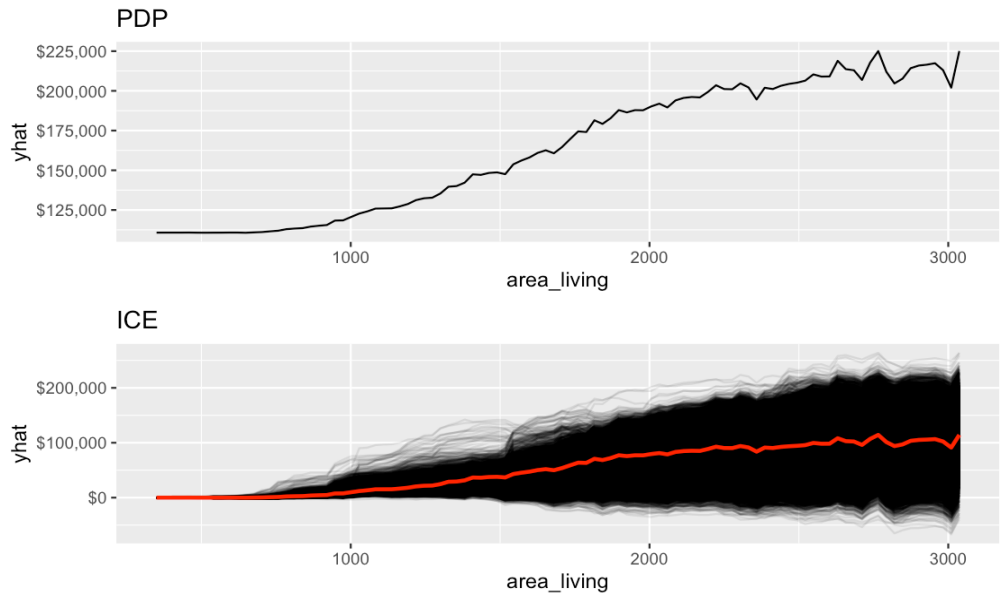

Figure 29 *PDP and ICE Plots: Living Area*

The heatmaps of predicted price with respect to daily infections and total living area show that both variables consistently increase XGBoost price predictions, with the Maximal Price Region (MPR) at the top 25th percentile of both variables.

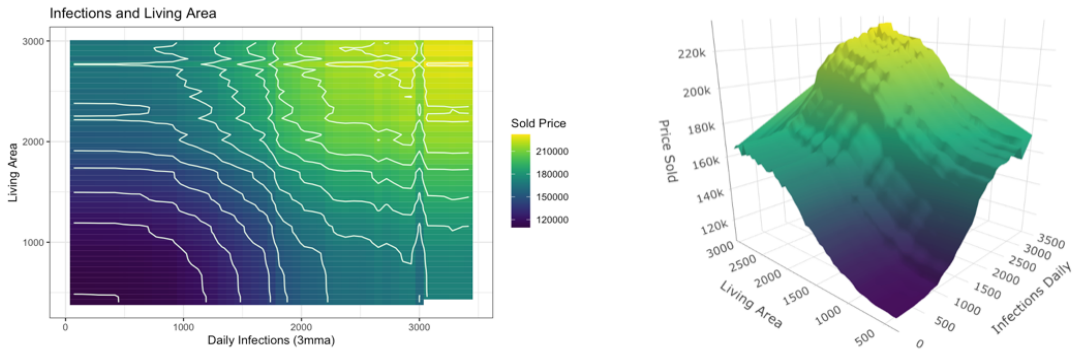

Figure 30 *PDP Heatmap and 3D: Living Area and Infections*

6.5 Corona: Premium for Age

6.5.1 Summary of Findings

A preliminary analysis of age shows that this feature is different in the pre- and post-infections periods, with the lower average age of homes sold in the post-infection period suggesting that homeowners prefer, on average, newer homes. The XGBoost model ranks this feature as the second most important variable in predicting prices.

In the fully controlled Alpha model, the coefficient for daily infections and age suggests that each additional daily infection is associated with an average premium *decrease of 0.06 USD* for each additional year of age, ceteris paribus. This result can be interpreted as an increase in the penalty per additional year of age, making older home even more undesirable post-infections period. This finding is significant at the $p < 0.00$ level.

Therefore, I reject the H_0: of *hypothesis VI* and conclude that The Corona crisis decreased the premium for property age.

Furthermore, I have run a separate Alpha model for the top and bottom 25th percentile of age. These results show that the premium for the oldest homes decreases per additional daily infection (-2.75 USD, $p < 0.00$), while the premium for being in the bottom 25th percentile of age (i.e., newest homes) is not significantly different pre- versus post-infection period. This suggests a distinct shift away from older homes in response to the Corona crisis. Therefore, I reject the H_0: of *hypothesis VII* and conclude that the Corona crisis impacted properties in the top 25th percentile of property age more than the bottom 25th.

6.5.2 Visual Review

The graph below highlights the differences in the age distributions between the pre and post infections period groups. We see a decrease in the average age in homes sold in the post-infection period.

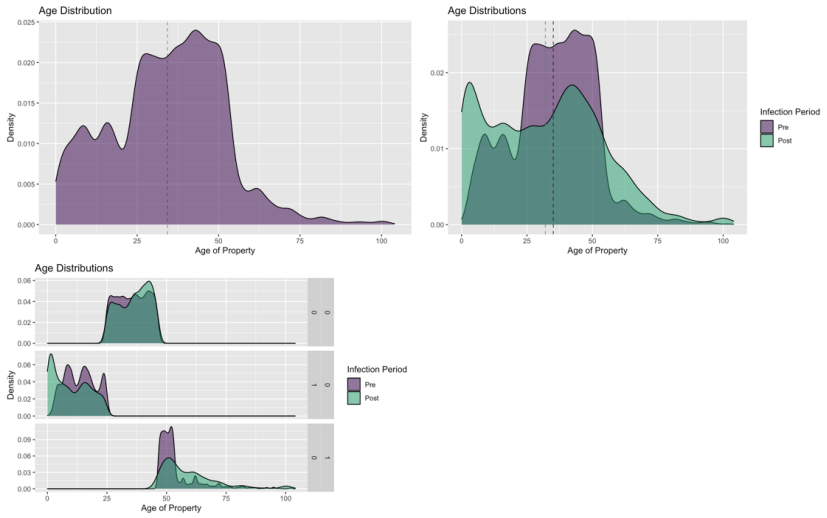

Figure 31 *Distributions of Age*

6.5.3 OLS Modeling

Alpha model finds that each additional daily infection is associated with an average premium *decrease of 0.06 USD* for each additional year of age, ceteris paribus. This result can be interpreted as an increase in the penalty per additional year of age, making older home even more undesirable in the post-infections period. This finding is significant at the $p < 0.00$ level.

Table 14 *OLS Results: Age and Infections*

	sold_price	
Predictors	*Estimates*	*p*
(Intercept)	80895.09 (-30282.40 – 192072.58)	0.154
age * data factor$infections 3mma	-0.06 (-0.10 – -0.02)	**0.002**
Observations	24394	
R^2 / R^2 adjusted	0.648 / 0.647	

Two separate Alpha models for the top and bottom 25th percentile of age show that the premium for the oldest homes *decreases* per additional daily infection (-2.75 USD, $p < 0.00$), while the premium for being in the bottom 25th percentile of age (i.e., newest homes) is *not significantly different* pre- versus post-infection period. This suggests that homeowners shifted distinctly away from older homes more than towards newer homes.

Table 15 *OLS Results: Top versus Bottom Age and Infections*

	sold_price	
Predictors	*Estimates*	*p*
(Intercept)	73702.52 (-37887.09 – 185292.12)	0.195
data factor$infections 3mma * top25 age	-2.75 (-4.61 – -0.89)	**0.004**
Observations	24394	
R^2 / R^2 adjusted	0.646 / 0.645	

	sold_price	
Predictors	*Estimates*	*p*
(Intercept)	126967.30 (16983.74 – 236950.87)	**0.024**
data factor$infections 3mma * bottom25 age	0.80 (-0.86 – 2.46)	0.346
Observations	24394	
R^2 / R^2 adjusted	0.654 / 0.653	

6.5.4 ML Modeling

The XGBoost model's PDP and ICE plots show the expected convened shape between age and price predictions with an inflection point at approximately 65 years. Since most properties sold in my time sample are between the ages of 0 and 55 years old (94%), the downward relationship between age and price prediction is most relevant for this paper.

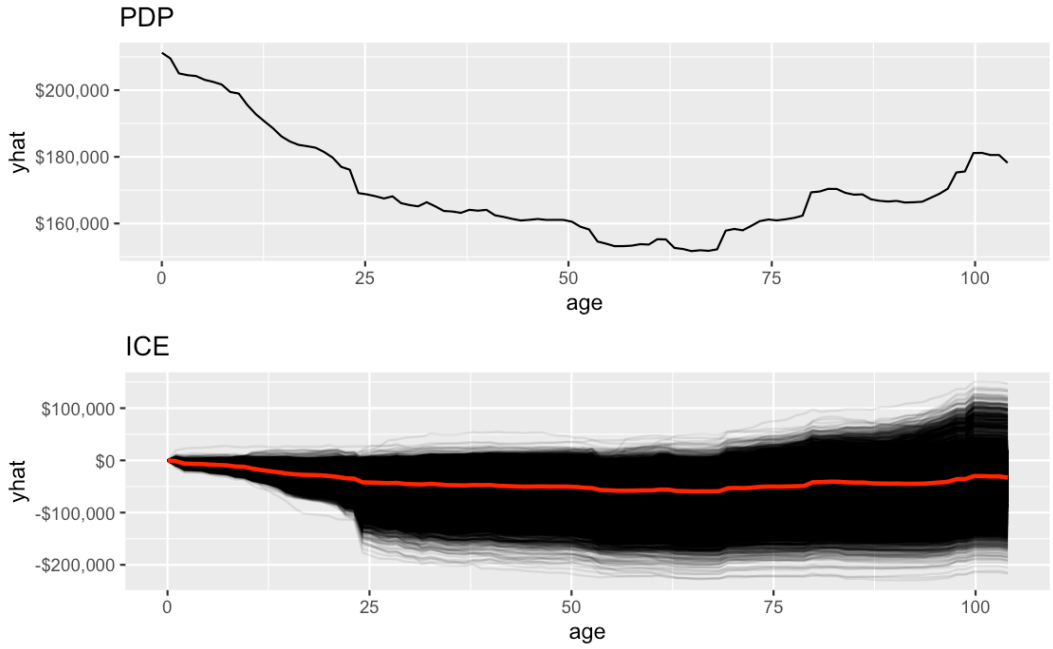

Figure 32 *PDP and ICE: Age and Price*

To better understand the marginal impact of age and daily infections on predicted home prices, we turn to the heat maps below. The highest marginal price additions happen in the region: 0 < age < 10 and 1800 < infections < 2300.

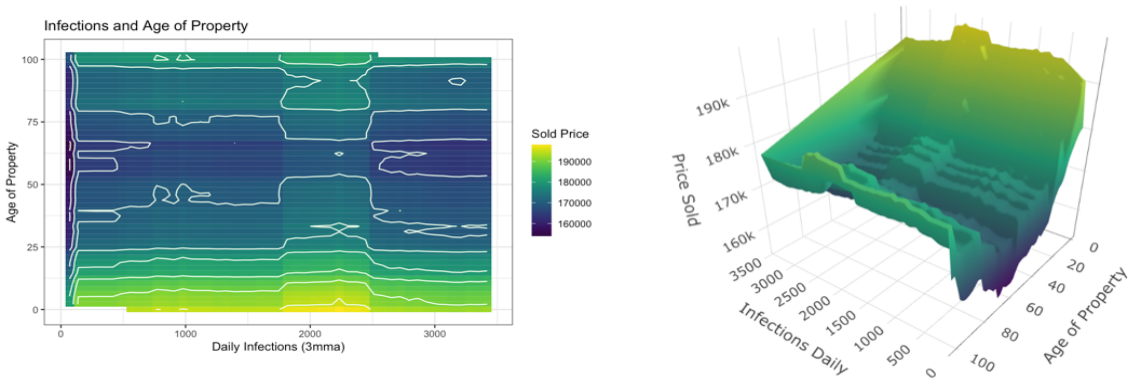

Figure 33 *PDP Heatmap and 3D: Age, Infections, and Price*

6.6 Corona: Change in Days on Market

Foreword for this section:

Days on market (DOM) is a strange variable as its relationship to the final selling price of a home is not unidirectional. This is because the number of days a home sits on the market is greatly determined by the relationship between the original listing price and the actual market value of the home, with over-priced homes taking longer to sell and underpriced homes taking less time. However, it is also the case that sellers who are willing to wait for the right buyer to come along will be rewarded by selling their home at a higher price. This

feedback loop makes standard econometric interpretations difficult and the following results in this section should be analyzed with this fact in mind.

6.6.1 Summary of Findings

A preliminary analysis of DOM shows that this feature noticeably shifted in the post-infection period, with the average DOM being lower when compared to the pre-infections period. Of the 104 variables used to train the XGBoost model, DOM is ranked as the 8th most important variable in predicting prices.

In the fully controlled Alpha model, the coefficient for daily infections and DOM suggests that each additional daily infection is associated with an average premium *decrease of 0.04 USD* for each additional day a property sits on the market before being sold, ceteris paribus. This result can be interpreted as an increase in the penalty per additional day on the market. This finding is significant at the $p < 0.00$ level. Therefore, I reject the H_0: of *hypothesis VIII* and conclude that the Corona crisis significantly decreased the premium for days on market

Furthermore, I have run a separate Alpha model for the top and bottom 25th percentile of DOM. These results show that the negative premium for the homes which sat on the market the longest decreased (-7.11 USD, $p < 0.00$), while the premium for being in the bottom 25th percentile (i.e., homes sold the fastest) is not significantly different pre- versus post-infection period. Therefore, I reject the H_0: of *hypothesis IX* and conclude that the Corona crisis impacted properties in the top 25th percentile of property age more than the bottom 25th.

6.6.2 Visual Review

The preliminary graphical analysis of DOM shows a clear average decrease of DOM for homes sold in the post-infection period.

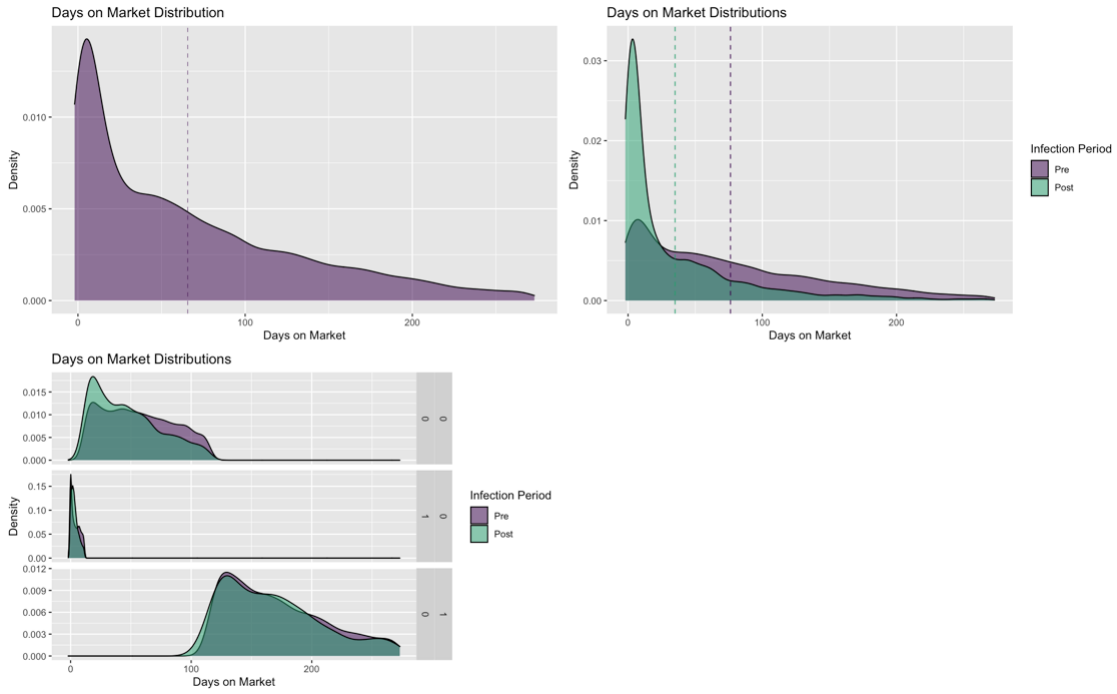

Figure 34 *Distributions: Days on Market*

6.6.3 OLS Modeling

Alpha model, the interaction coefficient for daily infections and DOM suggests that each additional daily infection is associated with an average premium *decrease of 0.04 USD* for each additional day a property sits on the market before being sold, ceteris paribus. This result can be interpreted as an increase in the penalty per additional day on the market. This finding is significant at the $p < 0.00$ level.

Table 16 *OLS Results: Days on Market and Infections*

	sold_price	
Predictors	*Estimates*	*p*
(Intercept)	160914.76 (51562.22 – 270267.29)	**0.004**
dom * data factor$infections 3mma	-0.04 (-0.05 – -0.02)	**<0.001**
Observations	24394	
R^2 / R^2 adjusted	0.661 / 0.660	

Separate Alpha models for the top and bottom 25th percentile of DOM. These results show that the negative premium for the homes which sat on the market the longest decreased (-2.75 USD, $p < 0.00$), while the premium for being in the bottom 25th percentile (i.e., homes sold the fastest) is not significantly different pre- versus post-infection period.

Table 17 *OLS Results: Top versus Bottom Days on Market and Infections*

	sold_price	
Predictors	*Estimates*	*p*
(Intercept)	165249.26 (55934.48 – 274564.03)	**0.003**
data factor$infections 3mma * top25 dom	-7.11 (-10.09 – -4.14)	**<0.001**
Observations	24394	
R^2 / R^2 adjusted	0.661 / 0.660	

	sold_price	
Predictors	*Estimates*	*p*
(Intercept)	171071.35 (61669.19 – 280473.50)	**0.002**
bottom25 dom [1] * data factor$infections 3mma	-2.19 (-3.86 – -0.51)	**0.010**
Observations	24394	
R^2 / R^2 adjusted	0.661 / 0.660	

6.6.4 ML Modeling

PDP and ICE graphs show a sharp initial decrease in prediction prices and then a consistent downward trend. The end behavior at the top of DOM is explained by a small number of observations at the extreme and can therefore be ignored.

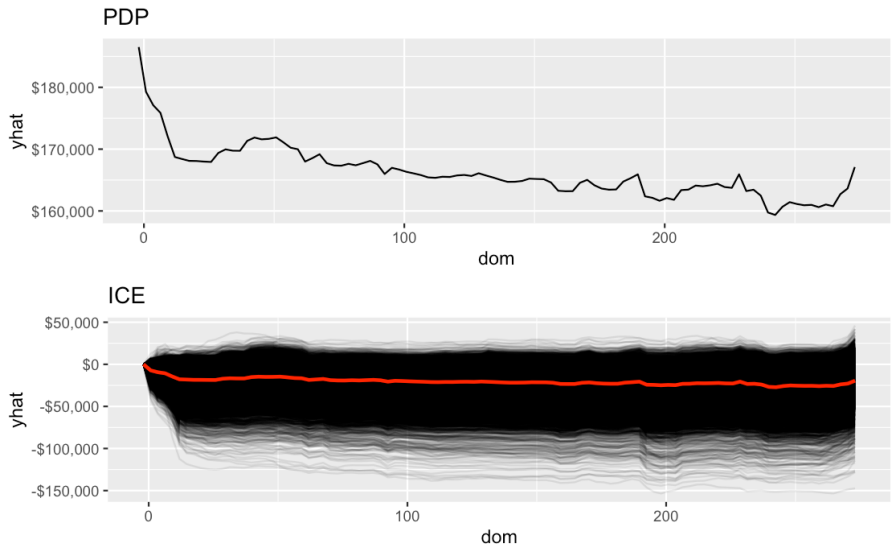

Figure 35 *PDP and ICE: Days on Market and Price*

The heatmaps for DOM and daily infections show that properties sold at approximately 0 DOM and within the infection-price ridge are predicted to have the highest price.

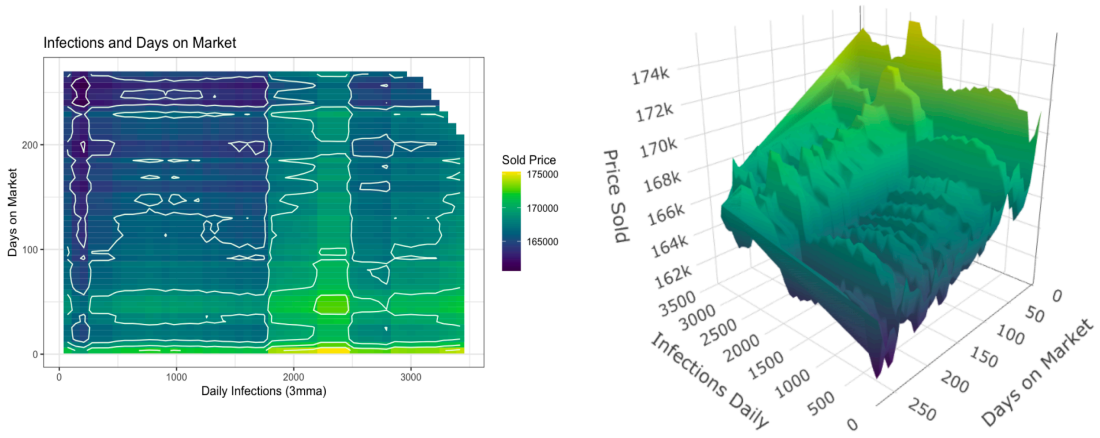

Figure 36 *PDP Heatmap and 3D: Days on Market, Infections, and Price*

7. Discussion

7.1 General Implications of Results

The results of this study show how market shocks not only change nominal price levels for housing markets but can also change the relative demand for certain features of homes. As social norms around how people live their lives, such as how they work, which home styles they prefer, the average number of children per household, etc., there is a measurable change relative demand for individual hedonic housing features. These finding are in line with the existing research discussed in the literature review of §2 and serve to further support the fundamental assumptions of the HPM.

7.2 Policy Implications

The results of this study suggest that homes which belong to the bottom quartile of the socio-economic scale (i.e., smaller, cheaper, etc.) can be impacted more by economic shocks than properties in the top quartile. This could suggest that families living in less expensive homes are disproportionately impacted by market shocks such as Corona and therefore it may be advisable for local and federal governments to consider special protections and provisions for owners of these homes.

7.3 Limitations

The scope of this paper, time and data availability lead to key limitations:

1. The variable used to measure the total economic impact of the Corona crisis is the 3-month moving average of recorded infections. Though this is a good single-variable to measure market response, it is obviously limited in its ability to capture the full impact of the crisis.

2. The data campaign conducted for this paper resulted in a large and rich data set, however, most variables were physical features of the properties and do not include very relevant temporal and spatial features such as locations, weather, general socio-economic characteristics of surrounding populations, etc. It is conceivable and even expected that a wider set of variables would produce more robust and informative results. A further data-related limitation of this data set is that it includes only properties in the state of Louisiana. To make broader claims about changes in housing market preferences, one would need to compare Louisiana housing market to other markets to measure similarities, or simply include a nation-wide or international data set. Lastly, this dataset does not consider interest rate changes, material-logistical issues, and other relevant information for modeling housing pricing dynamics. These are all extensions which could be considered for future research.

3. Though a very basic price index was created for this paper, it would be beneficial to have a true, repeated-sales index dataset to measure true changes across the same property at different times. This would allow for even stronger and more robust results.

4. The ML model used in this paper is very strong, however, it is expected that one could produce stronger and more robust results with more sophisticated modeling methods and stronger computers for extremely refined hyperparameter tuning.

8. Conclusion

The purpose of this paper is to understand the economic impact the Corona crisis had on housing prices and how these impacts changed the relative demand of homeowners for certain hedonic features of housing through modeling market data collected from Louisiana, USA. To determine which variables would be important to analyze, an *eXtreme Gradient Boost* (XGBoost) machine learning algorithm was trained on a large (i.e., 104×24412) dataset of hedonic features. An analysis of the XGBoost's variable-importance ranking shows that number of bedrooms, city centrality, total living area, age, and days on the market were among the topmost important variables in determining a home's market price. Therefore, a deeper dive into how the crisis impacted the relative demand for these features is critically important in understanding the bigger picture of Corona's impact on the housing market.

A panel of analyses shows that Corona, measured by the daily 3-month moving average of official infections at the time of sale, significantly increased average home prices by *8.97 USD* per additional daily infection (PADI). This finding establishes the Corona-specific reaction of housing market prices to daily infections. When analyzing how this price shift is explained by changes in relative demand for certain property characteristics, I find: the premiums for each level of number of bedrooms is significantly increased in response to daily infections, with an average increase of *32, 37, 27, and 47 USD* PADI for levels 2-through-5 bedrooms respectively; the premium for being central to a city increased by *5.17 USD* PADI while properties located outside the city were not significantly impacted by daily infections; premiums for home size increased by *0.02 USD* per square foot, PADI, with the premium for the smallest homes being significantly decreased by *3.15 USD* PADI; premiums

for property age, which are historically negative, decreased even further with the penalty for each additional year of age increasing by *0.06 USD* PADI with the oldest homes experiencing the largest loss of *2.75 USD* PADI, per year of age and; an increase in the penalty for each additional day a home sits on the market of *0.04 USD* PADI with homes sitting the longest period of time on the market experiencing the largest loss of *2.75 USD* PADI, per day. These findings describe the ideal home to own during the Corona crisis as a newly built, 3,000 sqft. living area, 5-bedroom home centrally located to the nearest city which is sold in 0 days once made publicly available for sale.

This paper establishes the HPM as an effective method of capturing changes in relative demand for differentiable home features across subgroups. The research framework established in this paper is rife with possible extensions, particularly for topics focused on temporal and spatial data. Furthermore, this paper is considerably wider than it is deep, with each variable of interest warranting an entire research topic of its own. To my knowledge, this is the only HPM study on the Corona crisis using ML algorithms on US data.

Real estate markets make up a significant part of every national economy and account for most of average household wealth, however, methods of analyzing price dynamics and changes in feature-specific preferences across time and space are largely lacking. These academic shortcomings have widespread market implications that generate avoidable market frictions such as an inaccurate understanding of the fundamental underlying utility which drives household-consumer behavior and newly constructed homes being built with suboptimal feature mixes. It is imperative that future research remedies these shortcomings, as doing so could nudge real estate markets into a new age of efficiency.

Bibliography

Anguita, Davide, Alessandro Ghio, Noemi Greco, Luca Oneto, and Sandro Ridella. 2010. "Model Selection for Support Vector Machines: Advantages and Disadvantages of the Machine Learning Theory." In *The 2010 International Joint Conference on Neural Networks (IJCNN)*, 1–8. IEEE. https://doi.org/10.1109/IJCNN.2010.5596450.

Anissanti, Meita. 2021. "The Link Between GDP Growth and the Real Estate Market." *Asia Green Real Estate*, November. https://www.asiagreen.com/en/news-insights/the-link-between-gdp-growth-and-the-real-estate-market.

Arnold, Mark J., and Kristy E. Reynolds. 2003. "Hedonic Shopping Motivations." *Journal of Retailing* 79 (January): 77–95. https://doi.org/10.1016/S0022-4359(03)00007-1.

BAO, Helen X. H., and Cynthia M. GONG. 2016. "ENDOWMENT EFFECT AND HOUSING DECISIONS." *International Journal of Strategic Property Management* 20 (December): 341–53. https://doi.org/10.3846/1648715X.2016.1192069.

Bernasconi, Michele, Luca Corazzini, and Raffaello Seri. 2014. "Reference Dependent Preferences, Hedonic Adaptation and Tax Evasion: Does the Tax Burden Matter?" *Journal of Economic Psychology* 40 (February): 103–18. https://doi.org/10.1016/j.joep.2013.01.005.

Berry, Brian J. L., and Robert S. Bednarz. 1975. "A Hedonic Model of Prices and Assessments for Single-Family Homes: Does the Assessor Follow the Market or the Market Follow the Assessor?" *Land Economics* 51 (February): 21. https://doi.org/10.2307/3145138.

Bingyang, L V, Mao Jie, and L V Yinhan. 2013. "Incentive Mechanism for Governments in the Market of Real Estate: Problems and Reforms." *Finance &Amp; Trade Economics*, 7.

Breusch, T. S., and A. R. Pagan. 1979. "A Simple Test for Heteroscedasticity and Random Coefficient Variation." *Econometrica* 47 (September): 1287. https://doi.org/10.2307/1911963.

Bruin, J. 2011. "Newtest: Command to Compute New Test @ONLINE." http://www.ats.ucla.edu/stat/stata/ado/analysis/.

Clapp, John M., and Carmelo Giaccotto. 1998. "Residential Hedonic Models: A Rational Expectations Approach to Age Effects." *Journal of Urban Economics* 44 (November): 415–37. https://doi.org/10.1006/juec.1997.2076.

Collett, David, Colin Lizieri, and Charles Ward. 2003. "Timing and the Holding Periods of Institutional Real Estate." *Real Estate Economics* 31 (June): 205–22. https://doi.org/10.1111/1540-6229.00063.

Cowling, Keith, and John Cubbin. 1972. "Hedonic Price Indexes for United Kingdom Cars." *The Economic Journal* 82 (September): 963. https://doi.org/10.2307/2230261.

Curcuru, Stephanie, John Heaton, Deborah Lucas, and Damien Moore. 2010. "Heterogeneity and Portfolio Choice: Theory and Evidence." *Handbook of Financial Econometrics: Tools and Techniques*, 337–82. https://doi.org/10.1016/B978-0-444-50897-3.50009-2.

Du, Hongyan, Yongkai Ma, and Yunbi An. 2011. "The Impact of Land Policy on the Relation Between Housing and Land Prices: Evidence from China." *The Quarterly Review of Economics and Finance* 51 (February): 19–27. https://doi.org/10.1016/j.qref.2010.09.004.

Dulberger, Ellen R. 1987. "THE APPLICATION OF AN HEDONIC MODEL TO a QUALITY ADJUSTED PRICE INDEX FOR COMPUTER PROCESSORS."

Dutta, Sourav. 2018. "An Overview on the Evolution and Adoption of Deep Learning Applications Used in the Industry." *WIREs Data Mining and Knowledge Discovery* 8 (July). https://doi.org/10.1002/widm.1257.

Epple, Dennis. 1987. "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products." *Journal of Political Economy* 95 (February): 59–80. https://doi.org/10.1086/261441.

FRED. 2021. "Gross Domestic Product." Federal Reserve Bank of the USA. https://fred.stlouisfed.org/series/GDP.

Gardiner, Brian. 1997. "Squatters' Rights and Adverse Possession: A Search for Equitable Application of Property Laws." *Ind. Int'l &Amp; Comp. L. Rev.* 8: 119.

Goodman, A. C. 1978. "Hedonic Prices, Price Indices and Housing Markets." *Journal of Urban Economics*, 471–84.

Gouriéroux, Christian, and Anne Laferrère. 2009. "Managing Hedonic Housing Price Indexes: The French Experience." *Journal of Housing Economics* 18 (September): 206–13. https://doi.org/10.1016/j.jhe.2009.07.012.

Guilkey, David, Mike Miles, and Rebel Cole. 1989. "The Motivation for Institutional Real Estate Sales and Implications for Asset Class Returns." *Real Estate Economics* 17 (March): 70–86. https://doi.org/10.1111/1540-6229.00474.

Herath, Shanaka, and Gunther Maier. 2010. "The Hedonic Price Method in Real Estate and Housing Market Research: A Review of the Literature."

Holbrook, Morris B., and Elizabeth C. Hirschman. 1982. "The Experiential Aspects of Consumption: Consumer Fantasies, Feelings, and Fun." *Journal of Consumer Research* 9 (September): 132. https://doi.org/10.1086/208906.

Hoy, Michael, and Emmanuel Jimenez. 1991. "Squatters' Rights and Urban Development: An Economic Perspective." *Economica* 58 (February): 79. https://doi.org/10.2307/2554976.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy* 98: 1325–48.

LaDH. 2022. "Louisiana Department of Health: Covid-19 Information @ONLINE." https://ldh.la.gov/Coronavirus/.

LeSage, James P., and R. Kelley Pace. 2004. "Models for Spatially Dependent Missing Data." *The Journal of Real Estate Finance and Economics* 29 (September): 233–54. https://doi.org/10.1023/B:REAL.0000035312.82241.e4.

LING, DAVID C., JOSEPH T. L. OOI, and THAO T. T. LE. 2015. "Explaining House Price Dynamics: Isolating the Role of Nonfundamentals." *Journal of Money, Credit and Banking* 47 (March): 87–125. https://doi.org/10.1111/jmcb.12194.

Locke, John, and Wolfram Engels. 1691. *Some Considerations of the Consequences of the Lowering of Interest and Raising the Value of Money*. Verlag Wirtschaft und Finanzen.

Manhertz, Treh. 2021. "The u.s. Housing Market Gained More Value in 2020 Than in Any Year Since 2005." *Zillow Research*. https://www.zillow.com/research/zillow-total-housing-value-2020-28704/.

Matas, Anna, and Josep-Lluis Raymond. 2009. "Hedonic Prices for Cars: An Application to the Spanish Car Market, 1981–2005." *Applied Economics* 41 (October): 2887–2904. https://doi.org/10.1080/00036840701720945.

McKibbin, Warwick J., and Roshen Fernando. 2020. "Global Macroeconomic Scenarios of the COVID-19 Pandemic." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3635103.

Mohd, Thuraiya, Nur Syafiqah Jamil, Noraini Johari, Lizawati Abdullah, and Suraya Masrom. 2020. "An Overview of Real Estate Modelling Techniques for House Price Prediction." *Charting a Sustainable Future of ASEAN in Business and Social Sciences*, 321–38. https://doi.org/10.1007/978-981-15-3859-9_28.

Moulton, Brent R. 1996. "Bias in the Consumer Price Index: What Is the Evidence?" *Journal of Economic Perspectives* 10 (November): 159–77. https://doi.org/10.1257/jep.10.4.159.

Nanda, Anupam, Sotirios Thanos, Eero Valtonen, Yishuang Xu, and Razieh Zandieh. 2021. "Forced Homeward: The COVID-19 Implications for Housing." *Town Planning Review* 92 (January): 25–31. https://doi.org/10.3828/tpr.2020.79.

Nicolaides, Phedon. 1988. "Limits to the Expansion of Neoclassical Economics." *Cambridge Journal of Economics* 12: 313–28.

Pace, R. Kelley, and Otis W. Gilley. 1998. "Generalizing the OLS and Grid Estimators." *Real Estate Economics* 26 (June): 331–47. https://doi.org/10.1111/1540-6229.00748.

Pagourtzi, Elli, Vassilis Assimakopoulos, Thomas Hatzichristos, and Nick French. 2003. "Real Estate Appraisal: A Review of Valuation Methods." *Journal of Property Investment & Finance* 21 (August): 383–401. https://doi.org/10.1108/14635780310483656.

Schultze, Charles L. 2003. "The Consumer Price Index: Conceptual Issues and Practical Suggestions." *Journal of Economic Perspectives* 17 (February): 3–22. https://doi.org/10.1257/089533003321164921.

Shimizu, Chihiro, Hideoki Takatsuji, Hiroya Ono, and Kiyohiko G. Nishimura. 2010. "Structural and Temporal Changes in the Housing Market and Hedonic Housing Price Indices." *International Journal of Housing Markets and Analysis* 3 (October): 351–68. https://doi.org/10.1108/17538271011080655.

Sirmans, G. Stacy, Lynn MacDonald, David A. Macpherson, and Emily Norman Zietz. 2006. "The Value of Housing Characteristics: A Meta-Analysis." *The Journal of Real Estate Finance and Economics* 33 (November): 215–40. https://doi.org/10.1007/s11146-006-9983-5.

Tajani, Francesco, 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 (November): 592. https://doi.org/10.3390/buildings11120592.

Turing, Alan Mathison. 1950. "Mind." *Mind* 59: 433–60.

Wakefield, Robin L, and Dwayne Whitten. 2006. "Mobile Computing: A User Study on Hedonic/Utilitarian Mobile Device Usage." *European Journal of Information Systems* 15 (June): 292–300. https://doi.org/10.1057/palgrave.ejis.3000619.

Wallace, Nancy E, and Richard A Meese. 1997. "The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches." *The Journal of Real Estate Finance and Economics* 14: 51–73.

Wheaton, William C. 1999. "Real Estate "Cycles": Some Fundamentals." *Real Estate Economics* 27 (June): 209–30. https://doi.org/10.1111/1540-6229.00772.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (May): 817. https://doi.org/10.2307/1912934.

WHO. 2021. "Covid19.who.int." *World Health Emergency Dashboard WHO*. https://covid19.who.int/table.