



저작자표시 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#) 

Master's Thesis of Geography

Real Estate Acquisition in
Bucheon-si by Foreigners:
Machine Learning Models & Interpretations

외국인의 부천시 부동산 취득:
기계학습모형과 해석

August 2021

Graduate School of Social Sciences
Seoul National University
Geography Major

Jiyong Bae

Real Estate Acquisition in
Bucheon-si by Foreigners:
Machine Learning Models & Interpretations

Keyho Park

Submitting a master's thesis of Geography

August 2021

Graduate School of Social Sciences
Seoul National University
Geography Major

Jiyong Bae

Confirming the master's thesis written by

Jiyong Bae

August 2021

Chair _____
Vice Chair _____
Examiner _____

Abstract

In 2020, the area and the number of lots owned by foreigners^① increased in Gyeonggi Province while the foreign population of the province decreased due to the COVID-19 pandemic. With soaring house prices and speculation emerging as a national issue, the province enforced a permission system for real estate acquisition by foreigners.

While absentee owners may lead to land grabbing in urban context, there should be caution on xenophobic sentiments that attribute the failure of stabilizing house prices to ordinary foreigners during the pandemic. In order to make a sound judgement on the issue, a proper quantitative analysis on the spatial characteristics of real estate acquisition should precede.

This research analyzes the spatial dynamics of real estate acquisition in the city of Bucheon by foreigners. Within Gyeonggi Province, Bucheon has been the most prominent city for real estate acquisition by foreigners during the last five years.

The analysis shows that Chinese nationals especially those with Korean ethnicity emerged as the main acquirers of real estate in Bucheon. The result is contrary to past studies which indicate Americans and Taiwanese as the main acquirers of real estate in Seoul while the Chinese were treated as tenants of old houses.

According to the Theil index, the spatial distribution of acquisition became even more uneven in both microscale and macroscale. During the COVID-19 pandemic, the distribution has become increasingly concentrated on apartments and officetels^② around Sosa station, which is the junction

^① People or corporates without South Korean nationality.

between a railroad connecting the cities of Incheon and Seoul, and another railroad connecting Bucheon and the city of Ansan.

Through interpretable machine learning, the apartments and officetels acquired by foreigners were characterized as newly-built stand-alone buildings that were about sixteen-stories tall. The newly enforced permission system was proved to be less effective in Bucheon because most of the acquisitions did not meet the conditions for being subject to the regulation.

This research implies a new formation of Chinatown in the Seoul Capital Area represented by new apartments with Chinese owners, which is different from other Chinatowns discussed in previous research. It focuses on the conversion of the Chinese as simple occupants of space to owners of space in the Seoul Capital Area, although the locations and characteristics of the apartments owned by them may be different from those typically favored by Koreans, such as big complexes in Sang and Sinjung.

Keywords : Exploratory data analysis, real estate, Korean Chinese, COVID-19 pandemic, interpretable machine learning, Theil index.

Student Number : 2019-27834

② A type of studio apartments that can be used as both office and dwelling.

Table of Contents

Chapter 1. Introduction.....	1
1.1. Research Background and Purpose	1
1.2. Definition of Terms.....	2
1.3. Organization of Chapters.....	3
Chapter 2. Literature Review.....	4
2.1. Controversy on Urban Land Acquisition by Foreigners	4
2.2. Spatial Mode of Foreigners in South Korea	5
2.3. Interpretable Machine Learning and Geography.....	6
Chapter 3. Research Methodology.....	7
3.1. Study Area Selection	7
3.2. Data Collection and Processing	10
3.3. Theil Index	11
3.4. Decision Tree based Machine Learning	12
Chapter 4. Research Result.....	14
4.1. The Origin of Foreigners Buying Real Estate	14
4.2. Comparison with Foreign Population	18
4.3. Dynamics of Real Estate Acquisition by Foreigners	26
4.4. Characteristics of Apartments Acquired by Foreigners	35
Chapter 5. Discussion and Conclusion.....	46
5.1. Main Findings	46
5.2. Future Research Direction	47
Bibliography.....	49
Appendix A. Decision Tree based Models and Interpretability.....	53
Appendix B. Derivation of Theil Index.....	56
Appendix C. Derivation of Spatial Filters.....	60
Abstract in Korean.....	64

List of Figures

[Figure 1] Trend in price index of apartments in Gyeonggi Province	1
[Figure 2] Foreign occupancy and ownership by subdivisions of Gyeonggi Province	8
[Figure 3] Trend of foreign occupancy and ownership before the COVID-19 pandemic	9
[Figure 4] Trend of foreign occupancy and ownership after the COVID-19 pandemic	10
[Figure 5] Number of lots in Bucheon owned by foreigners classified by major global regions	14
[Figure 6] Area owned by foreigners in terms of percentage compared to the whole area of Bucheon, classified by major global regions	15
[Figure 7] Foreign population vs. Chinese land in Bucheon.....	17
[Figure 8] Real estate acquisition in Bucheon by foreigners after adjusting for the variability of lots	18
[Figure 9] Population distribution by nationality for each district of Bucheon	20
[Figure 10] Foreigners' real estate acquisition by subdivisions of Bucheon.....	21
[Figure 11] Change in foreign population compared with new real estate acquisition by foreigners during 2020.....	23
[Figure 12] The distribution of 30 most sought after buildings in Bucheon.....	24
[Figure 13] Chinese signs around buildings with active acquisition by foreigners.....	25
[Figure 14] Time series of spatial distribution on real estate acquisition by foreigners in Bucheon	27
[Figure 15] Quarterly time series of real estate acquisition by foreigners in Bucheon	28

[Figure 16] Spatio-temporal distribution of Theil index on real estate acquisition by foreigners	30
[Figure 17] The type of real estate in Bucheon acquired by foreigners	31
[Figure 18] Boxplot on the top floor of buildings in Bucheon that have been acquired by foreigners	32
[Figure 19] The proportion of acquisitions by foreigners with land area larger than 18m ²	33
[Figure 20] Distribution of built year of buildings vs. acquired period of buildings by foreigners	34
[Figure 21] The illustration of the first spatial filter.....	35
[Figure 22] The illustration of the second spatial filter.....	36
[Figure 23] SHAP value for the district to which each building belongs to	38
[Figure 24] SHAP value for the value of the first spatial filter	39
[Figure 25] SHAP value for the value of the second spatial filter	40
[Figure 26] SHAP value for the proportion of officetels among the whole number of houses and offices in the complex	41
[Figure 27] SHAP value for the total number of floors in respective apartments or officetels	42
[Figure 28] SHAP value for the completion year of respective apartments or officetels	43
[Figure 29] SHAP value for the number of buildings in respective complexes of apartments or officetels	44
[Figure 30] The view of apartments and officetels which have been actively acquired by foreigners	45

List of Tables

[Table 1] Foreigners' real estate acquisition by subdivisions of Bucheon.....	22
[Table 2] Change in foreign population compared with new real estate acquisition by foreigners during 2020.....	23
[Table 3] The distribution of root mean squared error for each test set and each model.....	37
[Table 4] The distribution of coefficients of determination for each test set and each model.....	37

Chapter 1. Introduction

1.1. Research Background and Purpose

During the COVID-19 pandemic, while the population of foreigners in the Gyeonggi Province of South Korea decreased, the real estate owned by foreigners increased. The pandemic also coincided with the upsurge of housing prices in Gyeonggi Province (Figure 1), and the province tried to curb the suspected speculation by foreigners. Since November 2020, the province has enforced a policy for the first time in this century that obliged foreigners to get the permission of the local government when acquiring real estate (Kim and Kim, 2020). Before the enforcement of this policy, reporting their acquisition rather than asking permission for it sufficed.

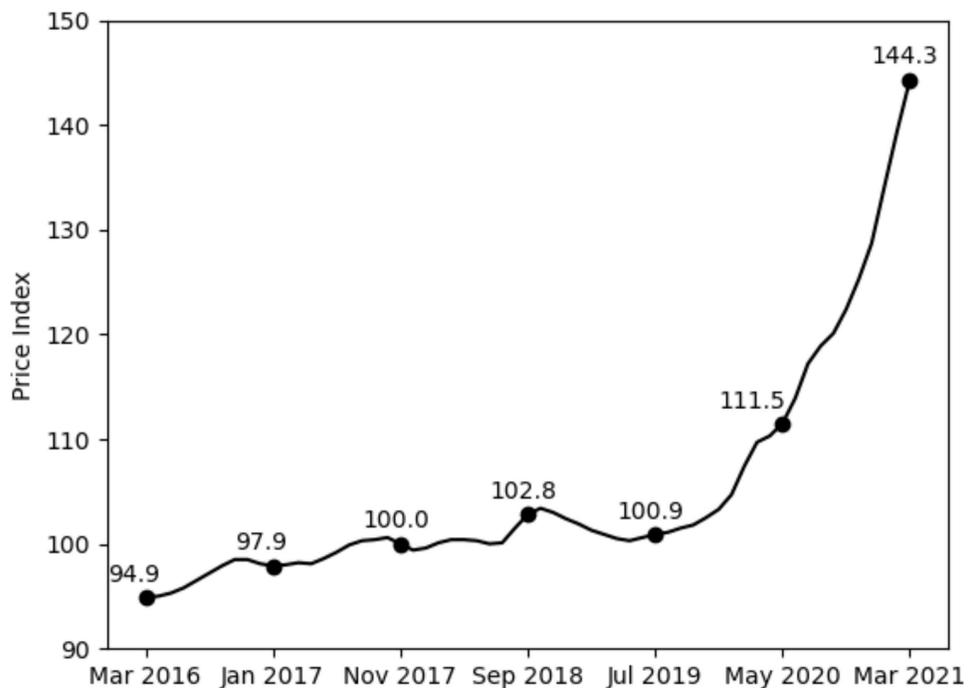


Figure 1 Trend in price index of apartments in Gyeonggi Province (Korea Real Estate Board, 2021). The value in November 2017 is 100.

Such divergence between the spatial occupancy represented as foreign population and the spatial ownership represented as foreign ownership of domestic real estate raises an issue of land grabbing and absentee landlords that make use of ample liquidity during the pandemic (Lee, 2020). On the other hand, there is also caution to such interpretation, blaming the new policy for attributing the failure of restraining house prices to foreigners by taking advantage of xenophobic sentiments during the pandemic (Pyeon, 2020).

In order to make a reasonable judgement on such a controversial issue of real estate acquisition by foreigners, a comprehensive analysis of data including the acquirer, foreign population, spatio-temporal dynamics and characteristics of the property acquired should precede. In this regard, this research demonstrates how the real estate acquisition by foreigners has evolved recently with a case study in Bucheon-Si, Gyeonggi Province.

The study tries to answer the following four questions: First, which country are the foreign owners of real estate in Bucheon coming from? Second, how does the population structure of foreigners relate to the magnitude of real estate acquisition by foreigners geographically? Third, how has the phase of acquisition changed before and after the COVID-19 pandemic? Fourth, how is the real estate acquired by foreigners characterized in terms of locations and property attributes?

1.2. Definition of Terms

In South Korea, the foreign population is classified into two categories: one is registered foreigners, and the other is ethnic Koreans with foreign nationality who have reported to stay in South Korea. As the latter group receives an F-4 visa, it will be called F-4 foreigners in this research.

In this thesis, Chinatown is defined based on the presence of Chinese people and their houses rather than the superficial landscape of Chinese stores and

architecture. Namely, Chinatown is interpreted as a residential ethnic enclave rather than a commercial site with a Chinese theme.

Among various types of residential buildings in South Korea, a conventional distinction between a villa and an apartment is made according to the scale of the building, such as occupying land area and height, in which case the latter is bigger. A villa is different from a single or multi-family house because the former is composed of multiple houses that can be owned separately while the latter is always owned by an individual. In addition, while the legal status of an officetel is an office, they are treated as high-rise residential buildings similar to apartments to the public. In this research, apartments and officetels are classified as a single type of residential building unless otherwise stated while villas are distinguished according to the two main online map services in South Korea: Kakao Map and Naver Map.

1.3. Organization of Chapters

In chapter two, controversy on urban real estate acquisition by foreigners in developed countries are discussed from which the importance of exploratory data analysis can be identified. Domestic research that dealt with the spatial characteristics of foreigners is also reviewed to point out its biased perspective. Also, the evolutionary history of interpretable machine learning is addressed for its application on geography. In chapter three, notes on processing raw data, on the selection of the study area, and on the application of the methodology are provided. Chapter four illustrates the result of analysis in detail, starting from the geographical origin of foreigners acquiring real estate in Bucheon, the relationship between real estate acquired by foreigners and the foreign population, the change in real estate acquisition around the time of the COVID-19 pandemic, and the characteristics of the real estate acquired by foreigners. The study ends in Chapter five with its significance and limitations.

Chapter 2. Literature Review

2.1. Controversy on Urban Land Acquisition by Foreigners

Because the landscape of Bucheon is mainly urban rather than rural, and because South Korea is a developed country, an approach that is different from the traditional view on land grabbing, which has been discussed in the rural context of developing countries, is required. On a theoretical background, building apartments is suggested as three-dimensional grabbing that deprives people on the ground floor of daylight and wide views in spite of its space-saving characteristic (Noorloos, 2019). More practically, increased acquisition of real estate by the Chinese is demonstrated in North America such as in Toronto (Ye, Shan et al., 2017) and in California (Liao, 2019) either through causality analysis or providing direct statistics. In particular, Wong (2017) illustrated new apartment construction in Sydney due to the demand of the Chinese. Such inflow of foreign capital and immigrants has been suggested to deprive local residents of housing affordability as in the case of Vancouver and Singapore (Rogers and Koh, 2017). However, there were also some reflections on the practices of the media that singled out the Chinese as the main culprit for the rising house prices (Wong, 2017), most of whom just aspired to ordinary local life for their children, as in the case of London (Brill and Raco, 2020). In the same vein, a skeptical review on the efficiency of policy that introduced additional tax on foreign purchasers of real estate in Vancouver exists (Huang, 2018).

To sum up, studies on the controversy show the necessity of backing up the arguments with various, well-processed data. Through refining raw data and applying diverse statistics, the exploratory data analysis provided in this research will serve as a case of proper understanding on the real estate acquisition by foreigners. Only based on such a reference could misinterpretation and instigation be avoided.

2.2. Spatial Mode of Foreigners in South Korea

Contrary to the massive amounts of research on foreign occupants in South Korea, research on real estate acquisition by foreigners is relatively scarce. As a case study in Seoul, Kim (2017) revealed that ethnic Koreans from North America and Oceania were the main drivers of real estate acquisition by foreigners from 2006 to 2010. On the one hand, Gangnam and Yeonhui were characterized as hot-spots due to ethnic Koreans from developed countries and ethnic Han of Taiwanese nationality, respectively. On the other hand, current Chinatown in Daerim was characterized by cold-spots in terms of real estate acquisition. However, Chinese investors were indicated as the main foreign acquirers of real estate in Jeju Island, which influenced the housing prices of the island (Han and Park, 2015; Xi, 2016).

As Brill and Raco (2020) pointed out, the focus of domestic studies on real estate acquisition by foreigners either highlighted prime areas, such as Seoul, or some high-class Chinese absentee landlords in Jeju. Knowledge on the real estate acquisition by more typical foreigners in a more typical area remains elusive. This research contributes by dealing with real estate acquisition by foreigners in Gyeonggi Province which is home to the largest foreign as well as domestic population in South Korea.

Former studies on the spatial distribution of the Chinese, who are the dominant group among the foreign population in South Korea, have

characterized them consistently as tenants of old houses that Koreans do not prefer to live in (Bhang, 2012; Kim, 2012; Lee, 2014; Kim and Moon 2018; Paek, 2018). Although such a perspective on foreigners in South Korea might have been true at the time of publication, the recent upsurge in house prices that coincided with the COVID-19 pandemic implies the need for an update such as the transition of foreigners to homeownership. Indeed, the COVID-19 pandemic has had a varying effect in terms of both real estate acquisition by foreigners and the foreign population across different regions (e.g., Tanrivermis, 2020; Huh and Jang, 2020). This research can play a role as a pioneer that suggests a new perspective on foreigners in South Korea.

2.3. Interpretable Machine Learning and Geography

The main drawback of applying machine learning models on explanation has been their poor interpretability. The model is more accurate than ordinary linear regression, but humans could not understand how much nor in which direction each explanatory variable contributed to the result. SHapley Additive eXplanations (SHAP) value was suggested as a solution to the problem (Lundberg et al., 2018). The value is the contribution of a selected single explanatory variable of a selected instance on the target variable of the instance. SHAP is getting popular in spatial studies (Oh et al., 2020; Chen et al. 2021) and it has also been introduced to geography, but with a caveat on the spatially related instances (Kema, 2020). The study can serve as a soft landing of interpretable machine learning on the field of geography by addressing the issue of spatial autocorrelation through adopting spatial filters (Dray et al., 2006) as explanatory variables for machine learning models.

3. Research Methodology

3.1. Study Area Selection

Bucheon was selected as a case study plot. Its location is special in that to the east is Seoul, the capital of South Korea where nearly 10 million people live, and Incheon to the west, a city with a population of nearly 3 million, and the venue of the main international airport and trading port. The Gyeongin railroad that connects Seoul and Incheon has passed through Bucheon since 1899. Besides the role as a bridge between the two megacities, Bucheon is connected by Ansan in the South, the most ethnically diverse city in the country. Ansan and Bucheon have recently been connected with Seohae railroad since June 2018. Bucheon itself is also populated by 0.8 million people with a population density that is higher than New York City's.

The reason for selecting Bucheon as the study site is simple; according to the provincial statistics (Figure 2) of real estate acquired by foreigners in other cities and counties of Gyeonggi Province, it is the place that stands out the most.

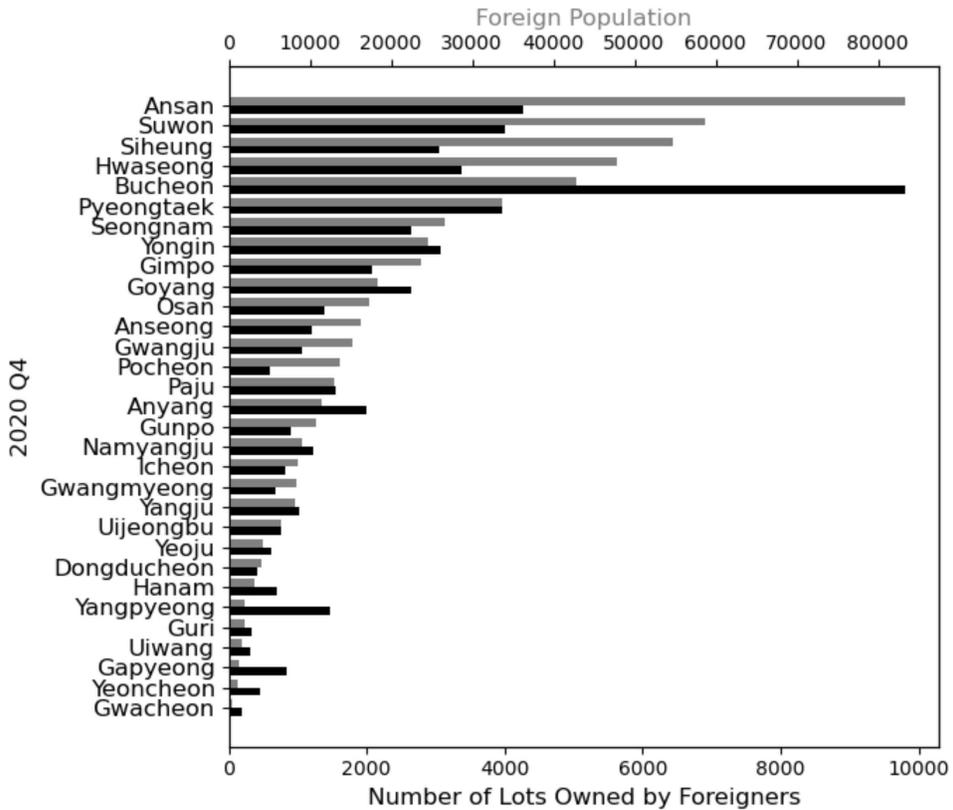


Figure 2 Foreign occupancy (Ministry of Justice Korea Immigration Service, 2021) and ownership (Gyeonggi Province, 2021) by subdivisions of Gyeonggi Province in the fourth quarter of 2020. Cities and counties of Gyeonggi Province are sorted by foreign population (grey bars) led by Ansan (83,410 foreigners). The number of lots owned by foreigners (black bars) is outstanding in Bucheon (9,807 lots owned by foreigners).

With respect to the dynamics of real estate acquisition by foreigners, Bucheon had been leading the increase of foreigners' real estate acquisition before the COVID-19 pandemic (a net increase of 3,570 lots), the value of which is very high even when compared to the increase of the foreign population (10,793 people, Figure 3).

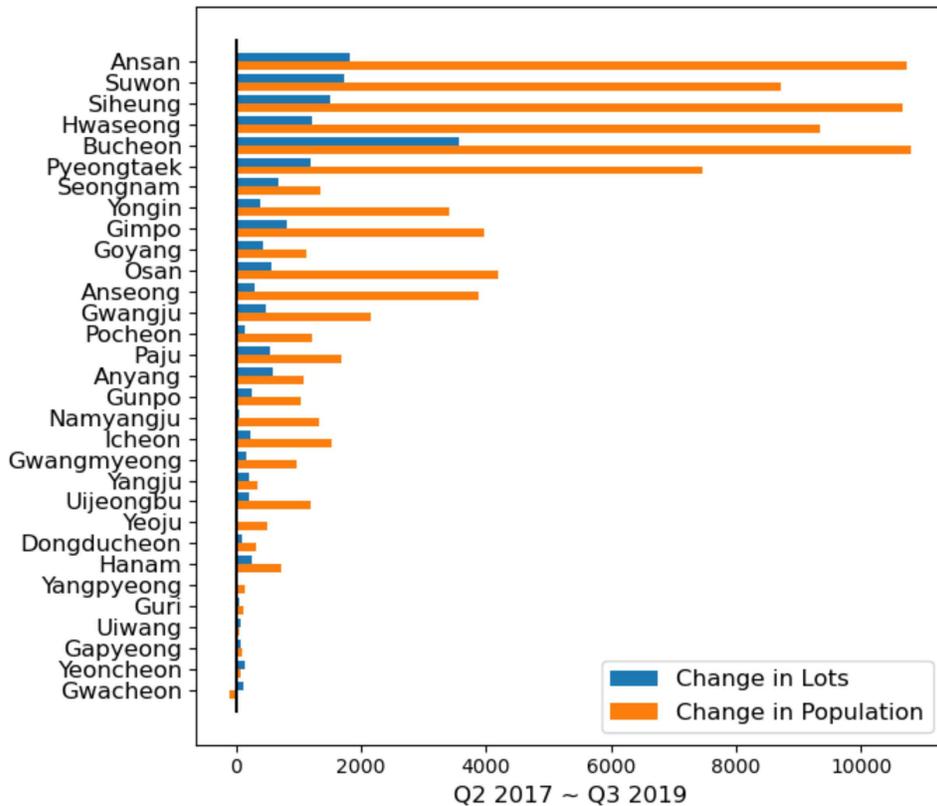


Figure 3 Trend of foreign occupancy and ownership *before* the COVID-19 pandemic. Change in foreign population (orange bars) and change in lots owned by foreigners (blue bars) for cities and counties of Gyeonggi Province from the second quarter of 2017 until the third quarter of 2019.

More surprisingly, the rate of real estate acquisition by foreigners even accelerated in Bucheon after the outbreak of COVID-19 (a net increase of 4,396 lots for 15 months) while real estate owned by foreigners also increased in most of the other cities (Figure 4). In contrast, most of the cities and counties suffered from a decrease in foreign population with a net increase of only 924 foreigners in Bucheon.

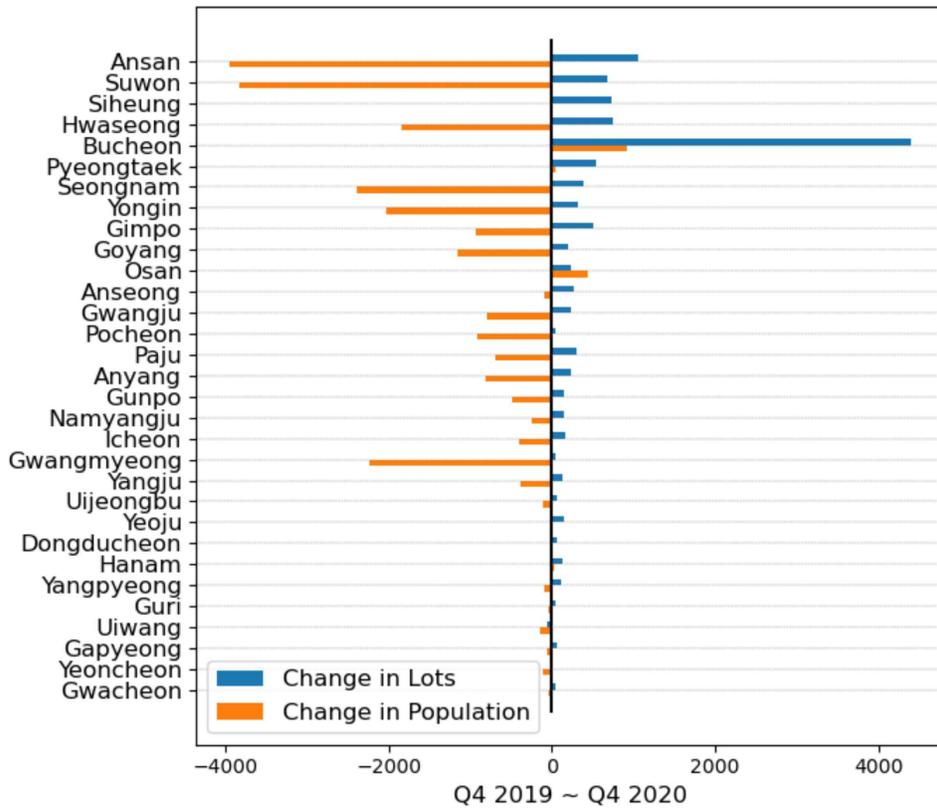


Figure 4 Trend of foreign occupancy and ownership *after* the COVID-19 pandemic. Change in foreign population (orange bars) and change in lots owned by foreigners (blue bars) for cities and counties of Gyeonggi Province from the fourth quarter of 2019 until the end of 2020.

3.2. Data Collection and Processing

The statistics of real estate acquisition by foreigners was acquired in a beforehand information disclosure service from both Gyeonggi Province (Gyeonggi Province, 2021) and Bucheon City (Bucheon City, 2021a). The attribute data of acquired real estate was retrieved from land registers and building registers of the government. The statistics of the foreign population by city and county was obtained from the Ministry of Justice. The foreign population of each district in Bucheon was retrieved from the official statistics website of Bucheon City Government. Land cover data of the city is retrieved

from Environmental Spatial Information Service. To check the relationship between spatial occupancy and spatial ownership by foreigners, web crawling was conducted on April 12, 2021 in ICNKR, the biggest online Chinese community in Korea. In the same vein, a field trip to Bucheon was conducted in May 2021.

It should be noted that the statistics from figure 2 to figure 4 are exaggerated as the unit of counting, *lot* (筆地), is highly variable, which needs further revision. For instance, Jeongdaun Greenville Apartment in Bucheon is composed of 25 lots. Therefore, if a foreigner acquires a house in the apartment, 25 lots are registered as acquired by that foreigner. As a result, the raw acquisition number of each real estate was divided by the number of lots the building actually occupies in such cases. The processed data was used from Section 4.2 to the end of the research result. The period for the processed data is four years, from the second quarter of 2017 until the first quarter of 2021. Although there was limited information on real estate disposal in Bucheon by foreigners, the number of disposals was overwhelmed by the number of acquisitions according to the provincial statistics, thereby ignored in the processing.

3.3. Theil Index

To get the information on the geographical concentration of real estate acquisition by foreigners in different scales, Theil index that has been used as a measure of inequality was applied (for details, please refer to Appendix B). Inequality is interpreted as a concentration of values in certain individuals and groups. In this regard, each individual represents each building acquired by foreigners, while values represent the number of transactions by foreigners in each building. Finally, each group represents a district to which buildings belong.

Theil index on real estate acquisition is higher when the acquisition is concentrated on certain buildings. It gets lower when either the acquisition is evenly distributed or when the absolute number of buildings acquired is low. Global Theil index is decomposed into a weighted sum of local Theil indices calculated for each district and a between-district Theil index. A higher between-district Theil index means that the number of acquisitions is concentrated within certain districts.

3.4. Decision Tree based Machine Learning

To describe the characteristics of the apartments and officetels that foreigners tend to acquire, interpretable machine learning was conducted to secure both accuracy and interpretability (for details, please refer to Appendix A). The target variable is the ratio of acquisitions by foreigners compared to the total number of houses and offices per apartment or officetel complex. The instances were 438 complexes in Bucheon with at least one acquisition by foreigners. In order to retrieve the spatial structure of apartments and officetels acquired by foreigners, spatial filters were created using the locations of the buildings (for details, please refer to Appendix C).

Explanatory variables, or features, are the two best spatial filters that describe the target variable, the district to which the buildings belong, the number of buildings per complex, the total number of floors per building, the completion year of buildings and the proportion of officetels compared to the total number of houses and offices in a complex. Unlike the first and second best spatial filters (eigenvector number 2 and 3 respectively), the third best spatial filter (eigenvector number 429) could represent only a few instances, thereby dropped off to prevent overfitting. The total number of houses per complex was highly correlated (correlation coefficient higher than 0.9) with the number of buildings per complex, so the former was removed.

Random forest and gradient tree boosting were trained five times respectively. The training sets were 80% of the total instances while test sets consisting of the remaining 20% of the total instances are mutually exclusive. Therefore, uniting the test sets of the five trained models led to the total dataset. Training sets and testing sets were split in a stratified fashion where the stratum was the target variable. To get the optimal hyperparameters for each model, each training set was split into five again and used four splits for calibration with the last split spared for validation. The split for cross validation was also stratified as the initial train-test split. Every split was used once for validation. Hyperparameters used exclusively for random forest were the proportion of explanatory variables used for each branching. On the other hand, those used exclusively for gradient tree boosting were shrinkage, the number of instances used for each sequence of trees, and the maximum depth of each tree. Hyperparameters common for both random forest and gradient boosting were the total number of trees and the minimum number of instances per group to branch further.

The performance was compared with ordinary linear regression to check the improvement. Furthermore, stacking the candidates from random forest and gradient boosting with different combinations of hyperparameters was conducted with lasso regularization to check for the further improvement. Finally, SHAP value was calculated for each instance in test sets using the trained random forest and gradient boosting. The R packages used for machine learning were tidymodels (Kuhn and Wickham, 2020) and DALEX (Biecek, 2018).

4. Research Result

4.1. The Origin of Foreigners Buying Real Estate

Figure 5 shows time series on the number of lots owned by foreigners in Bucheon. Note that the y-axis is in log scale. While Chinese lots soared from 426 at the end of the third quarter of 2015 to 9,338 lots at the end of 2020, lots owned by other foreigners did not show any meaningful increase. Although using lots as the unit of statistics can be misleading, on the one hand it can still be stated that almost all of the increase in real estate owned by foreigners is due to the Chinese, at least in terms of lots.

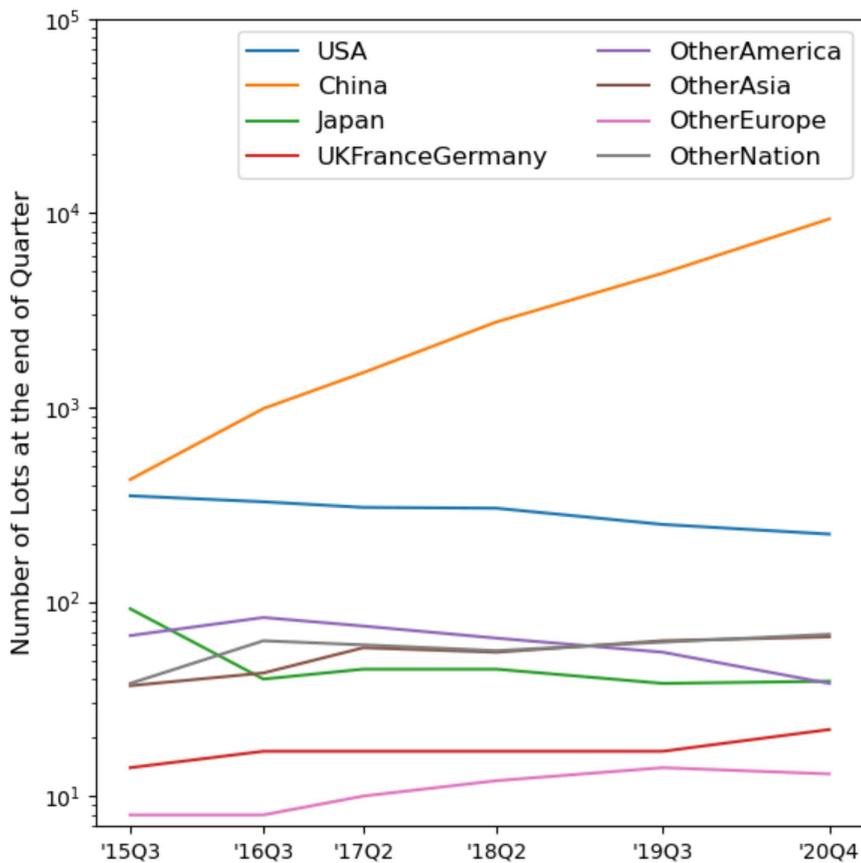


Figure 5 Number of lots in Bucheon owned by foreigners classified by major global regions (Gyeonggi Province, 2021).

On the other hand, in terms of land area, the proportion of Chinese in the total area owned by foreigners is smaller than that of Other Asians (Figure 6). However, as can be seen from the anomalous increase of the area owned by Other Asians from 2016 to 2017 (from less than 0.01% to 0.26% of the total city area), most of the Other Asian land is the newly built Bucheon Logistics Park and Bucheon Cold Chain Logistics Park which are owned by foreign companies. Likewise, a considerable portion of USA and Japanese land is due to the foreign factories in Bucheon, such as ON Semiconductor and Sinhanil Electronics. What can be inferred from here is that measuring the land area owned by foreigners has another limit. If a foreigner buys an apartment, the land area increases by only a tiny amount: that is, to put it simply, the area where the apartment building stands divided by the number of apartments in that building.

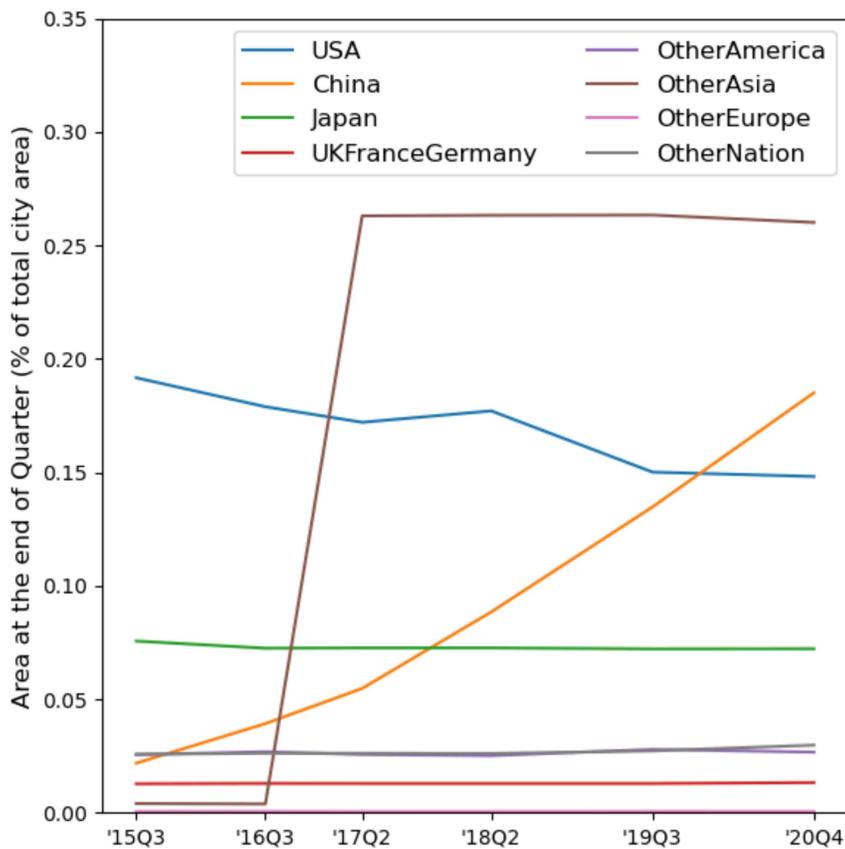


Figure 6 Area owned by foreigners in terms of percentage compared to the whole area of Bucheon, classified by major global regions.

However, the land owned by Other Asians is somewhat vague. For example, Bucheon Logistics Park was developed and owned by Bucheon Logistics Center PFV, Inc., identified as a Singaporean company (Ministry of Trade, Industry and Energy, 2015). The head of the company is Shen Jinchu (沈晉初; ESR REIT, 2019). Also, Bucheon Cold Chain Logistics Park is owned by Kendall Square LP Bucheon Cold Professional Investment that is identified as a Singaporean company as well (National Spatial Data Information Portal, 2021). However, ESR Kendall Square Co., Ltd. from Hong Kong is buying Bucheon Cold Chain Logistics Park (ESR Kendall Square, 2020). The parent company of the Hong Kong company is ESR Cayman Ltd.. The director of ESR Cayman Ltd. is also Shen Jinchu (ESR Cayman, 2020). Therefore, Bucheon Logistics Park and Bucheon Cold Chain Logistics Park can be regarded as owned by one real estate company. The main tenant of the two logistics centers is Coupang, a major e-commerce company in Korea.

The consistent increase of Chinese land since the fourth quarter of 2015 is unprecedented (from 0.02% at the end of Q3 2015 to 0.19% of the total city area at the end of 2020). The upsurging acquisition of Korean land by Chinese was proportional to relatively gradual increase in foreign population (from 31,075 at the end of Q1 2017 to 42,792 at the end of 2020), but the former was much steeper (Figure 7); Chinese area increased from 26,096 m² at the end of Q1 2017 to 98,885 m² at the end of 2020. Figure 7 shows that F-4 foreigners increased during COVID-19 pandemic from 16,452 to 18,387 people, while registered foreigners decreased from 25,416 to 24,405 people. It implies that ethnic Koreans with foreign nationality but staying in South Korea who are dominantly Chinese, may have maintained the boosting trend in land ownership by Chinese during COVID-19 pandemic.

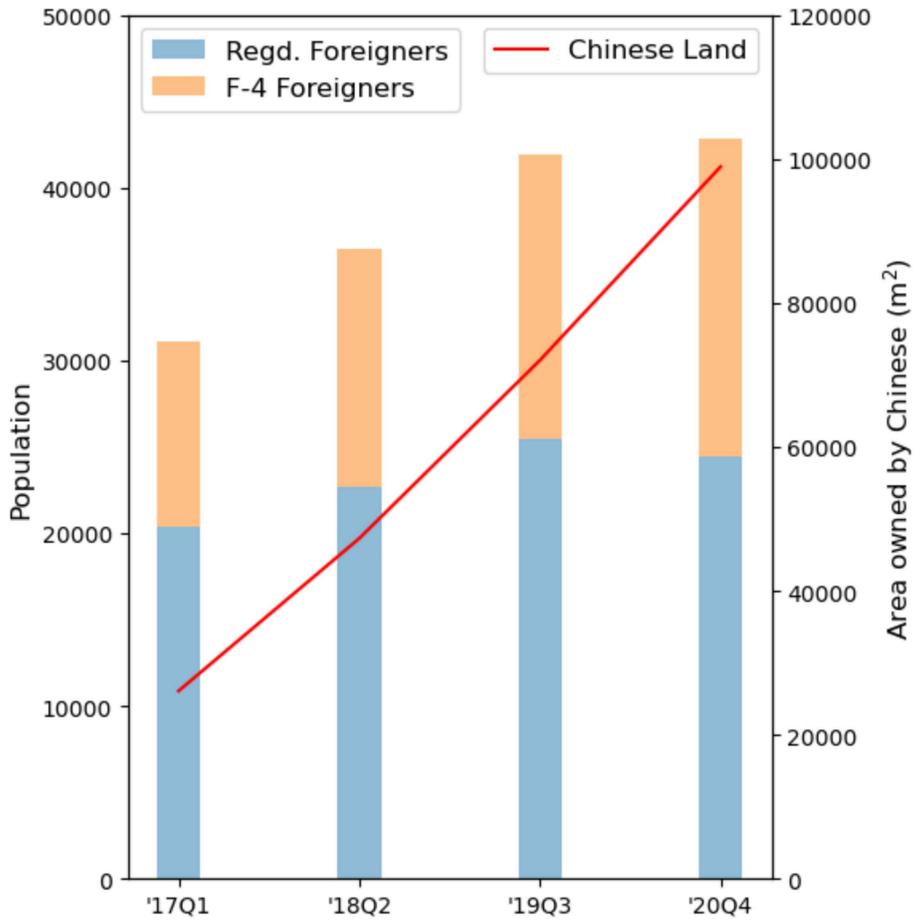


Figure 7 Foreign population vs. Chinese land in Bucheon. Comparison between the trend of foreign population and the trend of Chinese land acquisition in Bucheon.

As using lots for the unit of real estate acquisition by foreigners overestimates the real practice and using land area as the unit for real estate acquisition underestimates it, processed data is introduced in the following section.

4.2. Comparison with Foreign Population

The processed spatial information of real estate acquisition by foreigners was plotted in figure 8. The unit is neither lot nor land area because both units showed limits as mentioned above. The unit is adjusted as the proper number of real estate acquisitions by foreigners as described in Section 3.2. Most of the real estate acquired by foreigners was concentrated in urban dwelling areas rather than rural or industrial areas.

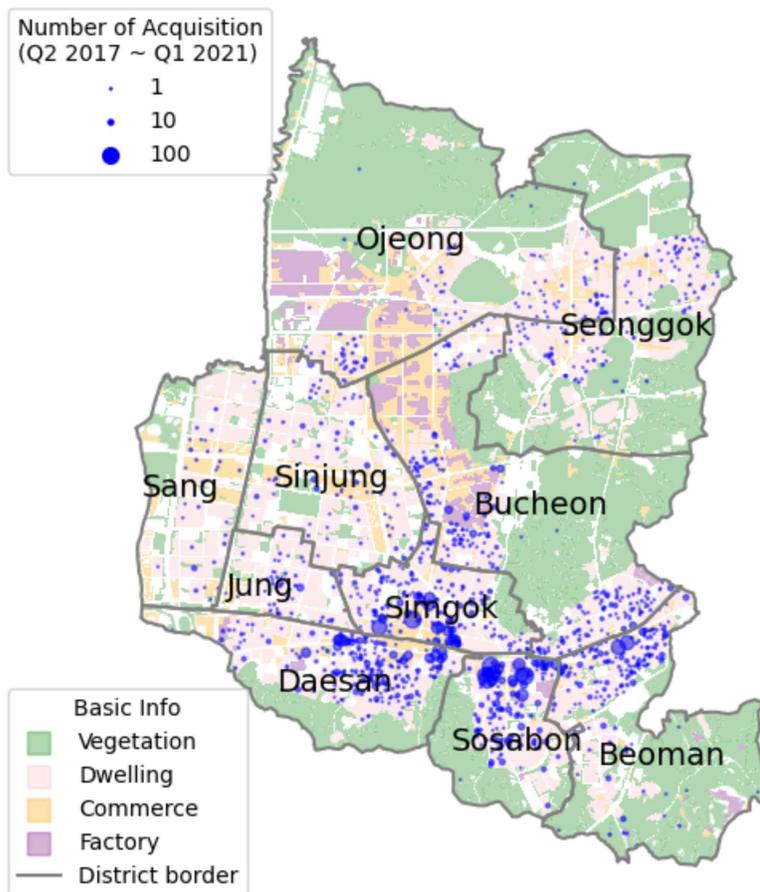


Figure 8 Real estate acquisition in Bucheon by foreigners after adjusting for the variability of lots.

Bucheon-si (*City*) is composed of ten districts (廣域洞), including Bucheon *District*. The northern borders of the three southern districts, Daesan, Sosabon, and Beoman, were characterized by the Gyeongin Line which is the oldest Korean railroad connecting Seoul and Incheon. The main place of real estate acquisition by foreigners was along the railroad.

Figure 9 describes the population distribution of Bucheon by nationality. Foreigners in Bucheon were dominantly Chinese, and they were especially concentrated along the Gyeongin Line except for Sang. To summarize the foreigners who came from discernible rich countries in the population statistics, American, Japanese, Taiwanese, F-4 Americans, F-4 Canadians, F-4 Australians and F-4 New Zealanders were grouped as Democratic. There were relatively many foreigners from the Democratic group in Sinjung and Sang, which are planned districts with big apartment complexes; among the 177 apartment complexes in Bucheon that passed some legal thresholds with regard to the number of households, 83 were in the two districts (Bucheon City, 2021c). Near the industrial area of Ojeong and Bucheon district, there were relatively many foreigners from the Other Asia group that included Vietnamese, Philipino, Myanmarese, Mongolian, Thai and Cambodian. Especially, Ojeong was the only district where the Chinese could not reach half of the total foreign population.

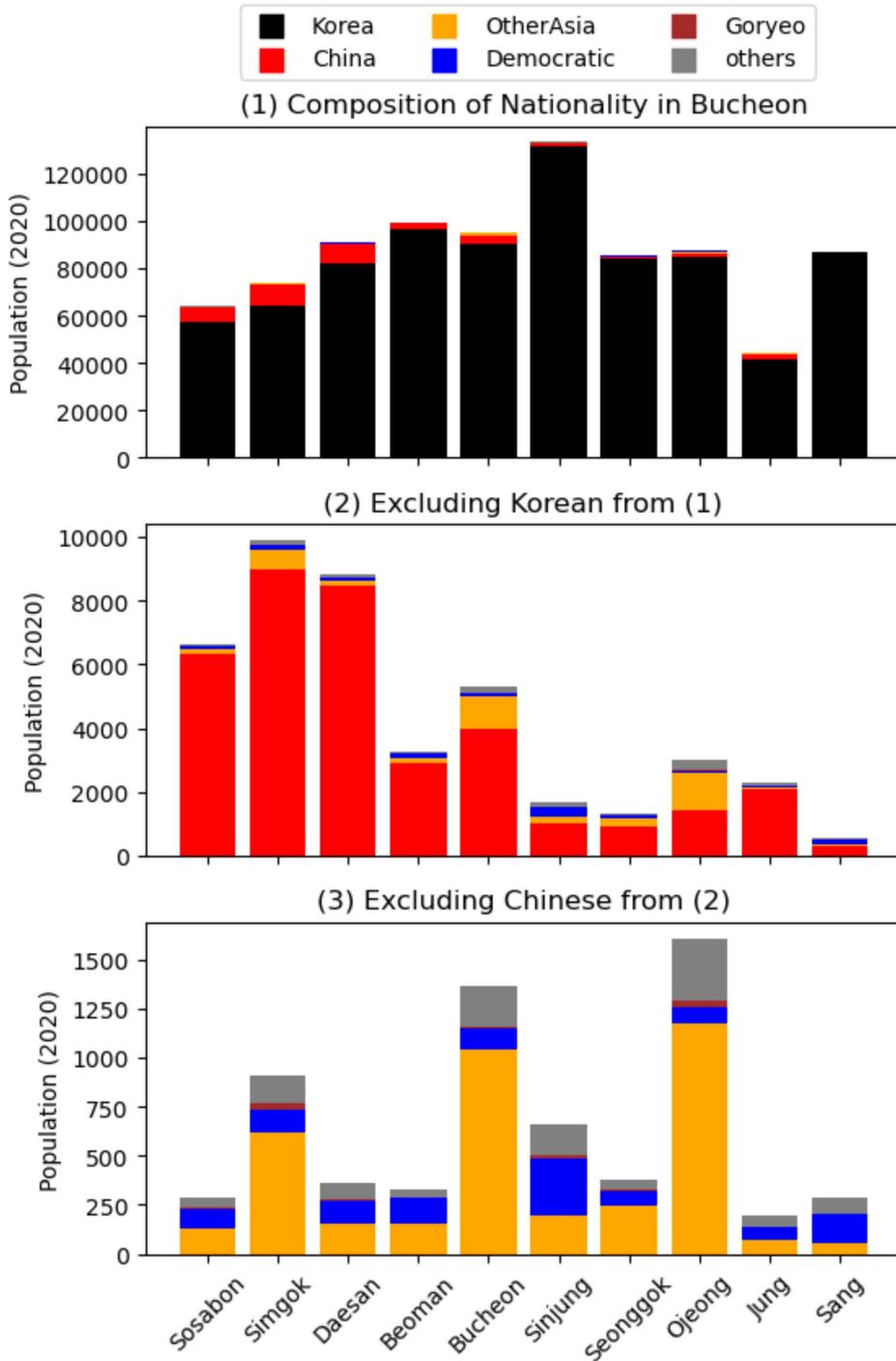


Figure 9 Population distribution by nationality for each district of Bucheon at the end of 2020. F-4 Uzbeks and F-4 Russians were categorized as Goryeo.

Based on the population data, real estate acquisition per foreign capita was calculated (Figure 10). The result was striking because not only the absolute amount of acquisition but also acquisition per capita was diverse among districts. First of all, the acquisition of real estate by foreigners was very active in Sosabon. In Sinjung and Sang, the number of acquisitions was not small relative to their small foreign population. On the contrary, real estate acquisition per foreign capita was very low in Ojeong.

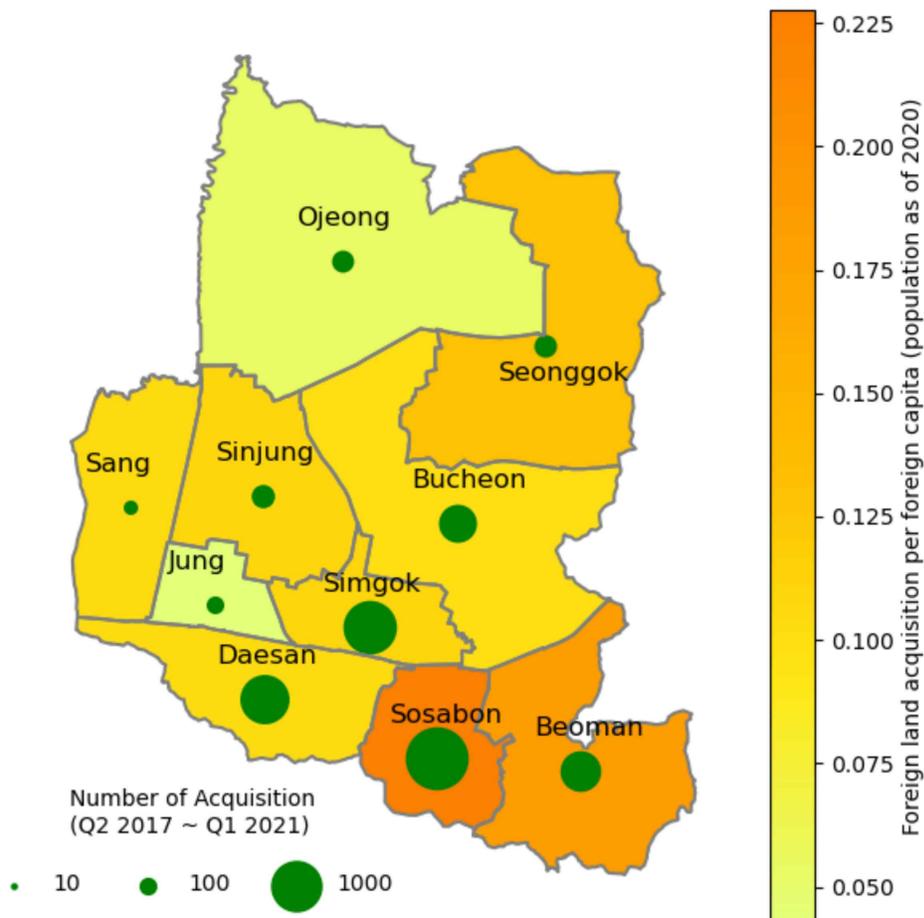


Figure 10 Foreigners' real estate acquisition by subdivisions of Bucheon: real estate acquisition by foreigners in absolute terms (circle size) and real estate acquisition normalized by foreign population (polygon color).

Table 1 Foreigners' real estate acquisition by subdivisions of Bucheon. The processed number of acquisitions are cumulated from the second quarter of 2017 to the first quarter of 2021 (Bucheon City, 2021a). The foreign population is represented by the statistics at the end of 2020 (Bucheon City, 2021b).

District	Number of Acquisitions	Acquisitions Per Foreigner
Sosabon	1509	0.23
Simgok	1075	0.11
Daesan	903	0.10
Beoman	606	0.19
Bucheon	525	0.10
Sinjung	181	0.11
Seonggok	171	0.13
Ojeong	161	0.05
Jung	98	0.04
Sang	60	0.10

Figure 11 compares new real estate acquisition by foreigners in 2020 and the population change of foreigners during 2020. Based on extremely high correlation (~ 0.98) between F-4 Chinese and new real estate acquisition by foreigners, F-4 Chinese are implied as the main group that acquired real estate during COVID-19 pandemic, while the population of other foreigners including Chinese without F-4 visa decreased in most districts.

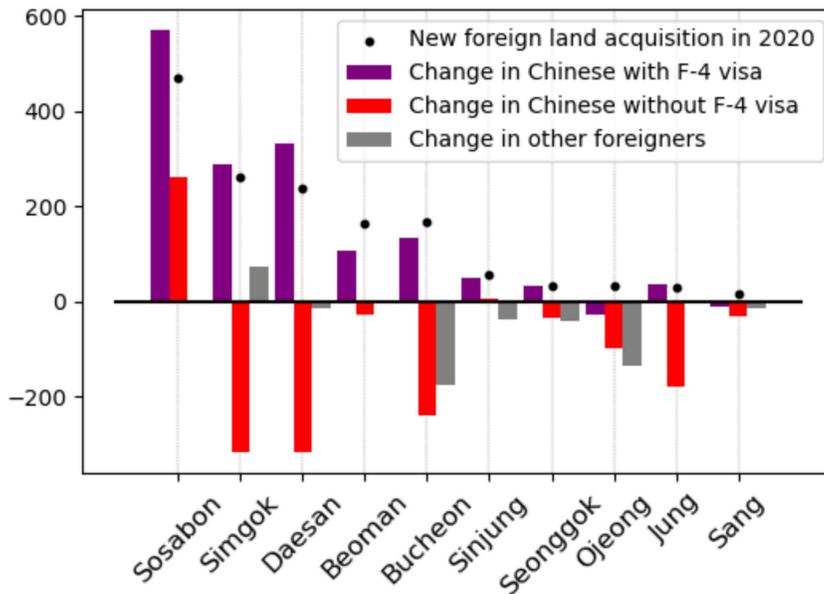


Figure 11 Change in foreign population compared with new real estate acquisition by foreigners during 2020. The districts are sorted by the total number of acquisitions during the whole study period (since Q1 2017 until Q1 2021).

Table 2 Change in foreign population compared with new real estate acquisition by foreigners during 2020. The unit of new acquisitions is the processed number of acquisitions according to Section 3.2.

District	New Acquisitions	Change in F-4 Chinese	Change in Non F-4 Chinese	Change in Other Foreigners
Sosabon	471	570	262	-4
Simgok	263	289	-317	72
Daesan	238	332	-315	-13
Beoman	164	108	-27	-5
Bucheon	168	133	-238	-175
Sinjung	57	51	6	-36
Seonggok	32	32	-35	-40
Ojeong	34	-28	-98	-134
Jung	29	37	-178	-5
Sang	17	-10	-31	-15

Figure 12 shows the distribution of the 30 most sought after buildings in Bucheon by foreigners. All of them are officetels or apartments, but composed of one to three buildings rather than a big complex, and all of them are near the Gyeongin Line, especially Sosa and Bucheon station.



Figure 12 The distribution of 30 most sought after buildings in Bucheon. The yellow lines indicate road and railroad.

The presence of many Chinese is implied in the landscape around the 30 most sought after buildings by foreigners. Some Chinese signs are noticeable around the buildings (Figure 13). Also, at least 316 out of the 399 real estates for sale in Gyeonggi Province were in Bucheon according to the result of web crawling on ICNKR, the biggest Chinese community for those staying in Korea. It implies the convergence of spatial ownership and spatial occupancy rather than speculation by the Chinese.



Figure 13 Chinese signs around buildings with active acquisition by foreigners.

4.3. Dynamics of Real Estate Acquisition by Foreigners

As the real estate acquisition by foreigners increased recently, it could have been hypothesized that the spatial extent of real estate acquisition expanded with time. However, the spatial extent did not show significant change during the study period at a glance (Figure 14). In contrast, big circles formed by massive acquisition in compact buildings such as apartments are getting more prominent with time.

Quarterly time series by districts of Bucheon is going up and down (Figure 15). However, it is remarkable to find a sudden peak in the third quarter of 2018. This quarter is right after the completion of the railroad connecting Ansan and Bucheon, which is called Seohae Line. Seohae Line ends at Sosa Station as of 2020, which is located on the border between Sosabon and Simgok District and is also the station of Gyeongin Line. As seen from figure 15, the new railroad might have affected the sudden increase in real estate acquisition by foreigners.

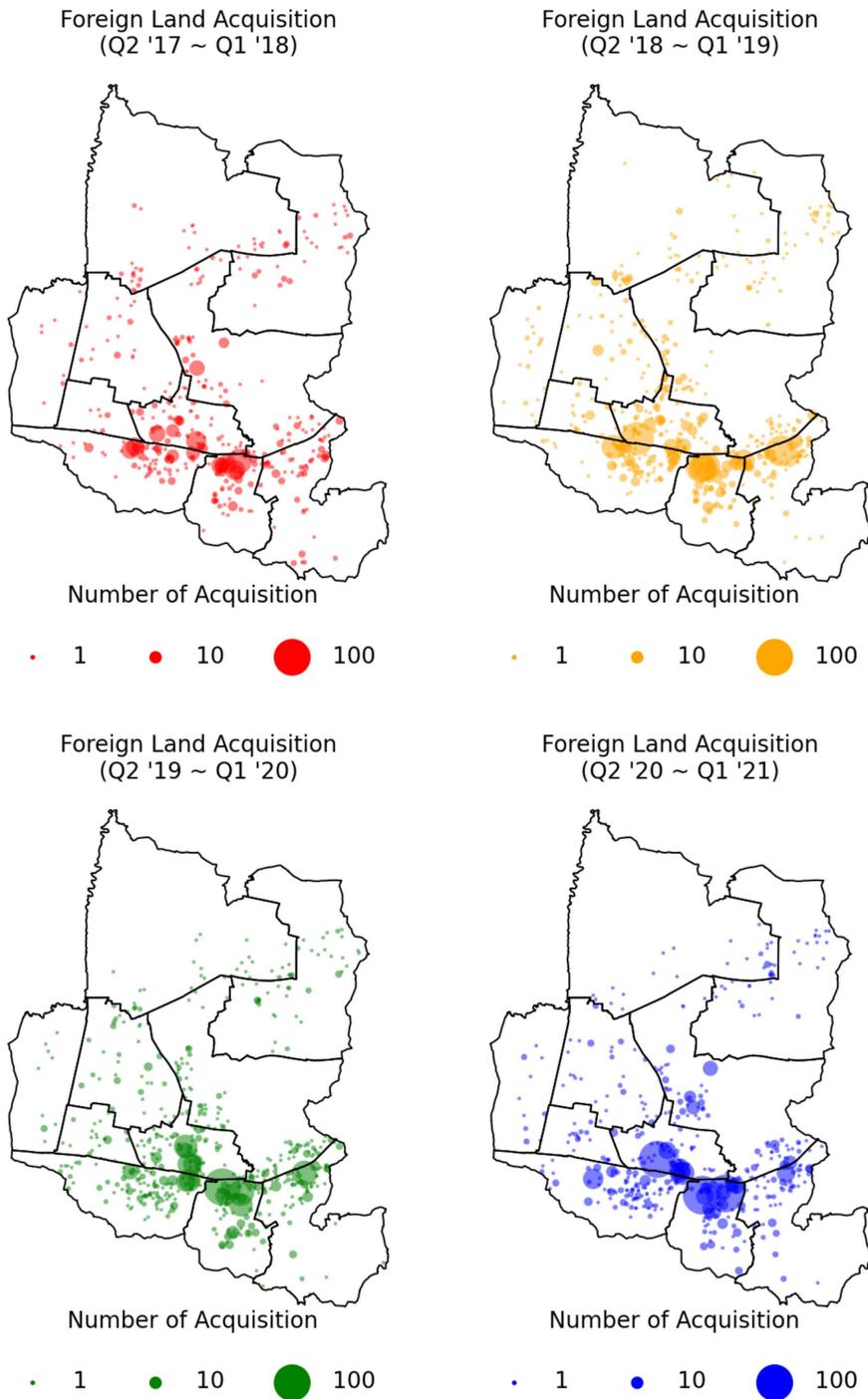


Figure 14 Time series of spatial distribution on real estate acquisition by foreigners in Bucheon. Period 1 (red, from Q2 2017 to Q1 2018), Period 2 (orange, one year after Period 1), Period 3 (green, one year after Period 2), Period 4 (blue, from Q2 2020 to Q1 2021).

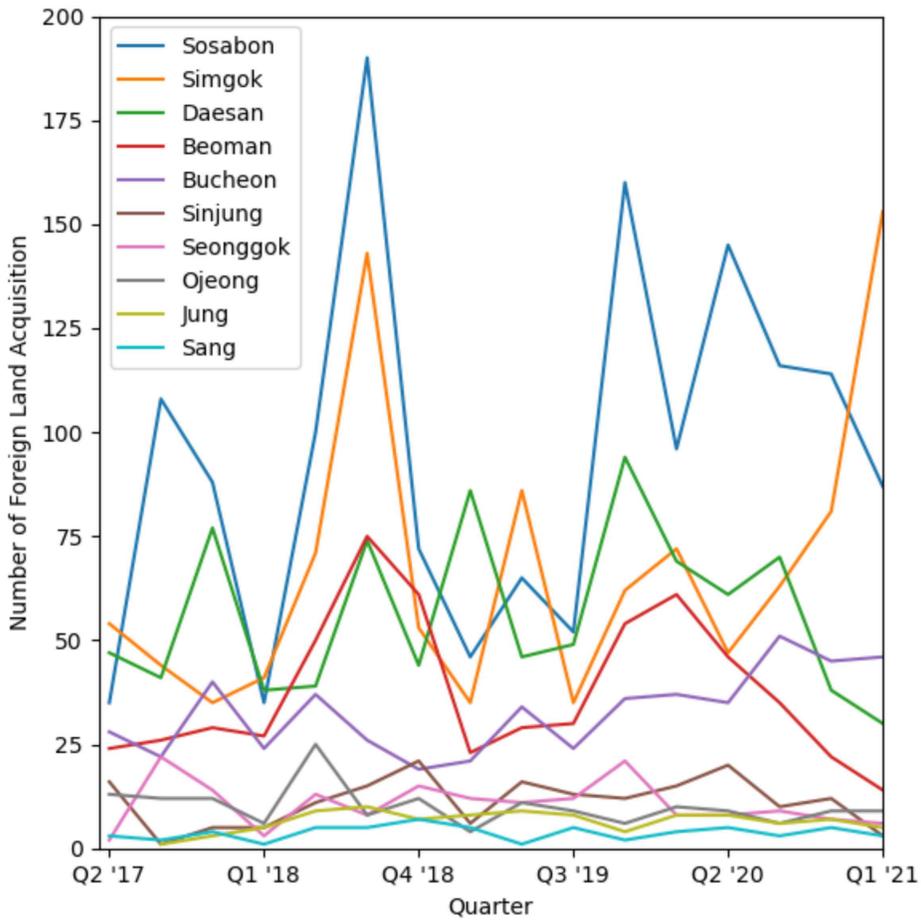


Figure 15 Quarterly time series of real estate acquisition by foreigners in Bucheon.

Figure 16 shows the result of Theil index analysis. To get a more detailed phase of real estate acquisition by foreigners, the districts are plotted by jurisdictional districts (法定洞) instead of administrative districts used in other figures. The global Theil index increased throughout the study period. However, the dynamics was different from period to period.

From Period 1 to Period 2, there was a dramatic increase in global Theil index that was mainly attributable to the increase in local Theil indices. This means there was an overall increase in the number of buildings acquired by foreigners throughout the districts with a concentration on certain buildings, such as officetels in each district, as can be seen from the warmer colors in Period 2.

Global Theil index decreased a little bit from Period 2 to Period 3, and the decrease was attributable to the decrease in the between-district Theil index. However, there is a caveat in interpreting this statistic; compared to Period 2, the divergence of local Theil indices between the northern and southern districts became more prominent but this spatial divergence was not taken into account in the calculation of the Theil index.

On the contrary, the increase of the Theil index from Period 3 to Period 4 is different in that it was mainly attributable to the upsurging between-district Theil index. During the COVID-19 pandemic, there was a divergence in real estate acquisition by foreigners among districts. In other words, real estate acquisition became more concentrated in the historical hot spots such as Simgok and Sosabon while it decreased in northern districts. It is also remarkable to see the increase of the local Theil index in eastern districts, which is expected to benefit from the northward extension of Seohae Line from Sosa Station in 2022.

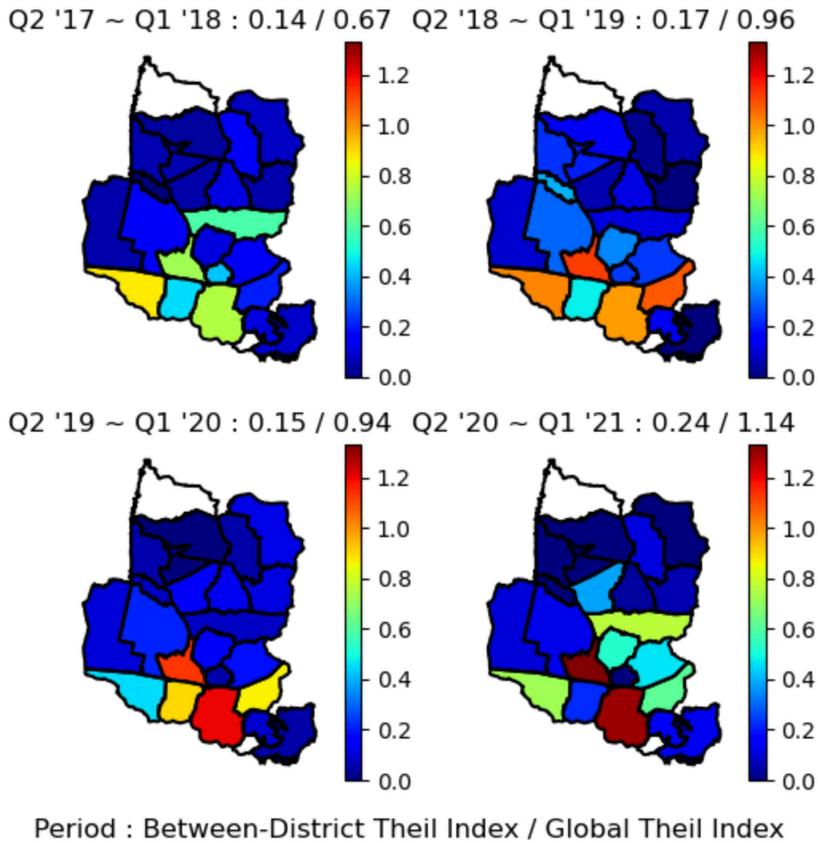


Figure 16 Spatio-temporal distribution of Theil index on real estate acquisition by foreigners. While the colors of each district indicate the local Theil index, white districts are not accounted for as there were few acquisitions there. The subtitles of each map indicate the period, between-district Theil index and global Theil index.

Most real estate acquired by foreigners in Bucheon was houses rather than factories or agricultural lands. Apartments and *officetels* became more popular among foreigners especially after the COVID-19 pandemic (Figure 17). As a result, the median of the top floor of the buildings acquired by foreigners is increasing (Figure 18). It should also be noted that after the enforcement of policy that obliged foreigners to get permission for acquiring real estate from

November 2020, the total number of acquisitions did not decrease significantly. After the enforcement of the policy, however, the acquisition of *villas*, a small building composed of multiple houses, was the lowest in 4 years.

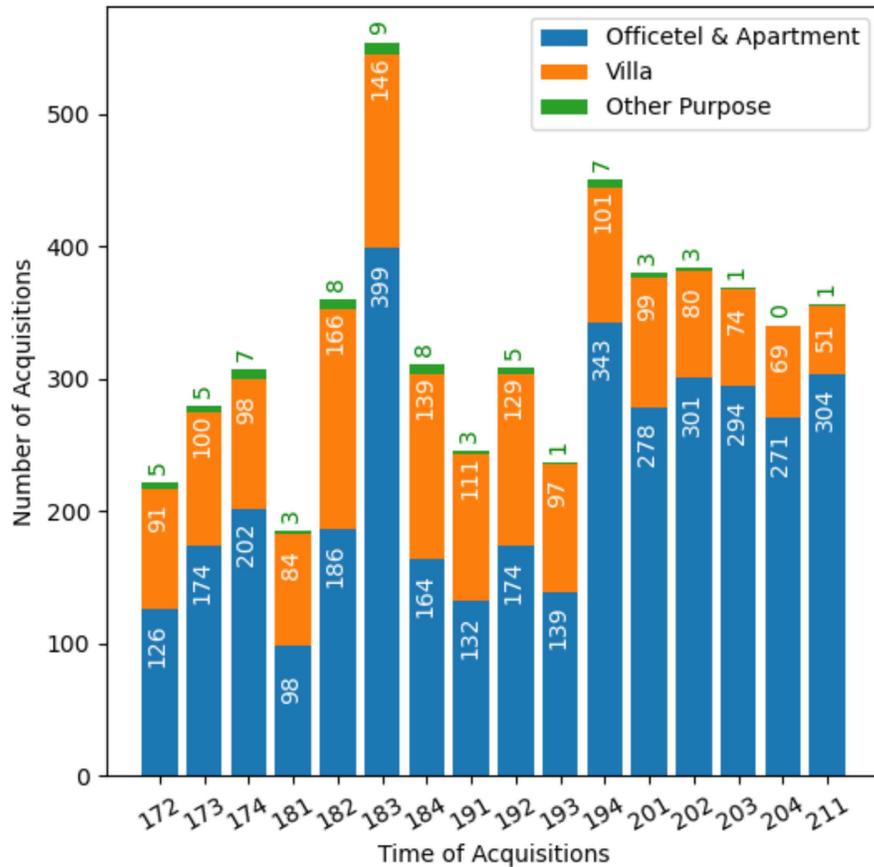


Figure 17 The type of real estate in Bucheon acquired by foreigners. For each label of x-axis, the first two numbers represent the year and the last number represents the quarter.

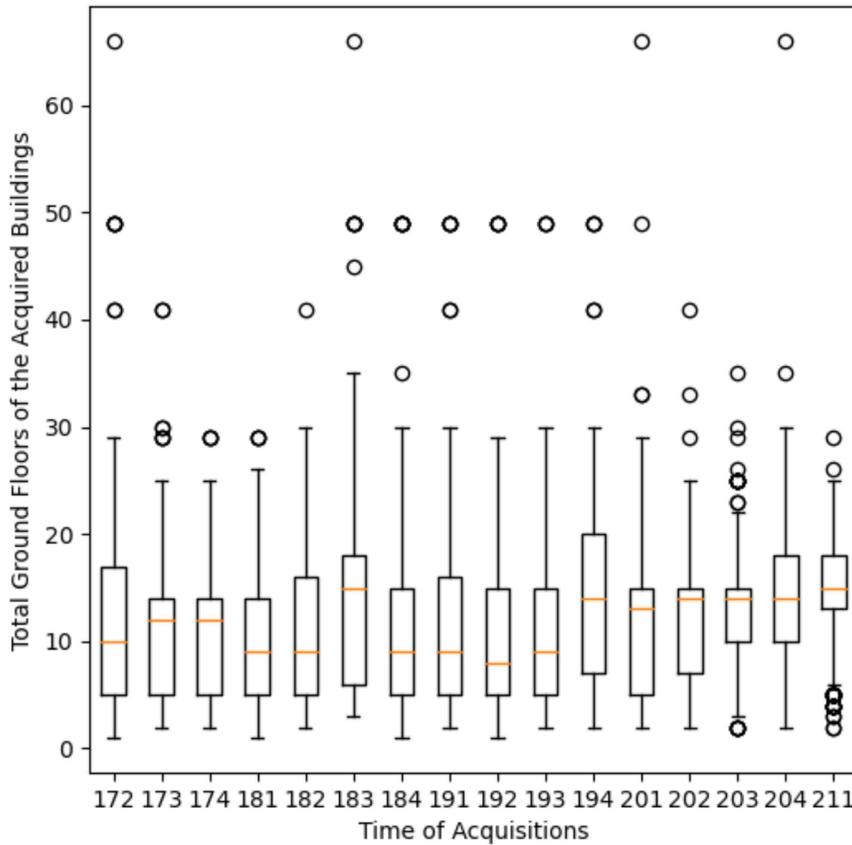


Figure 18 Boxplot on the top floor of buildings in Bucheon that have been acquired by foreigners. For each label of x-axis, the first two numbers represent the year and the last number represents the quarter.

The province announced that the number of permits to the acquisition of real estate by foreigners is much smaller than the whole number of acquisitions by foreigners before the enforcement of the new policy (Lee, 2021). It boasted that the policy effectively suppressed real estate acquisition by foreigners. However, there is a blind spot in the policy. When acquiring less than the threshold area, 18 m² for dwelling and 20 m² for commerce, there is no need to get permission (Gyeonggi Province, 2020). In the case of Bucheon during the study period, more than half of the acquisitions in Bucheon were below 18 m² (Figure 19). The ratio of acquisitions that are *not* subject to the

regulation even increased after the policy. This does not necessarily mean that foreigners are buying small houses because the *reported* area of acquisition is in terms of land area; the *actual* area of acquisition is much bigger in the case of high buildings such as apartments and officetels.

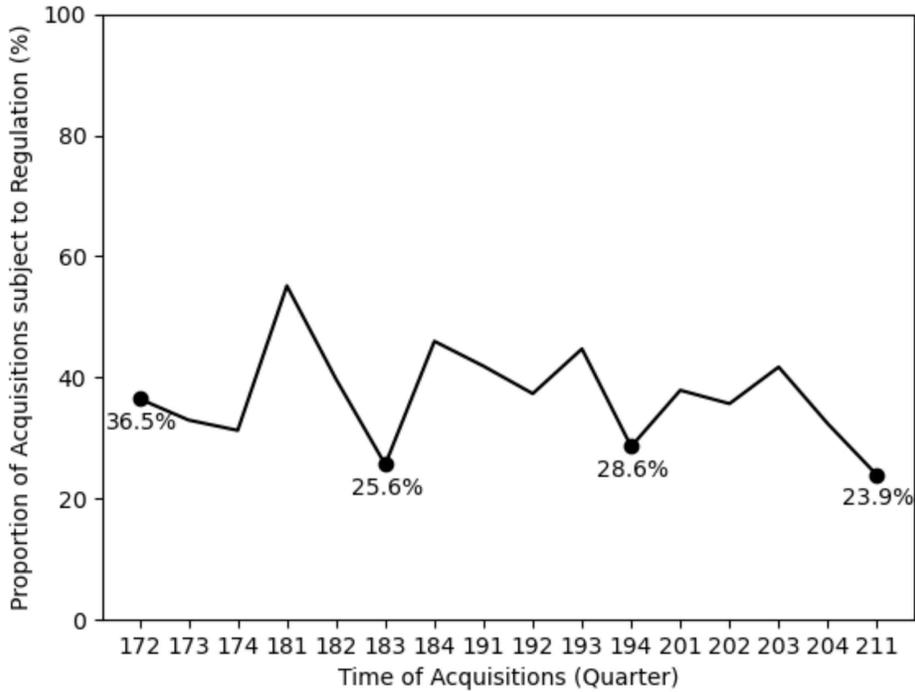


Figure 19 The proportion of acquisitions by foreigners with land area larger than 18m². For each label of x-axis, the first two numbers represent the year and the last number represent the quarter.

Contrary to past studies that described the Chinese in the Seoul Capital Area as residing in old houses, throughout the study period the real estate acquired by foreigners, most of whom were Chinese, was new buildings (Figure 20).

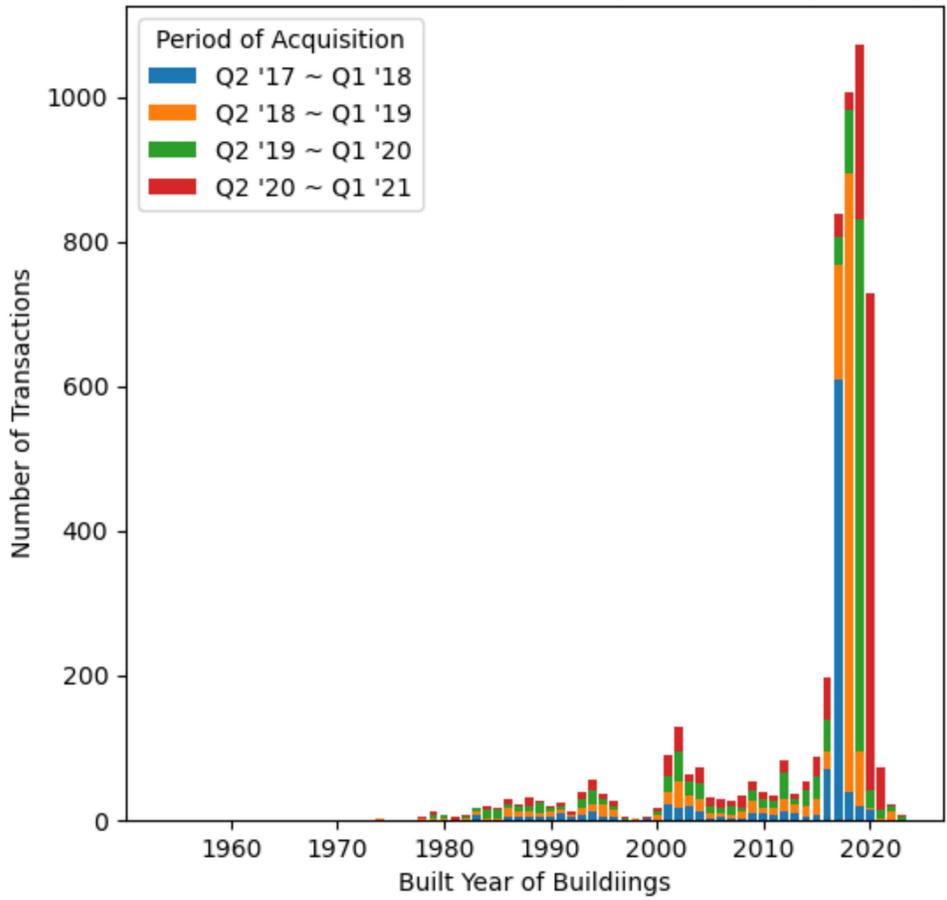


Figure 20 Distribution of built year of buildings vs. acquired period of buildings by foreigners.

4.4. Characteristics of Apartments Acquired by Foreigners

Apartments and officetels are the main properties acquired by foreigners, especially after the COVID-19 pandemic. This section distinguishes the apartments and officetels that were heavily acquired by foreigners from apartments and officetels that foreigners rarely acquired. The first spatial filter used as an explanatory variable can be interpreted as the proximity to Sosa Station, and its influence reached the eastern side of Bucheon station (Figure 21). The second spatial filter was related to Yeokgok station, and to a lesser magnitude, the western side of Bucheon station (Figure 22).

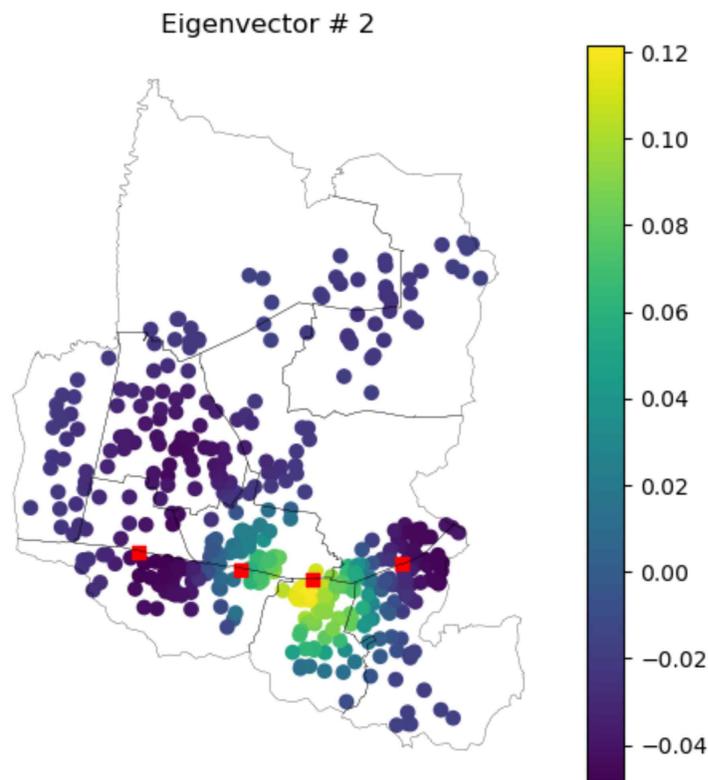


Figure 21 The illustration of the first spatial filter. The red dots indicate the location of Jungdong Station, Bucheon Station, Sosa Station and Yeokgok Station from left to right.

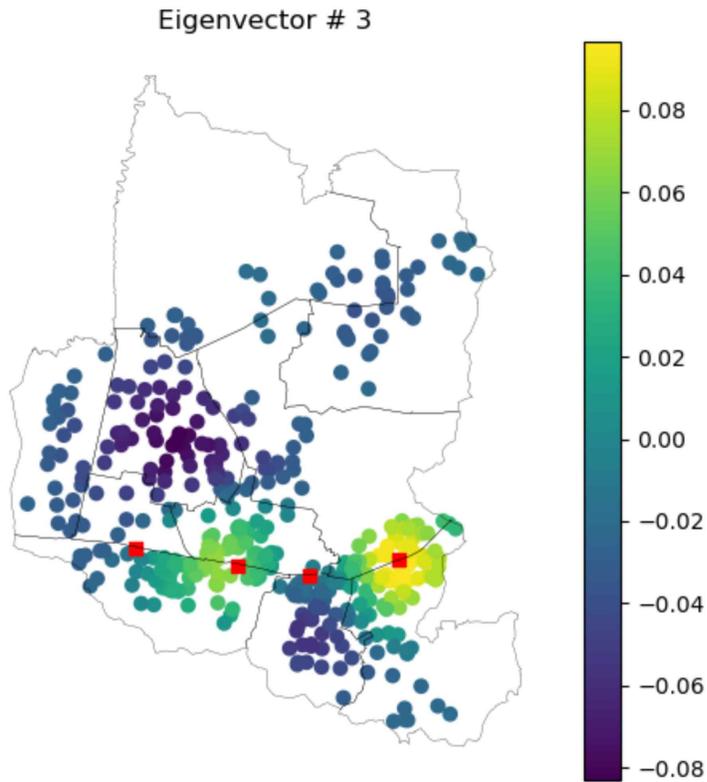


Figure 22 The illustration of the second spatial filter. The red dots indicate the location of Jungdong Station, Bucheon Station, Sosa Station and Yeokgok Station from left to right.

As can be seen from the tables below, the performance of random forest and gradient tree boosting significantly improved throughout the five test sets when compared with ordinary linear regression in terms of both root mean squared error and coefficient of determination. The performance comparison between gradient tree boosting and random forest showed different results from one test set to another. Stacking multiple models improved the performance, but to a lesser extent, when compared with the improvement from linear regression to random forest and gradient boosting.

Table 3 The distribution of root mean squared error for each test set and each best model. The unit is %p.

Test Set ID	Linear Regression	Random Forest	Gradient Boosting	Ensemble
1	15.9	12.2	11.3	11.7
2	13.5	7.3	10.1	7.2
3	14.9	8.6	8.5	8.3
4	13.1	9.8	9.6	9.5
5	13.7	9.0	11.2	9.2

Table 4 The distribution of coefficients of determination for each test set and each best model. The unit is %.

Test Set ID	Linear Regression	Random Forest	Gradient boosting	Ensemble
1	48.5	72.0	74.3	71.9
2	37.6	80.1	62.6	81.1
3	48.6	82.8	82.9	84.2
4	43.8	68.9	70.5	71.1
5	40.1	75.0	62.5	76.8

To characterize the apartments acquired by foreigners, SHAP was calculated with a best random forest model per set in case of test set ID 2 and 5, and SHAP was calculated with a best gradient boosting model per set in case of the remaining sets. The sum of SHAP values for a complex represents deviation from the expected value of a model, and the expected values for the five models were all around 12%. The respective SHAP values could be interpreted as the contribution over or below the expected values.

Generally, apartments located in Sosabon and Daesan districts contributed to the acquisition by foreigners (Figure 23). Especially, being approximate to Sosa Station contributed as high as 20%p on the target value (Figure 24). Among apartments and officetels near Sosa Station, officetels were more popular among foreigners than apartments. Proximity to Bucheon Station rather than Yeokgok Station showed more contribution to the target variable (Figure 25). These results illustrate the spatial concentration of real estate owned by foreigners, who are mainly Chinese, around Sosa and Bucheon Stations for the last four years.

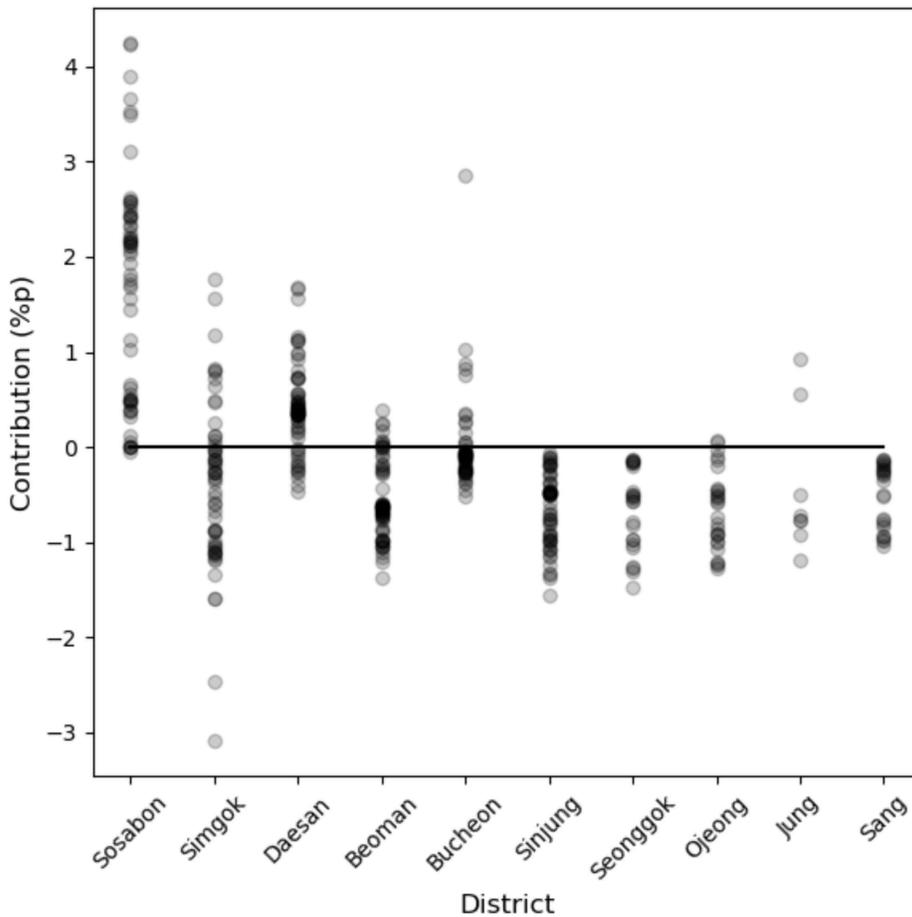


Figure 23 SHAP value for the district to which each building belongs to. Each dot represents a single complex of apartments or officetels.

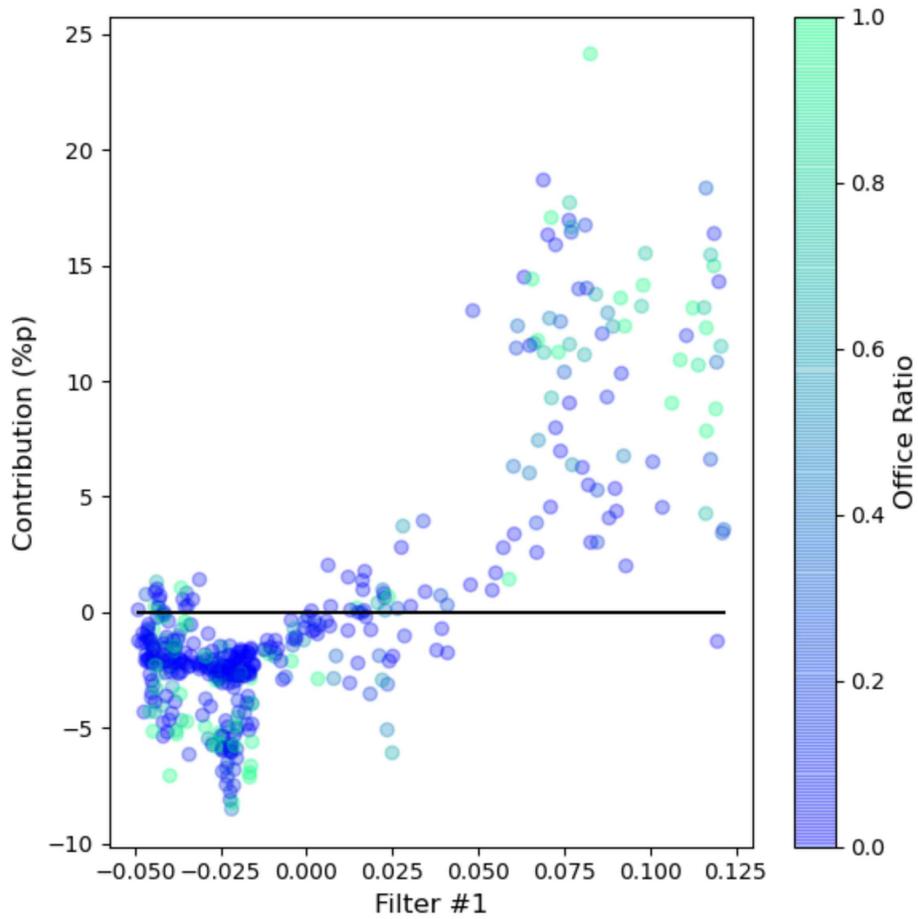


Figure 24 SHAP value for the value of the first spatial filter. A higher value of the filter indicates proximity to Sosa Station. Each dot represents a single complex of apartments or officetels. The color of each dot represents the proportion of officetels among the total number of homes in the complex.

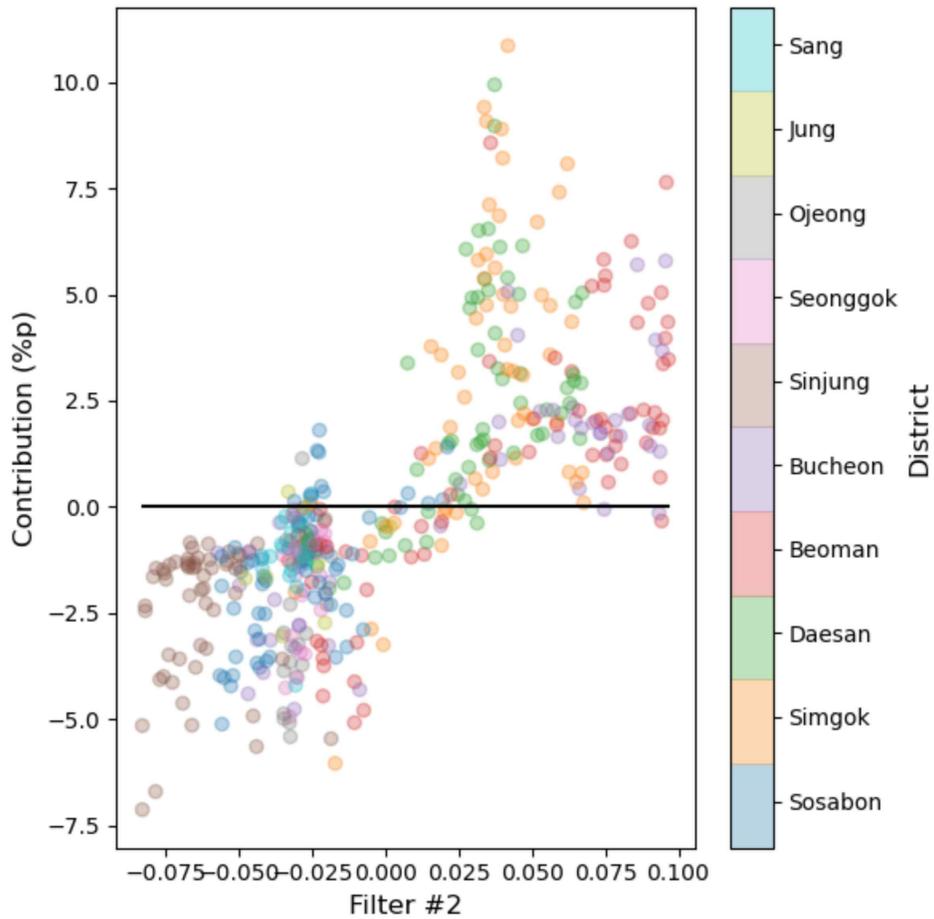


Figure 25 SHAP value for the value of the second spatial filter. The color of dots represents the district to which each apartment or officetel belongs to. Among the dots on the right half of the plot, dots in Bucheon and Beoman districts are near Yeokgok Station, while dots in Daesan and Simgok districts are near Bucheon Station.

Rather than pure officetels, foreigners favored acquiring a mixture of officetels with some apartments (Figure 26). Buildings with an average total of 16 floors were optimal for acquisition by foreigners (Figure 27). Foreigners acquired newly built buildings rather than used buildings; the contribution of newly built apartments to the target value was as high as 20%p (Figure 28). With regard to the number of buildings, stand-alone apartments rather than a group of apartments were acquired by foreigners (Figure 29). Big apartment complexes contributed negatively to the target variable, especially when they were newly built. Therefore, real estate acquired by foreigners is characterized as small-scale new apartments or officetels with moderate height.

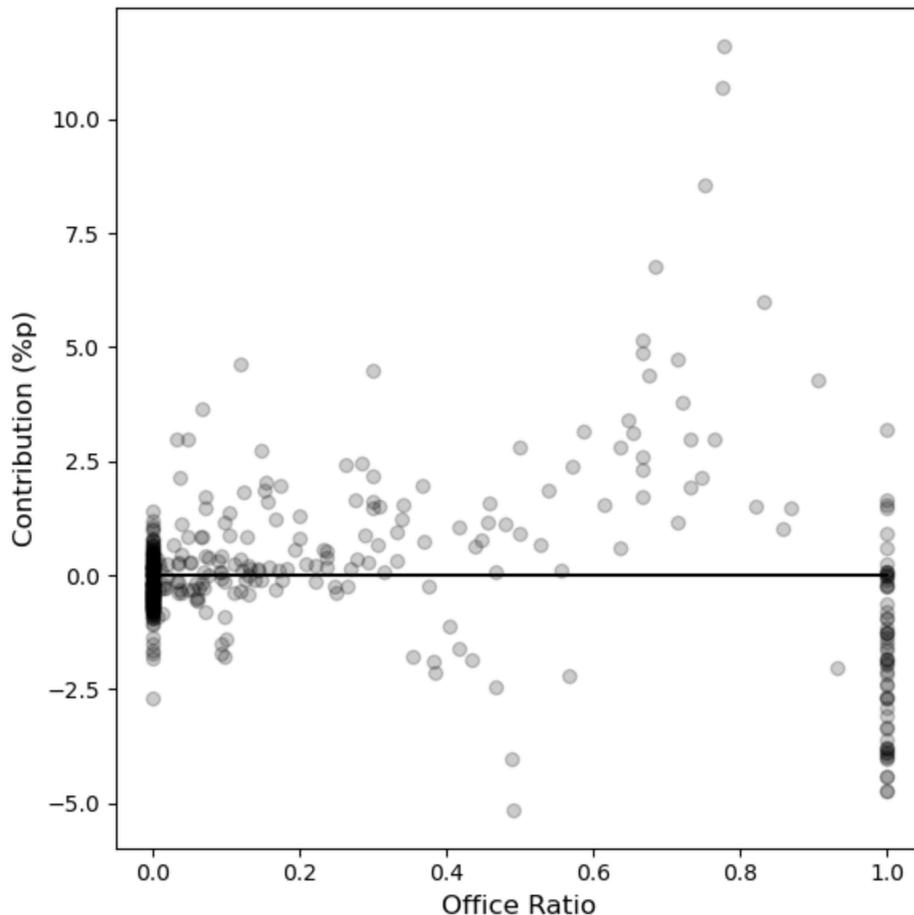


Figure 26 SHAP value for the proportion of officetels among the whole number of houses and offices in the complex.

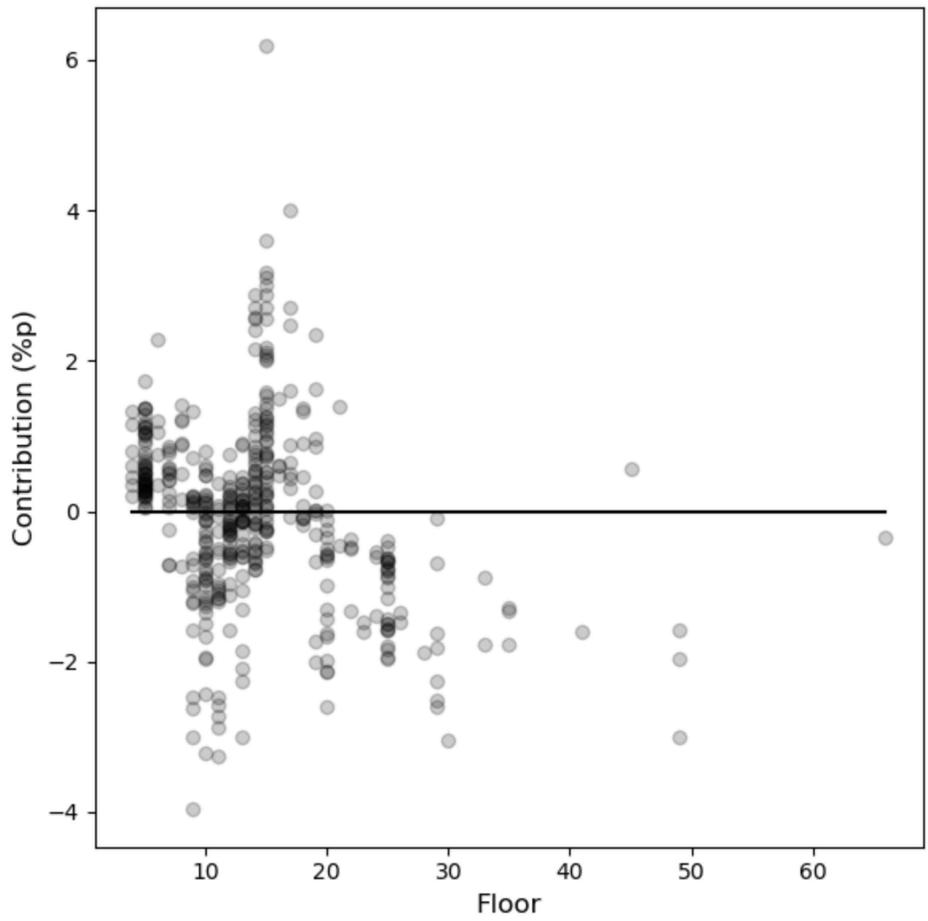


Figure 27 SHAP value for the total number of floors in respective apartments or officetels.

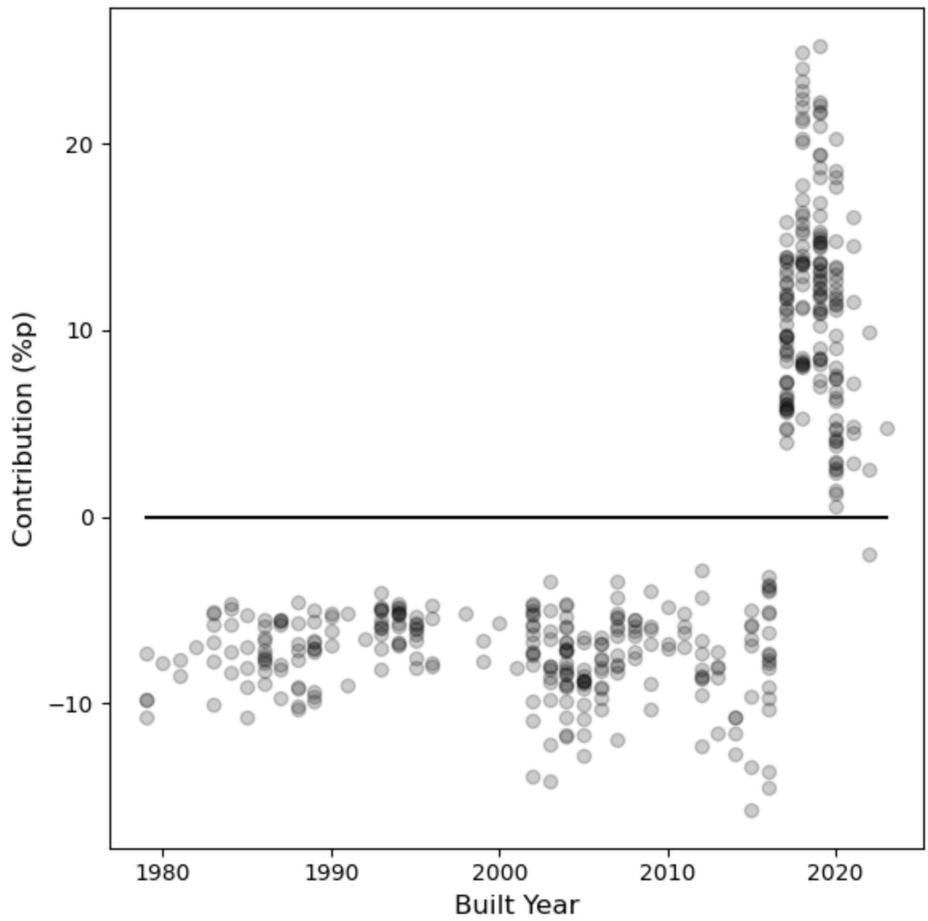


Figure 28 SHAP value for the completion year of respective apartments or officetels.

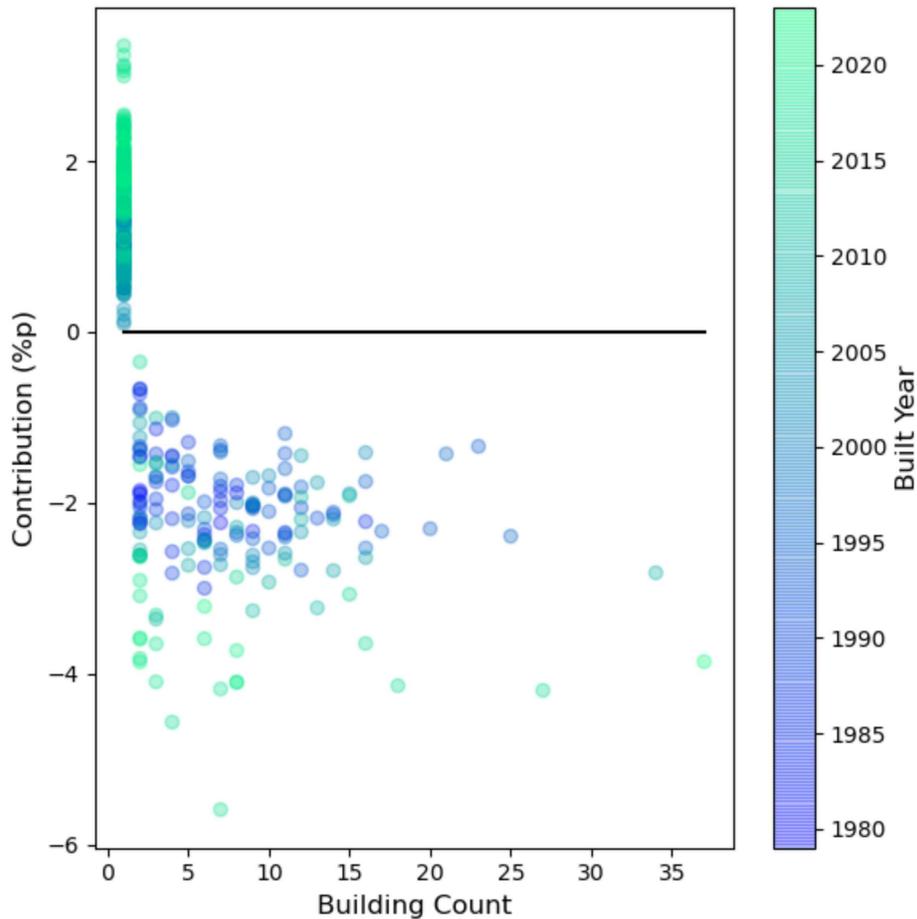


Figure 29 SHAP value for the number of buildings in respective complexes of apartments or officetels. The color of dots indicates the completion year of the complex.

Apartments and officetels in the pictures below are some of the 30 buildings described in figure 12. They accounted for more than 30% of the total number of acquisitions by foreigners during the study period, with each building acquired by foreigners at least 30 times. The characteristics illustrated by interpretable machine learning can be appreciated in the pictures below (Figure 30).



Figure 30 The view of apartments and officetels which have been actively acquired by foreigners.

5. Discussion and Conclusion

5.1. Main Findings

In Bucheon, Chinese are the main actors among foreigners who actively purchase real estate. This tends to be high and new officetels rather than small and old houses. Even in recent studies, Chinese in the Seoul Capital Area were treated as tenants in old houses and their influence on real estate acquisition was negligible compared to their population. This sudden change calls for a multifaceted approach that does not generalize the Chinese as a petty group. The steeper increase of land acquisition by foreigners compared to the increase of population can be interpreted as an evolution from the Chinese as a tenant of old houses to the Chinese as an owner of a new apartment.

Among the Chinese, ethnic Koreans of Chinese nationality were suggested as the drivers of such a trend because their population change coincided with the number of real estate acquisitions, especially during the COVID-19 pandemic, while the population of foreigners in general, except the Chinese, decreased. Even though the spatial ownership and the spatial occupancy by foreigners diverged in terms of the total foreign population during the pandemic, when focusing on the ethnic Koreans of Chinese nationality, the ownership and occupancy converged.

Throughout the study period, the mode of real estate acquisition by foreigners changed toward more compact buildings with a geographical concentration on districts related to Gyeongin and Seohae railroad rather than northwestern industrial districts where there were less Chinese. Namely, the mandatory permission policy for acquisition of real estates by foreigners could not surveil speculation effectively because the acquisition of compact buildings, such as apartments, is not subject to the regulation due to its small share on land area.

Interpretable machine learning could help characterize the apartments and officetels acquired by foreigners. As a result, it could be shown that the proportion of foreign owners in Bucheon is relatively low in big new apartment complexes or in skyscrapers of rich districts such as Sinjung and Sang. This implies that the reality is different from foreign speculators depriving Koreans of their dream home. That is, Chinese are different from the privileged in Sao Paulo who use helicopters to commute (Noorloos, 2019).

5.2. Future Research Direction

The first limit of this research is its lack of interaction with neighbors. For instance, after the completion of the Seohae Line, real estate acquisition by foreigners around Sosa Station increased. Correlating it with the change of the foreign population in Ansan will help understand the origins of acquisition. Also, real estate acquisition by foreigners in Bucheon should be further investigated by taking the effect of immigration from Gwangmyeong, a city under renovation southeast from Bucheon, into account. As Bucheon is near Daerim district, home of Seoul's most famous Chinatown, their relationship with each other should also be addressed.

Further access to data and refinement of the data might be required in the future. The lack of nationality on individual acquisition data was an obstacle to create a more thorough geographical analysis based on real estate acquisition by nationality. Also, the acquisition data adjusted for the variability of lots still has some limit. Units are represented by the number of reports in spite of normalization by the number of lots the reported real estate occupied. Therefore, multiple reports for a house can exist when several foreigners divide their share of the ownership. Moreover, the divergence among the Chinese could not be further illustrated than the divergence between F-4 Chinese and other Chinese during the COVID-19 pandemic. It is possible that only some rich Chinese have been buying multiple houses while most others are still just tenants. This issue must be addressed.

Finally, cause and effect relationships should also be revealed based on the data provided in this research. On the one hand, there is the possibility that foreigners buy their houses swiftly due to anxiety from the upsurge in housing prices, and the permission system could have suppressed them to compact apartments. On the other hand, it is possible that the increased demand of foreigners has contributed to the instability of housing prices, and the policy might succeed in preventing them from speculating in large-scale premium apartments favored by Koreans. Interviewing the acquirers may provide a clue to the issue.

Bibliography

- Bhang, S. (2012) A Study on the Characteristics and Residential Differentiation of the Korean Chinese Community: Focusing on Garibong-Dong and Jayang-Dong in Seoul, Korea (Master's thesis, Sejong University)
- Biecek, P. (2018) DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research*, 19(84), 1-5.
- Biecek, P. and Burzykowski, T. (2020) *Explanatory Model Analysis*. Available at: <https://ema.drwhy.ai/> (accessed April 14th 2021)
- Boehmke, B. and Greenwell, B.(2020) *Hands-on Machine Learning. With R* Available at: <https://bradleyboehmke.github.io/HOML/> (accessed April 14th 2021)
- Brill, F. and Raco, M. (2020) Towards a pluralistic understanding of Chinese homeowners: the case of 'ordinary' buyers. *Geoforum*, 117, 165-172.
- Bucheon City (2021a) Beforehand information disclosure on land acquisition by foreigners. Available at: <https://bucheon.go.kr/site/program/board/basicboard/list?boardtypeid=26716&menuid=148004003004&searchcategory=외국인%20토지거래%20신고%20및%20허가%20현황> (accessed April 13th 2021)
- Bucheon City (2021b) Statistics on population and household. Available at: <https://stat.bucheon.go.kr/site/program/board/basicboard/list?boardtypeid=6225&menuid=092002001> (accessed April 13th 2021)
- Bucheon City (2021c) Information on apartment complexes. Available at: http://apt.bucheon.go.kr/site/program/apt/new_list?menuid=160001001 (accessed June 17th 2021)
- Chen, H., Chen., H., Zhou, R., Liu, Z. and Sun, X. (2021) Exploring the mechanism of crashes with autonomous vehicles using machine learning. *Mathematical Problems in Engineering*, 2021, 5524356.
- Dray, S., Legendre, P. and Peres-Neto, P. (2006) Spatial modelling: A

- comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). *Ecological Modelling*, 196, 483-493.
- ESR Cayman Ltd. (2020) 1H 2020 Interim Report.
- ESR Kendall Square (2020) Investment Explanatory Document.
- ESR REIT (2019) Announcement of Resignation of Non-Executive Director.
- Gyeonggi Province (2020) Public announcement on new appointment of land transaction permit zone. Public Announcement No. 2020-1903.
- Gyeonggi Province (2021) Beforehand information disclosure on land acquisition and disposal by foreigners. Available at:
<https://www.gg.go.kr/opendata/openDataBoard.do?boardIdx=iXB>
 (accessed April 13th 2021)
- Han, S. and Park, K. (2015) The problem and solution on the Chinese's purchase of Jeju-do real estate. *Law and Policy Review*, 21(3), 405-437.
- Huang, C. (2018) Real estate price, and policy in Canada (Partial Fulfillment of Master's Degree, Saint Mary's University)
- Huh, J. and Jang, J. (2020) Mapping community-level mobility changes of Koreans and Immigrants using bigdata of 'de facto population of Seoul: Six ethnic enclaves in Seoul under COVID-19 epidemic. *Space and Environment*, 30(3), 99-137.
- ICNKR (2021) 论坛>房屋房产>房产买卖. Available at:
<http://bbs.icnkr.com/forum-1479-1.html> (accessed April 12th 2021)
- Kim, C. and Moon, T. (2018) Study on the determinants of residential location choice of foreigners in Seoul. *Housing Studies Review*, 26(4), 183-204.
- Kim, H. (2017) Ethnic connections, foreign housing investment and locality: a case study of Seoul. *International Journal of Housing Policy*, 17(1), 120-144.
- Kim, S. (2012) Study on the Residential Status and Perceptions for Real Estate of Multi-cultured Family (Master's thesis, Kyung Hee University)

- Kim, S. and Kim, Y. (2020) Issues and challenges of foreigners' real estate acquisition in Korea. *National Assembly Research Service Report*, 1768.
- Korea Real Estate Board (2021) Apartment price index. Available at : http://www.kab.co.kr/kab/home/trend/market_trend06.jsp (accessed June 17th 2021)
- Kuhn, M. and Wickham, H. (2020) Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. Available at: <https://www.tidymodels.org> (accessed April 12th 2021)
- Lee, M. (2020) 경기도가 외국인 법인 대상 '토지거래허가제' 꺼낸 이유는? Gyeonggi Province News Portal
- Lee, M. (2021) 외국인 법인 '토지거래허가제' 통했나... 거래량 대폭 감소. Gyeonggi Province News Portal
- Lee, S. (2014) A study on the Formation and Growth of Korean-Chinese Ethnic Clusters (Master's thesis, Seoul National University).
- Legendre, P. and Legendre, L. (2012) Multiscale analysis: spatial eigenfunctions. In *Numerical Ecology* (pp. 859-906), Elsevier.
- Liao, S. (2019) Chinese foreign real estate investment and voting in US presidential elections. *SSRN Electronic Journal*. Available at: <http://dx.doi.org/10.2139/ssrn.3504861>
- Lundberg, S., Erion, G. and Lee, S. (2018) Consistent individualized feature attribution for tree ensembles. *ArXiv*, abs/1802.03888.
- Ministry of Justice Korea Immigration Service (2021) Monthly Statistics. Available at: <https://www.immigration.go.kr/immigration/1569/subview.do> (accessed April 13th 2021)
- Ministry of Trade, Industry and Energy (2015) Investment Report Statistics System.
- National Spatial Data Information Portal (2021) Real Estate Development

Industry Search.

- Noorloos, F. (2019) Land in urban debates: Unpacking the grab-development dichotomy. *Urban Studies*, 56(5), 855-867.
- Oh, J., Lee, Y. and Kim, G. (2020) Improvement of solar power forecasting using interpretation of artificial intelligence. *The Transactions of the Korean Institute of Electrical Engineers*, 69(7), 1111-1116.
- Paek, Y. (2018) Foreigner's housing market and transnational housing filtering process: focusing on the case study of Korean-Chinese. *Journal of the Economic Geographical Society of Korea*, 21(4), 370-392.
- Pyeon, J. (2020) 외국인이 부동산 '씩쓸이'했다고? "실거주 목적 많아" The Kyeong-gi Shinmun.
- Rogers, D. and Koh, S. (2017) The globalization of real estate: the politics and practice of foreign real estate investment. *International Journal of Housing Policy*, 17(1), 1-14.
- Tanrivermis, H. (2020) Possible impacts of COVID-19 outbreak on real estate sector and possible changes to adopt: A situation analysis and general assessment on Turkish perspective. *Journal of Urban Management*, 9, 263-269.
- Wong, A. (2017) Transnational real estate in Australia: new Chinese diaspora, media representation and urban transformation in Sydney's Chinatown. *International Journal of Housing Policy*, 17(1), 97-119.
- Xi, W. (2016) The Effect of Land Transaction Volume with Chinese Investors on Jeju's Housing Sales Price (Master's Thesis, Seoul National University)
- Ye, X., Shan, Q. and Huang, H. (2017) Do house price spillover across main cities of Canada and China? - Evidence from causality analysis. *SSRN Electronic Journal*. Available at: <http://dx.doi.org/10.2139/ssrn.2958567>

Appendix A. Decision Tree based Models and Interpretability

As a type of regression, a decision tree tries to minimize the residuals through splitting the instances, and then estimating the target variable of each instance at the mean value of the instances that belong to the same group as the target instance.

Because a single decision tree cannot perform better than a team of decision trees, bagging makes the most of collective intelligence through the average result of numerous decision trees (Boehmke and Greenwell, 2020). Each decision tree is trained by the same number of instances as the original training set, but the instances are resampled with replacement from the original training set each time. The main drawback of bagging is that numerous decision trees could be so similar that using the team cannot create synergy as expected.

Random forest solved this issue through making diversity by introducing randomness (Boehmke and Greenwell, 2020). When a single decision tree makes a classification, it judges a randomly selected, predefined number of features rather than whole features. In order to predefine the number of features to select, cross validation is used: making a subset of the training set as another test set, using the rest of the training set to train the models with different candidates for the number of features to select, then check the performance of the models with the test subset.

Alternatively, gradient tree boosting is composed of a sequence rather than a mass of decision trees. Prediction is made through the sum rather than the average of these trees. Each decision tree tries to minimize the residuals made from the subsequent trees (Boehmke and Greenwell, 2020). In order to prevent overfitting on the training set that worsens the prediction power of the model on the testing set, a parameter called shrinkage should be determined beforehand through cross validation. Shrinkage plays a role as a

coefficient that attenuates the subsequent trees of the sequence. The term 'gradient' stems from defining each subsequent tree through differentiating the squared error between the true value and model estimation with a preceding tree, and then multiplying minus 1. Differentiation is closely related to the direction of slope. For example, the elevation of a concentric crater with its center at the origin can be described as $L = x^2 + y^2$. Partial differentiation of L by each coordinate gives the number of walks that should be taken eastward ($2x$) and northward ($2y$) to climb out of the crater with shortest distance. Minus 1 is multiplied to find a steepest descending road rather than an ascending road: the smaller the error, the higher the performance. Shrinkage can be understood as a scalar multiplied on the gradient ($2x, 2y$) to adjust the number of steps taken each time, consequently adjusting the speed to reach the bottom of the crater.

Randomness can intervene in gradient boosting too (Boehmke and Greenwell, 2020). For each sequence of decision trees, the tree is fit to the subset of the original training set. This is analogous to taking an eastward walk only to reach the very bottom of the crater rather than a pit that pretends to be a bottom.

In order to make a more precise model, ensemble modelling is introduced (Boehmke and Greenwell, 2020). The performance of a model can be improved by finding the best combination of candidate models, such as random forest and gradient tree boosting. Even among a single type of models, for instance a gradient tree boosting, multiple candidates with different hyperparameters such as shrinkage can be combined to improve the performance. To prevent overfitting, lasso regularization can be applied in case of making an ensemble by linear combination of candidates. Lasso regularization makes an upper limit for the sum of regression coefficients while minimizing the sum of squared errors.

Unlike a single decision tree, using multiple decision trees makes the interpretation of the model difficult despite improvement in prediction performance. For interpretability, SHAP value was suggested as a solution, which is based on Shapley value (Biecek and Burzykowski, 2020). When there are four features A_1, A_2, A_3, A_4 , Shapley value for an instance x , feature A_2 , and model g is defined as

$$\begin{aligned}
S(A_2, x, g) &= \frac{3!}{4!} \times [f(\{A_2\}) - f(\emptyset)] \\
&+ \frac{2!}{4!} \times [f(\{A_1, A_2\}) - f(\{A_1\})] \\
&+ \frac{2!}{4!} \times [f(\{A_2, A_3\}) - f(\{A_3\})] \\
&+ \frac{2!}{4!} \times [f(\{A_2, A_4\}) - f(\{A_4\})] \\
&+ \frac{2!}{4!} \times [f(\{A_1, A_2, A_3\}) - f(\{A_1, A_3\})] \\
&+ \frac{2!}{4!} \times [f(\{A_1, A_2, A_4\}) - f(\{A_1, A_4\})] \\
&+ \frac{2!}{4!} \times [f(\{A_2, A_3, A_4\}) - f(\{A_3, A_4\})] \\
&+ \frac{3!}{4!} \times [f(\{A_1, A_2, A_3, A_4\}) - f(\{A_1, A_3, A_4\})]
\end{aligned}$$

where A_i of x is a_i and $f(\{A_j, A_k\})$ is the expected value of g when $A_j = a_j$ and $A_k = a_k$. The numerators multiplied on each term count the number of permutations of features. For each permutation, features before the position of A_2 become the conditions for the conditional expectation. Each term means the contribution of a_2 to the respective conditional expectations. The weights are divided by the total number of permutations for normalization.

The advantage of Shapley value is that it accounts for every possible interaction effect through permutations and conditional expectations. Also, it

is theoretically sensible in that

$$\sum_{i=1}^4 S(A_i, x, f) = f(\{A_1, A_2, A_3, A_4\}) - f(\emptyset) = g(x) - (\text{expected value of } g) \text{ and}$$

$$S(A_i, x, g) + S(A_i, x, h) = S(A_i, x, \frac{g+h}{2}).$$

However, its disadvantage is that it even accounts for an interaction effect that is impossible due to correlated features, and that it is computationally expensive.

Appendix B. Derivation of Theil Index

The concept of Theil index is based on information theory, namely entropy. Suppose 8 individuals were sampled from the population who were asked his or her favorite color personally, and received 8 answers X_1, X_2, \dots, X_8 . The probability distribution of the favorite color of the population is unknown. Suppose a model that can figure out an individual's favorite color most efficiently was made assuming the probability distribution $q(X = x)$ based on training from X_1, X_2, \dots, X_8 .

$$q(X = \text{black}) = 4/8$$

$$q(X = \text{white}) = 2/8$$

$$q(X = \text{red}) = 1/8$$

$$q(X = \text{blue}) = 1/8$$

Assume that the constraint on the model is that the model can ask only yes-or-no questions. Then the most efficient model minimizes the number of questions asked for finding the answer. Based on this condition, the first question of our model should be:

Do you like black?

If the answer is yes, finding the answer in just one question is successful. According to training with the 8 individuals, this probability is 50% because $q(X = \text{black}) = 4/8$. If the answer is no, next question should be asked:

Do you like white?

The probability of saying no to the first question and yes to the second question is 25% because $q(X = \text{white}) = 2/8$. So the probability of asking two questions is 25%.

In the other 25% case, the final question should be asked to get the answer. The final question can be either, “Do you like red?” or “Do you like blue?” because the probability of answering yes(12.5%) to this question is the same as that of answering no(12.5%).

In sum, the average number of questions this model asks, in other words, *entropy*, is

$$\begin{aligned}
 & 0.5 \times 1 + 0.25 \times 2 + 0.125 \times 3 + 0.125 \times 3 \\
 &= -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{4} \log_2 \frac{1}{4} - \frac{1}{8} \log_2 \frac{1}{8} - \frac{1}{8} \log_2 \frac{1}{8} \\
 & \quad = 1.75 \\
 &= - \sum_x q(X=x) \log_2 q(X=x)
 \end{aligned}$$

The performance of this model can be tested from drawing 8 individuals again from the population and asking their favorite color according to this model of question sequences. Suppose the probability distribution of the testing group is $p(X = x)$. If $p = q$ (4 people like black, 2 people like white, etc.), the probability of finding the answer at once is 50%, finding the answer with two questions is 25%, finding the answer with the third question is 12.5% $\times 2$. If the performance of model is tested in terms of the average number of questions required to get the answer, or *cross entropy*,

$$\begin{aligned}
& 0.5 \times 1 + 0.25 \times 2 + 0.125 \times 3 + 0.125 \times 3 \\
&= -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{4} \log_2 \frac{1}{4} - \frac{1}{8} \log_2 \frac{1}{8} - \frac{1}{8} \log_2 \frac{1}{8} \\
&= 1.75 \\
&= -\sum_x p(X=x) \log_2 q(X=x)
\end{aligned}$$

The performance of this model should be interpreted as good if cross entropy is small because it means that asking many questions is not needed on average. When $p \neq q$, the model's performance will go down, or cross entropy will go up. Suppose,

$$p(X = \text{black}) = p(X = \text{white}) = p(X = \text{red}) = p(X = \text{blue}) = 2/8$$

then the probability of finding the answer at once with the model becomes only 25% because only two people will say yes to the first question. When calculating the average number of questions that should be asked to the test group in order to get the answer with this model,

$$\begin{aligned}
& 0.25 \times 1 + 0.25 \times 2 + 0.25 \times 3 + 0.25 \times 3 \\
&= -\frac{1}{4} \log_2 \frac{1}{2} - \frac{1}{4} \log_2 \frac{1}{4} - \frac{1}{4} \log_2 \frac{1}{8} - \frac{1}{4} \log_2 \frac{1}{8} \\
&= 2.25 \\
&= -\sum_x p(X=x) \log_2 q(X=x)
\end{aligned}$$

If the probability distribution of the population were $p(X = x)$ rather than $q(X = x)$, the ideal model should have asked the following questions in order:

Do you like achromatic or chromatic color?

(If achromatic,) Do you like black or white?

(If chromatic,) Do you like red or blue?

In that case, the entropy will be,

$$0.25 \times 2 + 0.25 \times 2 + 0.25 \times 2 + 0.25 \times 2$$

$$= - \sum_x p(X=x) \log_2 p(X=x) = 2$$

Therefore, the best model should minimize the difference between cross entropy and ideal entropy of the population model, in other words *relative entropy*, defined as,

$$[- \sum_x p(X=x) \log_2 q(X=x)] - [- \sum_x p(X=x) \log_2 p(X=x)]$$

$$= \sum_x p(X=x) \log_2 \frac{p(X=x)}{q(X=x)}$$

Based on these concepts of entropy, the Theil index can be derived. Suppose

x_{ij} = value of i th individual in j th group

n_j = population of j th group

\bar{x}_j = mean value of j th group

n = total population

Then the share of x_{ij} in its group becomes $p_{ij} = \frac{x_{ij}}{n_j \bar{x}_j}$, while its global share

becomes $q_{ij} = \frac{x_{ij}}{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \dots}$. Note that $\frac{q_{ij}}{p_{ij}} = y_j$ is the global share of j th group.

Then inequality of j th group is

$$T_j = (\text{ideal entropy, where } p_{ij} = 1/n_j \text{ for all } i) - (\text{real entropy})$$

$$= \log n_j - \sum_{i \in j} p_{ij} \log \frac{1}{p_{ij}}$$

Then inequality between groups is

$$\begin{aligned}
& T_{b/w} \\
& = (\text{cross entropy, assuming value is proportional to population of each group}) \\
& \quad - (\text{real entropy}) \\
& = (\text{relative entropy}) = \sum_j y_j \log \frac{n}{n_j} - \sum_j y_j \log \frac{1}{y_j}
\end{aligned}$$

Then global inequality is

$$\begin{aligned}
T & = (\text{ideal entropy, where } q_{ij} = 1/n \text{ for all } i \text{ and } j) - (\text{real entropy}) \\
& = \log n - \sum_{i,j} q_{ij} \log \frac{1}{q_{ij}} = \sum_{i,j} (q_{ij} \log n + q_{ij} \log q_{ij}) = \sum_{i,j} q_{ij} \log n q_{ij} \\
& = \sum_{i,j} \frac{q_{ij}}{p_{ij}} [p_{ij} (\log n_j p_{ij} + \log \frac{n q_{ij}}{n_j p_{ij}})] \\
& = \sum_j y_j (\sum_{i \in j} p_{ij} \log n_j p_{ij} + \log \frac{n}{n_j} y_j \sum_{i \in j} p_{ij}) \\
& = \sum_j y_j T_j + T_{b/w}
\end{aligned}$$

which is the weighted average of within-group inequality plus between-group inequality.

Appendix C. Derivation of Spatial Filters

In order to retrieve the spatial structure of apartments and officetels acquired by foreigners, spatial filters were created using the locations of the buildings. Suppose there are n points. Spatial filter is based on Moran's spatial autocorrelation statistic (I). Moran's I can be interpreted as the covariance with neighbors compared to the variance of x in the whole study region as follows:

$$I = \frac{\text{Cov}(x, x_{\text{neighbor}})}{\text{Var}(x)} = \frac{\sum_{i,j} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i,j} w_{ij}} / \frac{\sum_i (x_i - \bar{x})^2}{n}$$

where x_i is the attribute value at point i , \bar{x} is the mean of x in the whole study region and w_{ij} describes the degree of adjacency between point i and j . Moran's I can be expressed in matrix form as follows:

$$I = \frac{n}{\sum_{i,j} w_{ij}} \times \frac{(MX)^T W (MX)}{(MX)^T (MX)} = \frac{n}{\sum_{i,j} w_{ij}} \times \frac{X^T (M^T W M) X}{(MX)^T (MX)}$$

where X can be interpreted as a column vector where i th component of X equals x_i . M is defined as a $n \times n$ matrix where the identity matrix is subtracted by the matrix whose every component is $1/n$. MX centers the list of x into $x - \bar{x}$ and multiplying it with its transpose can be interpreted as the sum of squares. W is the matrix representing neighborhood relationship where the element of i th row and j th column equals w_{ij} . W should be symmetric for the reason that will be addressed later. This means that if point j is a neighbor of point i , point i should be a neighbor of point j , which can be a drawback because there exists one-way interaction between regions in the real world. The very right hand side of the equation is derived from the definition of matrix multiplication, namely $(MX)^T = X^T M^T$.

In order to extract spatial components, focusing on purely spatial parts is required. In this context, the matrix of attribute value X is ignored, and spatial structure is extracted only, namely S that is defined as follows:

$$S = \frac{n M^T W M}{\sum_{i,j} w_{ij}}$$

S is a symmetric matrix because M and W are also symmetric. Let the eigenvalues of the symmetric matrix S be $\lambda_1, \lambda_2, \dots, \lambda_n$ in descending order, and corresponding eigenvectors for each eigenvalue are V_1, V_2, \dots, V_n . From the definition of eigenvalue and eigenvectors, it follows that for any k , $SV_k = \lambda_k V_k$ and $V_k^T V_k = 1$. As a result,

$$\lambda_k = V_k^T S V_k = \frac{V_k^T (nM^T W M) V_k}{\sum_{i,j} w_{ij}} = \frac{n}{\sum_{i,j} w_{ij}} \times \frac{(M V_k)^T W (M V_k)}{\frac{n}{n-1} (M V_k)^T (M V_k)} \approx \frac{\text{Cov}(v, v_{\text{neighbor}})}{\text{Var}(v)}$$

In other words, λ_k becomes a Moran's spatial autocorrelation statistic (I) for a given spatial structure (W) when x_i equals i th element of V_k .

These eigenvectors can be used as additional explanatory variables representing spatial autocorrelation of the distribution of target variable. Because S was symmetric, its eigenvectors V_1, V_2, \dots, V_n are independent and multicollinearity among these spatial filters does not exist. That is why W should be symmetric above.

Like other explanatory variables, selection of significant eigenvectors is mandatory to prevent overfitting. One of the criteria for the selection is choosing eigenvectors that critically decrease the sum of squared error (Dray et al., 2006). And the selected eigenvectors are called spatial filters.

Moran's eigenvector maps take the distance of neighbors into account when constructing W . W is defined as according to the threshold distance for being the neighbor or not, d as below.

$$w_{ij} = 0$$

$$w_{ij} = 1 - \left(\frac{d_{ij}}{4d}\right)^2 \quad \text{if } 0 < d_{ij} \leq d$$

$$w_{ij} = 0 \quad \text{if } d_{ij} > d$$

The reason for dividing with the number $4d$ is to make the resulting eigenvectors V_1, V_2, \dots, V_n invariant even if using numbers greater than 4. Note that the minimal value of W except 0 will be $15/16$, which is almost 1. d is the maximum distance in the minimum spanning tree of the whole n points (Legendre and Legendre, 2012).

초록

2020년 외국인이 소유한 경기도 내 토지의 필지수와 면적은 증가한 데 반하여, 코로나-19 팬데믹으로 도내 외국인 인구는 줄었다. 급등하는 주택 가격과 부동산 투기가 국가적 문제로 대두된 가운데, 경기도는 외국인을 대상으로 토지거래허가제를 실시하게 되었다.

부재지주가 많아지면 도시의 맥락에서 토지횡령(land grabbing) 문제가 발생할 수 있겠지만, 코로나바이러스감염증-19가 유행하는 상황에서 주택 가격 안정화의 실패를 평범한 외국인의 탓으로만 돌리는 제노포비아 정서에도 주의가 요구된다. 이 문제에 올바른 판단을 내리기 위해서는 부동산 취득의 공간적 특성에 관하여 충분한 양적 이해가 선행되어야 한다.

이 연구는 부천시를 중심으로 외국인의 부동산 취득에 대한 공간적 역동성을 분석한다. 지난 5년간 외국인의 부동산 취득의 측면에서 부천은 경기도에서 가장 두드러지는 도시이기 때문이다.

부천시의 부동산을 취득하는 외국인은 주로 한국계 중국인으로 드러났다. 이 결과는 서울시의 부동산을 주로 취득하는 외국인이 미국인이나 대만인이었다는 선행 연구나, 국내 중국인을 낙후된 주택의 세입자로 바라보았던 선행 연구들과 대조된다.

테일 지수(Theil index)를 분석한 결과에 따르면, 외국인의 부천시 부동산 취득은 소축척으로도 대축척으로도 모두 공간적으로 불균등해지고 있다. 특히 코로나-19 팬데믹 기간에는 소사역 주변의 오피스텔과 아파트

로 취득이 집중되었는데, 소사역은 안산과 부천을 잇는 서해선 철도와 경인선 철도가 만나는 곳이라는 특징이 있다.

해석가능한 기계학습(interpretable machine learning)을 통해서도 외국인들이 16층 정도의 신축 나홀로아파트 또는 오피스텔을 활발하게 취득하는 것으로 밝혀졌다. 이러한 경우 토지거래허가대상이 아닌 경우가 많기 때문에, 토지거래허가제가 부천시에서는 그 효과가 떨어지는 것으로 나타났다.

이 연구는 신축아파트의 중국인 집주인으로 대표되는, 새로운 수도권 차이나타운의 형성을 다루었다는 점에서 기왕에 논의된 한국계 중국인 밀집지역과 결을 달리한다. 중국인의 공간점유에서 중국인의 공간소유로 전환되는 과정이 수도권에서 포착되었다. 그러나 중국인이 소유하는 아파트의 입지와 속성은 한국인이 선호하는 신중동 내지 상동의 대단지 아파트와 구별되었다.

주요어 : 탐색적 자료분석, 부동산, 한국계 중국인, 코로나19 팬데믹, 해석가능한 기계학습, 테일 지수.

학번 : 2019-27834