Retail Vacancy in the Neighborhood: the Association Between Retail Vacancy and Neighborhood Characteristics in the Netherlands

Martine Verwijs

ABSTRACT

This thesis analyzes the association between a neighborhood's retail vacancy rate and specific neighborhood characteristics. Kadaster and CBS data is combined to obtain information on retail vacancy, neighborhood characteristics, and the number of realized property transformations to housing. Multiple linear regressions are conducted in STATA with neighborhood's retail vacancy as the dependent variable. The research shows a positive association between retail vacancy and population characteristics, the urbanity level, and the number of realized property transformations to housing. Furthermore, construction before 2000 and the average housing value show a negative association with retail vacancy in the neighborhood. The relationship between retail vacancy and the number of realized property transformations to housing shows no significant difference between urban and non-urban areas.

Keywords: *real estate, retail vacancy, neighborhood characteristics, property transformations*

COLOFON

Title	Retail Vacancy in the Neighborhood: the Association Between Retail Vacancy and Neighborhood Characteristics in the Netherlands
Version	Final
Author	M.A. (Martine) Verwijs
Student number	S4767888
E-mail	m.a.verwijs@student.rug.nl
University	University of Groningen Faculty of Spatial Sciences Master of Science in Real Estate Studies Landleven 1, 9747 AD, Groningen
Supervisor	Dr. M. (Mark) van Duijn
Reviewer	Prof. Dr. Ir. A.J. (Arno) van der Vlist
Supervisor	Dhr. M. (Matthieu) Zuidema
Date	November 17, 2022

Disclaimer: *'Master theses are preliminary materials to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the author and do not indicate concurrence by the supervisor or research staff.'*

PREFACE

Over the past year I have enriched my financial, economic and geographic knowledge of real estate markets by following courses and actively participating in lectures on the MSc Real Estate Studies program at the Rijksuniversiteit Groningen (RUG). This thesis marks the end of this master's program. I thank my thesis supervisor, Mark van Duijn, for considering ways to improve this study and giving constructive feedback. He helped me complete this thesis and gave valuable suggestions.

The study was combined with a research internship at Kadaster. I thank Kadaster for giving me the chance to undertake my master's thesis in the research department and trusting me to work with the data. I want to show my deepest appreciation to my supervisor, Matthieu Zuidema, who guided me throughout this project and patiently answered all my questions. He helped me to better understand the topic and the underlying relationships. I also thank my other colleagues at Kadaster for helping me with the complex data processing. I extend my thanks to the people working at the CBS, especially at Microdata Services, for offering valuable data for this project.

Finally, I thank my family and friends for their unconditional support during the process. You kept me motivated, and I could not have achieved this without your support. I hope this thesis is interesting and educational and will add value to studies on the association between the neighborhood's retail vacancy and neighborhood characteristics.

Martine Verwijs

Groningen, November 17, 2022

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Motivation

Retail vacancy is a topical issue, especially when it is structural and geographically concentrated. Structural and geographically concentrated retail vacancy is not only an issue for the owners of such vacant premises but could be a societal issue as well as it causes a blister in the neighborhood.[1] The retail sector is having a difficult time after the worldwide rise in e-commerce and the COVID-19 pandemic. One may expect that retail vacancies have spiked due to these phenomena. Figure 1 presents the trend in retail vacancy rates in the Netherlands. When the COVID-19 pandemic started in 2020, vacancy rates started to increase. More bankruptcies in retail and fashion stores with permanent store closures were observed (Thompson, 2020).

However, due to state support, a remarkable number of Dutch retail companies survived the COVID-19 pandemic. Retail analyst Ward van der Stee states that “there is no trend break in retail vacancy because it is likely that the decrease of the number of stores continues now that the lockdowns are over and the state support for entrepreneurs will stop” (Stil, 2022). This is supported by the retail vacancy figures for 2021-2022 which decreased again and returned to pre-COVID19 figures (from 7.5% on January 1, 2021 to 6.8% on January 1, 2022).

Figure 1 *Retail vacancy in the Netherlands on January 1 from 2015-2021, measured by number of retail objects and surface in m2 of total retail supply (CBS, 2021a).*

A different perspective on the decline in vacancy rates in 2021-2022 is that almost 2,000 retail premises (2.39% of total) were withdrawn from the retail stock. The majority of these properties are transformed to housing accommodations (Slob, 2022). Dynamis (2019) state in their report that transformations of retail real estate to housing accommodations can be a solution to reduce retail vacancy and the housing shortage. In the

[1] It is well-known that structural and geographically concentrated vacancy has negative impacts on the neighborhood in terms of liveability and criminality (Remøy & van der Voordt, 2007; Ouwehand, 2018; Wang & Immergluck, 2018).

Netherlands around a million square meters of retail real estate is eligible for transformation into housing, according to RaboResearch (Aalders et al., 2021). The current measures to prevent and solve retail vacancy include plans of the government in cooperation with housing corporations, project developers, construction companies, and private investors to transform vacant retail premises.

1.2 Academic relevance

Multiple scholars have studied retail vacancy. Mallach (2018) provides an improved understanding of vacancy by stating that vacancy is a symptom of other problems, such as economic decline, market failure, and concentrated poverty. Benjamin et al. (1998) indicate that retail vacancy is a result of the difference between retail supply and demand as demand that does not meet the supply leads to vacancy. Earlier literature also studied the factors influencing (retail) vacancy. Park and Talen (2021) state that structural transformation of the retail industry, demographic change, and the increased cost of being a retailer are common explanations for the vacancy in retail.

The extent of retail vacancy differs between neighborhoods in the Netherlands due to economical and demographical developments. The location aspects of the retail premises are important (van Zweeden, 2009). For example, Jongsma (2020) finds that city centers in a shrinking region compared to a growing region is an important determinant for excessive retail vacancy. Remøy et al. (2007) identified object characteristics (such as the building period (year of completion), size, status, height, surface of the property in square meters) and location characteristics (such as the accessibility by car, parking facilities, proximity to city center) that impact vacancy. The population density and urbanity level seem to have an impact on vacancy as well (Hazelaar, 2019). The CBS (Centraal Bureau voor de Statistiek; Main Bureau of Statistics in English) identifies more neighborhood characteristics, such as the income, property values and, educational level (CBS, 2021b).

Few studies examined the association between retail vacancy rates and neighborhood characteristics and especially realized property transformations[2] to housing in the Netherlands. The aim of this study is to examine the association between retail vacancy on the neighborhood level and neighborhood characteristics, for which special attention has been paid to realized property transformations.

1.3 Research problem statement

The assumption is that the relationship is not unequivocal and strongly depends on the neighborhood characteristics. The central research question is as follows:

"*What is the association between neighborhood's retail vacancy rates and specific neighborhood characteristics in the Netherlands?"*

[2] Transformations can be operationalized as the reuse of existing premises where the function of the property is (partly) converted from a non-residential function to a residential function (CBS, 2021c). For this thesis, this includes the number of all transformed properties in the neighborhood, thus retail, but also offices, educational buildings, and commercial spaces to residential functions, measured between January 1, 2016 and January 1, 2021.

Multiple variables can influence retail vacancy, so the relationship between the dependent and independent variables is assumed to be probabilistic. Control variables are considered, which are held constant during the research so they do not impact the results. These control variables are the housing characteristics (property values and construction year). Figure 2 graphically shows the relationship in the conceptual model.

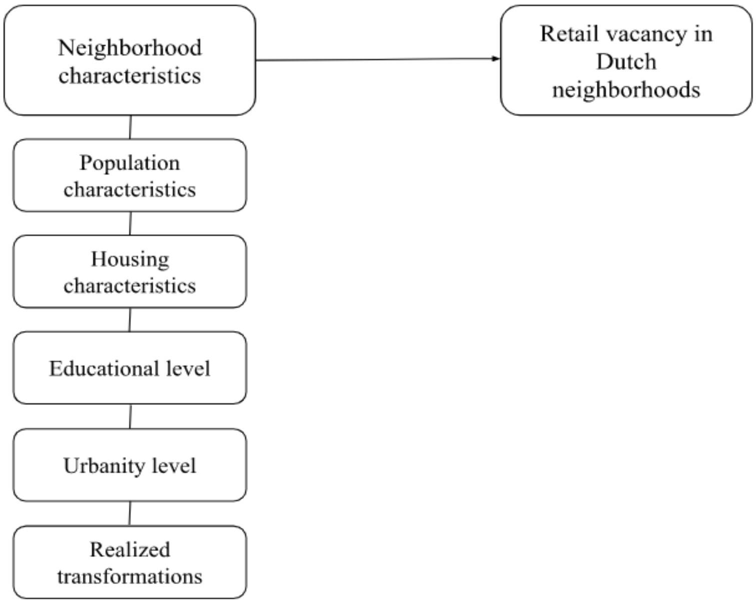

Figure 2 *Conceptual framework.*

This research studies the cross-sectional variation in vacancy across neighborhoods in the Netherlands. The first (theoretical) research sub-question is: *"Which factors influence retail vacancy according to the academic literature?".* Current literature and theories on this subject are studied to answer this question.

The second sub-question works toward the empirical model. This question is as follows: *"What is the empirical relationship between neighborhood's retail vacancy and specific characteristics of the neighborhood, while taking realized property transformations into account?".* The answer to this question adds meaning to the conceptual model by indicating the type of relationship between the two variables. Answering this question requires data on retail vacancy and transformed premises. During the author's research internship at Kadaster, register data was collected about retail premises in combination with CBS register data about retail vacancy, realized property transformations to housing and further information about the neighborhood.

The third sub-question concerns testing the robustness of the found relationship between a neighborhood's retail vacancy and neighborhood characteristics, especially realized property transformations to housing. Therefore, this third sub-question is as follows: *"How does the association between retail vacancy and neighborhood characteristics, especially realized property transformations, change depending on the urbanity level?".* For this question, register data on urbanity levels from the KWB (Kerncijfers Wijken en Buurten; key figures for districts and neighborhoods in English) is used to test whether a difference exists between different urbanity levels in the Netherlands (CBS, 2021b). Urbanity levels are classified into five categories: very urban, highly urban, moderately urban, minimally urban, and non-urban (CBS, 2019). For this analysis, a dummy variable was created for the urbanity level variable. The first two categories were combined for the dummy variable, being 1 if urban and the last three categories were combined, being 0 if non-urban. Answering these sub-questions enables the main research question to be answered.

1.4 Outline

The remainder of this paper is structured as follows. Chapter 2 provides a further explanation of the existing theory and a literature review on the factors and causes of retail vacancy. Moreover, the first sub-question is answered in this section. Chapter 3 gives a further explanation of the data collection and methods used. Chapter 4 presents the results and discusses these based on existing academic literature. Furthermore, this chapter describes the study’s limitations, and provides suggestions for future research. Chapter 5 concludes based on the theoretical and empirical findings and answers the main research question. Additionally, this chapter discusses policy implications.

2. THEORY, LITERATURE REVIEW AND HYPOTHESES

2.1 Vacancy

Van der Voordt (2007) gives a straight-forward definition of vacancy: "the non-let situation of an available real estate property". Rabianski (2002) defines vacancy in a property market as "excess supply at a given price in that property market". The CBS (Centraal Bureau voor de Statistiek; Main Bureau of Statistics in English) defines administrative vacancy as the concept that exists when there is:

- no occupation according to the BRP (Basisregistratie Personen; key registration of people in English);
- no business activity according to the HR (Handelsregister; commercial register in English);
- no (fiscal) use according to the registration of the WOZ (Waardering Onroerende Zaken; property valuation in English)(CBS, 2021a).

Keeris (2006) and van der Voordt (2007) distinguished similar classes of vacancy based on the economic effect. The category which has a low impact on the economy (less-harmful) is indicated as "normal" or "accepted" vacancy. It has three sub-groups:

- initial vacancy (occurs after a building is just completed);
- mutation vacancy (occurs between two different retailers in a location);
- frictional vacancy (vacant for less than one year).

The other category (harmful vacancy) largely affects the economic performance and is indicated as "problematic" or "excess" vacancy, with three sub-groups as well:

- structural vacancy (vacant for longer than two years);
- functional vacancy (caused by a less desirable location);
- technical vacancy (buildings do not meet the general market requirement) (van Zweeden, 2009; Keeris, 2006; van der Voordt, 2007).

The four quadrant diagram of DiPasquale and Wheaton (1992) provides insights into the market for real estate space and real estate assets. Figure 3 shows that the four quadrants are a representation of the real estate market and gives an explanation for the existence of vacancy. The focus will be on retail vacancy.

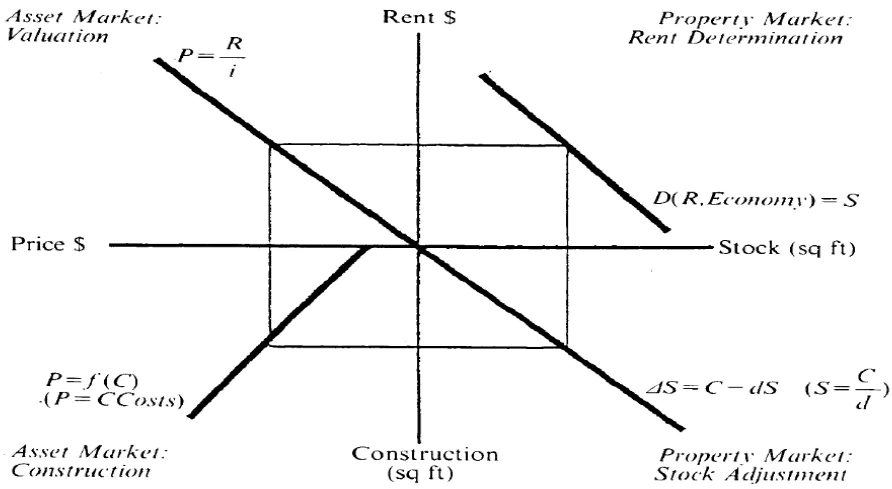

Figure 3: *Four quadrant model of the property and asset market (DiPasquale & Wheaton, 1992).*

For the retail market, the first (northeast; NE) quadrant represents the retail property market for the user of space and determines the rents in $ per unit of space. In equilibrium, the demand for space, D, is equal to the stock of space, S (in square feet). Given the stock, the rent, R, must be determined so that demand (rent and conditions in the economy) is equal to the stock (DiPasquale & Wheaton, 1992). The demand side is influenced by retailers and their location preference. Factors influencing the retail demand include the retail rents, retail sales, low retail vacancy, constraints on future development, population growth, and aggregate disposable income of locals (Benjamin et al., 1998).

The supply side of retail space is determined by retail real estate developers, as they decide to increase or decrease store space. The retail supply is influenced by the economic climate, land availability, capital market cycles, interest rates, tax law, demographic trends, sociological trends, and local conditions. The imbalance of the demand for retail space relative to the supply of retail space leads to retail vacancy (van Zweeden, 2009). This is supported by Pershio (1991) and Roulac (1994).

The NW (northwest) quadrant represents the capitalization rate for real estate assets: the ratio of rent (y-axis) to price (x-axis). It represents the current yield that investors demand to hold real estate assets; the rent determined in the previous quadrant is translated into retail property prices by investors (DiPasquale & Wheaton, 1992). It implicates that the investment value is directly related to the rental rates. The initial return has a positive relationship with the rents: if the capital market rent declines, the initial return should decline as well. This results in a declining line in the quadrant. The same rental rates results in a higher investment value (Besselaar, 2011).

The third SW (southwest) quadrant determines the construction of new assets. The curve represents the replacement cost of real estate. The construction cost is assumed to increase with greater building activity, thus the curve moves in a south-westerly direction. Given the price of real estate assets in the previous quadrant, new construction is generated. A line down to this replacement cost curve and over to the vertical axis determines the new level of construction where replacement costs equal asset prices (DiPasquale & Wheaton, 1992).

The last SE (southeast) quadrant is the one that represents the impact of the previous movements. The annual flow of new construction (determined in the third quadrant) is converted into a long-run stock of real estate space. The change in stock is equal to new construction minus losses from the stock measured by the depreciation rate (DiPasquale & Wheaton, 1992).

The retail property and asset markets are in equilibrium when the starting level of the stock (in NE) are the same as the ending levels of the stock (in SE). Figure 4 shows that an increase in demand for real estate due to market tightness results in a higher rental price and investment value. This leads to an increase in construction. The stock adjustment takes place in the long run due to the longer building period, so the stock stays behind at first. A decline in demand during the building period leads to an imbalance in the market with an oversupply of real estate and thus vacancy in the retail property market (Besselaar, 2011).

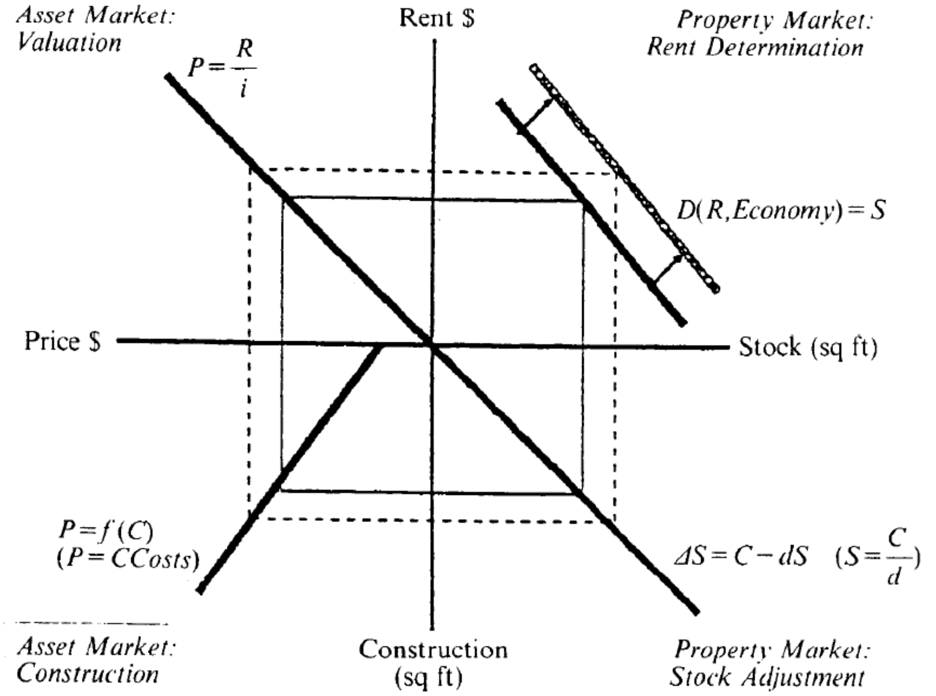

Figure 4: *Property demand shifts in the four quadrant model of the property and asset market (DiPasquale & Wheaton, 1992).*

DiPasquale and Wheaton (1992) do not include vacancy in their model, as they expect the market itself to restore the equilibrium over time. It is assumed that the price will decline due to the oversupply, which makes the developing of new real estate less attractive and the demand will grow. According to the four quadrant model, the effect is that the supply will decline and the market equilibrium will thus be restored (DiPasquale & Wheaton, 1992).

Colwell (2002) added vacancies to the four quadrant model. Natural vacancy can be defined as the propensity for tenants to move out of units and the time involved in filling vacated units associated with search and maintenance processes. Natural vacancy is developed in the SE quadrant and utilized in the NE quadrant. Figure 5 shows the determination of vacancy. Natural vacancy, vS1, is found by extending a 45 degrees line from the vacancy function at the current stock to the horizontal axis.

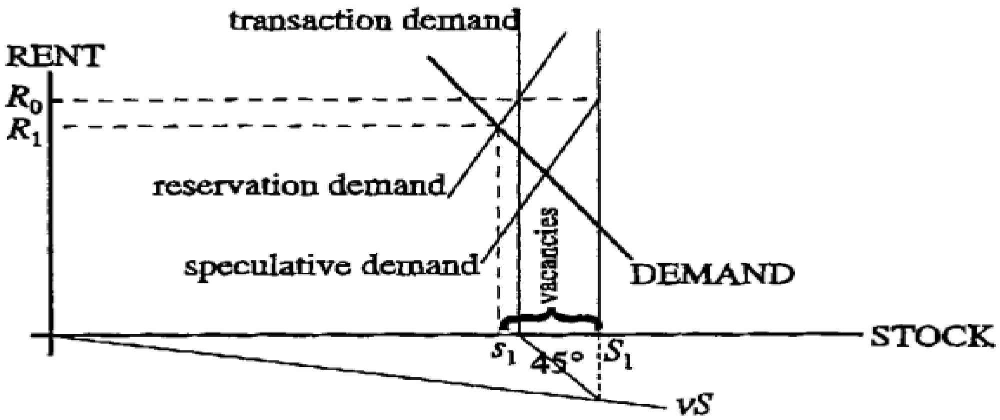

Figure 5: *Determining vacancies in the four quadrant model (Colwell, 2002).*

2.2 Determinants for vacancy

Writers have provided plausible theoretical explanations for retail vacancy. According to Brouwer and Tool (2018), a relationship exists between retail vacancy and the diversity of stores in a shopping area. This effect seems to be highest when measured on diversity in sectors. Retail vacancy is partly determined by an historical city center, filialisation rates, disposable income, Gross Regional Product, unemployment rates, ageing, and internet sensitivity (Brouwer & Tool, 2018).

The study implicates the necessity of a diverse retail supply to keep a shopping area alive. The more diversity, the less vacancy (Bouwer & Tool, 2018). This discovery is consistent with that of Elms and Teller (2010), who found that shopping areas with diverse stores are attractive to consumers.

Furthermore, research showed that anchor stores generate positive externalities compared to non-anchor stores by drawing customers to the shopping area, whereas many other stores benefit from this external mall traffic (Gould et al., 2005). Konishi and Sandfort (2003) define an anchor store as “a store that increases the traffic of shoppers at or near its location, through its name’s reputation”. A lack of anchor stores in a mall can affect a store’s sales and thus affect the vacancy rates.

In his policy focus report, Mallach (2018) provides an improved understanding of vacancy in general and ways to reduce (hyper)vacancy in the US. According to Mallach, vacancy is a symptom of other problems, such as economic decline, market failure, and concentrated poverty. He distinguishes seven categories of vacant housing units: vacant for rent, rented but not yet occupied, vacant for sale, sold but not yet occupied, maintained for seasonal/recreational/occasional use, maintained for migrant workers, and other vacant. Except for “other vacant”, all of the vacancies serve necessary functions because people switch houses regularly (Mallach, 2018).

According to Mallach, a difference is observed between vacant buildings in cities with a strong real estate market and those with a weak market. Vacant buildings in a city with a strong real estate market seldom stay vacant for long. They are mainly bought by someone who restructures them. The paper claims that, consequently, the market in cities that had serious problems with vacant premises in the past has improved significantly (Mallach, 2018).

Where Brouwer and Tool (2018), Gould et al. (2005), and Mallach (2018) discussed determinants for vacancy that are important explanations for retail vacancy, Park and Talen (2021) argue that structural transformation of the retail industry, demographic change, and the increased cost of being a retailer are common explanations for the vacancy in retail. Structural transformation concerns long-term changes in the retail industry, over which no control is possible. These changes relate to consumer behavior, government actions, and technological innovations (increase in e-commerce; Park & Talen, 2021).

Furthermore, demographic change concerns high versus low-income classes and consumption in certain areas (Bieniek et al., 2018; Kickert, 2021). Park and Talen (2021) contribute to research by stating that “the costs of owning and operating a small business are likely to contribute to retail vacancy". Particularly, small retailers have limits due to high rents, high taxes, minimum-wage laws, and many other additional costs (Park & Talen, 2021).

Remøy et al. (2007) identified office building characteristics (such as the building period, size, status, height, property price) and location characteristics (such as the accessibility by car, parking facilities, proximity to city center) that impact vacancy for offices within the region of Amsterdam. Hoekstra and Vakili-Zad (2011) supports that the building period impacts vacancy in Spain by stating that older buildings are more likely to be vacant than newer buildings. Yakubu et al. (2017) state as well that older buildings have higher vacancy rates. Furthermore, Ball (2002) supports that building size is a determinant for vacancy in the United Kingdom by stating that smaller buildings are more often vacant than larger ones and more suitable for redevelopment. Johnson et al. (2007) state that smaller vacancy rates are associated with lower property prices due to the speculative build-up in strong markets, which is also supported by Buttimer and Ott (2007). Retail vacancy in areas with lower property values is higher compared to areas with higher property values.

The population density and urbanity level seem to have an impact on vacancy in the Netherlands as well (Hazelaar, 2019). The more addresses in the surroundings of the property, the smaller the potential for vacancy. On the other hand, a decline in the number of citizens is associated with an increase in retail vacancy (Hollander et al., 2018). Also, the more urban the surroundings are, the more potential for vacancy and transformations of the premises. Oskam (2021) supports this finding by stating that retail stores are less likely to be vacant in lower urbanity areas (Level 5) compared to higher urbanity areas (Level 1), regardless of the number of realized property transformations in an area.

2.3 Property transformations

Transformations are defined as changes in function, from retail and other functions to residential accommodations. Considering the four quadrant model of DiPasquale and Wheaton, the retail supply is lowered due to property transformations. A decrease in the retail supply/stock results in a higher rental price. Investors get a higher return and are willing to pay more for new construction. This results in new construction and an increase in the retail supply.

Remøy and van der Voordt (2007) examined the opportunities, threats, and risks of converting vacant office buildings into housing. According to this article, vacant office buildings lead to financial problems for the owners of the buildings and social problems for the community in the building's neighborhoods (due to problems concerning insecurity and potential criminality). Vacancy also gives the surrounding area and buildings a negative image (Remøy & van der Voordt, 2007).

In addition to the opportunities, threats, and risks, the article distinguishes critical success factors. Overall, Remøy and van der Voordt discussed financial, aesthetic, technical, structural, and functional issues in this transformation process. The research questions studied in this paper concern why building conversion is a suitable option for addressing high vacancy levels, which buildings are suitable for conversion, and whether converting mediocre elements of the building stock makes sense. Building conversion is a suitable option for addressing vacancy because it is sustainable, building materials can be used again and the structure of an urban area is retained. Transforming vacant premises increases the historical value of a place. Additionally, transforming vacant premises saves construction time.

The transformations meter of Geraedts and van der Voordt (2004) is used to decide which buildings are suitable for transforming into housing. This meter comprises criteria to measure opportunities and risks. The article concludes that the short period during which a transformation can occur is a significant advantage. However, the market can be at risk from procedures in municipalities, which can take a long time, with the risk that the time gained through conversion can be lost (Remøy & van der Voordt, 2007).

Ossokina et al. (2017) found that property transformations are likely to be initiated by the real estate market. The paper states that "a location has transformation potential if it is more profitable for other land use than the current function and the difference between the residential and current function rent needs to be positive and larger than the expected cost of transformation". The article states that property transformations are often found at the edge of a shopping area where premises appear unprofitable for retail use due to a negative demand shock. These are the places where functions can easily be changed due to the multifunctional character of these premises. Transformations of functions is less likely to be done in declining cities and regions. These are the areas where premises are likely to remain vacant for a longer period of time. Ossokina et al. (2017) indicate these places as those where transforming would be a possible solution to vacancy.

Additionally, Ruigrok (2021) indicates that realizing property transformations to housing can be viewed as a solution to retail vacancy; however, this differs by location type. Transformed retail stores in shopping streets may not always lead to suitable housing due to the building dimensions and character of the public space. These are often unsuitable for a residential function. Households often prefer not to mix their housing accommodation with the public character of a shopping street. This may result in more vacancies in a shopping street. This preference is even stronger in planned shopping centers. Conversely, mixed shopping streets (for example, the streets leading to the city center) may provide interesting opportunities for transforming retail stores into housing accommodation (Ruigrok, 2021)

2.4 Hypotheses

Hypotheses are formulated to predict outcomes based on existing literature. This research studies the relationship between the neighborhood's retail vacancy rates and specific characteristics of the neighborhood. The assumption is that the relationship strongly depends on the population- and housing characteristics, i.e. the number of citizens, population density, construction years, average property values, and neighborhood characteristics, such as the number of highly educated people, and urbanity levels. It is expected that the retail vacancy rates will be higher in some neighborhoods compared to others. Table 1 provides an overview of the theoretical hypotheses based on the neighborhood's retail vacancy rate and neighborhood characteristics.

Table 1 *Hypotheses based on neighborhood's retail vacancy rate.*

Variable	Expected sign
Number of citizens	-
Population density	-
Average housing value	-
Construction before 2000	+
Urbanity level	+
Realized property transformations	+

To test the heterogeneity of the relationship in the first hypothesis, this research investigates whether the relationship differs between urbanity levels and the economic perspectives of areas. For urbanity levels, a distinction is made between urban and non-urban areas. The second hypothesis is specified as follows:

H2: The association between retail vacancy in a neighborhood and realized property transformations to housing is stronger in urban areas compared to non-urban areas.

3. DATA AND METHODS

This section describes the context and the collection of the data. The method of data linking and processing used for this study is explained, and the operationalization of variables used in this analysis is provided. Furthermore, an overview of the descriptive statistics is provided. The methodology is also explained in this chapter.

3.1 Data context

Table 2 shows the development of retail vacancy over time by providing the total number of retail objects and the total surface of retail supply from 2015-2021. The total number of retail objects declined over the past years, from 130,400 retail objects in 2015 to 128,570 objects in 2021. According to Slob (2022), this can be explained by the trend of transforming retail objects to housing accommodations. The number of vacant retail premises increased significantly in 2019 and decreased again in 2021, also due to the state support for retail owners during the COVID-19 pandemic and the decrease in the availability of retail objects (Slob, 2022; Stil, 2022).

Table 2 *The total number of retail objects and total surface of retail supply in the Netherlands from 2015-2021 (CBS, 2021a).*

	Total number of retail objects	**Number of vacant retail premises**	**Total retail supply in m^2**	**Surface of vacant retail premises**
2015	130,400	9,600	48,682,330	2,765,000
2016	130,300	9,610	48,122,680	2,765,770
2017	130,270	10,580	48,816,770	3,254,540
2018	130,130	10,230	47,803,940	2,815,500
2019	129,670	9,940	47,433,670	2,629,120
2020	129,200	11,040	47,220,700	2,922,820
2021	128,570	10,740	47,124,550	2,827,910

Retail premises are most common in shopping streets in larger cities, such as the Kalverstraat in Amsterdam, the Koopgoot in Rotterdam, and the Herestraat in Groningen. However, retail objects can also exist in smaller villages. The register data contains only the neighborhoods with ten or more retail objects due to privacy considerations. 2899 neighborhoods of the 8083 neighborhoods have ten or more retail stores. The neighborhood with the most retail stores (625) is the centre of Nijmegen (province of Gelderland, GD). A possible explanation why Nijmegen has the most retail stores in a neighborhood compared to Amsterdam for example, is that larger cities are divided into more and smaller neighborhoods and thus have smaller numbers of retail stores per neighborhood.

Figure 6 shows a map of the retail stores per neighborhood in the Netherlands. Figure 7 shows a zoom of the retail stores in neighborhoods in Rotterdam (province of Zuid-Holland, ZH). This shows that, for Rotterdam, most of the retail stores exist, as expected, in the city centre.

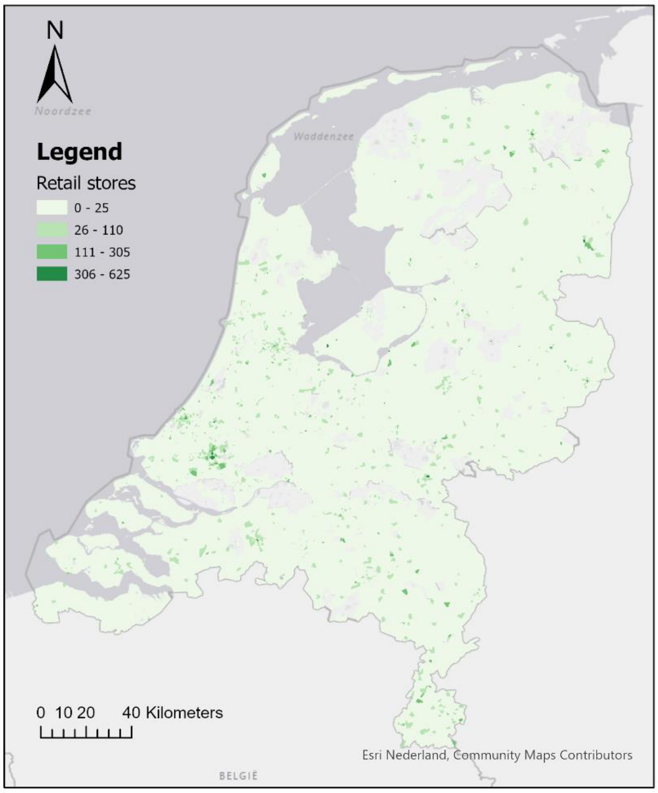

Figure 6: *Retail stores in the Netherlands per neighborhood (Kadaster, 2021).*

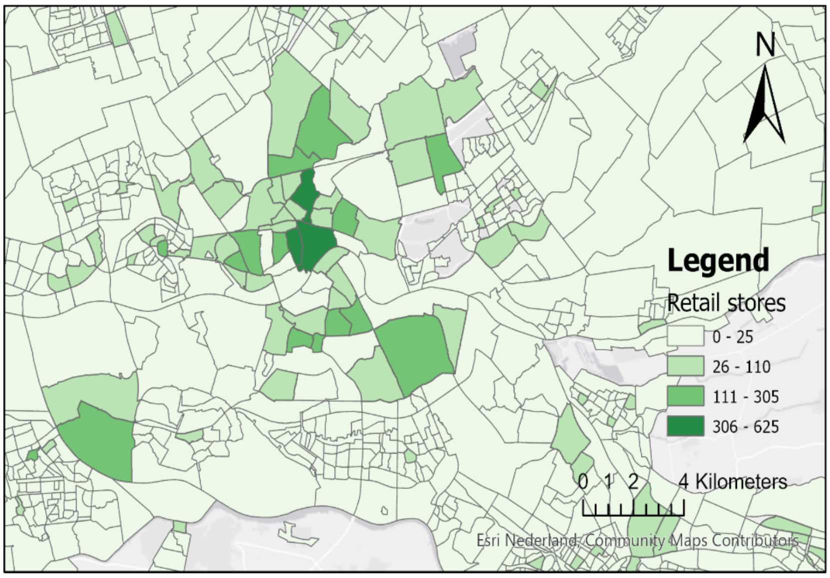

Figure 7: *Retail stores per neighborhood in Rotterdam (ZH), the Netherlands (Kadaster, 2021).*

3.2 Data collection and operationalization of variables

The data originates from different sources. Most of the datasets are derived from registered data from Kadaster and the CBS.

Kadaster

Kadaster is a public register maintained by the Dutch government. For this thesis, Kadaster made register data available for retail objects. This data is obtained from the BAG (Basisregistratie Adressen en Gebouwen; Fundamental Regression Addresses and Buildings in English). BAG data from Kadaster comprises key information on all the addresses and buildings in a municipality (Kadaster, 2021). With this information, the number of retail objects per neighborhood was included in the data. During the author's research internship at

Kadaster, the company facilitated a collaboration with the CBS (Centraal Bureau voor de Statistiek; Main Bureau of Statistics in English) to get register data on retail vacancy and realized property transformations.

Centraal Bureau voor de Statistiek (CBS)

The CBS is a statistics bureau in the Netherlands providing insights into societal issues with reliable statistical information and register data. The KWB (Kerncijfers Wijken en Buurten; key figures for districts and neighborhoods in English) dataset includes key figures and demographic information for each district and neighborhood in the Netherlands (CBS, 2021b). This research uses population data to measure the number of citizens, population density (number of citizens per km^2), average housing value (in € × 1,000), percentage of retail premises constructed before 2000, and the number of highly educated people in a neighborhood. Furthermore, the urbanity level is used as a possible explanatory variable in the study, distinguishing urban and non-urban areas.

As mentioned before, a collaboration was started with Kadaster and the CBS to access CBS register data. The dataset of Kadaster can be enriched with specific CBS information, such as historical vacancy of retail and realized property transformations to housing in an area. The national vacancy monitor provides insights into the administrative vacancy of houses for all municipalities, districts, and neighborhoods in the Netherlands, and for all municipalities, it provides insights into the administrative vacancy of non-houses on January 1, 2021 (CBS, 2021a).

The variable vacancy is used as the dependent variable in this research. The vacancy dataset comprises two vacancy variables: premises that are vacant now (measured on January 1, 2021) and premises that were already vacant in an earlier year. It was decided not to include the latter variable in the model because this would result in a time series vector autoregression model. The model would then become too extensive for this research. The micro data on retail vacancy is aggregated on the neighborhood level, so that both the dependent and independent variables are measured on the same spatial scale. For this research, the percentage of retail vacancy in a neighborhood is measured because the total number of retail stores per neighborhood is available in the data and percentages make it easier to see relationships. The percentage of retail vacancy per neighborhood is calculated by dividing the retail vacancy numbers by the retail stores in the neighborhood. Hence, the dependent variable vacancy is a continuous variable, indicating the percentage of retail vacancy on the neighborhood level.

CBS also provides registered data of realized property transformations to housing per neighborhood. Transformations can be operationalized as the reuse of existing premises where the function of the property is (partly) converted from a non-residential function to a residential function (CBS, 2021c). The statistics are based on changes in the functions of objects in the BAG. The transformation data comprises the total numbers of realized property transformations between January 1, 2016 and January 1, 2021. This includes the number of all transformed properties in the neighborhood, thus retail, but also offices, educational buildings, and commercial spaces to residential functions. For this variable, relative numbers were not available because there was no data on the total number of buildings in a neighborhood. Therefore, the variable of realized property transformations consists of absolute numbers of transformations in a neighborhood.

CBS only includes the neighborhoods in the data with ten or more records, both for vacancy and realized property transformations, due to privacy considerations. This is also the case for the variable of the number of retail objects from Kadaster. The “not-observed” (thus, less than ten records) realized property transformations, retail vacancy, and retail objects in a neighborhood are replaced by zero. If one does not make that assumption, one would lose many more neighborhoods. This data processing results in 327 neighborhoods in the Netherlands with retail vacancy rates measured on January 1, 2021, realized property transformations measured between January 1, 2016 and January 1, 2021, and other neighborhoods characteristics measured in 2020.

After merging all the data sources, the analysis data file comprises the demographic information per neighborhood, enriched with information regarding the vacancy situation and realized property transformations in the neighborhood. The complete table with the operationalization of all variables included in the analysis is provided in Appendix A, Table 7, as well as a stepwise summary description of the data management in Appendix A, Table 8.

3.3 Descriptive statistics

As follows from the research question in Chapter 1, this research examines the relationship between the dependent variable retail vacancy and neighborhood characteristics. From the CBS register data, the continuous variable “vacancy” is used in combination with retail information as the dependent variable in this study.

The CBS dataset is cleaned and narrowed down to answer the research question. The stepwise summary plan of the data management is provided in Appendix A, Table 8. Unnecessary variables are excluded from the dataset, resulting in neighborhoods in the Netherlands with retail vacancy rates, demographic information, and realized property transformations. 2,414 missing variables were deleted from the dataset. After cleaning, 11,394 observations remain in the dataset.

A correlation matrix is made with the variables included in the analysis, Appendix B, Table 9. Brooks and Tsolacos (2010) provide an explanation for such a matrix. Correlation is measured on a scale from -1 to 0 or 0 to 1, with -1 being a strong negative correlation between variables, and +1 being a strong positive correlation between variables (Brooks & Tsolacos, 2010). From the matrix, it appears that the number of highly educated people in a neighborhood is highly correlated with the number of citizens (0.8486). This can logically be explained, as the chance of highly educated people increases with more people living in a neighborhood. For this reason, the variable of the number of highly educated people is excluded for this analysis.

Furthermore, the variables number of citizens and population density are divided by 1,000 in the dataset. This is done to make better interpretations. An increase in 1 citizen in a neighborhood will most likely not have any significant effect on the retail vacancy in the neighborhood. However, an increase of 1000 citizens will most likely have a significant effect on the retail vacancy. The variable realized property transformations is divided by 100 in the data to see the effect more clearly.

Table 3 shows the descriptive statistics for the variables in the analysis.

Table 3 *Descriptive statistics for the variables in the analysis.*

Variable	Freq.	Percent	Cum.	Mean	St.dev.	Min	Max
Retail vacancy (in %)				.430	3.041	0	75
Neighborhood characteristics							
Population characteristics							
Number of citizens (× 1,000)				1.485	1.748	.035	28.87
Population density (number of citizens per km^2 × 1,000)				3.477	4.045	.002	35.53
Housing characteristics							
Average housing value (in € × 1,000)				313.775	138.643	48	2,065
Construction before 2000 (in %)				82.938	22.848	0	100
Urbanity level							
Urbanity level							
Non-urban	7,754	68.05	68.05				
Urban	3,640	31.95	100.00				
Realized property transformations							
Number of property transformations (× 100)				.0387	.222	0	8.60

Note: N = 11,394. Retail vacancy is the dependent variable. Number of property transformations consists of transformations from all functions to housing, measured between January 1, 2016 and January 1, 2021.

Furthermore, the data was checked for skewness by producing histograms of all independent variables. Some independent variables that were not normally distributed were log-transformed to manage skewed data. This stabilizes the variance and makes the data more normally distributed (Fang & Lütkepohl, 2012). The variable percentages of retail vacancy is skewed so the variable should be transformed into a natural logarithm. However, the numbers are too small and the variable is in percentages. The natural log is mainly for absolute numbers, thus this means that the variables should not be log-transformed. The variable for the building period is also skewed so it should be log-transformed to make a normal distribution. However, because the dependent variable retail vacancy in the neighborhood is in percentages no independent variable should be log-transformed. That would make the interpretation more difficult. It would result in an interpretation of percentages of percentage points.

For the analysis, a dummy variable was created for the urbanity level variable, which comprises five categories: very urban (1), highly urban (2), moderately urban (3), minimally urban (4), and non-urban (5).

The first two categories were combined for the dummy variable, which is 1 if urban and the last three categories were combined, being 0 if non-urban.

To examine the dependent variable retail vacancy, a GIS map is made with the data. Figure 8 shows the vacancy per neighborhood in Rotterdam. It shows that retail vacancy is highest in the city center.

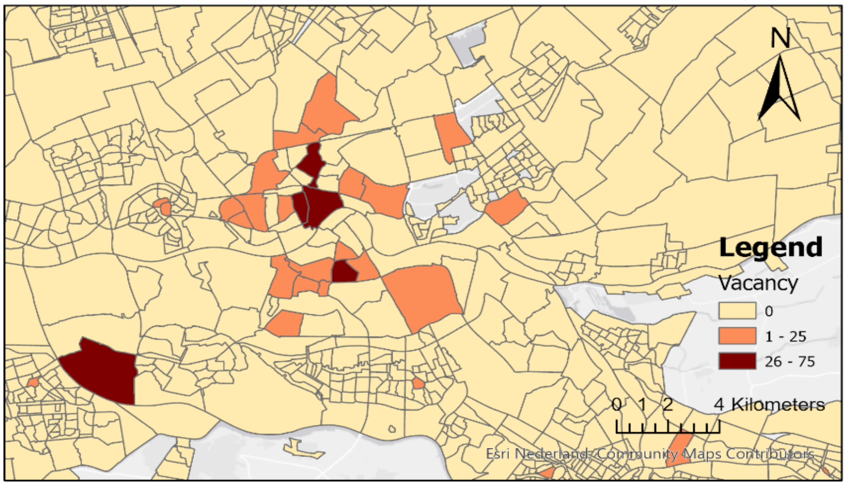

Figure 8: *Retail vacancy per neighborhood in Rotterdam (CBS, 2021a).*

For this research, the correlation with realized property transformations to housing is studied. To see the location of realized property transformations, a similar map is made in GIS. Figure 9 shows the property transformations per neighborhood in Rotterdam.

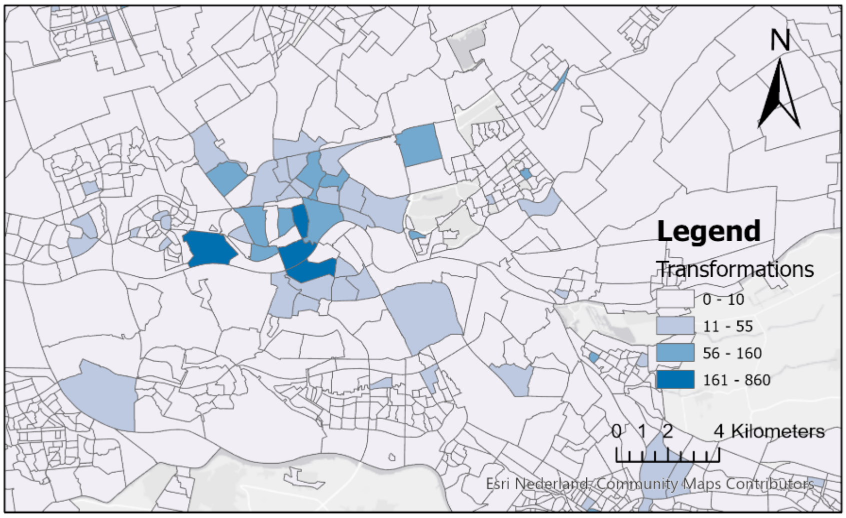

Figure 9: *Realized property transformations per neighborhood in Rotterdam (CBS, 2021c).*
All functions to housing from January 1, 2016 and January 1, 2021.

Comparing both maps for retail vacancy and realized property transformations in Rotterdam, some patterns can be seen in the two maps. The closer to the city centre, the higher the retail vacancy and realized property transformation numbers. It also shows that property transformations are realized in places with retail vacancy.

3.4 Methodology

The analysis comprises two models measuring the association between the retail vacancy and neighborhood characteristics. The first model includes the retail vacancy in a neighborhood and neighborhood characteristics. The second includes the transformation variable with the number of realized property transformations to housing in a neighborhood. This distinction between models was made to examine a significant difference in neighborhood characteristics (and realized property transformations) for neighborhoods with and without retail vacancy.

The dependent variable y is a continuous variable, indicating the percentage of retail vacancy on the neighborhood level. The independent variables are continuous as well and also measured on the neighborhood level. Due to the continuous dependent variable, a linear regression can be used. For this research, multiple explanatory variables are used to explain the dependent variable y. Therefore, the impact of the independent variables on the dependent variable was measured by using a multiple linear regression model.

In such a model, y is the response variable, β_0 is the intercept, β_1 is the slope, x is the predictor or regressor variable, and ϵ is the error term, which indicates the difference between the observed value of y and the straight line $(\beta_0 + \beta_1 x)$ (Montgomery et al., 2012). The multiple linear regression model results into the following regression equation, based on the variables in the analysis:

$$RV_j = \beta_0 + \beta_1 CI_{\mathrm{j}} + \beta_2 PD_{\mathrm{j}} + \beta_3 AHV_j + \beta_4 CB2000_j + \beta_5 UL_j + \beta_6 RPT_j + \epsilon_j \tag{1}$$

The dependent variable RV indicates the retail vacancy for each neighborhood j in percentages. CI denotes the number of citizens in an area j. PD indicates the population density, measured in number of citizens per km^2 for each neighborhood j. AHV denotes the average housing value in neighborhood j, measured in euros x 1000. CB2000 indicates the percentage of properties built before for each neighborhood j. UL denotes a dummy variable of the urbanity level for each neighborhood j, with either an urban area or a non-urban area. Finally, RPT denotes the realized property transformations to housing in neighborhood j.

The linear regression has some important assumptions that should be checked. The first assumption is that the independent variable should be independent, thus no problem of multicollinearity should exist. This could be tested by making a VIF (Variance Inflation Factor) table. A VIF score of greater than five indicates a multicollinearity problem and less than five means there is no problem of multicollinearity (Uyanik & Güler, 2013). The VIF table is presented in Appendix B, Table 10. None of the VIF scores is greater than five so no problem of multicollinearity exists.

The second assumption is about linearity between the outcome variable and the independent variables. This assumption can be checked by making a twoway scatterplot and fit plots to make a linear prediction (Uyanik & Güler, 2013). This graph is showed in Appendix B, Table 11. The graph shows a positive linear relationship so it fulfils the assumption of linearity.

The third assumption of multiple linear regression is about normality. The residuals in the models should be random or normally distributed. This assumption can be tested by conducting a Shaprio-Wilk test. If this test shows a result with a p-value of less than 5%, that means that the data is not normal. If the test shows a results of a p-value greater than 5%, the data is normally distributed (Choueiry, n.d.). The outcomes are presented in Appendix B, Table 12. It shows that the data is not normally distributed because in all cases it is less than 5% and expect for the urbanity level it is even less than 1%. This means that the assumption of normality can not be met.

A fourth assumption that must be met for linear regression is the constant error variance. Thus, the strength of the model cannot be more for some parts of the population and less for other parts. This is called the assumption of homoscedasticity. If this assumption is violated, it means that better estimates are available compared to this model (Chatterjee & Simonoff, 2013). The data should not be heteroskedastic but homogenous, which means that the variance of residual terms should be equal. This assumption is tested by Breusch Pagan heteroskedasticity test. If the p-value is less than the significance level or $p<0.05$, that means there is a problem of heteroskedasticity. If $p>0.05$ there is no problem of heteroskedasticity. The test is conducted and the outcomes are presented in Appendix B, Table 13. It shows a p-value of less than 0.05 so there is a problem of heteroskedasticity. This problem can be solved by running the regression model with robust standard errors.

Interpreting the outcomes for the multiple linear regression focuses on the change in one independent variable while keeping all other variables constant. A one unit change in the independent variable is associated with an increase of β_1 percentage points in the dependent variable (Kasza & Wolfe, 2013).

4. RESULTS AND DISCUSSION

This chapter describes and discusses the results of the conducted study in STATA with multiple linear regression using the existing academic literature.

4.1 Main results

The outcomes of the quantitative analysis of the multiple linear model are presented to test the hypotheses of this study. This provides insights into the relationship between the dependent variable retail vacancy and the independent variables. Table 4 presents the results of the regression for both models.

Table 4 *Results for multiple linear regression for the percentage of retail vacancy.*

	Model 1		**Model 2**	
Variable	**Coef.**	**SE**	**Coef.**	**SE**
Neighborhood characteristics				
Population characteristics				
Number of citizens (× 1,000)	.263***	.032	.230***	.033
Population density (number of citizens per km^2 × 1,000)	-.055***	.012	-.057***	.012
Housing characteristics				
Average housing value (in € × 1,000)	-.001***	.000	-.001***	.000
Construction before 2000 (in %)	-.002*	.001	-.002***	.001
Urbanity level				
Urban	.737***	.107	.616***	.106
(Reference category: Non-urban)				
Realized property transformations				
Number of property transformations (× 100)			1.841***	.412
Intercept	.473***	.124	.488***	.124
F-statistic	34.77***		30.78***	
R-squared	.039		.056	

*Note: Significance: *p < 0.1; **p < 0.05; ***p < 0.01. N = 11,394. Percentage of retail vacancy on the neighborhood level is the dependent variable. Number of property transformations consists of transformations from all functions to housing, measured between January 1, 2016 and January 1, 2021. SE = Robust Standard Error.*

The R-squared for model 1 is 3.90% and for model 2 is 5.60%. The R-squared is relatively low because not all relevant predictors are observed to explain retail vacancy on the neighborhood level (Neter et al., 1985). The F-statistic for both models is significant at the <0.01 level. Thus, the null hypothesis for the F-test, which is that there is no relationship between the independent and the dependent variables in the model, can be rejected (Moore & McCabe, 2006).

The first hypothesis states the effect between neighborhood characteristics and the retail vacancy on the neighborhood level. The first two columns of Table 5 are presented with the expected sign for each neighborhood characteristic. The regression statistics from Table 4 are included in Table 5 to compare the expected and actual signs of the relationships.

Table 5 *Hypotheses expected vs. actual signs based on neighborhood's retail vacancy rate.*

Variable	Expected sign	Actual sign
Number of citizens	-	+
Population density	-	-
Average housing value	-	-
Construction before 2000	+	-
Urbanity level	+	+
Property transformations	+	+

In both models, the association between retail vacancy on the neighborhood level and number of citizens in a neighborhood is significantly different from zero at the 1% significance level. The outcomes of model 1 (without realized property transformations) indicates that if the number of citizens changes with 1,000, it is associated with an increase in retail vacancy in the neighborhood of .263 percentage points, keeping all other variables constant. The outcomes of model 2 (with realized property transformations) indicates that if the number of citizens changes with 1,000, retail vacancy in the neighborhood increases with .230 percentage points, keeping all other variables constant. This is in contrast to the findings of Hollander et al. (2018) who state that a decline in the number of citizens is associated with an increase in retail vacancy.. The opposite relationship as expected can possibly be explained by the fact that a larger number of citizens is often found in cities, where also more retail stores exist and the more retail stores in a neighborhood, the higher retail vacancy rates.

The association between retail vacancy on the neighborhood level and population density in a neighborhood is significantly different from zero at the 1% significance level. Thus, an increase of 1,000 in population density is associated with a decrease of .055 percentage points in the neighborhood's retail vacancy rate, not taking realized property transformations into account. For model 2 (with realized property transformations), an increase of 1,000 in population density is associated with a decrease of .057 percentage points in the neighborhood's retail vacancy rate. Both findings are in line with the relationship as expected in the hypothesis and is supported by the study of Hazelaar (2019), who states that the more addresses in the surroundings of the property, the smaller the potential for vacancy.

The association between the neighborhood's retail vacancy rates and the average housing value in the neighborhood is also significantly different from zero on the 1% significance level, for both models. This indicates that an increase of €1,000 of the average housing value is associated with an decrease of retail vacancy in the neighborhood with .001 percentage points, which appears logical as one would expect that stores with higher property values are better maintained, have higher quality, and thus prevent vacancy. Johnson et al. (2007) studied the relationship between property prices and selling time, and, contrary to the findings of the present study, they state that smaller vacancy rates are associated with lower property prices. This seems counterintuitive but Buttimer and Ott (2007) explained this by pointing out that due to the speculative build-up in strong markets, both property price and vacancy are high.

Consequently, the association between retail vacancy on the neighborhood level and construction before 2000 is significantly different from zero on the 10% significance level in model 1 and on the 1% level in model 2. Both models show an influence of a house being constructed before or after the year 2000 on the retail vacancy rates in the neighborhood. An increase in the percentage of homes constructed before the year 2000 is associated with a decrease of a neighborhood's retail vacancy rates of .002 percentage points. This is in contradiction to the expected relationship, as one would logically expect that a higher percentage of homes constructed before the year 2000 would lead to higher retail vacancy rates. In contrast with this findings, Hoekstra and Vakili-Zad (2011) and Yakubu et al. (2017) studied the relationship as well and found that older buildings have significantly higher vacancy rates. A possible explanation for the relationship found in this research is that newer constructions for retail premises are also included in the dataset, and these affect the direction of the relationship. These new premises are vacant at completion and therefore count as vacant.

The association between the retail vacancy in a neighborhood and the urbanity level of the neighborhood is significantly different from zero at the 1% significance level for both models. Thus, urban areas compared to non-urban areas are associated with an increase in retail vacancy rates in the neighborhood of .737 percentage points. This is in line with the relationship as expected in the hypothesis and is supported by findings of Hazelaar (2019) and Oskam (2021) who state that retail stores are less likely to be vacant in lower urbanity areas (Level 5) compared to higher urbanity areas (Level 1),

The association between the neigborhood's retail vacancy rates and realized property transformations (from all functions to a housing function) is in both models significantly different from zero at the 1% significance level. This indicates that 100 property transformations in a neighborhood are associated with an increase of

1.841 percentage points for retail vacancy in the neighborhood. This relationship as expected is supported by Ruigrok (2021) who states that if an area gradually loses retail functions through property transformations and it becomes an increasingly strong residential location, it will be difficult for the remaining stores to continue, resulting in an increase of vacant retail stores.

4.2 Sensitivity analysis

In Chapter 2.4, the expectation was that the effect of realized property transformations to housing in a neighborhood on retail vacancy rates will not be the same whether it is an urban or non-urban area. A Chow F-test is conducted in order to check the differences in effect between the percentage of retail vacancy on the neighborhood level and realized property transformations (from all functions to a housing function) over the different urbanity levels. The outcomes are presented in Appendix C. The Chow F-test compares observations of separate groups and tests whether there are any differences in regression responses across these groups (Binkley & Young, 2020). For groups, the urbanity level of non-urban (ste_mvs=0) and urban areas (ste_mvs=1) is taken. First, the two separate restricted model regressions are runned (Appendix C, Table 14 and 15). Consequently, the pooled unrestricted model regression is runned (in Appendix C, Table 16).

It can be noted that the coefficients – particularly on the returns to realized property transformations per neighborhood – differ for the urbanity level sub-samples. The effect of realized property transformations on a neighborhood's retail vacancy rates turns out to be significantly different from zero at the 99% confidence level. The interpretation of the variable is that the variables are not log-transformed, so it is a growth effect. The interpretation is therefore than an increase of 100 in the number of property transformations per district, causes an increase of 1.974 in non-urban areas and 1.782 for urban areas. It is remarkable that the effect of number of property transformations in a neighbourhood on the retail vacancy in a neighborhood is positive for both urban and non-urban areas. However, the effect of realized property transformations on the retail vacancy in the neighborhood is larger in non-urban areas.

To test whether the differences are jointly significant, the Chow variant of the test of linear restrictions is used, by taking the F-value. For this research, filling in the formula gives:

$$F \;= \frac{(99{,}445.988-(25{,}871.656+73{,}327.873))\;/\;7}{(25{,}871.656+73{,}327.873))\;/\;11{,}394\,-\,2*7} \sim F\;[7, 11{,}394 \;-\; 2*7] \tag{2}$$

N_1 and N_2 are the numbers of observations in each group (in this case N1=7,754 and N2=3,640) and k is the number of parameters (in this case 7, i.e. 6 independent variables coefficients and the intercept). So the test statistic is F(7,11.380) = 4.039 and the 5% critical value F[7, ∞] = 2.0096. The null hypothesis that the coefficients are the same for the two sub-sets can be rejected on the 5% significance level.

However, the outcomes show no major differences, although the coefficients may be significant when controlling for other characteristics. An interaction variable urbanity level × property transformations dummy is created in a multiple linear regression (Table 6) to examine the influence of urbanity level on the association between retail vacancy on the neighborhood level and realized property transformations (from all functions to a housing function) in a neighborhood.

Table 6 *Multiple linear regression to test heterogeneity with interaction variable for transformations and urbanity level.*

Variable	**Coef.**	**SE**
Urbanity level		
Urban	.616***	.106
(Reference category: Non-urban)		
Number of property transformations (× 100)	1.855**	1.031
Urbanity level × Number of property transformations	-.015	1.115
Intercept	.488***	.124
F-statistic	96.55	
R-squared	.056	

*Note: Significance: *p < 0.1; **p < 0.05; ***p < 0.01. N = 11,394. Percentage of retail vacancy on the neighborhood level is the dependent variable. Number of property transformations consists of transformations from all functions to housing, measured between January 1, 2016 and January 1, 2021. SE = Robust Standard Error. A complete multiple linear regression is conducted, but only the output for the variables of interest is shown in this table.*

The R-squared for the model is 5.60%. The R-squared is relatively low because not all relevant predictors are observed to explain retail vacancy on the neighborhood level (Neter et al., 1985). The F-statistic is significant at the <0.01 level. Thus, the null hypothesis for the F-test, which is that there is no relationship between the independent and the dependent variables in the model, can be rejected (Moore & McCabe, 2006).

The outcomes of the regression show that the association between the neighborhood's retail vacancy rates and the interaction variable of urbanity level and number of property transformations is not significantly different from zero as p=0.989. Therefore, the model shows no influence of the urbanity level on the association between retail vacancy on the neighborhood level and realized property transformations (from all functions to a housing function) in a neighborhood.

5. CONCLUSION

This chapter answers the main research question. The relevant policy implications for this answer are discussed. Additionally, a recommendation is provided for future research on this topic.

5.1 Conclusion

This study investigates the association between retail vacancy on the neighborhood level and specific neighborhood characteristics, in particular realized property transformations to housing. The main research question of this research is, "*What is the association between neighborhood's retail vacancy rates and specific neighborhood characteristics in the Netherlands?"* For population characteristics (the number of citizens and the population density), the analysis finds significant influence on the neighborhood's retail vacancy rates. A higher number of citizens is associated with a higher percentage of retail vacancy in a neighborhood. On the contrary, a higher population density is associated with a lower retail vacancy. Furthermore, for the housing characteristics (average housing value and construction before 2000), the same direction is found. An increase in the average housing value and percentage of houses constructed before 2000 are both associated with a decrease in percentage in retail vacancy on the neighborhood level. The urbanity level also shows significant results on the retail vacancy rates in a neighborhood. When a neighborhood is in an urban area, retail vacancy rates are higher compared to neighborhoods in non-urban areas. Furthermore, the number of property transformations in a neighborhood show significant positive influence on a neighborhood's retail vacancy rates. However, an interaction variable with urbanity level × number of property transformations show no significant results, thus there seems to be no difference in the effect of realized property transformations in a neighborhood on retail vacancy rates in urban areas compared to non-urban areas.

5.2 Policy making advice

The introduction emphasizes the problem of high retail vacancy rates in the Netherlands. Chapter 2.2 explains the determinants for retail vacancy, but also the consequences in a neighborhood with retail vacancy. Ruizendaal (2018) studied the policy measures compiled by municipalities in the Netherlands to prevent retail vacancy. The policy measures are distinguished in four categories: 1) zoning expansion; 2) entrepreneurial support; 3) attracting visitors; and 4) visual interventions (Ruizendaal, 2018; Gelinck and Benraad, 2011). However, the study of Ruizendaal did not provide evidence for the effectiveness of these policy measures except for the zoning expansion. An expansion of a zoning plan appeared to have a positive effect on the decrease in retail vacancy.

Based on this thesis' research, some additional policy measures could be added to reduce retail vacancy in neighborhoods. This research shows that the neighborhood plays a role in retail vacancy rates. Since housing characteristics are important determinants for retail vacancy in the neighborhood, policies can be adopted to focus on building high-end properties with higher average housing values. The analysis finds that an increase in the average housing value is associated with a decrease in percentages in retail vacancy rates in the neighborhood. This will create more diverse neighborhoods which, in theory, might reduce vacancy. In addition, an increase in the percentage of constructed buildings before the year 2000 is associated with lower retail vacancy rates. Retail vacancy rates appear to be relatively low in older neighborhoods compared to older ones. This should be included in the policy making on the location of retail stores.

Furthermore, realized property transformations to housing accommodations in a neighborhood have a positive influence on the retail vacancy rates. This means that property transformations take place in neighborhoods with high retail vacancy rates. Policies can be adopted to stimulate the realisation of property transformations in such neighborhoods.

5.3 Limitations and future research

Follow-up research on the association between the neighborhood's retail vacancy and neighborhood characteristics is needed. According to Rabianski (2002), three types of vacancy can be distinguished: frictional, structural, and cyclical vacancy. Structural vacancy refers to unleasable places that will remain permanently vacant due to design flaws; frictional vacancy refers to easy movement in of users of space; and cyclical vacancy refers to spaces vacant due to a decline in demand due to economic and financial factors (Rabianski, 2002). This thesis does not distinguish between structural, frictional, and cyclical vacancy. Therefore, additional research is needed to examine the relationship between the types of retail vacancy, realized transformations, and neighborhood characteristics.

Furthermore, this research is limited to the vacancy of retail premises. However, the reasons for the vacancy are unknown. Future research could focus on reasons for vacancy and determine whether any differences exist between the reasons for vacancy and the relationship measured in this research. Also, the dataset for transformations is limited in the sense that it only consists of transformations from all functions to housing.

Moreover, the comparability of retail objects is limited because location fixed effects are not included in the analysis. Taking these location fixed effects into account could influence the results as retail vacancy in neighborhoods in Amsterdam are no longer compared to retail vacancy in neighborhoods in the countryside.

Additionally, this research shows that realized property transformations to housing accommodations in a neighborhood have a positive influence on the retail vacancy rates. This means that property transformations take place in neighborhoods with high retail vacancy rates. Future research is needed to examine the long-term effects of the transformations and to prove that property transformations as a neighborhood characteristic really can be a solution to decrease retail vacancy rates in the neighborhood.

Finally, this study examines the association between the likelihood of retail vacancy and realized transformations in the Netherlands and measures only the retail vacancy measured on January 1, 2021. However, the relationship might differ between countries and over time. Future research could involve a comparable study for countries other than the Netherlands and for a longer period.

REFERENCES

Aalders, R., Groot, S. & Ouwehand, A. (2021). *Transformatieatlas Retail: omvormen van winkels naar woningen draag bij aan oplossen woningtekort.* Retrieved on November 16, 2021 from Transformatieatlas Retail: omvormen van winkels naar woningen draagt bij aan oplossen woningtekort - RaboResearch (rabobank.com).

Ball, R.M. (2002). Re-use potential and vacant industrial premises: revisiting the regeneration issue in Stoke-on-Trent. *Journal of Property Research*, 19(2), pp. 93-110.

Benjamin, J.D., Jud, G.D. & Winkler, D.T. (1998). A Simultaneous Model and Empirical Test of the Demand and Supply of Retail Space. *Journal of Real Estate Research*, 16(1), pp. 1-14.

Besselaar, M. (2011). *Structurele leegstand: Voorkomen is beter dan genezen. Een onderzoek naar indicatoren van structurele leegstand in de kantorenmarkt van Amsterdam.* Amsterdam: Amsterdam School of Real Estate.

Bieniek, C., Lobaugh, K., Pincha, P. & Stephens, B. (2018). *The Great Retail Bifurcation – Why the Retail 'Apocalypse' Is Really a Renaissance*. Deloitte Insights Series. New York: Deloitte.

Binkley, J.K. & Young, J. (2020). The Chow Test with Time Series-Cross Section Data. *Faculty & Staff Research and Creative Activity*, 184. Murray: Murray State University.

Brooks, C. & Tsolacos, S. (2010). *Real estate modelling and forecasting.* Cambridge: Cambridge University Press.

Brouwer. A. & Tool, B. (2018). Eenzijdig aanbod van winkels leidt tot meer winkelleegstand. *Real Estate Research Quarterly*, 17(2), pp. 12-20.

Buttimer, R.J. & Ott, S.H. (2007). Commercial Real Estate Valuation, Development and Occupancy under Leasing Uncertainty. *Real Estate Economics*, 35(1), pp. 21-56.

CBS (2019). *Stedelijkheid.* Retrieved on November 18, 2022 from https://www.cbs.nl/nl-nl/nieuws/2019/44/meeste-afval-per-inwoner-in-minst-stedelijke-gemeenten/stedelijkheid.

CBS (2021a). *Landelijke Monitor Leegstand 2015-2021*. Retrieved on April 13, 2022 from https://www.cbs.nl/nl-nl/maatwerk/2021/47/landelijke-monitor-leegstand-2021.

CBS (2021b). *Wijk-en buurtstatistieken 2004-2022.* Retrieved on March 23, 2022 from https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/wijk-en-buurtstatistieken.

CBS (2021c). *Transformaties in de woningvoorraad 2020.* Retrieved on April 13, 2022 from https://www.cbs.nl/nl-nl/maatwerk/2021/45/transformaties-in-de-woningvoorraad-2020.

Chatterjee, S. & Simonoff, J.S. (2013). *Handbook of Regression Analysis.* Chicester: Wiley.

Choueiry, G. (n.d.). *Report the Shapiro-Wilk Test.* Retrieved on November 11, 2022 from https://quantifyinghealth.com/report-shapiro-wilk-test/.

Colwell, P.F. (2002). Tweaking the DiPasquale-Wheaton Model. *Journal of Housing Economics*, 11, pp. 24-39.

DiPasquale, D. & Wheaton, W.C. (1992). The Markets for Real Estate Assets and Space: A Conceptual Framework. *Journal of the American Real Estate and Urban Economics Association*, 20(1), pp. 191-197.

Dynamis (2019). *Sprekende cijfers: Winkelmarkten 2019.*

Elms, J. & Teller, C. (2010). Managing the attractiveness of evolved and created retail agglomerations formats. *Marketing Intelligence & Planning*, 28(1), pp. 25-45.

Fang, X. & Lütkepohl, H. (2012). The role of the log transformation in forecasting econoimc variables. *Empirical Economics*, 42, pp. 619-638.

Gelinck, S. & J.B. Benraad (2011). *Transformatie van kantoren gaat niet vanzelf*. Rotterdam: SBR.

Geraedts, R.P. & van der Voordt, T. (2004). Offices for living in: An instrument for measuring the potential for transforming offices into homes. *The Architecture Annual 2002-2003,* pp. 100-105. Rotterdam: 010 Publishers.

Gould, E.D., Pashigian, B.P. & Prendergast, C.J. (2005). Contracts, externalities, and incentives in shopping malls. *The Review of Economics and Statistics*, 87(3), pp. 411-422.

Hazelaar, M. (2019). *Transformatiepotentie van leegstaande kantoren; Een onderzoek naar het effect van omgevingskenmerken.* Utrecht: Universiteit Utrecht.

Hoekstra, J. & Vakili-Zad, C. (2011). High Vacancy Rates and Rising House Prices: The Spanish Paradox. *Journal of Economic and Human Geography*, 102(1), pp. 55-71.

Hollander, J.B., Hart, M.D., Wiley, A. & Vavra, S. (2018). Vacancy in Shrinking Downtowns: a Comparative Study of Québec, Ontario, and New England. *Journal of Housing and the Built Environment*, 33(4), pp. 591-613.

Johnson, K.H., Wiley, J.A. & Zhonghua, W. (2007). On the Relationship between Commercial Property Price and Its Selling Time. *The Journal of Real Estate Portfolio Mananagement*, 13(4), pp. 379-388.

Jongsma, M. (2020). *Leegstand in Nederland begint aan de randen.* Retrieved on September 21, 2022 from https://www.ewmagazine.nl/opinie/opinie/2020/10/leegstand-grootst-in-krimpregios-782209/.

Kadaster (2021) *Basisregistratie Adressen en Gebouwen*. Retrieved on February 2, 2022 from https://data.overheid.nl.

Kasza, J. & Wolfe, R. (2013). Interpretation of Commonly Used Statistical Regression Models. *Respirology*, 19(1), pp. 14-21.

Keeris, W.G. (2006). A Closer Look At Vacancy Levels In the Dutch Office Market. *Department of Real Estate and Housing, Faculty of Architcture.* Delft: Delft University of Technology.

Kickert, C. (2021). What's in Store: Prospects and Challenges for American Street-Level commerce. *Journal of Urban Design*, 26(2), pp. 159-177.

Konishi, H. & Sandfort, M.T. (2003). "Anchor stores". *Journal of Urban Economics,* 53, pp. 413-435.

Mallach, A. (2018). *The Empty House Next Door. Understanding and Reducing Vacancy and Hypervacancy in the United States*. Policy Focus Report. Cambridge, Massachussetts: Lincoln Insitute of Land Policy.

Montgomery, D.C., Peck, E.A. & Vining, G.G. (2012). *Introduction to Linear Regression Analysis*. 6th edition. Hoboken: John Wiley & Sons, Inc.

Moore, D.S. & McCabe, G.P. (2006). *Statistiek in de praktijk*. 5th edition. Den Haag: BIM Media B.V.

Neter, J., Wasserman, W. & Kutner, M.H. (1985). *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs.* 2nd Edition. Homewood, IL: Irwin.

Oskam, M. (2021). *Drivers of Retail Vacancy Pre-Covid19: The Effect of Building and Location Factors.* Groningen: Rijksuniversiteit Groningen.

Ossokina, I.V., Teulings, C.N & Svitak, J. (2017). Leegstand en transformatie van winkelvastgoed. *Real Estate Research Quarterly*, 16, pp. 12-20.

Ouwehand, A. (2018). *Drie argumenten om leegstand nú aan te pakken.* Retrieved on September 21, 2022 from https://www.vastgoedmarkt.nl/134771/drie-argumenten-om-leegstand-nu-aan-te-pakken.

Park, J. & Talen, E. (2021). Understanding Urban Retail Vacancy. *Urban Affairs Review*, pp. 1-27.

Pershio, D.B. (1991). The Coming Shakeout Among Regional Malls. *JMB Perspectives*, 1, pp. 3-5.

Rabianski, J.S. (2002). Vacancy in Market Analysis and Valuation. *Appraisal Journal*, 70(2), pp. 191-199.

Remøy, H.T., Koppels, P., Oel, C. van & Jonge, H. de (2007). *Characteristics of Vacant Offices, a Delphi-Approach.* ENHR International Conference.

Remøy, H.T. & Voordt, T.J.M. van der (2007). A New Life: Conversion of Vacant Office Buildings into Housing. *Emerald Insight*, 25(3/4), pp. 88-103. Delft University of Technology, Delft, The Netherlands.

Roulac, S.E. (1994). Retail Real Estate in the 21st Century: Information Technology + Time Consciousness + Unintelligent Stores = Intelligent Shopping? NOT! *Journal of Real Estate Research*, 9(1), pp. 125-150.

Ruigrok, A. (2021). *De vele gezichten van functieverandering als oplossing voor winkelleegstand.* Retrieved on June 16, 2022 from https://locatus.com/blog/over-de-vele-gezichten-van-functieverandering-als-oplossing-voor-winkelleegstand/. Utrecht: Locatus.

Ruizendaal, D. (2018). *Effectiviteit van gemeentelijk beleid om winkelleegstand tegen te gaan.* Groningen: Rijksuniversiteit Groningen.

Slob, G.J. (2022). *Spectaculaire daling winkelleegstand door afname retailpanden.* Retrieved on September 28, 2022 from https://locatus.com/blog/spectaculaire-daling-winkelleegstand-door-afname-retailpanden/#:~:text=Ondanks%20de%20maandenlange%20lockdown%20in,het%20eind%20van%20het%20jaar.

Stil, H. (2022). *Staatssteun heeft winkelbestand overeind gehouden, maar dat blijft niet zo.* Retrieved on April 25, 2022 from https://www.parool.nl/nederland/staatssteun-heeft-winkelbestand-overeind-gehouden-maar-dat-blijft-niet-zo~b7082f96/?referrer=https%3A%2F%2Fwww.google.com%2F. Amsterdam: Het Parool.

Thompson, D. (2020). *How the Pandemic Will Change American Streets*. The Atlantic. Retrieved on February 17, 2022 from https://www.theatlantic.com/ideas/archive/2020/04/how-pandemic-will-change-face-retail/610738/.

Uyanik, G.K. & Güler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia – Social and Behavioral Sciences*, 106, pp. 234-240.

Voordt, D.J.M. van der (2007). *Transformatie van kantoorgebouwen: thema's, actoren, instrumenten en projecten.* Rotterdam: Uitgeverij 010.

Wang, K. & Immergluck, D. (2018). The Geography of Vacant Housing and Neighborhood Healt Disparities After the U.S. Foreclosure Crisis. *The Housing-Health Connection*, 20(2), pp. 145-170.

Zweeden, J. van (2009). *Retail Vacancy in Dutch City Centers: How Can Differences in Retail Vacancy Between Cities be Explained?.* Rotterdam: Erasmus University Rotterdam.

APPENDIX

Appendix A – Operationalization of the variables

Table 7 *Operationalization of all variables included in the analysis.*

Variable type	*Variable*	*Definition*	*Indicator*	*Spatial scale*	*Source*
Dependent	Vacancy	No occupation according to the BRP, no business activity according to the HR, and no (fiscal) use according to the registration of the WOZ	Number	Neighborhood level	National vacancy monitor (CBS, 2021a)
Independent	Number of citizens	Number of Dutch citizens, who are included in the BRP	Number	Neighborhood level	Key figures for districts and neighborhoods (CBS, 2021b)
Independent	Population density	Dividing the number of inhabitants on January 1 by the land area	Number per km^2	Neighborhood level	Key figures for districts and neighborhoods (CBS, 2021b)
Control	Average housing value	The average property value of residential properties based on the WOZ value	In € × 1,000	Unit level	Key figures for districts and neighborhoods (CBS, 2021b)
Control	Construction before 2000	The number of dwellings built before 2000, expressed as a percentage of the total number of dwellings	In %	Neighborhood level	Key figures for districts and neighborhoods (CBS, 2021b)
Independent	Number of highly educated people	Number of persons aged between 15-75yrs on October 1, 2019 and were registered in a Dutch municipality, whose highest level of education attained was high education (level of HBO or WO)	Number	Neighborhood level	Key figures for districts and neighborhoods (CBS, 2021b)
Control	Number of retail objects	Objects with retail function	Number	Neighborhood level	Key registration (Kadaster, 2021)
Independent	Urbanity levels	Based on the ambient address density, each neighborhood has been assigned to an urbanity class	0. Very urban 1. Highly urban 2. Moderately urban 3. Minimally urban 4. Non-urban	Neighborhood	Urbanity (CBS, 2019); Key figures for districts and neighborhoods (CBS, 2021b)
Independent	Realized property transformations	Reuse of existing premises where the	Number	Neighborhood	Transformations in the housing

		function of the property is (partly) converted from a non-residential function to a residential function			stock (CBS, 2021c)

Table 8 *Stepwise description of the data collection and management.*

1. Kadaster provided a dataset “winkelbestand uitvoer” with retail stores in the Netherlands.
2. Only the first six numbers of the pht were generated to obtain the PC6.
3. The 2021PC6WBG dataset of Kadaster was merged on PC6 to obtain the neighborhood, district, and municipality codes.
4. The KWB dataset of the CBS was merged with the master dataset on these neighborhood codes.
5. The dataset was cleaned; only VBO IDs with retail function were kept, as well as VBO status with a retail object in use and retail objects in use (not measured). Duplicate VBO IDs were also removed.
6. The CBS provided register data on vacancy, realized property transformations, PC6 codes, and neighborhood, district and municipality codes.
7. The transformation dataset was merged on rinobjectnummer with dataset “rinobjectpc6crypt2021” from the CBS to obtain the PC6 in the transformation dataset.
8. The transformation dataset was merged on PC6 with dataset “PC6GWB2021” from the CBS to obtain the neighborhood, district, and municipality codes.
9. A new variable was created, called transformaties_buurt, counting the total realized property transformations per neighborhood.
10. The vacancy dataset was cleaned so that only the observations for retail vacancy remained.
11. A new variable was created, called leegstand_buurt, counting the total number of retail vacancy per neighborhood.
12. The dataset encrypted by the CBS was merged with the vacancy data on rinobjectnumber.
13. The dataset was merged with transformation data on neighborhood code.
14. The missing variables for vacancy and realized property transformations were renamed as 0 because these include the retail objects for which there is no vacancy of the object or property transformations in the neighborhood.
15. Missing variables were dropped, and the normal distribution of the variables was checked. A natural logarithm was generated for skewed variables to obtain a normal distribution.
16. Missing variables for log variables were removed, resulting in 14,776 observations remaining in the dataset for which the regression analyses were conducted.

Appendix B – Testing multiple linear regression assumptions.

Table 9 *Correlation matrix.*

```
. correlate leegstand_buurt a_inw bev_dich g_woz p_bjj2k a_opl_hg ste_mvs transformaties_buurt
(obs=11,394)
```

	leegst~t	a_inw	bev_dich	g_woz	p_bjj2k	a_opl_hg	ste_mvs	transf~t
leegstand_~t	1.0000							
a_inw	0.1922	1.0000						
bev_dich	0.0951	0.4223	1.0000					
g_woz	-0.0914	-0.2280	-0.2042	1.0000				
p_bjj2k	-0.0074	-0.0289	0.0246	-0.1175	1.0000			
a_opl_hg	0.1783	0.8486	0.4964	-0.0341	-0.0801	1.0000		
ste_mvs	0.1460	0.3272	0.6404	-0.1974	0.0417	0.3683	1.0000	
transforma~t	0.3104	0.1963	0.1695	-0.0654	-0.0025	0.2925	0.1982	1.0000

Table 10 *Multicollinearity assumption using VIF scores.*

```
. estat vif
```

Variable	VIF	1/VIF
bbev_dich	1.86	0.537566
ste_mvs	1.72	0.582000
aa_inw	1.26	0.793610
g_woz	1.09	0.914673
p_bjj2k	1.02	0.981493
Mean VIF	1.39	

Table 11 *Linearity assumption using a twoway scatter and linear predictions.*

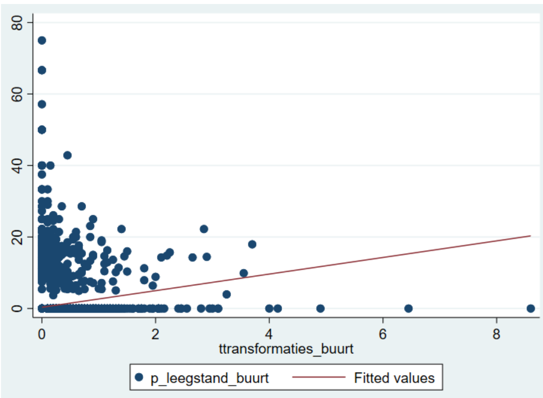

Table 12 *Normality assumption using Shapiro-Wilk test.*

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
p_leegstan~t	11,394	0.82831	956.412	18.437	0.00000
aa_inw	11,394	0.71817	1569.926	19.768	0.00000
bbev_dich	11,394	0.78965	1171.716	18.982	0.00000
g_woz	11,394	0.86079	775.488	17.873	0.00000
p_bjj2k	11,394	0.70609	1637.234	19.881	0.00000
ste_mvs	11,394	0.99981	1.048	0.126	0.44976

Note: The normal approximation to the sampling distribution of W'
is valid for 4<=n<=2000.

Table 13 *Breusch-Pagan heteroskedasticity test.*

```
Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of p_leegstand_buurt

H0: Constant variance

    chi2(1) = 9225.32
Prob > chi2 =  0.0000
```

Appendix C – Heterogeneity analysis

Table 14 *Chow F-test restricted model for urbanity level is 0: non-urban areas.*

```
. regress p_leegstand_buurt aa_inw bbev_dich g_woz p_bjj2k ste_mvs ttransformaties_buurt if ste_mvs==0
note: ste_mvs omitted because of collinearity.
```

Source	SS	df	MS			
				Number of obs	=	7,754
				F(5, 7748)	=	46.27
Model	772.582994	5	154.516599	Prob > F	=	0.0000
Residual	25871.6563	7,748	3.33913995	R-squared	=	0.0290
				Adj R-squared	=	0.0284
Total	26644.2393	7,753	3.43663605	Root MSE	=	1.8273

p_leegstand_buurt	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
aa_inw	.1813506	.017086	10.61	0.000	.1478574	.2148438
bbev_dich	-.0326968	.0120922	-2.70	0.007	-.0564008	-.0089928
g_woz	-.0004628	.0001662	-2.78	0.005	-.0007886	-.000137
p_bjj2k	-.00019	.0009547	-0.20	0.842	-.0020615	.0016815
ste_mvs	0	(omitted)				
ttransformaties_buurt	1.974388	.2478436	7.97	0.000	1.488548	2.460229
_cons	.1793494	.1144545	1.57	0.117	-.0450123	.403711

Table 15 *Chow F-test restricted model for urbanity level is 1: urban areas.*

```
. regress p_leegstand_buurt aa_inw bbev_dich g_woz p_bjj2k ste_mvs
note: ste_mvs omitted because of collinearity.
```

Source	SS	df	MS	
				Number of obs
				F(5, 3634)
Model	3726.26254	5	745.252508	Prob > F
Residual	73327.8728	3,634	20.1782809	R-squared
				Adj R-squared
Total	77054.1353	3,639	21.1745357	Root MSE

p_leegstand_buurt	Coefficient	Std. err.	t	P>|t|
aa_inw	.2675384	.0365733	7.32	0.000
bbev_dich	-.0511409	.0166799	-3.07	0.002
g_woz	-.0024582	.0005667	-4.34	0.000
p_bjj2k	-.0042704	.003132	-1.36	0.173
ste_mvs	0	(omitted)		
ttransformaties_buurt	1.781789	.2071634	8.60	0.000
_cons	1.585931	.3445275	4.60	0.000

Table 16 *Chow F-test pooled unrestricted model.*

```
. regress p_leegstand_buurt aa_inw bbev_dich g_woz p_bjj2k ste_mvs ttransformaties_buurt
```

Source	SS	df	MS			
				Number of obs	=	11,394
				F(6, 11387)	=	112.65
Model	5902.94935	6	983.824891	Prob > F	=	0.0000
Residual	99445.9876	11,387	8.73329127	R-squared	=	0.0560
				Adj R-squared	=	0.0555
Total	105348.937	11,393	9.24681269	Root MSE	=	2.9552

p_leegstand_buurt	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
aa_inw	.2299364	.0179272	12.83	0.000	.1947961	.2650767
bbev_dich	-.0569409	.0093356	-6.10	0.000	-.0752402	-.0386415
g_woz	-.0010931	.0002088	-5.24	0.000	-.0015024	-.0006838
p_bjj2k	-.0015215	.0012231	-1.24	0.214	-.003919	.0008761
ste_mvs	.615744	.0782898	7.86	0.000	.4622824	.7692056
ttransformaties_buurt	1.841344	.1285637	14.32	0.000	1.589337	2.093351
_cons	.4877845	.1379083	3.54	0.000	.2174605	.7581085

Appendix D – STATA Syntax

```
log using "C:\Users\marti\OneDrive\1. RUG\Master's thesis\Stata\log-file-thesis-final.smcl", replace
clear all
cd "C:\Users\marti\OneDrive\1. RUG\Master's thesis\Stata"
import delimited using "kwb-2020.csv"
save "kwb-2020.dta"
browse

drop if missing(gwb_code)

* Merge with CBS and Kadaster data on municipality level
rename gwb_code gemeentecode
sort gemeentecode
merge m:m gemeentecode using "20221006_gemeente 10_afgerond 5.dta"
drop if _merge == 2
rename _merge merge_gem
rename gemeentecode gwb_code

* Merge with CBS and Kadaster data on district level
rename gwb_code wijkcode
sort wijkcode
merge m:m wijkcode using "20221006_wijk 10_afgerond 5.dta"
rename _merge merge_wijk
drop if merge_wijk == 2
rename wijkcode gwb_code

* Merge with CBS and Kadaster data on neighborhood level
rename gwb_code buurtcode
sort buurtcode
merge m:m buurtcode using "20221006_buurt 10_afgerond 5.dta"
drop if _merge == 2
rename _merge merge_buurt
rename buurtcode gwb_code

* Remove unnecessary variables
drop v7 v8 v9 v10 gemcode
keep if recs == "Buurt"

* Change the variables format
split leegstand_wijk, parse(,)
replace leegstand_wijk = leegstand_wijk1
destring (leegstand_wijk), replace
drop leegstand_wijk1 leegstand_wijk2
split leegjaareerder_wijk, parse(,)
replace leegjaareerder_wijk = leegjaareerder_wijk1
destring (leegjaareerder_wijk), replace
drop leegjaareerder_wijk1 leegjaareerder_wijk2
split leegstand_buurt, parse(,)
replace leegstand_buurt = leegstand_buurt1
```

```
destring (leegstand_buurt), replace
drop leegstand_buurt1 leegstand_buurt2
split leegjaareerder_buurt, parse(,)
replace leegjaareerder_buurt = leegjaareerder_buurt1
destring (leegjaareerder_buurt), replace
drop leegjaareerder_buurt1 leegjaareerder_buurt2

* Replace missing variables for vacancy and property transformations with 0
replace transformaties_gemeente = 0 if transformaties_gemeente ==.
replace winkels_gemeente = 0 if winkels_gemeente ==.
replace leegstand_gemeente = 0 if leegstand_gemeente ==.
replace leegjaareerder_gemeente = 0 if leegjaareerder_gemeente ==.
replace transformaties_wijk = 0 if transformaties_wijk ==.
replace leegstand_wijk = 0 if leegstand_wijk ==.
replace winkels_wijk = 0 if winkels_wijk ==.
replace leegjaareerder_wijk = 0 if leegjaareerder_wijk ==.
replace transformaties_buurt = 0 if transformaties_buurt ==.
replace winkels_buurt = 0 if winkels_buurt ==.
replace leegstand_buurt = 0 if leegstand_buurt ==.
replace leegjaareerder_buurt = 0 if leegjaareerder_buurt ==.

* Make a distinction between urban and nonurban areas, ste_mvs 1 2 = 1 "Urban" and 3 4 5=0 "Non-
urban"
codebook ste_mvs
tabulate ste_mvs
drop if missing(ste_mvs)
replace ste_mvs = 1 if ste_mvs<=2
replace ste_mvs = 0 if ste_mvs>2
tabulate ste_mvs

* Create percentage variables for vacancy and property transformations
gen p_leegstand_buurt = 100 * leegstand_buurt / winkels_buurt
replace p_leegstand_buurt = 0 if p_leegstand_buurt ==.

* Check missing variables and drop these
codebook leegstand_buurt
codebook a_inw
codebook transformaties_buurt
codebook ste_mvs
codebook p_bjj2k
codebook bev_dich
drop if missing(bev_dich)
codebook g_woz
drop if missing(g_woz)
codebook a_opl_hg
drop if missing(a_opl_hg)

* Summarize
summarize p_leegstand_buurt aa_inw bbev_dich g_woz p_bjj2k a_opl_hg ste_mvs ttransformaties_buurt
ssc install fre
fre ste_mvs

summarize leegstand_buurt
display r(sum)
```

```
* Correlation matrix
correlate leegstand_buurt a_inw bev_dich g_woz p_bjj2k a_opl_hg ste_mvs transformaties_buurt

* Check if the variables are normally distributed, if not: check if generating a natural logarithm to make it more normally distributed instead of skewed would work
histogram p_leegstand_buurt, normal
histogram lnp_leegstand_buurt, normal
histogram a_inw, normal
histogram bev_dich, normal
histogram g_woz, normal
histogram p_bjj2k, normal
histogram lnp_bjj2k, normal
histogram ste_mvs, normal
* Note: the dependent variable p_leegstand_buurt is in % so do not make any log-transformations

*Transform the variables with 100 and 1000
gen aa_inw = a_inw / 1000
gen bbev_dich = bev_dich / 1000
gen ttransformaties_buurt = transformaties_buurt / 100

* Testing multiple linear assumptions
regress p_leegstand_buurt aa_inw bbev_dich g_woz p_bjj2k ste_mvs
estat vif
twoway (scatter p_leegstand_buurt ttransformaties_buurt) (lfit p_leegstand_buurt ttransformaties_buurt)
swilk p_leegstand_buurt aa_inw bbev_dich g_woz p_bjj2k ste_mvs

*Multiple linear regression: model 1
regress p_leegstand_buurt aa_inw bbev_dich g_woz p_bjj2k ste_mvs, robust

* Multiple linear regression: model 2
regress p_leegstand_buurt aa_inw bbev_dich g_woz p_bjj2k ste_mvs ttransformaties_buurt, robust

* Chow test for hypothesis 2
* pooled unrestricted model
regress p_leegstand_buurt ttransformaties_buurt

* pooled restricted model non-urban
regress p_leegstand_buurt ttransformaties_buurt if ste_mvs==0

* pooled restricted model urban
regress p_leegstand_buurt ttransformaties_buurt if ste_mvs==1

* Interaction variable included
regress c.p_leegstand_buurt c.aa_inw c.bbev_dich c.g_woz c.p_bjj2k i.ste_mvs c.ttransformaties_buurt
i.ste_mvs#c.ttransformaties_buurt, robust
```