Rijksuniversiteit Groningen

Faculty of Spatial Sciences

Thesis MSc Real Estate Studies

Living next to the cloud
The impact of data center development on housing prices

Author:
Timber van Tilburg S4454448
Email:
timbervantilburg@gmail.com

Supervisor:
Dr. E. Margaritis
Assessor:
prof. dr. A.J. van der Vlist

Abstract

Data centers are an essential link in our digital infrastructure. Together they form the backbone of our digitalised lifestyles and economies. Since our lifestyles and economies keep pace for further digitalisation, our digital infrastructure needs to lead the way. Therefore, additional data center development is necessary, however not-in-my-backyard (NIMBY). Lately, data center development can count on opposition from residents and public institutions. Data centers would contribute to the transition of the Dutch agricultural landscape, cause environmental damage, and lead to further scarcity of available development land for housing and employment. This master thesis will investigate whether data center development has an external impact on the transaction prices of nearby residential units. By using a difference-in-differences regression method, the results show that the transaction price of a house located nearby a data center increases, on average, by 5.27% after development. This would suggest that Dutch spatial policies can be successful in their attempt to ensure data center development has no external negative impact on nearby residential units.

August 23, 2021

Contents

1 Introduction **3**

1.1 Motivation . . . 3

1.2 Data centers . . . 4

1.3 Research problem statement . . . 7

1.4 Research method and data . . . 8

1.5 Conceptual model . . . 8

1.6 Guide . . . 9

2 Theory **10**

2.1 Location theories . . . 10

2.2 Spatial planning context and NIMBYism . . . 12

2.3 Impact of comparable sites and functions on housing prices . . . 13

2.4 Determination of housing prices . . . 14

3 Data & methodology **17**

3.1 Hedonic regression model . . . 17

3.2 Data selection . . . 20

3.3 Descriptive statistics . . . 23

3.4 Regression transformation . . . 27

3.5 Regression models . . . 27

4 Results **30**

4.1 Impact of data center development on transaction prices . . . 30

4.1.1 Interpretation of regression results . . . 30

4.2 Sensitivity analysis . . . 32

4.2.1 Discussion on results and literature . . . 33

4.3 Impact of data centers across building types . . . 34

5 Conclusion & discussion **36**

5.1 Main findings . . . 36

5.2 Limitations & future research . . . 37

5.3 Recommendations . . . 38

6 Appendices **44**

6.1 Appendix I . . . 44

6.2 Appendix II . . . 45

6.3 Appendix III . . . 46

6.4 Appendix IV . . . 47

6.5 Appendix V . . . 49

6.6 Appendix VI . . . 49

List of Figures

1 The function of a data center . . . 5
2 Map of all data centers in the Netherlands . . . 6
3 Conceptual model housing price determinants . . . 9
4 Data center polygon . . . 19
5 Flow chart of data selection procedure . . . 21
6 Histogram of dependent variable *transaction price* . . . 23
7 Hyperconnectivity in MRA . . . 44
8 Overview of the selected data centers . . . 46
9 Development of average *transaction price* relative to data center development . . . 47
10 Development of the average *transaction price* . . . 47
11 Development of average *transaction price per squared meter* relative to data center development . . . 48
12 Development of the average *transaction price per squared meter* . . . 48
13 Histogram of *log transaction price* . . . 49

List of Tables

1 Determinants of housing prices according to comparable studies . . . 16
2 Data center characteristics of the selected data centers . . . 22
3 Descriptive statistics . . . 25
4 Descriptive statistics for the target area 0 - 3,000 m . . . 26
5 Descriptive statistics for the control area 3,000 - 6,000 m . . . 27
6 Overview regression models . . . 28
7 Difference-in-differences regression results . . . 31
8 Regression results sensitivity analysis: Target area 0 - 2,500 m . . . 33
9 Regression results sensitivity analysis: Target area 0 - 3,500 m . . . 33
10 Heterogeneity test on variable Building type . . . 35
11 Regression analysis target area . . . 45
12 Chow test . . . 49

1 Introduction

1.1 Motivation

In May 2021, a letter signed by 3,077 residents was delivered to the mayor of Hollands Kroon. Each signature represented one vote against new data center developments in 'De Wieringermeer', a polder located in the north of The Netherlands. The municipality Hollands Kroon is home to the data centers of the Big Tech companies Google and Microsoft and is willing to welcome others (Trouw, 2021). However, not if it is up to its inhabitants. They are opposing new development because, in their opinion, it will lead to various negative external (environmental) impacts, such as landscape devastation, excessive energy and water consumption, water pollution, and the sprawl of windmills and solar panels to fulfil the energy requirements of the data centers (De Stendor, 2021). The 'De Wieringermeer' is not the only place where resistance against data center development is growing. In December 2020, a local municipal party 'Leefbaar Zeewolde' send an open letter, 'The Achilles heel of the data center', to its Province state to speak out its worries about the negative impact of the development plans for Europe's biggest data center on its agricultural fields (Leefbaar Zeewolde, 2020). On July 12th, 2019, the municipalities Amsterdam and Haarlemmermeer even announced a moratorium on further data center development. Within a short period, the Metropol region Amsterdam (MRA) became 'the Digital nexus of Western Europe' (Copenhagen Economics, 2020).

Data center development in the Netherlands is interesting because of its geographical location around one largest the internet nodes in the world, the Amsterdam Internet Exchange (AMS-IX), limiting the latency in data exchange between users. Furthermore, with the digitisation of our economies and lifestyles, more and better digital infrastructure is needed, and data centers fulfil an essential role in the provision. A strong position in the data center market is essential for the Netherlands, the MRA, and local regions. Google's data centers in the Eemshaven (Groningen) and Agriport ('De Wieringermeer') alone already contributed EUR 3.6 billion in GDP and 3,400 jobs per year on average between 2014 – 2019 (Copenhagen Economics, 2020). Therefore, municipalities still participate in the development of data centers.

The sprawl of data center development begins to form a point of 'not-in-my-backyard' (NIMBY) discussion, *'While some appreciate the industrial design of these facilities, most just see a box that blocks the open landscape'* (DDA, 2021), nourished by landscape devastation and contributing to the transition of agricultural land ('verdozing van het landschap') (College van Rijksadviseurs, 2019; NH Nieuws, 2020). Moreover, data centers put too much pressure on the existing electricity grid (Gemeente Amsterdam, 2019). The largest data centers claim to use sustainable energy, but they require at least 100 wind turbines (Leefbaar Zeewolde, 2020). Lastly, developing data centers contribute to further pressure on the residential market since available building space is already scarce (Gemeente Amsterdam, 2019). However, municipalities attempt to take all negative external effects into account in their spatial planning policies. They mainly emphasise the positive external impacts, among which employment and local economic growth, as a consequence of data center development. Residents are questioning that strongly.

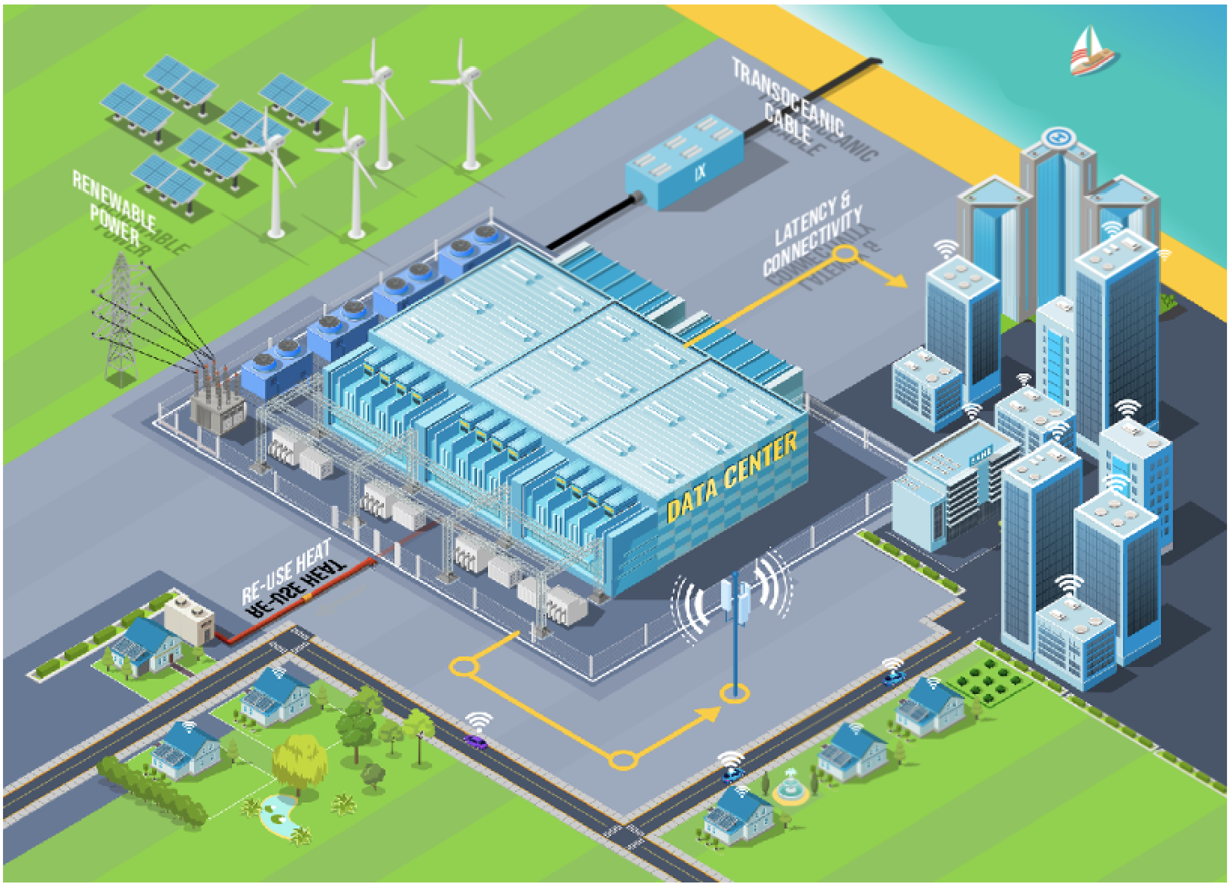
This master thesis investigates the external impact of data center development on nearby residential property prices. As far as known, the impact that data center development has on the nearby residential property prices is unclear and not investigated before. Thus, an insight into the potential impact is noteworthy to residents and both the public and private sectors.

1.2 Data centers

Simplified, a data center is a building stocked with computer servers used to store, process, and disseminate data and applications. A data center represents the physical 'cloud', where companies store the data of their users. Due to data centers, companies as Apple, Facebook, and Google can run their programs 24/7 so that everyone can store, share and collect information from the internet as quickly as possible. Besides these Big Tech companies, the Dutch government, municipalities, hospitals, universities, and small and medium-sized enterprises lease computer servers in data centers to store their data safely.

Data centers have gained importance due to the digitalisation of our economies and lifestyles. With business processes becoming further digitalised, data centers are an essential link within the digital infrastructure. However, some argue that data centers are already footloose and can function independent of location (Trouw, 2021). The most recent example of the importance of data centers is visible at the start of the global COVID-19 pandemic, where a transition to 'being online' has been made. Due to all the data centers and the digital infrastructure in the Netherlands, this transition to digital forms of education, working remotely, and social interactions went efficiently (DDA, 2021). It emphasises the importance of extensive digital infrastructure regarding the expectation of further digitisation. Figure 1 shows a simplified indication of the function of a data center in our society.

Figure 1: The function of a data center



Note: Data centers are an essential link in the digital infrastructure. Transoceanic cables make data able to travel the world. A data center stores data to give users quick access with negligible latency. Data centers use (renewable) energy to run their processes. The residual heat of a data center could be used to heat nearby households (Source: Lünendonk Hossenfelder GmbH, 2021).

Within this research, 125 different independent data centers in the Netherlands are identified and visualised in figure 2. Data centers can be categorised into three types of independent data centers: Multi-tenant international data centers, Multi-tenant regional data centers, and Hyperscale data centers (Buck Consultants International, 2020).

Figure 2: Map of all data centers in the Netherlands

Note: This GIS map indicates all external data centers in the Netherlands (125). The red triangles show the selected data centers for this research.

The first data center type is the Multi-tenant data center for the international market. These data centers can take up to 20,000 square meters, and host larger companies and institutions, often with global reach. They require locations with the best connection since their data needs to travel longer distances and as fast as possible. Therefore, these data centers often located at 'hyperconnective' places. Hyperconnectivity is available at locations where an internet exchange and multiple network cables come together and form the fastest and most stable connection, which minimises the latency of data processing (Buck Consultants International, 2020). Hyperconnectivity is currently available at three locations in the Netherlands, all located in the MRA: Amsterdam Science Park, Amsterdam Southeast, and Schiphol. A visible consequence of hyperconnectivity is the clustering of data centers (CE Delft, 2020). Appendix I shows the cluster forming within the hyperconnectivity areas of the MRA. Due to spatial development constraints on existing industrial and business sites and the aim of holding on to a leading position of the MRA as 'Digital hub', spatial planning policies point at Almere-Zeewolde to become the fourth location for a hyperconnectivity cluster (CE Delft, 2020).

The second data center type is the Multi-tenant data center lays its focus on the regional market. These data centers provide computer servers for multiple tenants, usually small or middle-sized companies or institutions with regional reach. Compared to the other data centers, these data centers are smaller, often between 500 and 5,000 square meters, and do not necessarily locate in a hyperconnectivity location (CE Delft, 2020). Moreover, it is noticeable that these data centers frequently occupy a multi-tenant building.

The last type of data center is the Hyperscale data center. These data centers are single-tenant data centers which mean they accommodate data space for one company. In the Netherlands, there are three Hyperscale data centers in operation: Google (Eemshaven and 'De Wieringermeer') and Microsoft ('De Wieringermeer'), and a few are in preparation ('De Wieringermeer' and Zeewolde). Compared to the other types of data centers, Hyperscale data centers are much larger (>40,000 m^2) (Buck Consultants International, 2020) and can take up a plot of 330,000 square meters (Microsoft 'De Wieringermeer'). Therefore, developing a hyperscale in an urban area is not suitable. However, developing a hyperscale data center in a rural area has a significant spatial impact (DDA, 2021). Consequently, hyperscale data centers are the instigator of the public debate on the development of data centers in rural areas.

1.3 Research problem statement

As far as known, there is no theory nor academic literature that emphasises specifically the impact of data centers (development) on the price of a nearby residential unit. Therefore, this master thesis investigates the external impact of data center development on nearby residential property prices. The central question to be answered is, *'To what extent does a newly developed data center impact nearby housing prices?'*

The following four research questions are formulated to substantiate the central question.
RQ 1: *'What are the important factors for determining the development location of a data center?'*
RQ 2: *'What is the relationship between housing prices and the proximity to an existing industrial or business site, on which data centers are zoned?'*
RQ 3: *'What is the quantitative impact of the development of a data center on the transaction price of a nearby residential unit?'*
RQ 4: *'To what extent does the quantitative impact of the development of a data center change across building types?'*

1.4 Research method and data

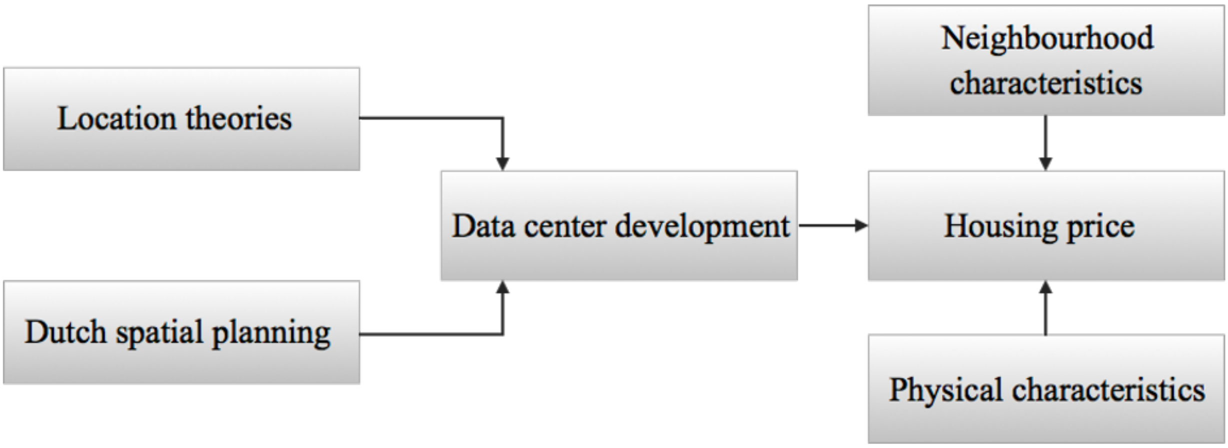
To formulate an answer to these research questions, the hedonic pricing method of Rosen (1974) is used. In this theory, Rosen explains that a combination of various characteristics determines the value of a residential unit. More specifically, within this research, a variant of this technique is used, namely, the difference-in-differences regression method. The method, initially derived by Schwarz et al. (2006), explains the impact of a single 'event' by comparing the before and after situation between a target group and a control group. In this research, the 'event' will comprise the development of a data center. The technique of the difference-in-differences method will be further explained in Chapter 3, 'Data'. The analysis will be executed within the statistical program STATA, and geographical information is visualised by the geographical program Geographic Information System (GIS).

Transaction observations from the Dutch Association of Real Estate Agents (NVM) are used to determine the potential impact of data center development on housing prices. The NVM collects a large part (± 70%) of all housing transactions in the Netherlands, together with details such as property type, building period, and house size and parcel size. Lastly, specific data of the data centers is provided by various sources, among which the Dutch Datacenter Association (DDA) and merged to form an all-included dataset for data centers in the Netherlands.

1.5 Conceptual model

Based on the theory that derives the housing price, a conceptual model is established and visualised in figure 3. Within this research, it is emphasised that location theories and spatial planning policies both are determining factors for the location for data center development. The aim is to quantify the relationship between data center development and the housing price (transaction price). Based on the research of Visser & Van Dam (2006), the conceptual model is completed with the categories *physical characteristics*, which represent the characteristics of an individual house, and *neighbourhood characteristics*, which represent the subcategories, *physical neighbourhood characteristics, social neighbourhood characteristics*, and *functional neighbourhood characteristics*. Eventually, these characteristics together determine the housing price.

Figure 3: Conceptual model housing price determinants



Note: Based on the determining factors for data center development, location theories explain where data center development happens from the private sector perspective. The public sector's spatial planning policies determine if a location is appropriate for data center development by statutory regulations. This research investigates the impact of data center development on housing prices. *Physical characteristics* and *neighbourhood characteristics* determine the housing price too (Visser & Van Dam, 2006).

1.6 Guide

The remainder of this research is organised as follows. In chapter 2, the theoretical background on location theories, spatial planning policies and NIMBYism, and comparable spatial functions are explained to derive a potential impact of data center development and location decisions. Chapter 3 describes the data and the methodology used in this research. The results of all regression analyses are reported in chapter 4. Chapter 5 concludes the research findings, provides the limitations of this research and recommendations for spatial planning policies.

2 Theory

2.1 Location theories

In economic geography, location theories address the questions of what economic activity is located where and why. The location theories are economically orientated, where they aim for a financial cost optimum. The work of Von Thunen (1826), Weber (1909) and Marshall (1890) form the foundation of the location theories in the determination of location choice. Applying these location theories on data center development is previously executed by Goorman (2008) and Oxford Economics (2018) in the USA.

The base of the classical location theory comes from the land use theory of Von Thünen (1826), which focuses on agricultural land use. The theory determines land use relative to the market, based on transportation costs. Von Thünen explains that the land nearby a central market (consumers) is high in value because of lower transportation costs. Therefore, (time and labour) intensive production processes are located near the market. Moreover, the land value decreases when the distance to the market increases because of higher transportation costs.

Von Thünen's location theory is comparable to modern data center site selection. Where Von Thünen refers to the 'market' in previous times in modern times and applied to data center this is respectively a network hub; a place where internet cables come onshore or where network cables come together, such as the MRA, Eemshaven, and the 'De Wieringermeer'. These 'network backbones' are an essential determinant in the selection of a location for data center development (Giori et al., 2011; Lünendonk Hossenfelder GmbH, 2021). Data centers locate close to these network hubs to minimise the distance that data must travel over fibre optic lines between the point of originating and sending and receiving (Lünendonk Hossenfelder GmbH, 2021). Minimising that distance will reduce data latency (Watkins, 2019; Giori et al., 2011).When located further away from a hub location, additional investments to extend the digital infrastructure, and enhance latency, must be made. This is in line with land use theory, in which Von Thünen argues that transaction costs increase when located further away from the market. However, there will be compensation in the form of lower land values since there is less demand for these locations.

Alfred Weber (1909) shifts the focus of this theory to the economic activities of industrial use and adds the importance of resources to location choice. The industrial production process will locate where transportation costs between resources and the market are the lowest. Therefore, the theory divides two possible cases: The weight gaining process and the weight losing process. In the weight gaining process, an industry that gains weight in the production process locates relatively closer to the market. That is because the final product is now more expensive to distribute. A reversed effect occurs in the weight losing process, where weight is reduced. This stimulates the production process closer to its resources to reduce transportation costs.

Data centers are dependent on resources also, namely the supply of energy and water. Locating close to these resources is cost-efficient due to reduced transportation costs. Within the location decision of a data center, the supply of resources and the price cost optimalisation are determining factors (Giori et al. 2011; Yang Ye, 2011; Ounifi et al. 2015; Depoorter et al., 2015). The considerations arise from the fact that data centers consume 3% of the total consumed electricity in the Netherlands (REOS, 2019). Therefore, energy costs are a significant component within the cost operationalisation of a data center (Ounfini, 2015).

Another perspective of the location choice comes from Marshall's agglomeration theory (1890). Marshall argues in his agglomeration theory that firms can increase their returns by agglomeration economies and, consequently, opt to cluster. Firms will experience higher returns because of agglomeration benefits, such as a pool of skilled workers, knowledge spill-overs, and supplier linkages. According to Oxford Economic (2018), Marshall's theory applies to data center location decisions in the USA. The study reports that the data center development of Google works as a magnet, causing the clustering of related firms and competing businesses. In the end, Google and the other companies benefit from local spill-over effects, such as improvement of employment rates and increasing level of education.

In addition to Marshall, Porter (1998; 2000) adds to the cluster theory that comparative advantages and innovations are consequences of interlinkages with nearby located partners, other (competing) companies, suppliers, and institutions in globalised economies. Cluster locations will achieve a faster pace of innovations and higher comparative advantages (Porter; 2000). Therefore, Porter's cluster theory emphasises the importance of local factors in the location decision paradox of the globalised economy. Also, Porter's cluster theory (1998; 2000) shows similarities with data center development since data centers cluster in the MRA (CE Delft, 2020). In a study of Eickelpasch et al. (2007) on Porter's cluster theory, the local factors that cause comparative advantages are distinguished into hard and soft factors because they expect significant differences in the importance of the factors. On the one hand, hard factors that directly impact a company, such as the supply of qualified labour and proximity to a network hub, are more critical for location decisions than soft location factors. On the other hand, soft factors exert an indirect effect, such as local government support or financial support (Eikelparsch, 2007).

However, these location theories are not determinative in the location choice of a data center development. The location theories follow a monocentric city model, where the cost of transportation is the most determining factor. Nowadays, the Dutch city model has a polycentric structure with multiple markets. In addition, transportation technologies improved enormously, which impacts the location decision of businesses. And most importantly: Dutch spatial planning determines where developments take place.

2.2 Spatial planning context and NIMBYism

The not-in-my-backyard (NIMBY) movement against data center development seems to be the result of the unsuccessful execution of spatial planning policies. In the Netherlands, spatial planning policies are determinative in the location decision for data centers. These policies are legally binding documents drafted by public institutions to organise the available space and minimise the experience of negative externalities from data centers development. Particularly for residents, the restrictions that spatial planning policies impose on data centers, are important since not everyone desires a data center nearby. Therefore, spatial planning policies should be the answer against NIMBYism. However, the opposition against data center development is currently going on, causing the questioning of the effectiveness of Dutch spatial planning policies.

In the planning process for new development, local government make use of tools as Business Location Monitor (BLM) to forecast the expansion demand for industrial or business sites in the area (CPB, 2005). When the BLM indicate promising forecasts, the local government is encouraged to zone industrial sites since industrial sites stimulate the local employment growth (Louw Bontekoning, 2006). The downside of existing industrial sites is that they have a negative price impact on the nearby houses (De Vor & de Groot, 2011). Therefore, the local government makes use of spatial planning policies to minimise negative external effects. For example, by developing industrial zones away from residential areas, no negative external effects should be experienced. But with the current city sprawl and data center development strategies aiming to use the residual heat of data centers to warm nearby houses, the industrial and residential areas become more intertwined (PBL, 2019; REOS, 2019). This can complicate the process of minimising negative external effects.

The underlying motivation for NIMBYism comes from residents who are willing to protect their neighbourhood against an unwelcome development (van der Horst, 2007). The occurrence of NIMBYism is most substantial by residents who live close to the development location, felt ignored in previous stages of the planning process, and where sites without industrial history (agriculture) are transformed drastically (Turner, 2021, as cited in Swinhoe, 2021). Moreover, NIMBYism under residents is strong at the planning phase, peaks at the construction phase, but when the operation phase starts, the resilience weakens (Boyle et al., 2019). This illustrates the complexity of NIMBYism and how hard it is to quantify the impact of NIMBYism in its relation to public planning.

According to Van der Horst (2007), NIMBYism in the UK is a problem for its public planning system since NIMBY'ers are allowed to oppose developments without unfounded arguments or the inconvenience of thinking how else the underlying societal objectives of a development can be better achieved. Compared to the Netherlands, NIMBY'ers seem to have less influence here, since spatial planning policies are the decisive factor to allow certain developments. As long as the zoning plan adheres to strict regulations that permit a maximum level of negative external effects as noise, light and air pollution a building permit ('Bouwvergunning') should be issued. Based on location theories and the Dutch spatial planning context, an answer to the first research question can be formulated: ***'Which factors are important for determining the location of data centers?'***

According to the location theories, data centres preferably locate where the benefits are the highest and costs are the lowest. This can refer to the proximity to a network hub to minimise latency, proximity to resources to minimise transportation cost, and hard location factors to achieve agglomeration benefits. However, the criteria for a location decision seem to be influenced by the size of a large data center. With larger (hyperscale) data centers more depending on energy supply and available land, smaller data centers on minimising latency by locating near a network hub, resulting in clustering, agglomeration benefits, and innovations. However, as mentioned before, the location theories simplify the (modern) reality. Eventually, public institutions' spatial planning policies are the decisive factor to allow data center development. Therefore, data center developers do not freely choose the development locations. The local government establishes a zoning plan, in which it has investigated the external effects coming with development. For data center development, reducing the negative external effect seems the most important. As long as the zoning plan adheres to strict regulations that permit a maximum level of negative external effects as noise, light and air pollution a building permit ('Bouwvergunning') should be issued.

2.3 Impact of comparable sites and functions on housing prices

The impact of comparable sites and functions on housing prices is, contrary to data center development, known to have an external impact. Therefore, the second research question is formulated: *'What is the relationship between housing prices and the proximity of an industrial site or business park, on which data centers can be located?'*

In a study by De Vor & De Groot (2011), the impact of existing industrial sites on nearby property values in the Netherlands is researched. The study concluded that industrial sites cause various negative externalities, such as traffic disturbance, noise pollution, and obstruction of view. These externalities result in a negative dichotomous effect on nearby housing prices, meaning that a house within proximity to an industrial site shows a strong negative impact that convexly decreases until a certain distance. Beyond this distance, the effect diminishes until it fades away. According to their findings, a house located within 250 meters from an industrial site has a 14.9% lower sales price relative to a comparable house situated at a distance of 2,250 meters or beyond an industrial site. The impact of industrial sites on nearby residential prices is various times investigated outside the Netherlands (Kain & Quigley, 1970; Li & Brown, 1980; and Grether & Mieszkowski, 1980). Despite the differences in time and locations, the findings of De Vor & De Groot (2011) are in line with the results of the other studies, reporting a decreasing negative relation between the distance to industrial sites and housing prices. Aydin et al. (2010) even argue that there is an almost universal deleterious impact. Besides a negative relation, De Vor & De Groot (2011) found that the size of an industrial site intervenes with the impacted area. The larger the size of an industrial site is, the larger the area is that measures a negative impact on housing prices.

Similar to industrial sites, the distance to an existing business site in the Netherlands also negatively impacts housing prices, which can be caused by noise pollution, traffic, or view (Visser & Van Dam, 2006). Compared to industrial sites, business sites often have a lower environmental code since they do not contain heavy industry. Therefore, the occupiers of business sites are more focused on the logistic, office, or data center functions. Especially the distribution centers show high similarities to the data centers. Firstly, both a distribution center and a data center can occupy land that is intended for *'Business site'*. Secondly, similar to data centers, distribution centers have seen an upward trend in investment popularity. Therefore, in 2019 and 2020, respectively 19 and 22 'mega' distribution centers (>40,000 m^2) were developed in the Netherlands (Buck Consultants International, 2020). Thirdly, data center shows strong exterior comparisons with the development of large distribution centers, making them both responsible for the public discussion about the transition of an agricultural to an industrial landscape ('verdozing van het landschap') (College van Rijksadviseurs, 2019; NH Nieuws, 2020). According to a qualitative study by Van der Veen (2019), inhabitants of the rural areas experience negative impacts (identity change of their rural located villages) of the development of XXL (mega) distribution centers. Since mega distribution centers require large plots of land, development often takes place in rural areas on former agricultural lands. Thus, data centers have a comparable significant spatial impact as distribution centers (DDA, 2021).

Based on the above literature review, the second research question could be answered, ***'What is the relationship between housing prices and the proximity to an existing industrial or business site, on which data centers are zoned?'***

According to the literature, the proximity of an existing industrial or business site in the Netherlands negatively affects housing prices. The negative externalities of industrial or business sites overrule the positive externalities, causing the average house price to increase when the distance to an existing industrial or business site increases. Nevertheless, the impact fades away quickly with distance. Comparable studies outside the Netherlands find similar results, indicating a universal impact (Aydin et al., 2010). Important to note is that in this research newly developed data centers are studies, which could have divergent relations towards housing prices compared to existing industrial sites.

2.4 Determination of housing prices

The value of a house is determined by various house characteristics. The theory behind this comes from Rosen (1974), who emphasises that the value of a house is based on various characteristics together, among which housing characteristics and neighbourhood characteristics.

To acquire a potential and significant impact, adding various relevant housing characteristics and neighbourhood characteristics to the hedonic model is essential. When implying more pertinent characteristics to the regression analysis, the total explanatory value (adjusted R^2) will increase, and the coefficient stabilise. Firstly, *physical housing characteristics* are determinants of housing price. In comparable studies, the characteristics, *house size, number of (bed)rooms, building type*, and *building period* are frequently used to determine the value of an individual house.

Secondly, *neighbourhood characteristics* are determinators of the housing price. Visser & Van Dam (2006) distinguish three neighbourhood characteristic categories: *social, physical*, and *functional neighbourhood characteristics*. The *social neighbourhood characteristics* category includes characteristics among which, *population composition, the share of affordable housing*, and *unemployment rate*. The *physical neighbourhood characteristics* category includes characteristics among which, *the amount of water and greening, building quality*, and *quality of public space*. Lastly, the *functional neighbourhood characteristics* category includes characteristics among which, *accessibility* and *distance to* various amenities and the nearby employment opportunities. Comparable studies show that amenities could also have a non-linear impact or a different impact because of their geographical location. Having an amenity in close distance, such as a supermarket or elementary school, could impact a house price positively. However, being too close to such an amenity could adversely impact house prices due to negative externalities, such as noise or traffic (Visser & Van Dam, 2006). Moreover, the impact of *neighbourhood characteristics* is not always generalisable to every geographical location. For example, the *distance to a highway* negatively impact housing prices in urban environments because of various forms of pollution but have a positive impact in rural environments because of accessibility benefits (Visser & Van Dam, 2006).

To give an indication of which variables impact the determination of housing values, Table 1 provides an overview.

Table 1: Determinants of housing prices according to comparable studies

Variable	Impact	Author(s)	Included
Property characteristics			
House size	+	van Duijn et al. (2014)	Yes
Parcel size	+	Knaap & Song (2004)	Yes
Building type	+	Visser & Van Dam (2006)	Yes
	NS	van Duijn et al. (2014)	
Building period	-	van Duijn et al. (2014)	Yes
Number of (bed)rooms	+	van Duijn et al. (2014)	No
Garage	+	van Duijn et al. (2014)	No
Building condition	+	van Duijn et al. (2014)	No
Neighbourhood characteristics			
Population density	NS	van Duijn et al. (2014)	No
Population composition	-	van Duijn et al. (2014)	No
	NS	Knaap & Song (2004)	No
Average income	+	Knaap & Song (2004)	No
Distance to CBD	+	Knaap & Song (2004)	No
Distance to park	+	Visser & Van Dam (2006)	No
Distance to supermarket	- \+	Visser & Van Dam (2006)	No
Distance to highway	- \+	Visser & Van Dam (2006)	No

Note: This table denotes an overview of determinants frequently used in hedonic price models to determine the housing price. *+* : indicates that the literature has found a positive relation to housing price. - : suggests that the literature has found a negative link to housing price. *NS* : suggests that the literature has found a not statistically significant impact on housing price. CBD (Central Business District).

3 Data & methodology

3.1 Hedonic regression model

In this research, a hedonic pricing model is used to analyse the external effects of the development of a data center on the transaction prices of nearby residential units. The hedonic framework is shaped as follows:

$$H = f(P, L, T) \tag{1}$$

Within equation (1), the *transaction price* of the property (H) is determined by various characteristics (f), which include *property characteristics* (P), *location characteristics* (L), and *year characteristics* (T). *The property characteristics* within this research are *house size, parcel size, building period, and property type.* For *location characteristics*, only the variable *city* is used. For *year characteristics*, *transaction year* is used.

To determine the impact of an event on the housing price, a hedonic pricing model can be executed in three ways. Important is that a *characteristics* in a hedonic regression model does not have an individual value but is measured indirectly based on changes in the housing price.
The first manner to execute a hedonic regression analyse is by comparing transaction prices before and after a decisive event. Within this research, the decisive event is the development of a data center. In the year(s) of construction and operationalisation, the development can cause external effects, affecting the transaction prices. However, other developments and trends could contribute to the external effects on property prices too, making this manner inaccurate.

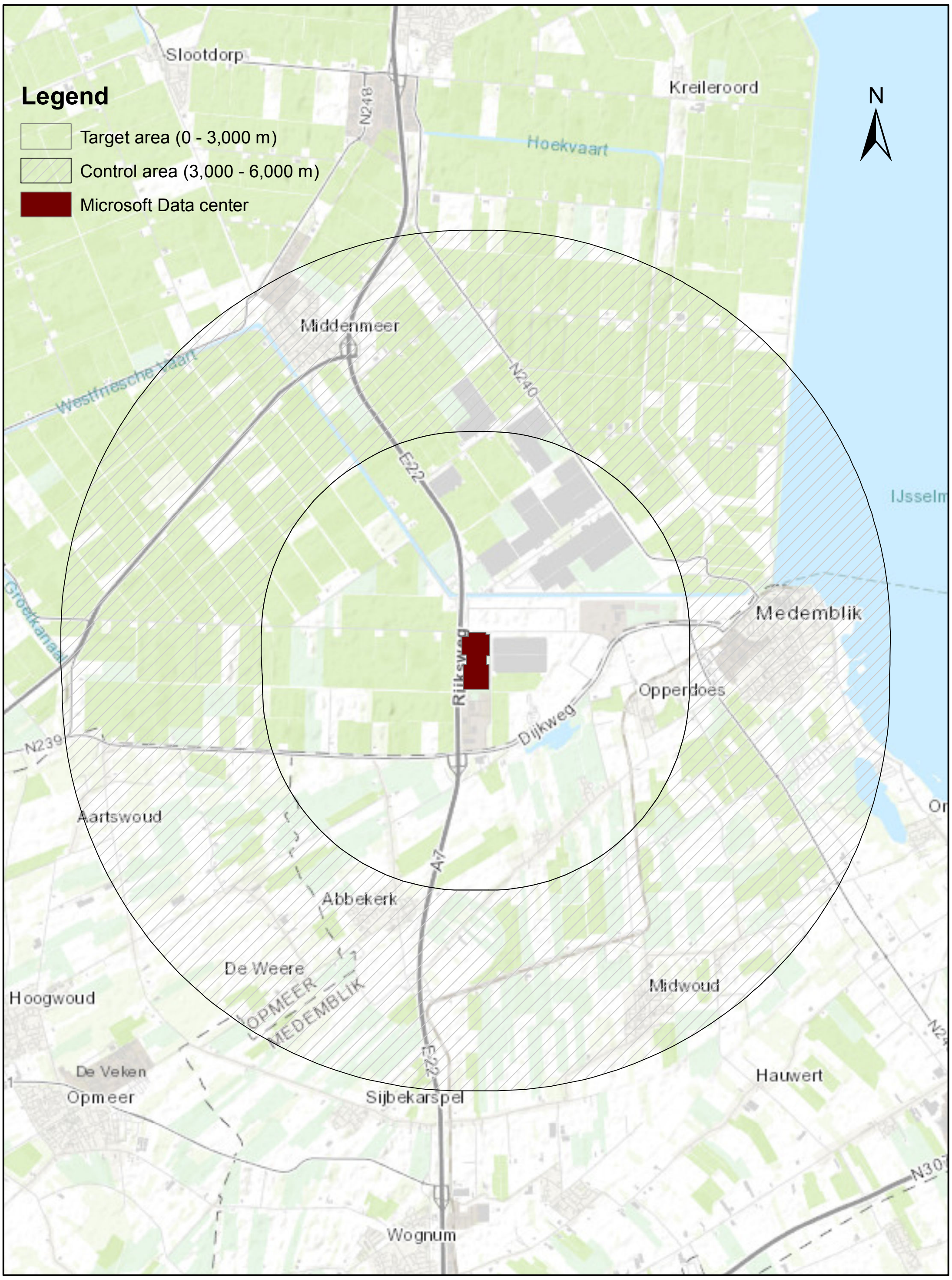
The second manner to execute a hedonic price model is by comparing transaction prices between areas, such as a target area and a control group. The target area is a bundle of observations within an area that experiences the impact of the decisive event. The control area is identical to the target group, except for the event, and consequently does not share an effect on transaction prices. This control area could be a comparable city or the area directly beyond the target area, the outer ring. Using the outer ring as a control area is a frequently used technique to determine the external effect on housing prices (Zhang et al., 2019; Schwartz et al., 2006; van Duijn et al., 2014). That would mean in this research that a target group in which a data center is located is compared with a control area, in which no data center is located. Important to note is that besides the decisive 'event', everything else is identical between the target and control group.

The third manner is to combine the first and second manner, to execute a difference-in-differences analysis. The difference-in-differences technique is used by Schwartz et al. (2006) and compares the selected target group with a control group before and after the event has occurred. The difference-in-differences analysis is frequently used to assess the external impact of an event (an independent variable) on housing prices (a dependent variable), among which transformation of industrial heritage (van Duijn et al., 2014), wind farm development (Sunak & Madlener, 2016), and shopping center redevelopment (Zhang et al., 2019). Besides an independent variable, other control variables are added in the building up process of a regression model. These control variables are usually *property characteristics, location characteristics, transaction characteristics, and asset-specific characteristics.* Idem, the target area and the control area must be comparable, except for the event.

To determine the potential impact of a data center development on housing prices, the outer ring technique is used. Consequently, the selected range for target and control areas are derived by creating five different distance categories *(1: $\leq$ 1000m, 2: $\leq$ 2000m, ..., 5: $\leq$ 5000m).* By performing a regression on these five distance categories, with category 5 as reference category, the border between target and control area could be determined. The regression results in Appendix II, show that the categories *1*, *2* and *3* all reported significant values, suggesting that the distance to a data center plays a role in impact on the dependent variable *TP*. However, the coefficient of category 4 ($\geq$ 3000m - $\leq$ 4000m) is not statistically significant, suggesting that the distance to a data center does not impact the dependent variable *TP*. Therefore, the target area was selected for all transaction observation until 3,000 meters. All transaction observations located beyond the range of 3,000 meters from the data center are pooled in the control area, ending at a 6,000 meters distance.

The variable distance to the data center measures if an individual transaction observation belongs to the target area or the control area. This is done by measuring the Euclidean distance between every individual transaction observation and the edge of a selected data center's polygon by using the geographical program GIS. In figure 4, the shape of the target area and control area is feasible around the selected data center. In figure 4, the first ring around the Microsoft data center in the 'De Wieringermeer' indicates the target area, the second ring *(striped)* indicates the control area. Using a polygon instead of coordinates reduces the measurement errors in the distance from a data center to individual transaction observations. Within our selection of data centers, this is especially important for the Microsoft data center in 'De Wieringermeer' due to its large size and stretched shape. A similar technique is used in the study of Zhang et al. (2019).

Figure 4: Data center polygon



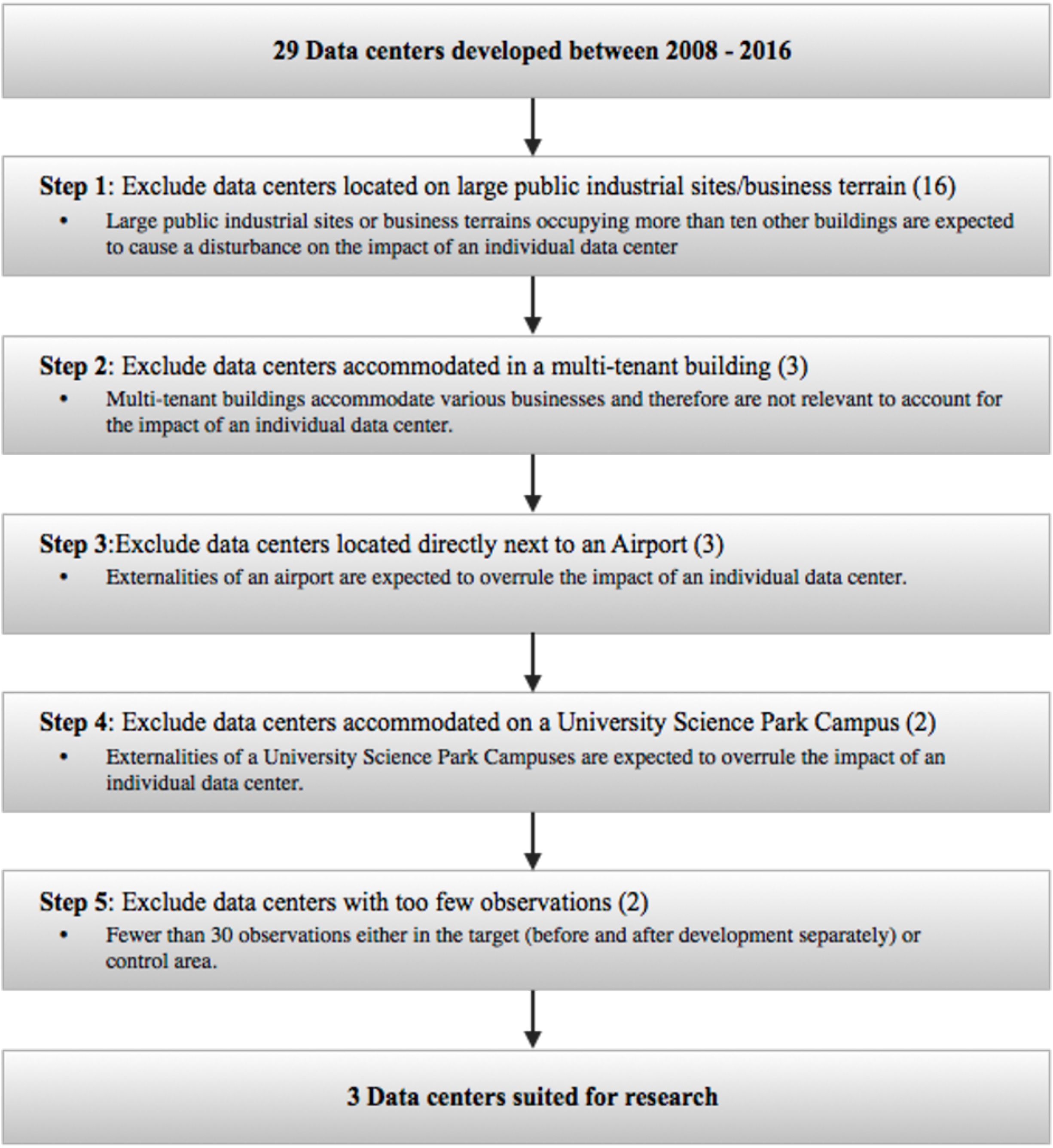
Note: Polygon around the selected data center Microsoft 'De Wieringermeer' with a buffer for target area (0 - 3,000 meter) and control area (3,000 - 6,000 meter) *(striped)*.

3.2 Data selection

The data used for this research comes from multiple sources. First, the residential transaction observations come from the NVM database (NVM, 2021). Access to the NVM database was granted by an NVM member. Secondly, all data centers in the Netherlands are listed in a database by the Dutch Datacenter Association (DDA), Datacenterplatform, and Data Center Map. All databases are combined to control for each other since the DDA only reports data centers that are bound to their association (113) (DDA, 2021), and Datacenterplatform (111) (Dutchdatacenterplatform, 2021) Data Center Map (101) (Data Center Map, 2021) had some missing observations. Combining all three databases shows that the Netherlands holds 125 external data centers (excl. Cellnex media towers (24)). The 24 media towers of Cellnex are excluded from this research since their visual appearance is significantly different from the usual data centers.

The selection of data centers is based on the following: Among these 125 data centers, all data center development before 2008 and after 2016 are excluded (96) because there are not enough years of residential transaction observations to measure an impact before and after the data center development. However, not all 29 data centres are suited for this research. Figure 5 shows the five-step selection procedure to derive a final list of data centers appropriate for this research. The selection process is focused on including the data centers that provide an individual impact. In the end, this will contribute to the relevance of this research. Consequently, all data centers where the spatial environment could cause a disturbance of the individual impact of a data center on housing prices are excluded from this research.

Figure 5: Flow chart of data selection procedure



Note: The selection procedure of the data selection is derived in five steps. The selection aims only to include the data centers that have an individual impact on housing prices and are not disturbed by other effects. Each step briefly indicates the reason for exclusion. The number of data centers that are excluded during each step is shown between parenthesis. Eventually, 3 data centers are selected to include in this research.

Table 2 presents the 3 data centers selected for this research, namely *COLT Roosendaal, Microsoft 'De Wieringermeer'* and *NorthC Groningen.* Since only 3 data centers made the selection, Appendix III provides more insides on each data center.

Table 2: Data center characteristics of the selected data centers

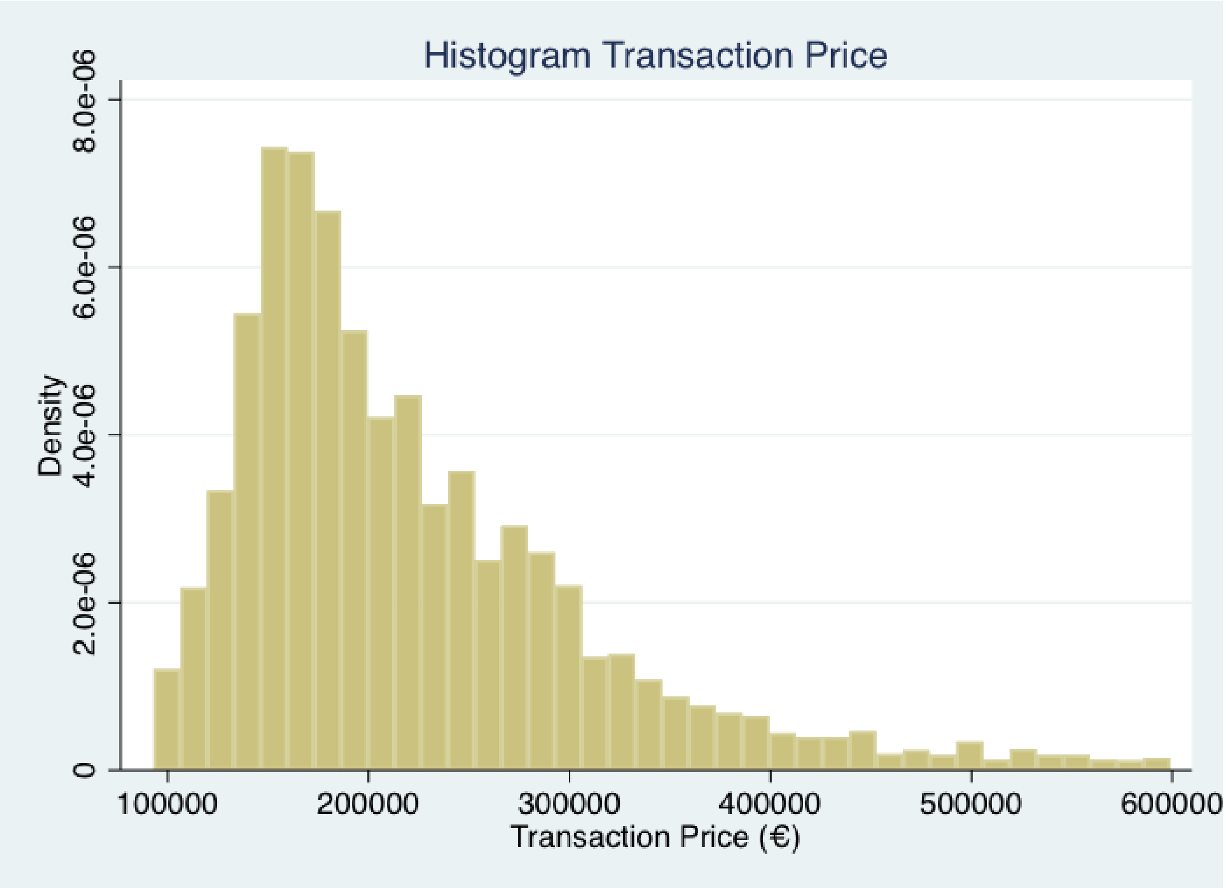
Data center	Year of operation	Size (m^2)	Trans. years incl.	% of obs.
COLT Roosendaal	2013	17,900	2008 - 2017	42.31%
Microsoft Wieringermeer	2014	330,000	2005 - 2020	21.40%
NorthC Groningen	2015	4,000	2005 - 2020	36.29%

Note: This table denotes the overview of the selected data centers Source: Dutch Datacenter Association.

The selection of these 3 data centers led to a total of 6,811 individual residential transaction observations between 2005 – 2020. However, not all observations are suited to use in this research since the data set of NVM includes a small number of missing values or incorrect observations. Incorrect or missing values can disturb the research results and therefore be deleted. However, transaction observations with incorrect or missing property characteristics were adjusted manually (10) if complete transaction information was available of an identical house next door. Ultimately, 57 transactions are deleted due to uncertain missing values.

Moreover, outliers are removed from the data set since they do not represent the population (Hair et al., 2010). The difference between outliers and incorrect observation is that an outlier is a correct observation. Nevertheless, outliers can be removed since they do not represent the population and can cause a disturbance of the results. In the study of Zhang et al. (2019), outliers on the dependent variable, *transaction price*, smaller than the 1st and larger than the 99th percentile, are deleted from the data set. Here, transactions with sales price below €92,988 are deleted (67) and above €599,001 (68) are removed. Figure 6 is a visualisation of the dependent variable *transaction price, (TP)* after deleting the outliers. The distribution of the variable is skewed to the right. Therefore, the dependent variable *transaction price (TP)* needs to be transformed into a logarithm.

Figure 6: Histogram of dependent variable *transaction price*



Note: Histogram of the dependent variable *transaction price* in Euro after deleting outliers. The distribution is skewed to the right.

Additionally, two other control variables show non-normal distributions too. The variable *house size* shows non-normal distribution (rightly skewed). Therefore, transaction observations larger than 300m^2 are deleted (24). Besides *house size*, the highest 1% ($\geq$ 3610m^2) of the variable *parcel size* are deleted too (67) because of a non-normal distribution (rightly skewed). Eventually, 6.528 observations are included within the hedonic regression analysis.

3.3 Descriptive statistics

Table 3 contains the descriptive statistics of the dependent variable and independent variables. The variables are displayed by: *Name*, *number of observations* (Numb. of obs.), *the mean*, *the standard deviation* (St. dev.), *minimum value* (Min.), and *maximum value* (Max).

The dependent variable is *TP* and comprises the transaction price for an individual residential unit. *TP* is a continuous variable, measured in Euro and collected from the data set of the NVM. Table 3 shows that the mean of *TP*, the average transaction price, is equal to €217,779.60. In appendix IV, the development of the average transaction price for both the target and control area is visible. Based on Zhang et al. (2019) and van Duijn (2014), *TP* is the dependent variable instead of the *TP per square meter*. The dependent variable *TP* is consciously not corrected by the *House Price Index* (HPI) of the CBS (2021a) because by adding the variable, *transaction year* (Model 4), time-fixed effects are controlled.

The first independent variable is the *distance*. The variable is measured by the Euclidean distance between an individual transaction observation and the edge of the polygon of the nearest data center. The variable is a continuous variable, measured in meters and derived by the geographical program GIS. The mean distance to a data center is 3,533.56 meters. The variable determines if an individual transaction observation belongs to the target area or the control area.

The following variable is the independent variable *house size*. *House size* comprises the size of the house. The variable is continuous, is measured in square meters and belongs to the category of the *property characteristics*. The average house size is equal to 119.58 m^2, equal to the Dutch average of 119 m^2 (CBS, 2021b). However, the majority of transactions included in this research are Single-family houses located in rural areas.

Parcel size comprises the size of a parcel of an individual property. The variable is a continuous variable, measured in square meters and belongs to the category of the property characteristics. The average parcel size is 258.57 m^2 and obtained from the NVM data set.

The *building type* comprises the property category, according to the NVM. The variable is categorical, meaning that each observation can take one of a limited, fixed number of possible values. The original NVM data set included ten different property types. However, for this research, the highly comparable categories are merged to limit the variety. Therefore, the following categories are defined: *Apartment, Detached, Townhouse*, and *Single-family*. The categorisation of the variable *building type* is used frequently in comparable studies (Zhang et al., 2019). Every category is a dummy variable, which means they can either be equal to 0 or 1. For example, if an observation belongs to the category Apartment, the dummy *building type Apartment* equals 1. Automatically, all other building types are equal to 0 for that observation. Table 3 clarifies that the *building type Single-family* is highly represented (76%) in this research. The selected data centers are located in rural areas, where Single-family houses are built more frequently.

The *building period* comprises the period of construction of an observation. The variable is derived from the NVM data set as building year. However, building year is not suited to use within a regression analysis, and therefore the variable is transformed into the categorical variable *building period*. Therefore, the *building period* variable is divided into the following category dummies: < *1905, 1905 – 1944, 1945 – 1969, 1970 – 1989, 1990 – 2000,* > *2000*. Noticeable in the descriptive table is that the building period dummies are evenly distributed, except for the category < *1905*, which corresponds to 2.1% of the transaction observations.

For *location characteristic*, the variable *city* is included. The variable is a category variable, including 18 different cities. By incorporating *location characteristics*, location-fixed effects are controlled.

For *transaction characteristics*, the research includes the variable *transaction year*. The variable is a category variable, including 11 different transaction years (2005 – 2020). By incorporating *transaction year*, time-fixed effects are controlled. The *transaction year* comes from the NVM data set by every individual transaction observation. The variable is for similar reasons as *city* not included in Table 3.

Table 3: Descriptive statistics

Variable	Mean	St. Dev.	Min.	Max.
Dependent variable				
Transacation price	217,779.6	84,610.48	92,988	599,000.3
Independent variable				
Distance to data center	3,533.56	955.088	448.74	5,998.18
House size	119.58	32.96	40	300
Parcel size	258.57	317.37	0	3610
Building type apartment	0.1438	0.3510	0	1
Building type detached	0.0450	0.2074	0	1
Building type townhouse	0.0493	0.2166	0	1
Building type Single Family	0.7618	0.4260	0	1
Building period $\leq$ 1904	0.0211	0.1439	0	1
Building period 1905 - 1944	0.1236	0.3292	0	1
Building period 1945 - 1969	0.2169	0.4122	0	1
Building period 1970 - 1989	0.2384	0.4227	0	1
Building period 1990 - 2000	0.1999	0.4000	0	1
Building period $\geq$ 2001	0.2056	0.4042	0	1
Numb. of obs.	6,528			

Note: This table denotes the descriptive statistics of the dependent variabele *TP*, independent variables *property characteristics*. Other variables are not reported. *Mean*: Average value; *St. Dev.*: Standard deviation; *Min.*: Minimum value; *Max.*: Maximum value. The mean reported for *Building type* and *Building type* indicate the percentage relative to 1.0 = 100%

In Tables 4 & 5, the descriptive statistics are divided between the target area (0 – 3,000 m) and the control area (3,000 – 6,000 m) and analysed. In a difference-in-differences method, the observations within the target and control group are identical, except for the specific event (data center development). Comparing Table 4 with Table 5 shows that there are differences between the target and control area. So is the dependent variable *TP* is higher in the target area. This could be explained by the *building types* and the *parcel size*. The target area shows a high percentage of single-family units and a larger average parcel size than the control area. The difference in *TP* is not even a quarter of the standard deviation of the target area, so the target and control area are still comparable. However, the target area has significantly less *building type apartments* compared to the control area. This difference can have a stimulating impact on *TP* since apartments usually have lower transaction prices. In addition, the target area holds significantly more pre-war homes (*building period 1905 - 1944*). Pre-war homes could be in a worse condition, impacting the *TP* of the target area negatively. Another difference is that the target area has fewer observations than the control areas. This can be explained by the fact that data centers develop on sites categorised as *'Business terrain'* or *'Business terrain - Industrial.* These sites lay, in general, further away from denser populated residential areas because of statutory regulations in the zoning plan to reduce negative external effects, such as noise, light and air pollution. Despite the differences between the target and control areas, the available data is appropriate for this research. However, the differences must be noted.

Table 4: Descriptive statistics for the target area 0 - 3,000 m

Variable	Mean	St. Dev.	Min.	Max.
Dependent variable				
Transacation price	229,766.6	9,7476.68	9,3520	598,999.5
Independent variable				
Distance to data center	2,298.103	504.07	448.74	2,999.90
House size	123.2768	36.5129	44	300
Parcel size	349.4667	403.2973	0	3610
Building type Apartment	0.0701	0.2553	0	1
Building type Detached	0.0712	0.2581	0	1
Building type Townhouse	0.0429	0.2028	0	1
Building type Single-family	0.8153	0.3882	0	1
Building period $\leq$ 1904	0.0305	0.1720	0	1
Building period 1905 - 1944	0.2288	0.4201	0	1
Building period 1945 - 1969	0.1525	0.3596	0	1
Building period 1970 - 1989	0.2729	0.4456	0	1
Building period 1990 - 2000	0.1746	0.3797	0	1
Building period $\geq$ 2001	0.1407	0.3478	0	1
Numb. of obs.	1,770			

Note: This table denotes the descriptive statistics of the target area which hold all transaction observation between 0 and 3,000 meters distance to a data center. The dependent variabele *TP*, independent variables *property characteristics*. Other variables are not reported. *Mean*: Average value; *St. Dev.*: Standard deviation; *Min.*: Minimum value; *Max.*: Maximum value. The mean reported for *Building type* and *Building type* indicate the percentage relative to 1.0 = 100%

Table 5: Descriptive statistics for the control area 3,000 - 6,000 m

Variable	Mean	St. Dev.	Min.	Max.
Dependent variable				
Transacation price	213,320.4	78,840.8	92,988	599,000.3
Independent variable				
Distance to data center	3,993.294	614.6457	3,000.67	5,998.184
House size	118.2074	31.4304	40	300
Parcel size	224.7619	271.1059	0	3,565
Building type Apartment	0.1713	0.3768	0	1
Building type Detached	0.0351	0.1840	0	1
Building type Townhouse	0.0517	0.2214	0	1
Building type Single-family	0.7419	0.4376	0	1
Building period $\leq$ 1904	0.0177	0.1317	0	1
Building period 1905 - 1944	0.0845	0.2781	0	1
Building period 1945 - 1969	0.2409	0.4276	0	1
Building period 1970 - 1989	0.2179	0.4129	0	1
Building period 1990 - 2000	0.2093	0.4069	0	1
Building period $\geq$ 2001	0.2297	0.4207	0	1
Numb. of obs.	4,758			

Note: This table denotes the descriptive statistics of the control area which hold all transaction observation between a 3,000 and 6,000 meters distance to a data center. The dependent variabele *TP*, independent variables *property characteristics*. Other variables are not reported. *Mean*: Average value; *St. Dev.*: Standard deviation; *Min.*: Minimum value; *Max.*: Maximum value. The mean reported for *Building type* and *Building type* indicate the percentage relative to 1.0 = 100%

3.4 Regression transformation

All variables must be in their best form to guarantee the most significant coefficient within our regression analyses. If a variable performs a not statistically significant value, transforming the variable can help. Moreover, transformation ensures that variables are in line with the assumptions for regressions methodology.

The variables are controlled for normal distributions by executing a histogram. When the histogram visualises critical skewness, the variable is not normally distributed. Therefore, transforming the variable is necessary. The dependent variable, *TP*, is the only variable with critical skewness and, therefore, not normally distributed. Consequently, *TP* is transformed into a logarithmic function. A logarithmic transformation of the dependent variable frequently occurs when determining house prices (Zhang et al., 2019; van Duijn et al., 2014). In appendix V, a histogram is plotted, showing a normal distribution after transformation.

3.5 Regression models

Within this research, various models are used to identify the external impacts of a data center by considering housing values. Multiple difference-in-difference models are used to investigate the external impacts of data center development on housing values.

Table 6: Overview regression models

Variable	Model 0	Model 1	Model 2	Model 3	Model 4
Target		X	X	X	X
Post		X	X	X	X
Target×Post		X	X	X	X
Target×Post×Distance		X	X	X	X
Property characteristics	X		X	X	X
Location characteristic	X			X	X
Transaction characteristic	X				X

Note: This table denotes which variables are included in the models 0, 1, 2, 3 & 4. The X denotes if a variable is included in the model.

In Table 6, the structure of the different models is shown. In totality, there are four models examined with a difference-in-differences method. Besides the four models that make use of a difference-in-differences method, there is also a Model 0. The model is built up by adding variables, which has an increasing effect on the explanatory impact of the dependent variable. Model 0 is there to control for the explanatory impact of *property characteristics, location* and *time characteristics.* In Model 1, only the key variables, *Target, Post, Target×Post* and *Target×Post×Distance* are included. In Model 2, the property characteristics, *property characteristics*, namely *house size, parcel size, building type*, and *building period* are added. In Model 3, the *location characteristic city* is added. In Model 4, the complete model, the *transaction characteristic transaction year* is added. The formula of Model 4 is as follows:

$$\log(P_i) = \beta_0 + \beta_1 Housesize_i + \beta_2 Parcelsize_i + \beta_p \sum_{p=1}^{4} Buildingtype_{ip} + \beta_t \sum_{t=2}^{6} Buildingperiod_{it} + \beta_c \sum_{c=1}^{17} City_{ic} + \beta_T \sum_{T=1}^{16} Transactionyear_{iT} + \beta_3 Target_i + \beta_4 Post_i + \beta_5 Target_i \times Post_i + \beta_6 Target_i \times Post_i \times Distance_i + \epsilon_i \quad (2)$$

Within formula (2), the $log(P_i)$ is the log of the is the log of the transaction price of the residential transaction i; $Housesize_i$ is a continuous variable, and measures the house size of the residential transaction i in meters; $Parcelsize_i$ is a continuous variable, and measures the house size of the residential transaction i in meters; $Buildingtype_{ip}$ is a vector dummy with 4 categories (*Apartment, Detached, Townhouse, Single-family*), with a value equal to 1 if residential transaction is categorised as building type p and otherwise a value equal to 0; $Buildingperiod_{it}$ is a vector dummy with 6 categories (*<1905, 1905 – 1944, 1945 – 1969, 1970 – 1989, 1990 – 2000, >2000*), with a value equal to 1 if residential transaction i is built in period t and otherwise a value equal to 0; $City_{ic}$ is a vector dummy with 17 categories, with a value equal to 1 if residential transaction i is located within city c and otherwise a value equal to 0; $Transactionyear_{iT}$ is a vector dummy with 16 categories (2005 – 2020), with a value equal to 1 if residential transaction i is sold in transaction year T and otherwise a value equal to 0; $Target_i$ is a dummy variable which indicate if residential transaction observation i is located in the selected target area (0 – 3,000m) and therefore has a value equal to 1, or if located in the control area (3,000 – 6,000m) valued equal to 0; $Post_i$ is a dummy variable which indicate if residential transaction observation i is sold after the year of operation of data center, and therefore has a value equal to 1, or if residential transaction i was sold before year of operation of data center valued equal to 0; $Target_i \times Post_i$ is the main variable of interest. $Target_i \times Post_i$ is a dummy variable of the interaction between $Target_i$ and $Posti$, which has a value of 1 if residential transaction i were both located in the selected target area and sold after development of nearby data center, zero otherwise. The variable measures the external impact of the development of a data center on transaction prices in the selected target area; $Target_i \times Post_i \times Distance_i$ is a continuous variable of the interaction between $Target_i$, $Post_i$ and $Distance_i$ and measures the distance of a residential transaction i when located within the selected target area after development; ϵ_i is the error term; and $\beta_{(0-6,p,t,c,T)}$ are the coefficient that will be estimated.

4 Results

4.1 Impact of data center development on transaction prices

This chapter reports the regression results of the difference-in-differences hedonic regression analysis. The coefficients interpret the possible external impact of the development of data centers on nearby housing values and the magnitude of the external impact across space and time. In all models, the dependent variable is the natural logarithm of the transaction price. The interpretation of the coefficients following Halvordsen & Palmquist (1980), $(e^{\text{coefficient}} - 1) \times 100)$. The result of Table 7, Model 4, help to answer RQ 3: *'What is the quantitative impact of developing a data center on the transaction prices of nearby residential units?'*

4.1.1 Interpretation of regression results

In Table 7, it is visible that Model 4 includes all variables. Starting at the top, the coefficient on $Target_i$ reports a positive but not statistically significant value. This indicates that there is no price difference measurable between the target area and the control area. The following variable is $Post_i$ which shows a substantial decrease in its coefficient when controlled for time-fixed effects. However, the coefficient of 0.0497 is positive and significant, at 1 percent. This means that transactions of residential units after the development of a data center are on average sold $(e^{0.0497} - 1) \times 100) = 5.10\%$ higher than residential units sold before the data center development. This suggests that data center development results in positive external impacts, which can be caused by employment and local economic growth. However, this variable says nothing over the distance to a data center. Therefore, variables need to interact.

The key variable is the interaction variable $Target_i \times Post_i$. The variable is equal to 1 if the property i is located in the selected target area and sold after the development of a data center, otherwise zero. The variable measures the external impact of the development of the data center on the transaction prices of residential units in the target area. The variable shows a positive coefficient of 0.0514, at a 10 percent significance level. This indicates that the development of a data center generates a 5.27% increase in transaction prices on average relative to residential units in the control area. This positive impact on the transaction price suggests that Dutch spatial planning policies have worked successfully to ensure no negative externalities occur for nearby residents and even stimulate the transaction prices of nearby residential units. As explained, in the research phase of the zoning plan, the local government investigates the external impacts of land uses. This research aims to stimulate positive external impacts and minimise negative external impacts. Consequently, a building permit will only be issued when a development is in line with strict statutory regulations. In addition, data centers development could spur additional investment and development. Examples can vary from infrastructural improvements, installation of optical fibre, and connection to the residual heat network. Therefore, residential areas nearby a data center could experience a positive impact on transaction prices.

The last variable, $Target_i \times Post_i \times Distance_i$, determines the impact of a data center over distance. The variable's coefficient is negative, indicating that the price difference becomes greater for properties located further from the data center development within the target area. However, the coefficient shows a not statistically significant value. Thus, it cannot be assumed that there is a price difference measurable within the target area as distance increases.

Subsequently, Table 7 shows the adjusted R^2 for Model 4, which is 74.49%. The adjusted R^2 indicates the percentage to which the dependent variable is explained by other variables, controlled by the number of variables used. Desired is that the adjusted R^2 is as high as possible, meaning that all control variables together explain as much of the dependent variable *TP*. A percentage of 74.49% is satisfying but could be increased when adding more relevant variables to the model. Table 7 shows that by adding relevant variables to the model, the adjusted R^2 increases in value.

Table 7: Difference-in-differences regression results

Variable	Model 1	Model 2	Model 3	Model 4
Target	0.0593***	0.0116	0.0203**	0.0025
	(0.0152)	(0.0088)	(0.0091)	(0.0085)
Post	0.0614***	0.1091***	0.1060***	0.0497***
	(0.0102)	(0.0058)	(0.0058)	(0.0144)
Target×Post	0.0385	-0.0379	0.0062	0.0514*
	(0.0525)	(0.0113)	(0.0298)	(0.0102)
Target×Post×Distance	-2.0e-05	-6.3e-06	-1.1e-0.5	-1.8e-05
	(2.1e-05)	(1.2e-05)	(1.2e-05)	(1.2e-05)
Property characteristics	No	Yes	Yes	Yes
Location characteristic	No	No	Yes	Yes
Year characteristic	No	No	No	Yes
Adjusted R^2	0.0121	0.6823	0.7041	0.7449

Note: This table denotes difference-in-differences regression results. The dependent variable is the log of Transaction Price, *TP*. Standard errors are in parentheses. Property characteristics includes *house size, parcel size, building type* and *building period*. Location characteristic includes *city*. Year characteristic includes *transaction year*. $^*p\leq0.1$, $^{**}p\leq0.05$, $^{***}p\leq0.001$

Based on the results reported in Model 4, an answer can be formulated on RQ 3: ***'What is the quantitative impact of developing a data center on the transaction prices of nearby residential units?'***

The development of a data center on the transaction prices of nearby residential units shows a positive impact of 5.27% relative to residential units in the control area. The result stays contrary to the used literature, which assumed a negative impact on residential units. A possible explanation for this impact could be related to the Dutch spatial planning policies. A positive impact suggests that the Dutch spatial planning policies worked successfully and contributed to this impact by careful decision location and strict statutory regulations. Moreover, data centers development could spur additional investment and development, which result in a positive impact on transaction prices of nearby residential units.

4.2 Sensitivity analysis

To test the robustness of the above regression results, a sensitivity analysis can be executed. In the initial difference-in-differences regression analysis, the range of the target area (0 - 3,000 m) was selected based on a regression analysis to control the distance of a data center on the transaction prices of residential units. Now, a regression is performed for two alternative target areas. Table 8 shows the regression results for a target area of 0 – 2,500 meters, and in Table 9, the regression results of a target area of 0 – 3,500 meters are visual. This specific sensitivity analysis is based on a study of Den Hertog (2019), which uses, similar to this study, relatively large target areas compared to other, more urban area focused difference-in-differences regression analysis. By shifting the target area, the regression results determine if the results in the initial regression (Table 7) are robust.

In Table 8, Model 4, the key variable $Target_i \times Post_i$ shows a not statistically significant impact. An interesting finding since the original target area showed a positive significant impact (5.27%). An explanation for a not statistically significant coefficient can be limited transactions observations in the target area. Since data centers are located in more rural areas, fewer transaction observations are included (995 compared to 1,770) when the range of the target area is reduced. Because of a not statistically significant coefficient for this target area (0 - 2,500 m), it can not be concluded that there is a price difference between the target and control area after data center development. In addition, a not statistically significant impact could indicate that spatial planning policies ensured that there are no negative externalities measured on housing prices.

In Table 9, Model 4, the key variable $Target_i \times Post_i$ shows a positive and significant impact of 0.0461 at 5 percent. This result is in line with Table 7. It indicates that the development of a data center generates on average a 4.72% increase in transaction prices relative to residential units in the control area. This suggests that Dutch spatial planning policies work correct and ensure no negative external impact on nearby housing prices. Moreover, the magnitude decreased slightly relative to Table 7. The target area increased, and therefore the positive impact decreased. In addition, the variable $Target_i \times Post_i \times Distance_i$ become significant, indicating that transaction price decreases when distance increases. This suggests that located further away from the data center, positive external impact decrease. This could be explained since residential units further away from a data center experience less positive externalities of additional investments and development near the data center.

Based on coefficients for the sensitivity analyses in Table 8 & 9, it can be concluded that the regression results for the target area (0 - 3,000 m) are quite robust.

Table 8: Regression results sensitivity analysis: Target area 0 - 2,500 m

Variable	Model 1	Model 2	Model 3	Model 4
Target	0.0250	0.0205*	0.0403***	0.0138
	(0.0196)	(0.0112)	(0.0111)	(0.0104)
Post	0.0621***	0.1089***	0.1076***	0.0531***
	(0.00945)	(0.0054)	(0.0054)	(0.0143)
Target×Post	0.2299***	-0.0391	0.0121	0.0520
	(0.0710)	(0.0142)	(0.0397)	(0.0370)
Target×Post×Distance	-1.2e-04***	-9.9e07	-2.8e-0.5	-2.8e-05
	(3.4e-05)	(1.9e-05)	(1.9e-05)	(1.8e-05)
Property characteristics	No	Yes	Yes	Yes
Location characteristics	No	No	Yes	Yes
Year characteristics	No	No	No	Yes
Adjusted R^2	0.0096	0.6825	0.7046	0.7449

Table 9: Regression results sensitivity analysis: Target area 0 - 3,500 m

Variable	Model 1	Model 2	Model 3	Model 4
Target	0.1088***	0.0037	0.0273***	0.0029
	(0.0131)	(0.0077)	(0.0081)	(0.0076)
Post	0.0864***	0.1198***	0.1176***	0.0535***
	(0.0118)	(0.0067)	(0.0068)	(0.0147)
Target ×Post	-0.0073	-0.0142	-0.0069	0.0461**
	(0.0397)	(0.0100)	(0.0230)	(0.0215)
Target×Post×Distance	-1.9e-05	-8.3e-06	-1.1e-0.5	-1.9e-05***
	(1.4e-05)	(6.1e-06)	(6.1e-06)	(7.2e-06)
Property characteristics	No	Yes	Yes	Yes
Location characteristic	No	No	Yes	Yes
Year characteristic	No	No	No	Yes
Adjusted R^2	0.0208	0.6833	0.7047	0.7450

Note: Both Table 8 & 9 table denotes the results of a difference-in-differences regression analysis. The dependent variable is the log of *Transaction Price*, *TP*. Standard errors are in parentheses. Property characteristics include House size, Parcel size, Building type and *building period*. Location characteristic includes *city*. Transaction characteristic includes *transaction year*. *p≤0.1, **p≤0.05, ***p≤0.001

4.2.1 Discussion on results and literature

The regression results show a positive impact on transaction prices, contradicting the literature used in chapter 2, theory. First, the regression results are adverse to De Vor & De Groot (2011). Their study shows a negative impact of industrial sites on housing prices, indicating that housing prices increase when located further away from an industrial site. The negative externalities that frequently appear with industrial sites, such as noise, traffic and obstruction of view, cause a negative impact on housing prices. This suggests that data centers are not valued equally to industrial sites.

Secondly, the results are adverse to the literature of Sims (2002) and Vyn & McCullough (2014) on the impact of NIMBYism on housing prices. On the one hand, Sims (2002) results indicate that a negative stigma around power lines could have a negative impact on housing prices. On the other, Vyn & McCullough (2014) found a not statistically significant impact, despite the public perception against windmills. Similar to both studies, NIMBYism is visible to data center development. Nevertheless, it seems that it does not negatively impact housing prices in the case of data center development in the Netherlands.

4.3 Impact of data centers across building types

To further analyse the robustness of the regression results a heterogeneity test can be executed. A heterogeneity test is another form of sensitivity analysis and examines if data center development has the same impact across a category compared to the *pooled model (Model 4)*. For this test, the variable *building type* has been chosen as a variable to test for heterogeneity since it was impossible to test other, more categories, such as development size or urban and rural location, due to too few observations. Nevertheless, performing a heterogeneity test on building type is previously performed by Xiao et al. (2016). Therefore, the fourth research question is formulated: *'To what extent does the quantitative impact of the development of a data center change across building types?'*

First, to test for heterogeneity, a null hypothesis is formulated: *'The quantitative impact of the development of a data center remains equal across different building types?'*. The null hypothesis argues that all categories will report an equal impact, so there is no heterogeneity visible. The second step is to test the hypothesis by performing a 'Chow test'. In appendix VI, the Chow test shows a significant value at 1 percent, meaning that the null hypothesis, which assumes no heterogeneity across building types, can be rejected. Therefore, it can be concluded that there is a significant difference in the impact of data center development on building types. This suggests that for building types, the coefficients of separate models are rather interpreted than the coefficient of the pooled model (Model 4).

In Table 10, the coefficients clearly emphasise the heterogeneity since the coefficient across building types vary strongly. For the key variable $Target_i \times Post_i$, all coefficients report a positive value, indicating that the transaction prices have increased after the development of a data center in the target is, relative to the control area. However, no building category reports a significant value. Therefore it can be assumed that there is no evidence for a difference in transaction price of a building type individually, before and after the development of a data center. A similar impact is visible for the interaction variable, $Target_i \times Post_i \times Distance_i$, where all building categories report a negative coefficient, indicating a decrease in transaction price when the distance to a data center increases. However, no building category shows a significant value.

Table 10: Heterogeneity test on variable Building type

Variable	Pooled Model	Apartment	Detached	Townhouse	Single-Family
Target	0.0025	-0.0633**	0.0820*	0.0273	0.0076
	(0.0085))	(0.0267)	(0.00492)	(0.0283)	(0.0092)
Post	0.0497***	0.0498	0.05142	-0.1702*	0.0493***
	(0.0144)	(0.0331)	(0.0974)	(0.0899)	(0.0155)
Target×Post	0.0514*	0.0512	0.0998	0.1045	0.0409
	(0.0278)	(0.1024)	(0.1010)	(0.1551)	(0.0301)
Target×Post×Distance	-1.8e-05	-2.9e-05	-5.5e-0.5	-4.4e-05	-9.8e-06
	(1.1e-05)	(4.5e-05)	(3.5e-05)	(6.0e-05)	(1.2e-05)
Property characteristics	Yes	Yes	Yes	Yes	Yes
Location characteristic	Yes	Yes	Yes	Yes	Yes
Year characteristic	Yes	Yes	Yes	Yes	Yes
Observations	6,528	939	294	322	4,973
Adjusted R^2	0.7449	0.8035	0.6570	0.6069	0.6996

Note: The dependent variable is the log of Transaction Price, TP. Standard errors are in parentheses. Property characteristics includes House size, Parcel size, Building type and building period. Location characteristic includes city name. Transaction characteristic includes Transaction year. *p≤0.1, **p≤0.05, ***p≤0.001

Based on the Chow test and the coefficient in Table 10, an answer could be formulated on the fourth research question: ***'To what extent does the quantitative impact of the development of a data center change across building types?'***

With the significant results of the Chow test, the hypothesis, assuming no heterogeneity across building types, could be rejected. All building types report a positive impact of data center development on nearby residential transaction prices relative to the control area. However, none of these coefficients is significant.

5 Conclusion & discussion

5.1 Main findings

This research investigates the external impact of data center development on the transaction prices of nearby residential units in the Netherlands by exploiting information on data center development. Data center development is a topic of discussion since residents and institutions make obligations against further development. According to the opposition, the development will contribute to further landscape transition, several negative environmental impacts, and residents felt uninvolved and unheard in the development plans. However, it is unknown what the impact of data center development on housing prices is. Investigating this would be of great interest to all stakeholders in the process of future data center development. Therefore, the following central research question is formulated: 'Does the proximity of a newly developed data center impact nearby housing prices?'

First, the location determinants of a data center are clarified to answer the central research question. Based on the location theories, the private sector prefers a compromise between low latency, availability of space, cost efficiency regarding resources, and agglomeration benefits due to data center clustering. However, the commitment of the public sector is the determining factor. No building permit will be issued when development is not in line with the zoning plan and the statutory regulations. Thereby, the local government ensures that negative externalities for the built environment are minimised.

Secondly, data centers have strong similarities in their zoning plan with industrial sites and business terrains. Contrary to data centers, the external impact of industrial sites is known and frequently researched. The literature reports a negative effect between industrial sites and housing prices. This indicates that the price of a house increases when the distance to an industrial site rises, due to negative externalities, like noise, traffic and visual pollution. Nevertheless, this does not automatically indicate that this impact is generalisable to data centers.

A difference-in-differences regression method is executed to quantify the impact of the development of a data center on the transaction price of nearby residential units. According to the results, the development of a data center has, generates on average, a 5.27% increase in transaction prices relative to residential units in the control area. This suggests that Dutch spatial planning policies work successfully. Spatial planning aims to ensure no negative external impact of data center development is experienced on nearby residential transaction prices. Only then a building permit would be issued by the responsible public institution.

Finally, the results are tested for their robustness by a heterogeneity test across building types since heterogeneity of external impact could occur. According to the executed Chow test, the hypothesis which assumed no heterogeneity is rejected. This means that there is heterogeneity in the impact of data center development across building types. All building types report a positive impact of data center development on nearby residential transaction prices. However, none of these building types shows a significant coefficient.

Thus, to answer the central research question. This research can conclude that proximity to a newly developed data center positively impacts housing prices. Important to note is that this research is based on a small selection of rural located data centers in the Netherlands. Therefore this conclusion is not generalisable to all data centers globally. It can be expected that more urban locations, smaller/larger/higher, or more energy-efficient data centers in other countries generate other external impacts on nearby housing prices. This can be the cause because of the lack of planning policies or NIMBYism. Therefore, this research functions as a starting point to establish the impact that data center development could have on housing prices. Dutch spatial planning policies seem to work successfully to ensure no negative externalities are experienced on residential transaction prices, coming with data center development. Moreover, it even stimulated the positive externalities. This suggests that data center development, related to the prices of a nearby house, does not always have to be associated with NIMBYism.

5.2 Limitations & future research

Throughout this research, multiple constraints have complicated the process of investigating the external impact of data center development on the transaction prices of nearby residential units. This chapter will elaborate on the main limitations to advise future research.

First, there are too few data centers included within this research to generalise the findings. A substantial part of the data centers is excluded in the selection procedure to circumvent other external impacts on housing prices. The consequence is that three selected data centers do not represent the population of data centers in the Netherlands. Thus, the results are not generalisable. To generalise an impact in future research, including more data centers is essential. This seems possible regarding the current development trends of data centers, focusing on larger and rural located data centers since the available space in urban areas is shrinking. These data centers could be included in the upcoming years when there is enough transaction data available over the years.

Secondly, an endogeneity issue since spatial planning policies already accounts for externalities. The public institutions only issue a building permit when a development is in line with the zoning plan, and no negative externalities for nearby residents occur. Therefore, the public institutions already consider their place relative to other factors, which could cause endogeneity in the observations.

Thirdly, the variable Transaction after development is taken on a yearly base equal to the year of operationalisation of a data center, making it unable to capture all anticipation effects. Depending on the complexity and objections, the overall development process can take years, and negative externalities could peak in the construction phase (Boyle et al., 2019). When including the variable Transaction during development, this covers the anticipation effect. Therefore, it is recommended for future research to implement the development announcement and the start of construction as independent moments.

Fourthly, the used NVM data set only contained a few property characteristics. Desired is to add more property characteristics, including garage and energy label, to increase the adjusted R-squared and stabilise the coefficients in the regressions. In turn, that would give more certainty to argue the external impact of a variable. Therefore, advised is to ask the NVM for a more inclusive data set for further research.

The last limitation of this research is that the transaction price is driven by an excess demand of the previous years. Appendix IV shows that both the target and control areas have seen an increasing TP. This increase could bias the results since the average transaction price has increased enormously, independently from data center development.

5.3 Recommendations

NIMBYism against data center development seems to have increased since data centers appear more in the news. Consequently, their performances are constantly under a public magnifying glass, varying from their high energy use to their contribution to reducing agricultural lands. It is not remarkable that data center development incites NIMBYism. However, there are two recommendations for spatial planning policies to turn NIMBYism into YIMBism (yes-in-my-backyard) or a more neutral attitude.

First, since available space becomes scarce and data centers are not desired in the scenery of agricultural lands, spatial planning policies need to focus on bringing data centers out of sight. This could be done by permitting data center development underground — the benefits of developing underground are various. For example, there will be no more obstruction of view, energy costs are reduced, and the upper ground could be reserved for other functions. Underground data centers are already in function in the USA, where security and energy efficiency were the greatest initiators (Iron Mountain, 2021). On the other hand, the construction cost will be higher, which could deter investors. However, bringing the data center out of sight could lead to less resilience from residents and positive contribution to the built environment could outweigh that. Especially in areas where residents are more protective against new development and NIMBYism gains a following, preventing public discussion could save time and money (Van der Horst, 2007).

Secondly, spatial planning policies should focus more on ensuring that data centers generate their own electricity. Currently, large data centers use energy generated by subsidised windmill farms, sometimes located insight of residents. Most residents are in the first place not encouraging windmill farms development in their sight. However, when the generated energy goes straight to the data center, they feel robbed, stimulating NIMBYism. This is one of the reasons why NIMBYism is so strong in the 'Wieringermeer'. In 'De Wieringermeer', Vattenfall developed the largest windmill farm in the Netherlands, however, 100% of the generated energy goes to Microsoft his data center (Vattenfall, 2017). To prevent this from happening again, spatial planning policies need to reserve more space for less visible, sustainable energy generators on a data center site. It will be complex to achieve the 'Climate Neutral Data Center Pact' in 2030 if data centers could not use all the windmills capacity onshore. However, to turn NIMBYism into a neutral attitude, making the residents contribute to windmill farms is essential.

References

Aydin, R., Smith, B. Crawford, E. (2010). Commercial Development Spillover Effects upon Residential Values. *Southwestern Economic Review*. 37. 47-61.

Boyle, K. J., Boatwright, J., Brahma, S. Xu, W. (2019). "NIMBY, not, in siting community wind farms," *Resource and Energy Economics*, Elsevier, vol. 57(C), pages 85-100.

CBS (2021a). Existing own homes; purchase prices, price indices 2015=100 [Online]. Available at: https://www.cbs.nl/nl-nl/cijfers/detail/83906ENG (Accessed: 15-04-2021)

CBS (2021b). Voorraad woningen; gemiddeld oppervlak [Online]. Available at: https://opendata.cbs.nl/statline//CBS/nl/dataset/82550NED/table?fromstatweb (Accessed: 08-07-2021)

CE Delft (2020) MRA-brede Strategie Datacenters Eindrapport. [PDF]

College van Rijksadviseurs (2019). (X)XL-verdozing, Minder, compacter, geconcentreerder, multifunctioneler.

Copenhagen Economics (2020). Inside the Netherlands, Google's European hyperscale data centres and infrastructure ecosystem.

Depoorter, V., Oró, E. Salom, J. (2015). The location as an energy efficiency and renewable energy supply measure for data centres in Europe. *Applied Energy*. 140. 338–349

De Stendor (2021). Dit zijn De Grootste zorgen over datacenters: van waterslurpers tot chemicaliëndumpers [Online]. Available at: https://www.destentor.nl/zeewolde/dit- zijn-de-grootste-zorgen-over-datacenters-van-waterslurpers-tot-chemicaliendumpers1ffb78/ (Accessed: 02-01-2021)

De Vor, F. De Groot, H. L. F. (2011). The Impact of Industrial Sites on Residential Property Values: A Hedonic Pricing Analysis from the Netherlands. *Regional Studies*, 45(5), 609-623.

Droes, M., Koster, H. (2019). Windturbines, zonneparken en woningprijzen: Rapport in opdracht van het Ministerie van Economische Zaken en Klimaat. Vrije Universiteit.

Gemeente Amsterdam (2019). Meer regie op vestiging van datacenters in Amsterdam en Haarlemmermeer [Online]. Available at: https://www.amsterdam.nl/nieuwsarchief/persberichten/2019/persberichten-mariekedoorninck/regie-vestiging-datacenters-amsterdam/ (Accessed: 21-03-2021)

Grether, D. M. Mieszkowski, P. (1980). The Effects of Non-residential Land Uses on the Prices of Adjacent Housing: Some Estimates of Proximity Effects. *Journal of Urban Economics* 8 (1): 1-15.

Knaap, G. J. Song, Y. (2004). Measuring the effects of mixed land uses on housing values. *Regional Science and Urban Economics*, Volume 34, pp. 663-680.

Koster, H. R. A., Van Ommeren, J. (2019). Place-based policies and the housing market. *The Review of Economics and Statistics*, 101(3), 400–414.

Leefbaar Zeewolde (2020). De Achillieshiel van datacenters.

Li, M., Brown, H. (1980). Micro-Neighborhood Externalities and Hedonic Housing Prices. *Land Economics*, 56(2), 125-141.

Ministerie van Binnenlandse zaken. (2021) Staat van de Woningmarkt Jaarreportage 2020 [Online]. Available at: https://www.rijksoverheid.nl/documenten/rapporten/2020/06/15/staat-van-de- woningmarkt-jaarrapportage-2020 (Accessed: 19-04-2021).

NH Nieuws. (2021). Hoe de Wieringermeer onvermoeibaar strijdt tegen softwaregiganten Google en Microsoft [Online]. Available at: https://www.nhnieuws.nl/nieuws/279921/hoe-de-wieringermeer-onvermoeibaar-strijdt-tegen-softwaregiganten-google-en-microsoft (Accessed: 02-03-2021).

NH Nieuws. (2020). Het landschap 'verdoost': "Maak een datacenter een icoon van deze tijd" [Online]. Available at:https://www.nhnieuws.nl/nieuws/276383/het-landschap-verdoost-maak-een-datacenter-een-icoon-van-deze-tijd (Accessed: 07-04-2021).

Ounifi, H.A., Ouhimmou, M., Paquet, M. Montecinos, J. J. I. (2015). Datacenter location for Internet services in the province of Quebec, *IFAC-Papers Online* 48. 2266–2272.

PBL. (2019). Putting Dutch 'urban sprawl' in a European perspective [Online]. Available at: https://www.pbl.nl/en/blogs/putting-dutch-urban-sprawl-in-a-european-perspective (Accessed: 07-07-2021).
REOS. (2019). Ruimtelijke Strategie Datacenters Routekaart 2030 voor de groei van datacenters in Nederland.

Rijkswaterstaat (2021). Bedrijven en mileuzonering [Online]. Available at: https://www.infomil.nl/onderwerpen/ruimte/functies/bedrijven/milieuzonering/ (Accessed:

01-06-2021).

Sunak, Y. Madlener, R. (2016). The Impact of Wind Farm Visibility on Property Values: A Spatial Difference-in-Differences Analysis. In: *Energy Economics* 55, p. 79–91.

Swinhoe, D. (2021). Data center NIMBYism: How to engage with local communities properly during data center projects [Online]. Available at: https://www.datacenterdynamics.com/en/analysis/data-center-nimbyism-how-to-engage-with-local-communities-properly-during-data-center-projects/ (Accessed: 08-04-2021).

Trouw (2021). De polder wordt als wingewest verkocht aan databedrijven, hoog tijd dat het parlement ingrijpt [Online]. Available at: https://www.trouw.nl/verdieping/de-polder- wordt-als-wingewest-verkocht-aan-databedrijven-hoog-tijd-dat-het-parlement- ingrijpt bedc6174/

Van der Horst, D. (2007). NIMBY or not? Exploring the relevance of location and the politics of voiced opinions in renewable energy siting controversies. *Energy Policy*, 35 (5) Pages 2705-2714, ISSN 0301-4215, https://doi.org/10.1016/j.enpol.2006.12.012.

Van der Veen, G. (2019). Rethinking XXL Distribution, a case study in North-Brabant.

Van Duijn, M., Rouwendal, J., Boersema, R. (2014). Transformations of Industrial Heritage: Insights into External Effects on House Prices. (TI Discussion Paper; No. 14-122/VIII). *Tinbergen Institute.*

Vattenfall (2017). Nuon tekent megaovereenkomst met windenergie met Microsoft [Online]. Available at: https://group.vattenfall.com/nl/newsroom/persbericht/2017/nuon-tekent-mega-overeenkomst-voor-windenergie-met-microsoft (Accessed 28-05-2021).

Warren, C.R., Lumsden, C., O'Dowd, S. Birnie, R. V. (2005). Green on green: public perceptions of wind power in Scotland and Ireland. *Journal of Environmental Planning and Management*, 48 (6), pp. 853-875

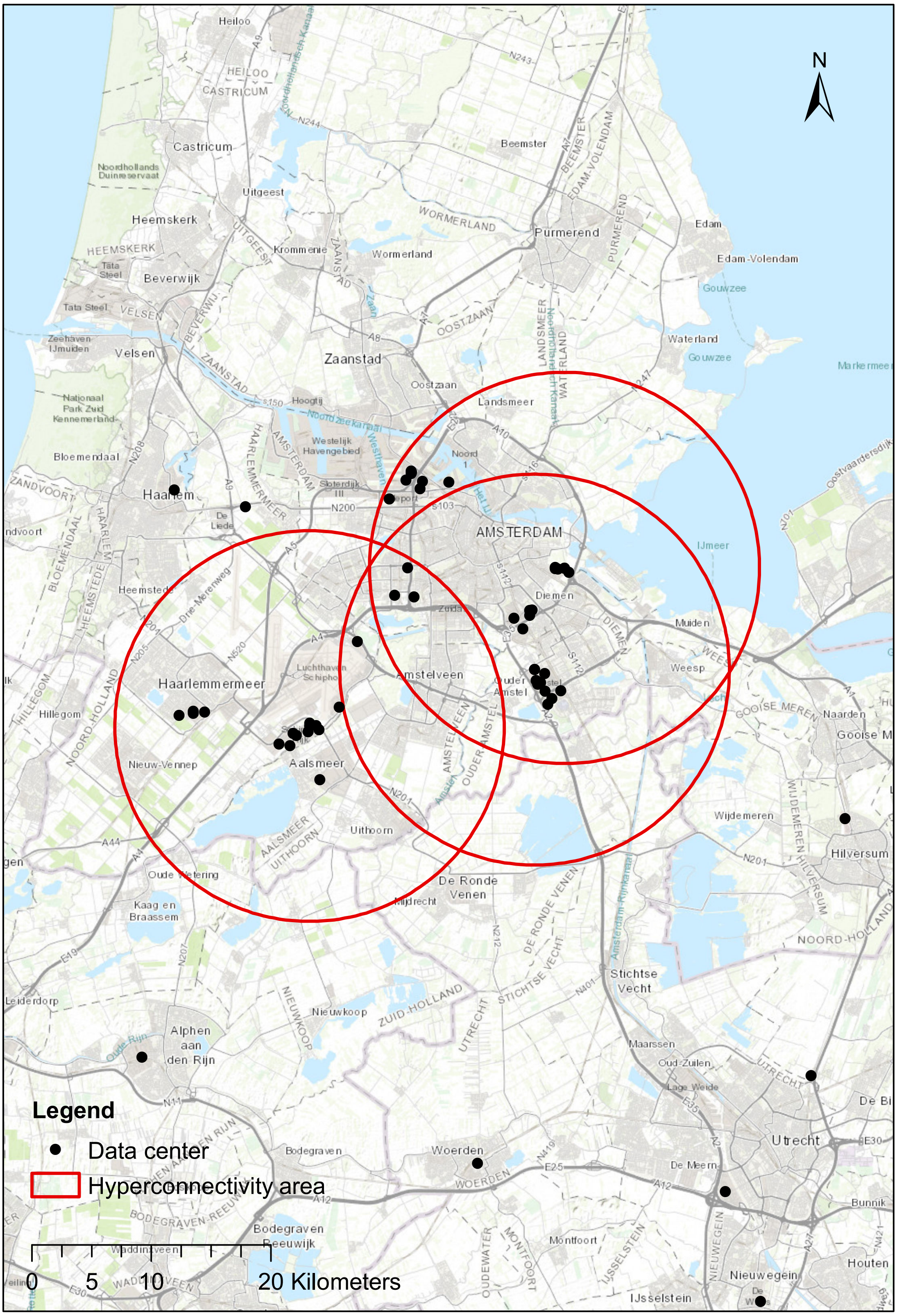
Xiao, Y., Webster, C. Orford, S. (2016) Can street segments indexed for accessibility form the basis for housing submarket delineation?, *Housing Studies*, 31:7, 829-851

Zhang, S., van Duijn, M., van der Vlist, A. J. (2020). The external effects of inner-city shopping centers: Evidence from the Netherlands. *Journal of Regional Science*, 60(4), 583-611.

6 Appendices

6.1 Appendix I

Figure 7: Hyperconnectivity in MRA



Note: Locations in the MRA with access to hyperconnectivity.

6.2 Appendix II

Table 11: Regression analysis target area

Target area dummy (ref. $\geq$ 4,000 - $\leq$ 5,000 m)	Coefficient
0 - 1000 m	0.1266(***)
1000 - 2000 m	0.1436(***)
2000 - 3000 m	0.1602(***)
3000 - 4000 m	-0.0042
Transaction characteristics	Yes
Adjusted R^2	0.1195

Note: This table denotes the regression analysis to determine the target area. The dependent variable is log Transaction price. Transaction characteristics includes textittransaction year

6.3 Appendix III

Figure 8: Overview of the selected data centers

Data center: COLT Roosendaal
Location: Argonweg 9, Roosendaal
Occupier: COLT Data Center Services
Size: 17,900 sqm
Year of operationalisation: 2013
Zoning: Business terrain - 4
Forms of resilience: No
Site: Previous agricultural land
Connected to residual heat network: No
Energy consumption: Unknown
Numb. of observations (Target area): 2753 (1,069)

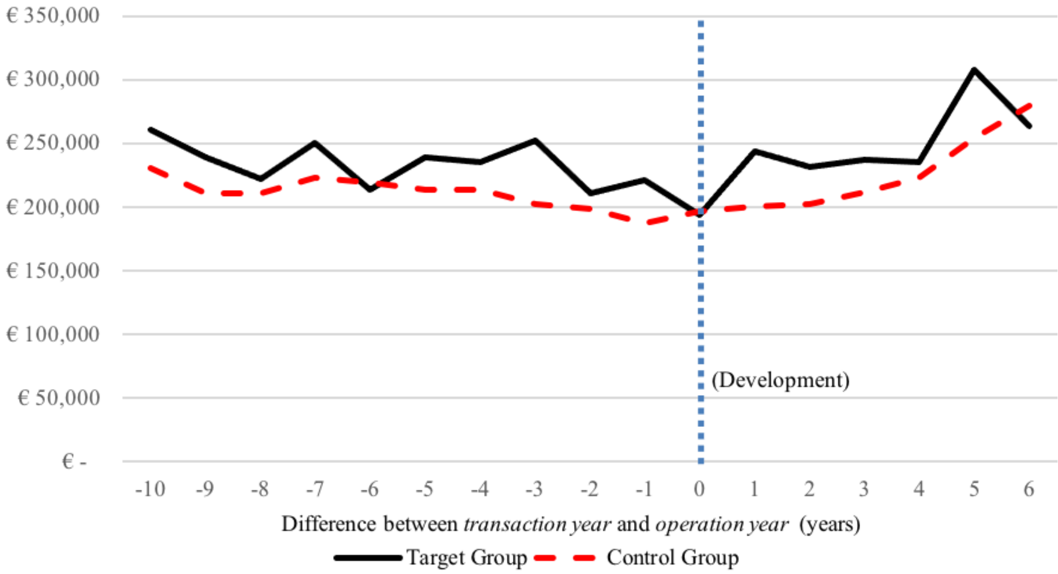
Data center: Microsoft 'De Wieringermeer'
Location: Agriport 601, Middenmeer
Occupier: Microsoft
Size: 330,000 sqm
Year of operationalisation: 2014
Zoning: Business terrain 1 - 5
Forms of resilience: Petition & protest
Site: Previous agricultural land
Connected to residual heat network: No
Energy consumption: Uses 100% of electicity generated by Windmill farm Wierinermeer (Vattenfall)
Numb. of observations (Target area): 1,390 (262)

Data center: NorthC Groningen
Location: Liverpoolweg 10, Groningen
Occupier: COLT Data Center Services
Size: 4,000 sqm
Year of operationalisation: 2015
Zoning: Business terrain
Forms of resilience: No
Site: Previous agricultural land
Connected to residual heat network: No
Energy consumption: 100% green energy
Numb. of observations (Target area): 2385 (439)

Note: An overview of the most important information on the three selected data centers.

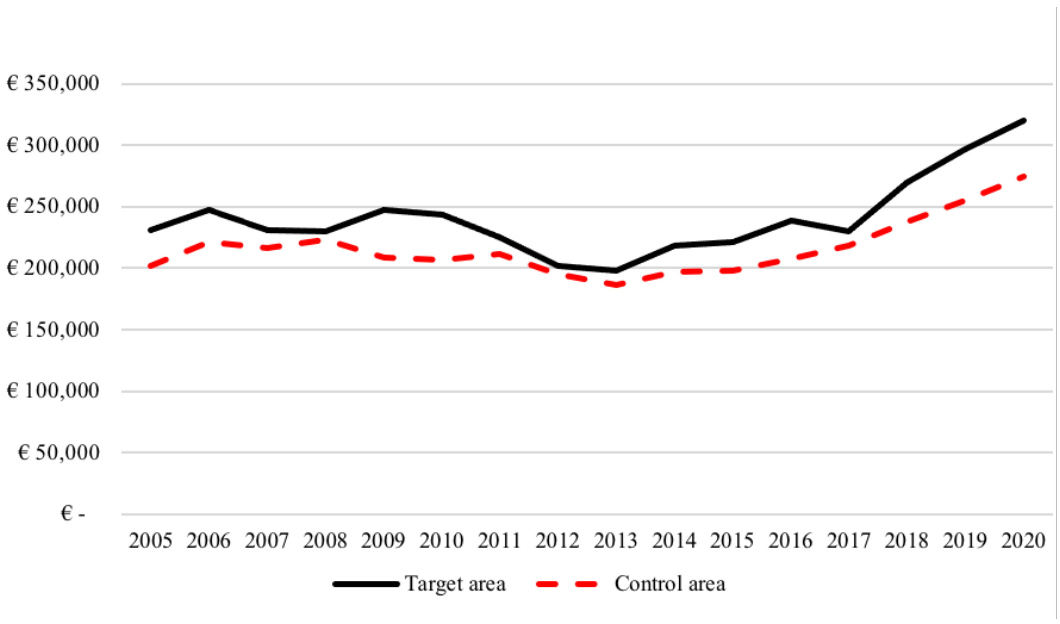
6.4 Appendix IV

Figure 9: Development of average *transaction price* relative to data center development



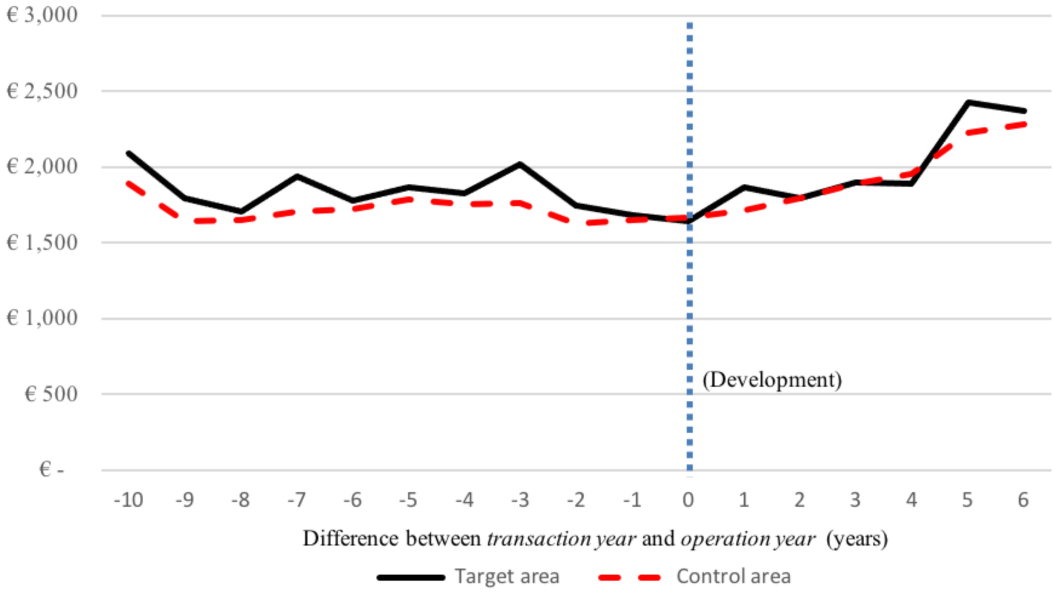
Note:Graph denotes that the average *transaction price* of the target area and control area are comparable and follow the same trend. The target area line is more erratic due to less observations. The blue lines indicates the moment of operationalisation of data center.

Figure 10: Development of the average *transaction price*



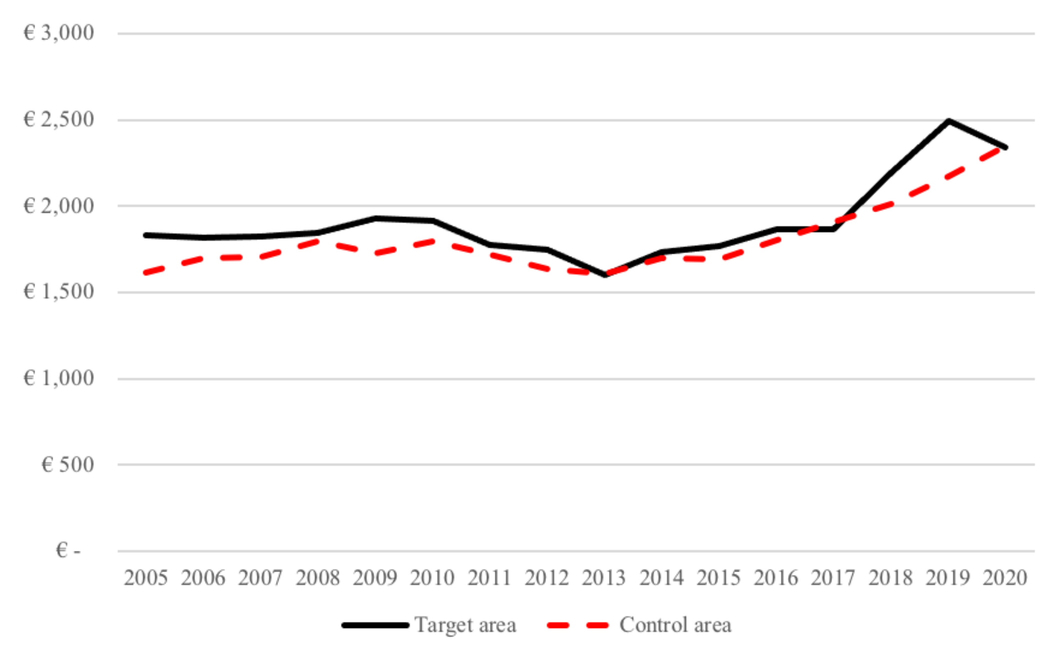
Note:Graph denotes that the average *transaction price* over the year 2005 - 2020 of the target area and control area are comparable and follow the same trend.

Figure 11: Development of average *transaction price per squared meter* relative to data center development



Note: Graph denotes that the average *transaction price per square meter* of the target area and control area are comparable and follow the same trend. The target area line is more erratic due to less observations. The blue lines indicates the moment of operationalisation of data center.

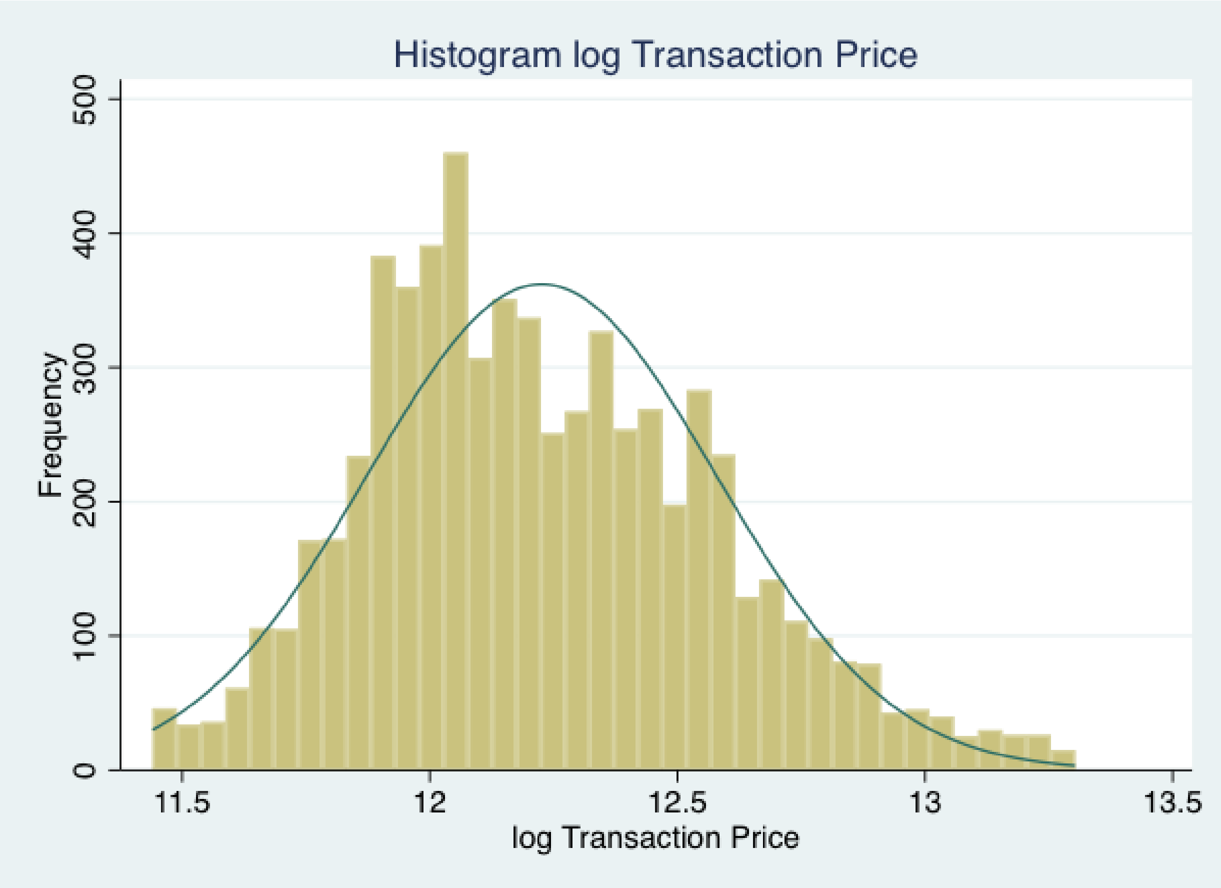
Figure 12: Development of the average *transaction price per squared meter*



Note: Graph denotes that the average *transaction price per square meter* over the year 2005 - 2020 of the target area and control area are comparable and follow the same trend.

6.5 Appendix V

Figure 13: Histogram of *log transaction price*



Note: The dependent variable *transaction price* is transformed to a logarithmic function. The result is that the dependent variable has a normal distribution.

6.6 Appendix VI

Table 12: Chow test

Variable	df	F	P >F
Building type	3	56.83	0.0000

Note: This table denotes the results from the Chow test on building types: *Building type Apartments*,*Building type Detached*, *Building type Townhouse*, *Building type Single-family*