Notater

Documents

2017/31

Statistics Norway

Statistisk sentralbyrå

Svein Johan Reid

Mapping attractive urban areas

Documentation of a Eurostat-supported project under the "Merging statistics and geographic information grant programme"

Documents 2017/31

Svein Johan Reid

Mapping attractive urban areas

Documentation of a Eurostat-supported project under the “Merging statistics and geographic information grant programme”

Statistisk sentralbyrå • Statistics Norway
Oslo–Kongsvinger

In the series Documents, documentation, method descriptions, model descriptions and standards are published.

Published 18 September 2017

ISBN 978-82-537-9595-9 (electronic)

Symbols in tables	**Symbol**
Category not applicable	.
Data not available	..
Data not yet available	…
Not for publication	:
Nil	-
Less than 0.5 of unit employed	0
Less than 0.05 of unit employed	0.0
Provisional or preliminary figure	*
Break in the homogeneity of a vertical series	—
Break in the homogeneity of a horizontal series	|
Decimal punctuation mark	.

Preface

A main purpose of The European Commission is economic -, social - and environmental development in Europe. One of the approaches to these issues is the “Quality of life in cities - Perception survey”, carried out in 2004, 2006 and 2009, and lastly in 2013, with 79 cities in Europe including Oslo. In the last 2013-survey, in total 41 000 people were interviewed answering questions about various aspects of urban life. Interviewees were asked to identify important issues for their city.

Making interviews of 41 000 persons is however resource and time consuming. The “Mapping urban attractivity” project probes a methodology of testing out alternative data resources first, specifically statistical registers and georeferenced data, with the aim of potentially providing input on quality of life/attractivity that could contribute to the survey.

The project has received funding from the European Commission – Eurostat, under the program “Merging statistics and geographic information”.

Author of this publication is Senior Advisor Svein Johan Reid, Division for natural resources and environmental statistics.

Statistics Norway, 30 August 2017

Lise Dalen Mc Mahon

Abstract

The “Quality of life in cities - Perception survey” of The European Commission takes a qualitative approach to issues of urban attractivity, with interviewees asked to identify important issues for their city.

Instead of asking what parameters that are important for the population, the “Mapping urban attractivity” project takes a quantitative and geographical approach to these questions, testing a methodology which at a European level might be used as a supplement to the survey. The aim is to aid the survey in “which questions to ask?”, “where should we ask them?” and interpretation of results.

The probed methodology uses house prices as a proxy for attractivity, as they are a reflection on a kind of attractivity. Both “Total sales prices” and “Price per m^2” are explored, with all point georeferenced dwelling sales throughout a year as the data which we wish to explain. We have focused on Norway’s largest cities, using Ordinary Least Square Regression analysis tools to correlate price and place with factors such as m^2 floor space, mean income or education level of the adult population in a buffer zone around each dwelling, or other types of variables. The variable types tested for are: *1. Intrinsic characteristics of a dwelling, 2. Population characteristics 3. Employment, 4. Distance to geographic entities*, *5. Distance to buildings*, *6. Environmental*.

A focus of the project is variation within cities, making comparisons between cities on to which degree our variables explain variation. This brings city intrinsic differences into an equation to a large degree lacking in the “Perception survey”. Our approach touches therefore into whether city planners have been successful in distributing important city services and amenities in an even fashion.

Dwelling-intrinsic characteristics such as m^2 floor space clearly count for a vast amount of price variation. There is however variation in how true this is throughout different cities, leaving more explanatory power to non-dwelling-intrinsic characteristics. Of these variables, our findings are that “Education level” and “Household income” are the best indicators of variation in neighbourhood attractivity. Also “urban pull” variables can count heavily, with variables on distances to town centre, restaurants and higher education facilities.

The strengths of these correlations vary between the cities, playing in to a general picture that capital and largest city Oslo is the most socioeconomically divided of Norway’s largest cities. However, results show that size isn’t all, history and socioeconomic issues clearly matter. The potential in calculating these same correlations for a different year, or for creating a time series, is apparent. Results would pick up on nuances of correlation values in and between Norway’s cities.

The OLS-analysis produces results on which combination of variables that best correlate to our house price attractivity variable, producing coefficients on the strength of each variable. In the project, we reuse this output to produce predicted attractivity datasets, generating 500m X 500m attractivity grids for each city, scaling from “least attractive” to “most attractive”. The variation in these predictions are by definition an expression of variation in attractivity. Potential lies in locking these coefficients and creating a time series, mapping expected changes in attractivity, correlating this again to observed house price changes.

A lot of common sense and logic can be read from the resulting correlations. The conclusion of the project group is that findings and methodology definitely have a potential as supplement to the Europe’s “Perception survey”, making more out of time and resources invested in this important undertaking.

Contents

1. Summary

1.1. Background

The urban population in Europe is increasing and currently more than two thirds of the European population live in cities and towns. In Norway, about 80 per cent of the total population lived in densely populated areas in 2012. Almost 20 per cent of the total population lived in the urban settlement of Oslo.
The European Commission is following the economic -, social - and environmental developments in European urban areas. One example of this work is the European Commission's survey on how citizens perceive quality of life in their home cities.

1.1.1. Quality of life in cities

Quality of life in cities - Perception survey in 79 European cities was published in October 2013[1] and is a new edition of the same surveys carried out in 2004, 2006 and 2009. The previous surveys were conducted in 75 cities in EU27, Croatia and Turkey. The latest survey comprised 79 cities including Oslo.

The 2013-survey included all European capitals (except for Switzerland) as well as between one to six more cities in the larger countries. In each city, around 500 citizens were interviewed. In total 41 000 people were interviewed answering questions about various aspects of urban life as how do they assess the quality of services such as public transport, health care, education, cultural and sport facilities.

The interviewees were also asked about employment opportunities, housing situation, safety and environmental variables (air quality, noise, green spaces, fight against climate change).
Out of these variables the interviewees were asked to identify the three most important issues for their city, and on an overall European level health services, unemployment, education facilities were found to be the most important. For Oslo, the respondents answered that health services, education facilities and public transport were found to be the most important issues.

1.2. Description of action

1.2.1. General objective of the action

The general objective of this project is to combine relevant statistical registers and georeferenced data in order to determine attractive urban areas. This in turn, may complement the questions in "Quality of life in cities - Perception survey" of the European Commission. This action aims for developing an innovative procedure for assessing how changes in population and land use in urban settlements relate to "Quality of life" parameters.

1.2.2. Initial objectives, foreseen methodology

In testing for relevant indicators of attractivity, we set out to look at migration, on the working theory that people tend to move to somewhere they perceive as more attractive. Other angles we wished to explore were housing prices, location of new buildings, income and education.

In general, we found migration within a city to be problematic as an indicator for attractivity, as the availability of housing in an area quite often is more a function of planning processes within the city than of the attractivity of the area. New dwellings are not necessarily built in the city's most attractive areas, due to for

[1] http://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/urban/survey2013_en.pdf

example space issues, or that city planners wish to offer more affordable housing. Location of new building permits is therefore for the same reason not necessarily in the city's perceived most attractive areas. Chapter 8 gives an overview of our findings on migration.

Leaving migration and building permits out of our basis for generating attractivity datasets shortened our initial scope to house prices, income and education, making it necessary to reconsider our choice of methodology.

In the foreseen activities, we planned to test the "Quality of life - Perception survey" parameters in relation to generated grid based attractiveness datasets, using statistics and georeferenced data in the production of these. Each grid cell (for example 500m X500m) would receive a numeric value indicating its place on a scale from "least attractive" to "most attractive". The brunt of the project focus would be on how to create these attractiveness datasets.

We found it necessary to deviate from this, opting out of creating predefined attractiveness datasets.

The following three objectives outline the foreseen activities of the project, which were deviated from:

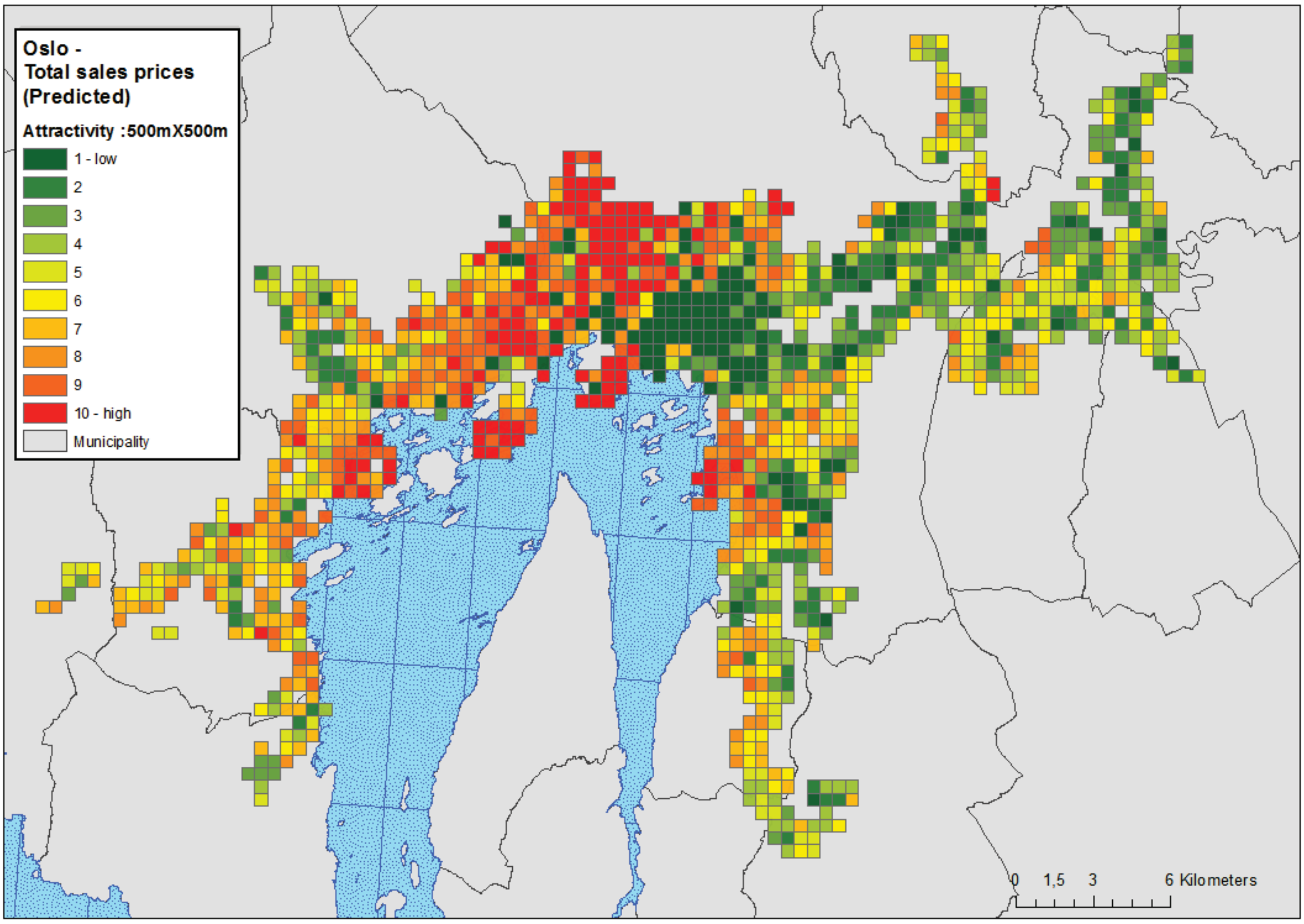
> Specific objective 1: Describing the quality of data sources and the possibilities for combining these into an urban area attractiveness dataset. Produce a conceptual model of the data structure and data format of an urban area attractiveness dataset.
> Specific objective 2: Develop a methodology for producing attractiveness dataset for urban settlements.
> Specific objective 3: Determine "Quality of life" parameters of importance based on the location of attractive urban areas.

1.2.3. Revision of foreseen methodology

Instead we found it a better choice to include our statistics, "Quality of life" parameters and georeferenced data in an Ordinary Least Square -regression analysis, using housing prices as a proxy for attractivity (that which we wish to explain), exploring whether any of our variables (or combination thereof) can be said to correlate to variation in our attractivity proxy.

A second part of our project involves visualising city-intrinsic variations in attractivity, as well as all other variables we have wished to check for correlation. The OLS-analysis produces results on which combination of variables that best correlate to our house price attractivity variable, producing coefficients on the strength of each variable. We have used this output to produce predicted attractivity datasets, reusing the concept of a 500m X 500m attractivity grid, scaling from "least attractive" to "most attractive".

Figure 1.1. Urban attractivity index on 500m X 500m grid for the city of Oslo: index 1-10, (Proxy for attractivity: Total sales prices)



Two attractiveness datasets are generated for each of the four largest cities, and are basically predicted mean "Total sales price" and "Price per m^2" within each grid cell.
The variation in these predictions are by definition an expression of variation in attractivity. It is also this variation which is essential, not whether the predicted Kroner price is totally precise. Herein lies also the potential for a time series (not explored in this project), mapping expected changes in attractivity, which again could be compared against actual changes in attractivity (house prices).

1. Generate the same statistics for a different year, e.g. + 2 years
2. Calculate attractiveness dataset year+2 with same coefficients as year 0
3. Overlay between attractiveness datasets year 0 and +2
4. Highlight locations with change in values, suspected change in attractivity

Chapter 3 outlines all steps of the methodology followed in this project.

1.3. General findings of the project

1.3.1. Testing indicators of attractivity

We have used housing prices as a proxy for attractivity, as they are a reflection on supply and demand, a numerical representation of some kind of attractivity. Intrinsic characteristics of a dwelling such as square meter dwelling area, need for renovation and so on dictate a certain sales price, but they do not dictate neighbourhood attractivity. The project addresses whether there is price variation dependent on location within the city, seeking to explain this by correlating price and place with factors such as income and education in the neighbourhood surrounding a sold dwelling, or to “Quality of life survey”-variables such as distance to public transport, health services, education facilities, or other variables such as coast and lakes, recreational areas and so on.

If there is a significant correlation, the relevant variable/variables can be said to be indicators for how dwellings are pushed up or down the demand scale of “supply and demand”, location pushing up or down the price people are willing to pay, within a price scope for that specific area. Intrinsic characteristics of the dwelling (floor space, age of building, etc.) contribute then to placement within a specific scope.

Finding a variable not significant does however not mean that it doesn’t matter, and is something we find to be true for several services to the population. Not finding health services significant would rather suggest that city planners have been successful in distributing it evenly, making access close enough to not matter.

Finding a significant correlation between housing prices and neighbourhood socioeconomics (as household income and education levels of the population surrounding a given sale) suggest that socioeconomics matter, and the strength of that correlation for a city. In the project, the same variables are tested for all Norwegian urban settlements > 50 000 inhabitants, with interesting differing strengths. These differences can be said to point to how more or less “divided” Norway’s larger cities are, giving interesting comparison and insights.

Testing for socioeconomic variables tell us something about who lives in the neighbourhood, of the population characteristics therein. In our project population characteristics data is georeferenced to mainly address points, allowing estimation of average household income, educations levels and so on by collecting population characteristics within a buffer zone around each georeferenced real estate dwelling. The approach gives more accurate and more continuous results than an approach based on gathering area based population characteristics. We have probed different sized buffer zones, and found 250 metres to be sufficient for a balance between enough address points to make valid averages, and detail.

In the project, we might or might not then find a correlation between these variables on population characteristics and house prices, an indication of attractivity. Why it is so, is a more complicated field. It is also a politically highly important field, as it gives insight to general trends on people’s actual preferences; by what they choose to do.

A common perceived problem for many larger cities is “dividedness” based on socioeconomics, for example underprivileged more concentrated to certain areas, and in Norway and Europe there is considerable will to iron out these differences, through initiatives targeting different aspects seen as critical: For example, better integration of non-western immigrants into society is in Europe by many seen as a political goal, attempts for better integration being manifold. They can range from national initiatives addressing how immigrants are perpetrated in school books or media, to neighbourhood-specific initiatives at bettering access to health care, public transport, or schools.

Following this relationship over time may therefore also point to whether general or neighbourhood specific policies to alleviate “dividedness” have a positive effect.

Our project looks collectively at all house sales throughout 2014, intrinsically for each city in the project, but also making comparisons between these cities. The project group sees the potential for comparison over time as apparent, within same cities, looking at whether specific variables have a strengthening or weakening effect/reflection on attractivity, or if there is a tendency to a greater geographic divide in city specific attractivity[2].

This is a potential for further work, building on the findings and proposed methodology in our project.

Variables as intensity of traffic noise, or distance to water, recreational areas, restaurants or town centre are a different type compared to those describing who live in a neighbourhood. Why they matter (if we find they do) is often more directly understood. Having a shoreline on your property is quite universally seen as more attractive than not, affecting the potential sales price of that house.

Chapter 4 gives an overview of all variables tested for. We have grouped these by following variable types:

Variable type
Dwelling
Distance to geographic entities
Distance to buildings
Intensity-environment
Population characteristics
Employment

1.3.2. Limitations

Many variables have been have tested for, and one could always include more. We have in the project tried to not be too expansive, keeping in mind that this is a suggested methodology for national statistics agencies, and that there are limits to what data Statistics Norway or statistics agencies in Europe can be expected to hold or acquire.

The main variables available in our sold dwellings dataset (real estate sales data) is “Total sales price”, “Floor space m^2”, “Age of building” and xy-location by centroid of the property the dwelling lies upon. The dataset holds nearly all sales of dwellings throughout 2014 for the entire of Norway.

Other intrinsic characteristics on the condition of a dwelling, such as last renovated or building faults is data Statistics Norway do not hold, or are difficult to obtain, and are therefore not included in the dataset.

[2] Western parts of Oslo are for example generally perceived as more affluent than eastern parts. A relevant question is whether this geographic divide in attractivity is widening.

Other building intrinsic variables such as "has garage" or "has lift" is readily available in the Cadastre. However, the real estate sales data lacks necessary variables to make a sufficient one-to-one join from all house sales to the Cadastre. 84 per cent of Norwegians own the home they live in, as owners in a housing cooperative or as free-holders. To encompass all sales (2/3 of all sales in Oslo were in housing cooperatives), it was necessary to georeference with the Cadastre's property register, rather than building register, which would give more intrinsic variables on buildings.

The project focus is neighbourhood attractivity throughout the city, so taking all housing cooperative sales out of the analysis would have distorted results severely. We realize that these missing intrinsic variables affect sales price, and that a perfect correlation between individual sales prices and variables we test for is not possible. One of the conclusions of the project is also that we do not have a model without clustering of residuals; there is spatial autocorrelation. We have had to accept this, that we lack variables which fully can predict variation in price, there are locations where our models will over- or underestimate.

On the other hand, leaving housing cooperative sales out is a much larger trade-off, leaving large swaths of the cities without any data at all, biasing also by population characteristics. To do so would be the lesser choice in explaining and predicting neighbourhood attractivity, by looking at "what matters?".

For Oslo, this choice gave us a georeferenced dwelling sales dataset of approximately 23 000 sales, encompassing nearly all sales of dwellings throughout 2014.

1.3.3. Choice of proxy for attractivity: Total sales price or Price square meter

Seeking to explain housing prices also raises the question whether it is "Total sales price" or "Price per m^2" that expresses attractivity best. It is possible to make a case for both approaches, that they both say something about attractivity, but different aspects. A potential buyer might tolerate a higher "Price per m^2" if being close to education facilities, restaurants and theatres is more important than amount of floor space. At a different stage in life the same buyer might prefer or need more space, trying with his means to optimize on space, in an as attractive location as he has means to.

Instead of choosing one of these two approaches we have in the project explored both, finding that several variables are only significant in one of them.

A third approach where similar sized dwellings are compared is also explored, where square metres floor space is "baked" in to the variable we to seek to explain.

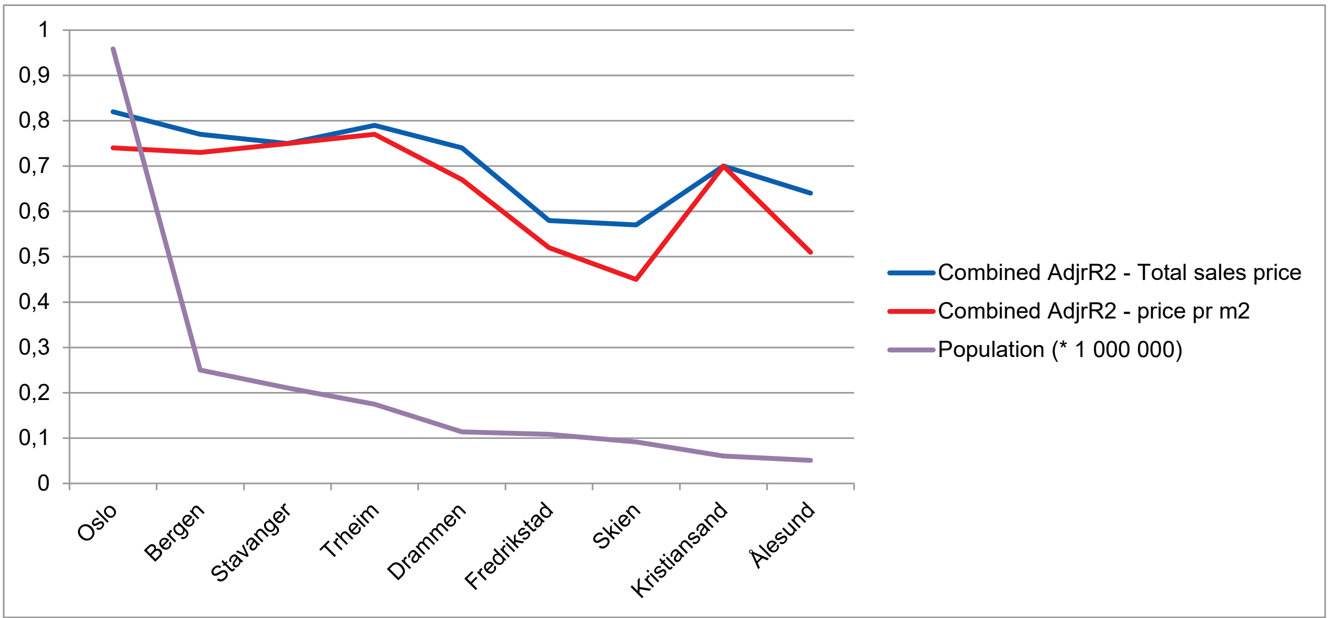
1.3.4. Ordinary Least Squares regression – Oslo and the rest

As we have chosen to use dwelling prices as a proxy for attractivity, using OLS-regression to explain and predict variation in dwelling prices, we also gain information on which attributes that matter in the different cities. Our basis is Oslo, variables are chosen on whether they are significant for Oslo, and the same variables are tested for the other cities. It is difficult to make sound predictions on a small scale within the cities, but the OLS analysis give good indications on what is important for the city as a whole.

The cities in our project vary in size, with Oslo as by far being the largest. There are differences between the cities in variable significance and strength, which partly can be understood and explained. Having results for several cities help in understanding general truths in relation to attractivity.

How much of dwelling price variation we can explain varies by city. For Oslo, we can explain 82 per cent ($AdjR^2 = 0.82$) of "Total sales price" variation, and 74 per cent of the variation in "Price per m^2". There are four cities with populations of more than 150 000, which all have $AdjR^2$ scores between 0.73-0.82, for both approaches (chart and table below). Below this population size we see decreasing values, our non-dwelling-intrinsic variables explain less of price variation, or their values are more erratic, making interpretation more difficult. Distance/cost of access to amenities such as town centre, coast, restaurants matter understandably less, as they maybe cross "ease of transport" thresholds. As city size falls we are able to explain less of sales price variation, meaning that prediction of sales prices for these cities will be equally "off the mark". It is relevant to set a threshold value for meaningful prediction, which we have set to $AdjR^2 = 0.70$. This translates to being able to explain 70 per cent of sales price variation. The five smallest cities in the project fall under this threshold.

Figure 1.2. How much of price variation we are able to explain in Norway's 9 largest cities, correlated with size of population. $AdjR^2$ of 1 = 100 per cent



How much of price variation we are able to explain in Norway's 9 largest cities, correlate with size of population. $AdjR^2$ of 1 = 100 per cent

$AdjR^2$									
Urban settlement	Oslo	Bergen	Stavanger	Tr.heim	Drammen	Fredrikstad	Skien	Kr.sand	Ålesund
Population	*958 378*	*250 420*	*210 874*	*175 068*	*113 534*	*108 636*	*91 737*	*60 583*	*50 917*
Combined $AdjR^2$ - Total sales price	0.82	0.77	0.75	0.79	0.74	0.58	0.57	0.7	0.64
Combined $AdjR^2$ - Price per m^2	0.74	0.73	0.75	0.77	0.67	0.52	0.45	0.7	0.51

1.4. Conclusions

In this project, real estate dwelling sales prices are used as a proxy for attractiveness within cities, exploring if and how georeferenced statistics and geography can be used to explain variations.

We have looked at variables telling us about population characteristics in the vicinity around individual sales, and “asked” (through regression analysis) whether there is any correlation between sales price and variables as education levels, household income, immigration levels and mean age. We have found especially education levels and household income to be strong indicators of price variation. Calculating the strength of these correlations in all nine cities in our study give us knowledge on how socioeconomically divided the different Norwegian cities are.

Mean price levels vary most within Oslo (compared to other cities), and our findings are that mean price levels of any neighbourhood to a large degree can be predicted by these statistics. Socioeconomics matter most for predicting price variation (attractivity) in Oslo. Comparing to other cities we see that city size definitely matters, but that this is a rule with interesting exceptions, which we can use statistics to better understand.

Effects of geographic or dwelling intrinsic variables are a different type of variable, where effect on attractivity is more directly understood. “Distance to water” and “Floor space m^2” are both indicators of attractivity, and a more tangible reason for that attractivity

Reasons for differing neighbourhood attractivity is clearly a complicated issue. Closeness to water and parks clearly counts somewhat, but to a large degree it is also the sum of preferences, decisions and practices in play over many years, the history of the city itself. The smells from the tanneries and the smoke from the factory chimneys shaped preferences a 100 years ago, contributing to a neighbourhood which tends to persist, shaping present attractivity.
In our project we test for present day variables: Do they matter and how much do they matter? Our approach has looked at house prices in two ways, “Total sales prices” and “Price per m^2”, testing for “what matters?”. How can we by statistics and geography best indicate variations for these?

We take a broad approach, and encompass also variables answered most important in the ”Quality of life in cities - Perception survey” for Oslo. These are “Public transport”, “Education facilities”, and “Health services”. Of the variables found most important in other European cities, we also specifically test for “Recreational areas”, “Noise and “Employment opportunities”.

The following pages give conclusions for each the two approaches ***Total sales prices*** and ***Price per m^2***:

Total sales prices

Looking at "Total sales prices", which variables tested for are the best indicators of price variation within Oslo and the other cities? Our findings are that there are three main variables. One of them is amount of "Floor space m^2" in the dwelling, as very much expected. The two others are population variables, being: "Mean education level of population aged more than 26 years old within 250 metres of a dwelling", and "Mean income level (before tax) of population aged more than 26 years old within 250 metres of a dwelling".

The strength of these two population variables vary by city. They count most for Oslo, generally falling by city size. This can be said to indicate that Oslo is more socioeconomically divided than Norway's smaller cities, and that there is an element of scale in this. People are more similar to their neighbours (by education and income) in Oslo than in all other cities. Especially these variables correlate strongly to "Total sales prices", and the expression of attractivity that lies within this.

Of the other population variables we have tested, we found that level of immigration also expresses some of the variation in "Total sales prices", but not consistently, and not in any way that education and income better cover.

The table below specifies how much each variable isolated can explain of variation in "Total sales prices" ($AdjR^2$ of 1 is 100 per cent), and how much all variables combined can explain (82% for Oslo). Of all our tested variables, these are the eight we for Oslo find significant and consistently contributing to price variation in the expected direction (= how often it is true that distance to water correlates to an expected higher total sales price). The colour of the figures indicates whether the eight variables behave the same way in the other cities. Type 1-variables (black) do so, they are significant and contribute consistently in the same direction to price variation. Type 2-variables (blue) contribute to this same main direction, but not consistently. Type 3-variables (brown) contribute oppositely to the main direction.

Total sales sums - How much of price variation we are able to explain in Norway's 9 largest cities. $AdjR^2$ for each variable isolated, and total combined $AdjR^2$.

$AdjR^2$ of 1 = 100 per cent

$AdjR^2$ *Urban settlement*	Oslo	Bergen	Stavanger	Trheim	Drammen	Fredrikstad	Skien	Kristiansand	Ålesund
Population	*958 378*	*250 420*	*210 874*	*175 068*	*113 534*	*108 636*	*91 737*	*60 583*	*50 917*
RESTAURANT-DISTANCE	.00	.00	-.01	-	.01	.01	-	.00	.02
CITY CENTRE-DISTANCE	.00	*.02*	-	-	-	.01	-	-	.14
WATER-DISTANCE	.01	.01	-	.00	.00	-	.01	.03	.02
FLOOR SPACE M^2	**.60**	**.62**	**.67**	**.64**	**.53**	**.45**	**.48**	**.45**	**.53**
EDUCATION LEVELS - POPULATION	**.20**	**.09**	**.04**	**.10**	**.23**	**.14**	**.14**	**.18**	**.08**
HOUSEHOLD INCOME - POPULATION	**.39**	**.23**	**.26**	**.24**	**.32**	**.25**	**.23**	**.22**	**.24**
AGE – MEAN OF POPULATION	.02	*.00*	-	.00	.00	.00	-	-	.01
BUILDING AGE	.02	.04	.00	.00	.05	.04	.06	.08	.01
COMBINED	**.82**	**.77**	**.75**	**.79**	**.74**	**.58**	**.57**	**.70**	**.64**

Type 1 (black) = significant variable, contributes consistently in same direction (+ OR –) to price variation
Type 2 (blue) = contributes in same direction (+ OR –) as type 1-figures*, but NOT consistently or as a non-significant variable*
Type 3 (brown) = contributes in opposite direction (+ OR –) as type 1-figures

A forth variable found significant in explaining variation throughout all cities is "Age of building". It is an adjusted variable taking into account how buildings built before the 1950s might be perceived as more attractive than buildings built in the era afterwards. We set all pre-WW2 buildings to value of 5, and found the variables performance change from never significant to always significant, in all cities.
Distance to town centre, restaurants, water are significant contributors in Oslo, as well as mean age of population (price rises with mean age). Their explanatory strengths are lower, and not always significant in all cities.

No other of our variables are found significantly important in Oslo, which also encompass the three "perception survey" variables. These are 1. Distance to "Education facilities", 2. Distance to "Health services" 3. Distance to "Public transport". These three variables have been specifically tested in all cities, with similar results. In general, the findings are that distance to these are close enough within our cities to not matter in terms of "Total sales prices". This does not mean they don't matter, but that their existence within the city is distributed satisfactory enough for the population, at least enough to not effect "Total sales prices".

As for other survey variables "Recreational areas", "Noise" and "Employment opportunities", we find similar results. We suspect "Recreational areas" and "Noise" to be relevant on a very small scale. However, on the city scale we cannot prove a correlation between their values and price variation.

Price per m²

Looking at "Price per m²", we find that income levels here is not a significant variable. Amount of "Floor space" and "Education level" of population are most important variables also here.

Differences in neighbourhood attractivity are larger in Oslo than the other cities, resulting in "Floor space m²" being able to explain much less of the city-wide variation in "Price per m²" than the 3 next largest cities ($AdjR^2$ for Oslo at 0.32, Bergen 0.52, Stavanger 0.66 and Trondheim 0.62).

Education levels weigh highly in Oslo ($AdjR^2$ at 0.35), with 0.30 for Bergen and then generally lower levels. This suggests again greater a socioeconomic divide in Oslo, that there is an element of scale to this, but not only. The value for 3rd largest city Stavanger is just 0.04, lower than many smaller cities.

As would be expected, there is a correlation between "Price per m²" and variables reflecting centrality/urban "pull". Distance to main centre zone is the most important, but also distance to restaurant buildings weighs highly, picking up on the "pull" of local centres within the cities. Distance to university buildings is another. Their location is for all cities a positive in relation to "Price per m²", but many coincide with town centres (centre zone). It still is a general truth that locations of universities are a positive indicator of "Price per m²".

For hospitals, we have instances of locations being a negative indicator of "Price per m²", other as positives. A conclusion being that hospitals not necessarily are perceived as attractive neighbours, or at least that they not always are placed in attractive areas.

Price per m² - How much of price variation we are able to explain in Norway's 9 largest cities. $AdjR^2$ for each variable isolated, and total combined $AdjR^2$. **$AdjR^2$ of 1 = 100 per cent**

$AdjR^2$									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kr.sand**	**Ålesund**
Population	*958 378*	*250 420*	*210 874*	*175 068*	*113 534*	*108 636*	*91 737*	*60 583*	*50 917*
HOSPITAL - DISTANCE	.30	.17	.01	.18	.20	.09	.01	.20	.08
RESTAURANT - DISTANCE	.24	.21	.01	.17	.17	.05	.08	.31	.13
EDUCATION LEVELS - POPULATION	.35	.30	.06	.10	.03	.01	.00	.05	-
CITY CENTRE - DISTANCE	.43	.30	.04	.35	.14	.03	-	.30	.02
WATER -DISTANCE	.04	.03	.04	.14	.08	.03	-	.03	-
FLOOR_SPACE M²	.32	.52	.66	.62	.39	.37	.28	.42	.37
AGE -MEAN OF POPULATION	.01	.00	.00	.01	.15	.14	.08	.19	.07
HIGHER EDUCATION -DISTANCE	.19	.27	.02	.12	.13	.02	.03	.23	.02
BUILDING AGE	.16	.10	.13	.12	.11	.06	.05	.09	.00
COMBINED	**.74**	**.73**	**.75**	**.77**	**.67**	**.52**	**.45**	**.70**	**.51**

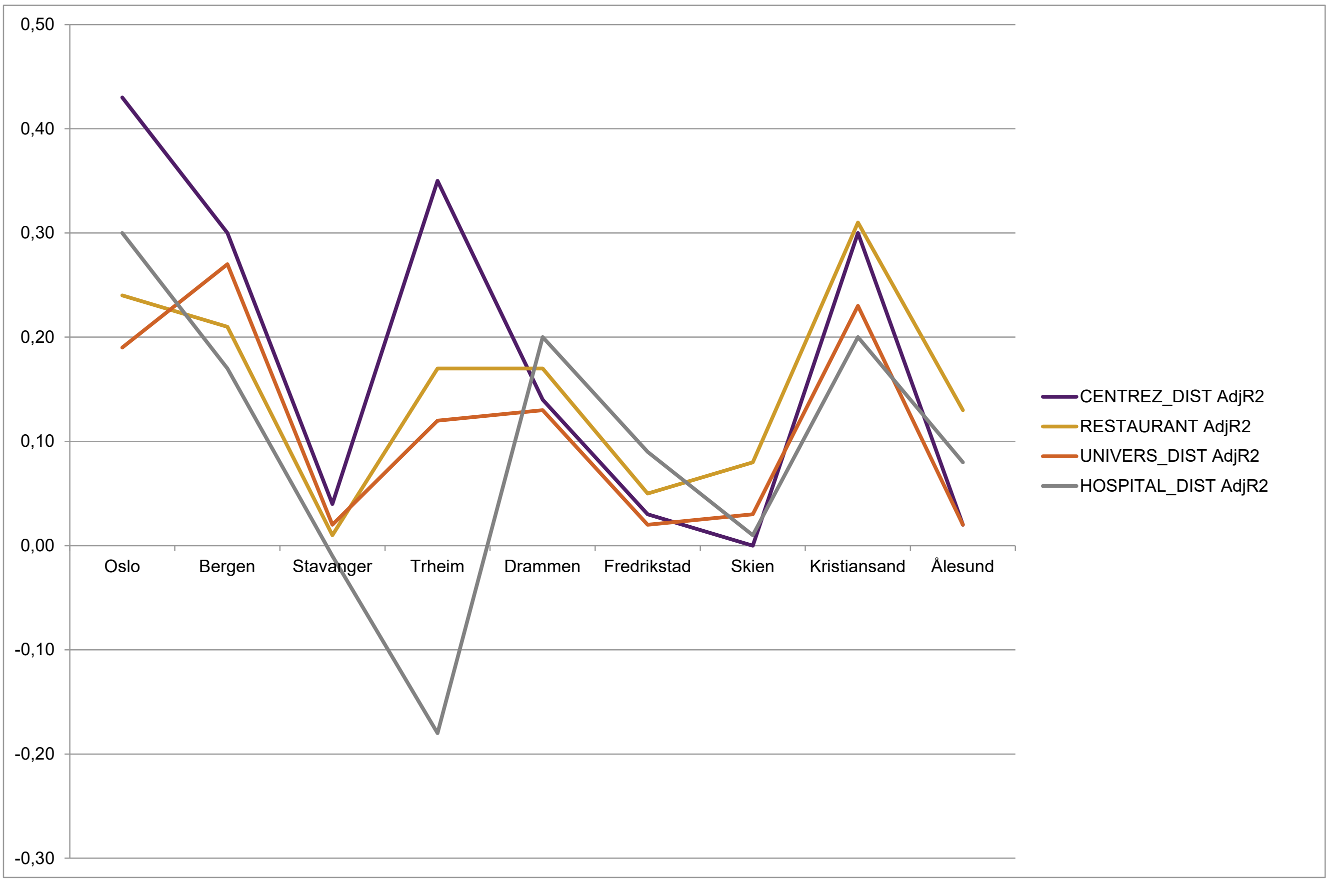
Type 1 (black) = significant variable, contributes consistently in same direction (+ OR –) to price variation
Type 2 (blue) = contributes in same direction (+ OR –) as type 1-figures*, but NOT consistently or as a non-significant variable*
Type 3 (brown) = contributes in opposite direction (+ OR –) as type 1-figures

An interesting finding is that urban "pull" is not all about city size, as is also true for "Education level".

Oslo "pulls" most, but 3rd largest city Stavanger scores significantly lower than the others on these distance variables, being city centre (CENTREZ_DIST), restaurants (RESTAURANT DIST, higher education facilities(UNIVERS_DIST) and hospitals (HOSPITAL_DIST).

How AdjR^2 for these four variables are linked to each other become very clear when plotted together in the diagram below:

Figure 4.5. Price per m^2 - isolated AdjR^2 for 4 distance variables. Distances to: city centre, restaurant buildings, higher education buildings and hospitals

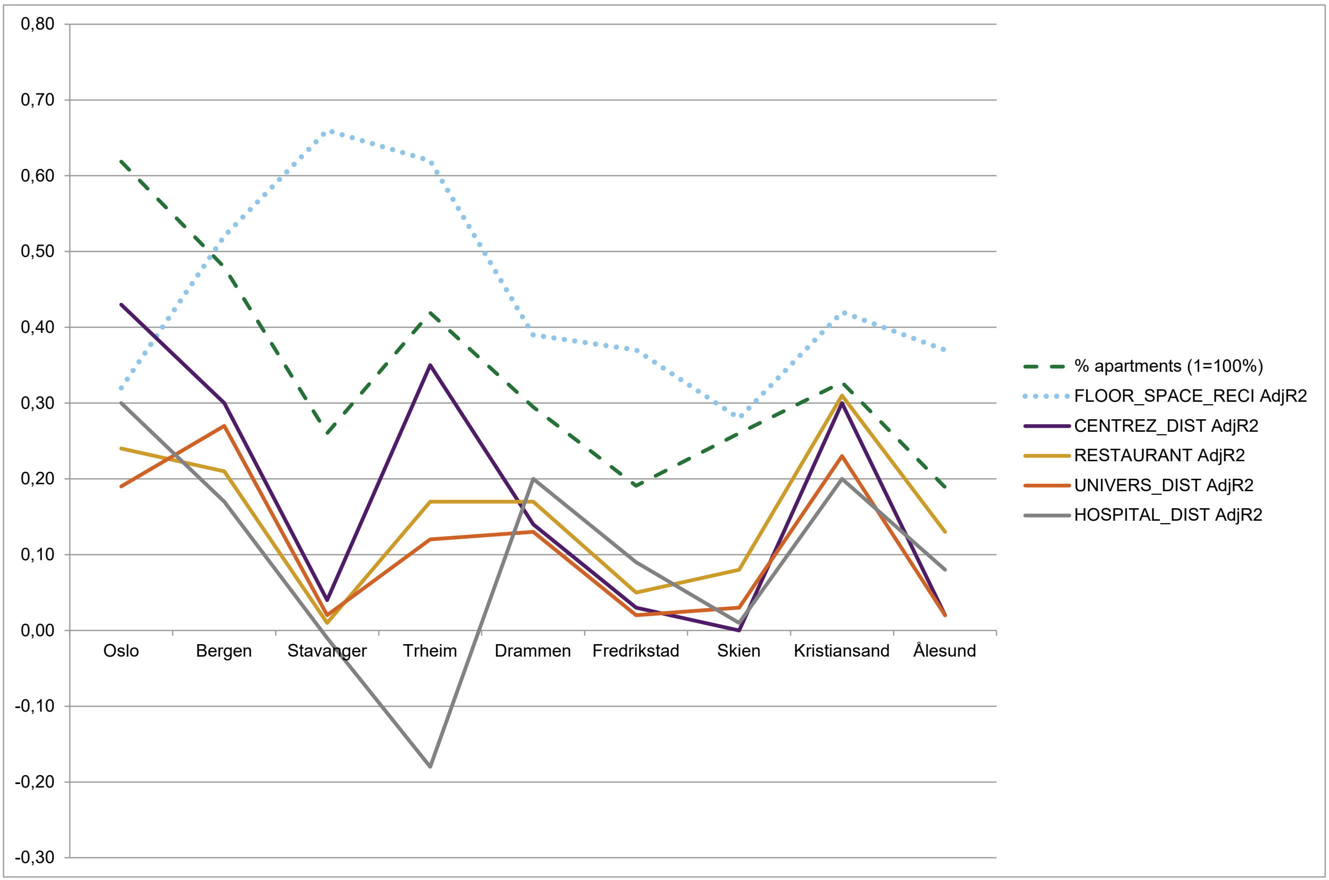


Below we add two more variables into the same diagram. One of them is the percentage of apartments[3] found in each of the urban settlement (dotted green line). There is a clear correlation between percentage apartments in a city and our four variables. We might speculate that this variable in Norway expresses a degree of urbanity, which is then not just a function of city size, effecting a "pull" on "Price per m^2".

We have also added AdjR^2 for *m^2 floor space (*FLOOR_SPACE_RECI, dotted light blue line*)*. For the largest 4-5 towns, we see that rising centrality effects (on "Price per m^2") leave less space for *m^2 floor space* as explanatory variable for "Price per m^2":

[3] Not a detached house, semi-detached house or free-standing house.

Figure 4.6. Price per m² - in addition to variables in figure 4.5 : Percentage of apartments found in each city, and AdjR² for variable floor space.



Oslo, Bergen and Trondheim are historically Norway's university towns, and have the highest amounts of students. Stavanger is the odd one out, and is a newer large city with its growth based on the oil industry. It does not share the same "higher education" traditions and has far fewer students. The "pull" of higher education facilities and its town centre count less. Also, education levels indicate less price variation in Stavanger.

The following table correlates "pull" of city centre (CENTREZ_DIST) with number of students studying in the main municipalities in our four largest cities. Both values are relatively small for Stavanger.

4 largest urban settlements: Number of students in main municipality & isolated AdjR2 for distance to city centre

Municipality	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**
Number of students	74 169	33 555	12 305	35 474
$AdjR^2$ - CITY CENTRE-DISTANCE	.43	.30	.04	.35

Adjusted "Age of building" weighs also highly in this approach, as well as "Mean age of population", at low levels. "Distance to water" is also found a significant variable. Concluding then for our three "perception survey" variables:
1. Using "Distance to university" as a proxy for "Education facilities", our conclusion is that the variable generally is a positive indicator of "Price per m²" throughout all cities. The variable "Distance to school" was not found significant.

2. For hospitals, as a proxy for "Health services", our conclusion is that hospitals not necessarily are perceived as attractive neighbours. There are instances of locations being a negative indicator of "Price per m^2". Our other variable on health services encompassed "Local medical centres", which we did not a find significant variable.
3. As in the "Total sales prices" approach, we did not find distance to "Public transport" to be significant in indicating price variation. The variable has been tested specifically in all cities, with similar results. Again, this does not mean that it doesn't matter. It rather suggests that city planners have achieved to distribute this in a fashion so access is close enough to not matter, pricewise.

No other variables are found significantly important in Oslo, which also encompass our "other" survey variables "Recreational areas", "Noise" and "Employment opportunities", with the same considerations as above for the "Total sales price" approach. Actually though, this is not totally true, as we on the city scale see correlation between higher price per m^2 and both more noise and further distance to recreational areas. City centres are noisiest and farthest from recreational areas, but their urban "pull" weighs more, giving highest price per m^2.

1.5. Potential for a time series – further work

The conclusions of chapter 1.3. point to how our spread of variables and variable types correlate to our proxy of attractivity, being dwelling prices for the year 2014. The same variables are tested in each city isolated, with resulting variation in correlation strengths.
The methodology gives an exact numeric on the correlation between dwelling price and for example neighbourhood' "Mean education level", for each of Norway's largest cities. This again allows for direct numeric comparisons between the cities, where nuances might be large or slight. These nuances can be said to be indicators of the status on socioeconomic divide between Norwegian cities, at the given time oy year 2014.

The potential in calculating these same correlations for a different year, or for creating a time series, is apparent. Results should pick on nuances of correlation values within each city, and between Norway's cities. What are the trends, and can we see that for example initiatives to alleviate socioeconomic divide in specific cities are having any effect? This is within the scope of this methodology, as long as data quality issues are in general equal for all larger cities within the project, in our case for Norway.

A second potential for a times series lies within the grid based attractiveness datasets generated for each of the four largest cities (chapter 6).
The variation in these predictions are by definition an expression of variation in attractivity. It is also this variation which is essential, not whether the predicted Kroner price is precise. Herein lies also the potential for a time series, mapping expected changes in attractivity, which again could be compared against actual changes in attractivity (house prices).

1. Generate the same statistics for a different year, e.g. + 2 years
2. Calculate attractiveness dataset year+2 with same coefficients as year 0
3. Overlay between attractiveness datasets year 0 and +2
4. Highlight locations with change in values, suspected change in attractivity

2. Definitions, variables and abbreviations

2.1. Definitions

Urban settlement / city The concept "city" is used in this report, and is meant as interchangeable with the concept "urban settlement". They differ from municipality by not being delineated by administrative boundaries. See appendix B on urban settlements

Regression analysis is a commonly used statistic in the social sciences. "Regression is used to evaluate relationships between two or more feature attributes. Identifying and measuring relationships lets you better understand what's going on in a place, predict where something is likely to occur, or begin to examine causes of why things occur where they do.

Ordinary Least Squares (OLS) is the best known of all regression techniques. It is also the proper starting point for all spatial regression analyses. It provides a global model of the variable or process you are trying to understand or predict; it creates a single regression equation to represent that process" [4].

"OLS a method for estimating the unknown parameters in a linear regression model, with the goal of minimizing the sum of the squares of the differences between the observed responses (values of the variable being predicted) in the given dataset and those predicted by a linear function of a set of explanatory variables. Visually this is seen as the sum of the squared vertical distances between each data point in the set and the corresponding point on the regression line – the smaller the differences, the better the model fits the data. In regression analysis, dependent variables are designated on the vertical Y axis and explanatory variables are designated on the horizontal X axis. These designations will form the equation for the line of best fit, which is determined from the least squares method."[5]

A **dependent variable** represents the quantity we wish to explain variation in, or the thing we are trying to explain

An **explanatory variable** represents a quantity whose variation will be used to explain variation in the dependent variable

R^2 (R-squared) is the coefficient of determination indicating goodness-of-fit of the regression. This statistic will be equal to one if fit is perfect, and to zero when the explanatory variables have no explanatory power whatsoever. This is a biased estimate of the population R^2, and will never decrease if additional explanatory variables are added, even if they are irrelevant.

AdjR^2 (Adjusted R-squared) is a slightly modified version of R^2, designed to penalize for the excess number of explanatory variables which do not add to the explanatory power of the regression. This statistic is always smaller than R^2 and can decrease as new regressors are added, and even be negative for poorly fitting models

The **coefficient** for each explanatory variable reflects both the strength and type of relationship the explanatory variable has to the dependent variable. When the sign associated with the coefficient is negative, the relationship is negative (for example, the larger the distance from the urban core, the smaller the number of residential burglaries). When the sign is positive, the relationship is positive (for

[4] http://resources.arcgis.com/en/help/main/10.1/index.html#/How_OLS_regression_works/

[5] https://en.wikipedia.org/wiki/Ordinary_least_squares

example, the larger the population, the larger the number of residential burglaries). Coefficients are given in the same units as their associated explanatory variables (a coefficient of 0.005 associated with a variable representing population counts may be interpreted as 0.005 people).

Statistically significant. An explanatory variable associated with a statistically significant coefficient is important to the regression model if theory/common sense supports a valid relationship with the dependent variable, if the relationship being modelled is primarily linear, and if the variable is not redundant to any other explanatory variables in the model.

Multicollinearity/VIF. Multicollinearity is a phenomenon in which two or more explanatory variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a substantial degree of accuracy. VIF (variance inflation factor) measures this degree of redundancy. As a rule of thumb, explanatory variables associated with VIF values larger than about 7.5 should be removed (one by one) from the regression model. Large VIF values indicating that two (or more) variables are telling the same story; one of them should be removed from your model.

Residuals - the observed/known dependent variable values minus the predicted/estimated values

The **Jarque-Bera** statistic indicates whether or not the residuals are normally distributed. If they are not, the model is biased, suggesting that a key variable is missing from the model.

Spatial autocorrelation assesses whether the regression residuals are spatially random. Statistically significant clustering of high and/or low residuals (model under- and overpredictions) indicates a key variable is missing from the model (misspecification).

Exploratory Regression-tool (ArcGis)
"Finding a properly specified OLS model can be difficult, especially when there are lots of potential explanatory variables you think might be important contributing factors to the variable you are trying to model (your dependent variable). The Exploratory Regression tool can help. It is a data mining tool that will try all possible combinations of explanatory variables to see which models pass all of the necessary OLS diagnostics. By evaluating all possible combinations of the candidate explanatory variables, you greatly increase your chances of finding the best model to solve your problem or answer your question. While Exploratory Regression is similar to Stepwise Regression (found in many statistical software packages), rather than only looking for models with high Adjusted R^2 values, Exploratory Regression looks for models that meet all of the requirements and assumptions of the OLS method"[6]

Passing models (specific to Exploratory Regression-tool)
Specific to output report for the ArcGis Exploratory Regression-tool :"…summaries give you an idea of how well your models are predicting (AdjR^2), and if any models pass all of the diagnostic criteria you specified. If you accepted all of the default Search Criteria (*Minimum Acceptable AdjR Squared***,** *Maximum Coefficient p-value Cut-off***,** *Maximum VIF Value Cut-off***,** *Minimum Acceptable Jarque Bera p-value***,** and *Minimum Acceptable Spatial Autocorrelation p-value*

[6] http://resources.arcgis.com/en/help/main/10.1/index.html#/How_Exploratory_Regression_works/

parameters), any models included in the Passing Models list will be properly specified OLS models. “[7]

2.2. Variables

See Appendix C for lists of variables used in this project.

2.3. Abbreviations

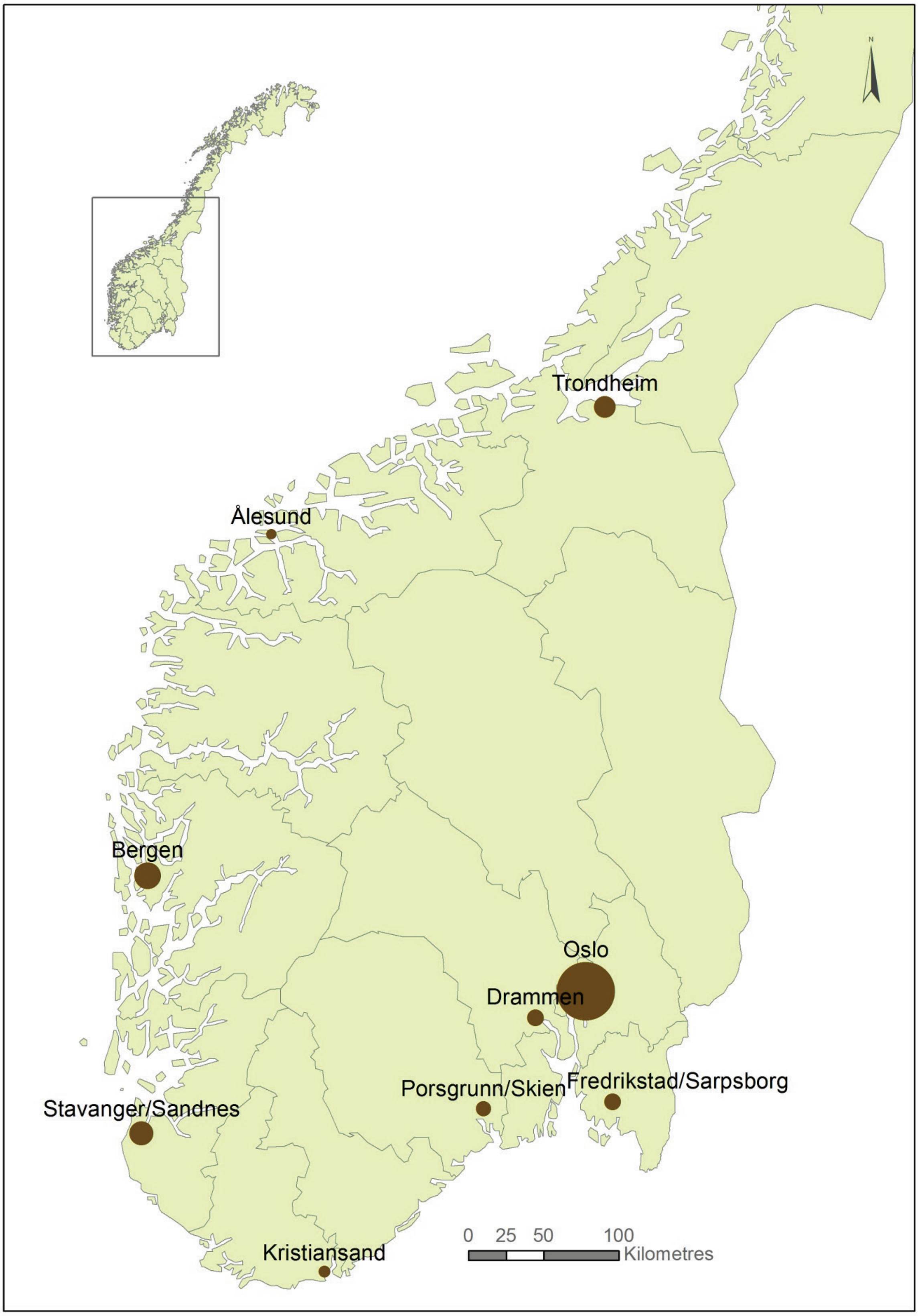
NOK = Kr

2.4. Overview - Urban settlements included in the project

Urban settlement	Short name	Population
Oslo	Oslo	*958 378*
Bergen	Bergen	*250 420*
Stavanger/Sandnes	Stavanger	*210 874*
Trondheim	Tr.heim	*175 068*
Drammen	Drammen	*113 534*
Fredrikstad/Sarpsborg	Fredrikstad	*108 636*
Porsgrunn/Skien	Skien	*91 737*
Kristiansand	Kr.sand	*60 583*
Ålesund	Ålesund	*50 917*

[7] http://resources.arcgis.com/en/help/main/10.1/index.html#/Interpreting_Exploratory_Regression_results

Figure 2.1. Location of urban settlements in the project. The area of each circle is proportional to population size



3. Methodology

3.1. Step-by-step overview

The following steps 1 - 8 illustrate the different parts of the work packages.

Step 1

Literature overview

Carrying out a literature overview of previous studies of how to combine the various datasets.

Step 2

Data structuring and georeferencing of statistical register data

Real estate data, georeferenced to Property centroid, geo-derived parameters

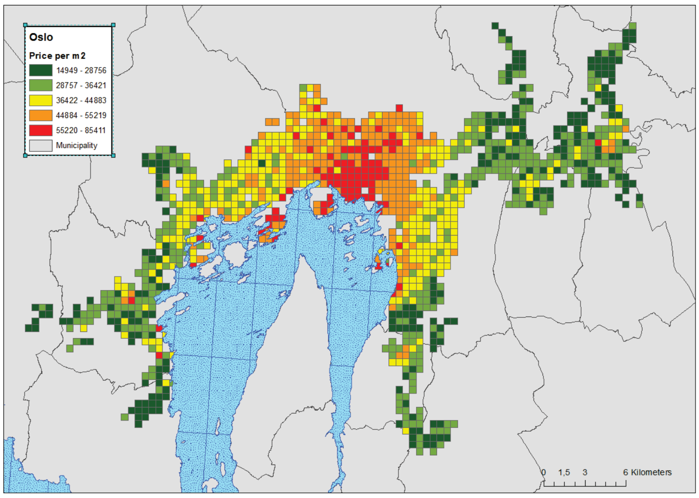
Data from **real estate agencies** by dwelling	
Variable type	**Variable**
Dwelling	DwellingId
	Floor space
	Age of building
	Total Sales Price
	Price per m^2
Property centroid	X coordinate
	Y coordinate
Distance to geographic entities	CentreZones, Recreational areas, Coast, etc..
Distance to buildings	Health institutions, Schools, Restaurants, etc..
Intensity-environment	Noise, Sun hours
Population characteristics within 250m radius	Household income, Education levels, Immigration, etc..
Employment	Employees within 5/10 km

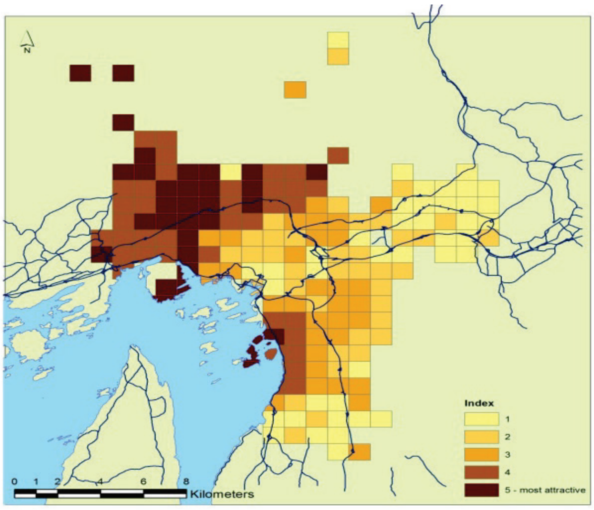
Step 3

<u>Examining and determine the best suitable final output format for attractive areas</u>

Based on the source data study, how the largely point based data can be presented in line with regulations concerning data protection, confidentiality and INSPIRE directive. Several options for dissemination (step 8) explored, where our chosen option is:

Predefined geographical grid dataset 500m X 500m





Step 4

<u>Production of attractive urban areas</u>

Exploratory regression analysis in order to obtain insights in the relationships – OSLO (urban settlement). Three separate approaches for Dependent variable:

1. Total Sales Price
2. Price per m^2
3. Comparable sizes

Extend scope of regression analysis to all urban settlements > 50 000: BERGEN, STAVANGER/SANDNES, TRONDHEIM, DRAMMEN, FREDRIKSTAD/SARPSBORG, KRISTIANSAND, PORSGRUNN/SKIEN, ÅLESUND

Step 5

<u>Production of attractive urban areas</u>

1. Ordinary Least Squares (OLS) linear regression with chosen explanatory variables from step 3 for OSLO, gaining coefficients and basis for prediction for OSLO.
2. Ordinary Least Squares (OLS) linear regression for all other urban settlements separately, with chosen explanatory variables from step 3 for OSLO. Calibration of coefficients specific to each urban settlement. Gaining coefficients as basis for prediction.

Two separate approaches for Dependent variable:

1. Total Sales Price
2. Price per m^2

Step 6

Production of attractive urban areas

Calculate and join chosen explanatory variables (step 4 & 5) to Norway's georeferenced building register (Cadastre)

Data from **building register** by dwelling	
Variable type	**Variable**
Building *(type Dwelling)*	*BuildingId*
	Floor space
	Age of building
Building centroid	X coordinate
	Y coordinate
Distance to geographic entities	CentreZones, Recreational areas, Coast, etc..
Distance to buildings	Health institutions, Schools, Restaurants, etc..
Intensity-environment	Noise, Sun hours
Population characteristics within 250m radius	Household income, Education levels, Immigration, etc..

Step 7

Production of attractive urban areas

1. Calculate predicted **Total Sales Price** for each building (type dwelling) in point based building dataset from step 6 → coefficients separately weighted, by urban settlement

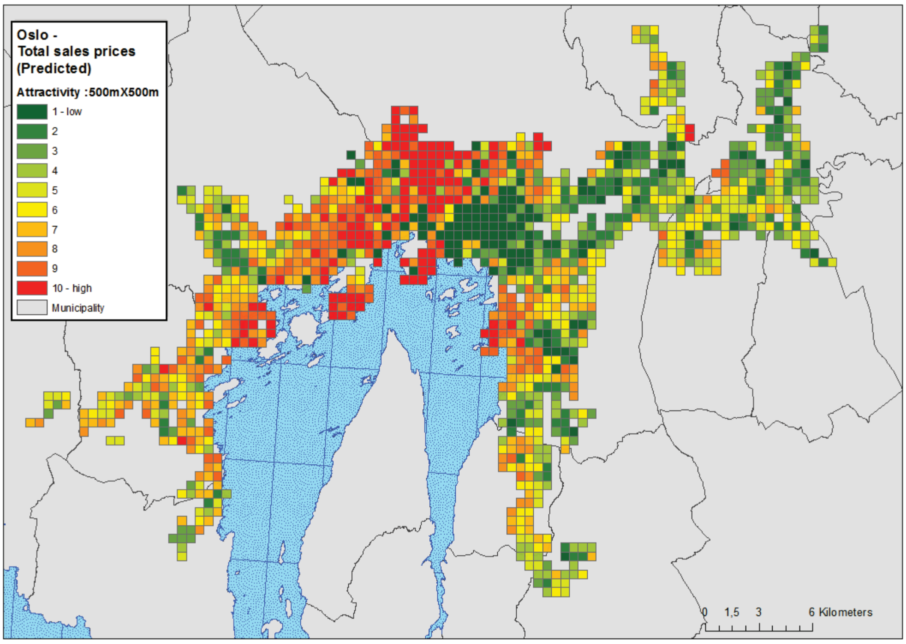
2. Calculate predicted **Price per m^2** for each building (type dwelling) in point based building dataset from step 6 → coefficients separately weighted, by urban settlement

Step 8
Production of attractive urban areas

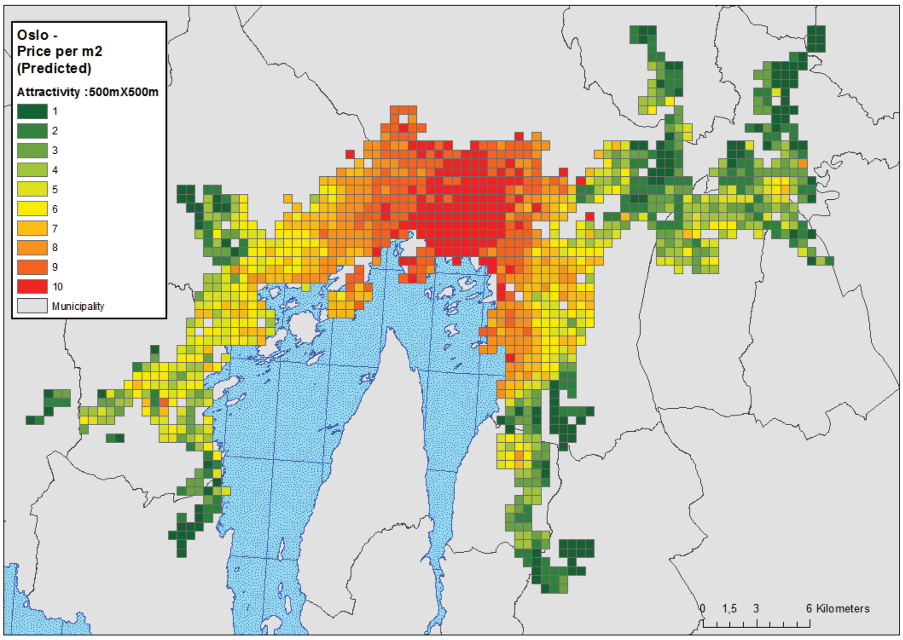
Calculate two grid-based attractivity indexes – overlay between Building points and 500m X 500m grid dataset. For each urban settlement:

1. Calculate a **Total sales price - Attractivity index** from medium predicted "Total sales price" in grid cell → index range from 1 to 10, Quantile grouping, where 10 is 10 per cent highest priced dwellings

2. Calculate a **Kr per m² - Attractivity index** from medium predicted "Kr per m²" in grid cell → index range from 1 to 10, Quantile grouping, where 10 is 10 per cent highest "Kr per m²- dwellings"



Predicted
Total sales price
Attractivity index 1-10



Predicted
Price per m²
Attractivity index 1-10

3.2. Regression models and coefficients

In the foreseen activities, we planned to test the “Quality of life - Perception survey” variables in relation to generated attractiveness datasets. As basis for generating these datasets, we set out to look at migration, housing prices, location of new buildings, income and education.

As described in chapter 1, we have deviated from this, as we found migration within a city to be problematic as an indicator for attractivity, as well as location of new buildings. Our focus areas were then housing prices, income and education, and with this shortened scope we found it wise to also reconsider our choice of methodology.

Our choice fell on **Ordinary Least Squares** (OLS) regression analysis. The choice of using OLS, allows us to use house prices as the variable we seek to explain (dependent variable), with income and education as external variables (explanatory variables) in a hedonic price function. This also accounts for the “Quality of life - Perception survey” variables, as public transport, education facilities and other variables do or do not affect price variation within cities. Not using these variables in this part of the project would be an odd choice, as we would be assuming that they do not affect price variation.

Statistics Norway have a long tradition for using hedonic methods and linear regression in their price indexes for dwellings and buildings, and have since 1992 utilized these methods as basis for calculating their **House price index**[8]**.**
The method builds on the assumption that the market price of a dwelling can be described as the function of internal (for example dwelling floor space) and external (for example quality of schools) characteristics. By estimating a hedonic price function, one can construct quality adjusted price indexes for the property market
OLS is also internationally a very well-known regression technique, also found described as “the proper starting point for all spatial regression analyses”[9]. See chapter 2.1 for description of OLS.

In the project, we also use the ArcGis tool **Exploratory Regression.** As described in chapter 2.1 - “*Finding a properly specified OLS model can be difficult, especially when there are lots of potential explanatory variables you think might be important contributing factors to the variable you are trying to model (your dependent variable). The Exploratory Regression tool can help. It is a data mining tool that will try all possible combinations of explanatory variables to see which models pass all of the necessary OLS diagnostics. By evaluating all possible combinations of the candidate explanatory variables, you greatly increase your chances of finding the best model to solve your problem or answer your question.*”[10]

[8] http://www.ssb.no/a/publikasjoner/pdf/notat_201210/notat_201210.pdf
[9] http://resources.arcgis.com/en/help/main/10.1/index.html#/How_OLS_regression_works/
[10] http://resources.arcgis.com/en/help/main/10.1/index.html#/How_Exploratory_Regression_works/

4. Description of the action

4.1. Data structuring and georeferencing of statistical register data

4.1.1. Literature overview

Throughout the first year of the project the project participants have met a range of persons with knowledge about quality of life and urban planning. Apart from the "Quality of life in cities Perception survey in 79 European cities" produced by the European Commission, Directorate-General for Regional and Urban Policy the following literature have been relevant for the results in the project.
See Literature overview (Appendix A).

4.1.2. Identifying comparable data

17 data sources have been utilised in the project.
See Appendix B for details on data sources.

4.1.3. Structuring and georeferencing data

See Appendix C for how data sources have been structured, georeferenced and combined to obtain datasets and variables directly used in the project, creating datasets prepared for analysis.

4.1.4. Real estate dwellings, prepared for analysis

Variables described in above sections (Appendix B) are added in to the real estate dwelling-dataset, prepared for analysis.
Real estate Dwelling data, georeferenced to Property centroid, geo-derived parameters

Real estate dwellings, prepared for analysis		
Variable type	**Variable**	**Comments**
Dwelling	DwellingId	
	Floor space	= Usable floor area (UFA) or Net internal area (NIA). UFA is the Gross internal area less the floor areas taken up by lobbies, enclosed machinery rooms on the roof, stairs and escalators, mechanical and electrical services, lifts, columns, toilet areas (other than in domestic property), etc..[11]
	Age of building	2014 minus Construction year. Buildings<1945=5
	Total Sales price	Adjusted for any debt connected to the dwelling (usual for Housing cooperatives)
	Price per m^2	
	Number of bedrooms	
Property centroid	X coordinate	
	Y coordinate	
	MunicipalityId	
	Urban settlement-id	
Distance to geographic entities/areas	CentreZoneId	
	Recreational areas	
	Lakes&Rivers & Coastline	Closest of Lakes&Rivers AND Coastline
	Distance to public rail transport	
	Distance from road with speed limit 60 km/h	
Distance to buildings	Primary Health Institutions	
	School	Primary/Secondary
	Hospital	
	Kindergarten	
	University/Higher Education	
	Restaurant	
	buildings built pre 1900	
Intensity-environment	Noise 2011 (day equivalent level in dba)	
	Number of Sun hours	
Population	Household income – before taxes	As mean of population within 250 metres
	Household income –after tax	As mean of population within 250 metres
	Level of education	As mean of population>25 years old within 250 metres
	Immigrants	As percentage of population within 250 metres
	Population with non-western ancestry	As percentage of population within 250 metres
	Age – mean if population	As mean of population within 250 metres
	Percentage below 18 years old	As percentage of population within 250 metres
Employment	Employees within 5 km	“as the crow flies”
	Employees within 10 km	“as the crow flies”

[11] https://en.wikipedia.org/wiki/Floor_area_(building)

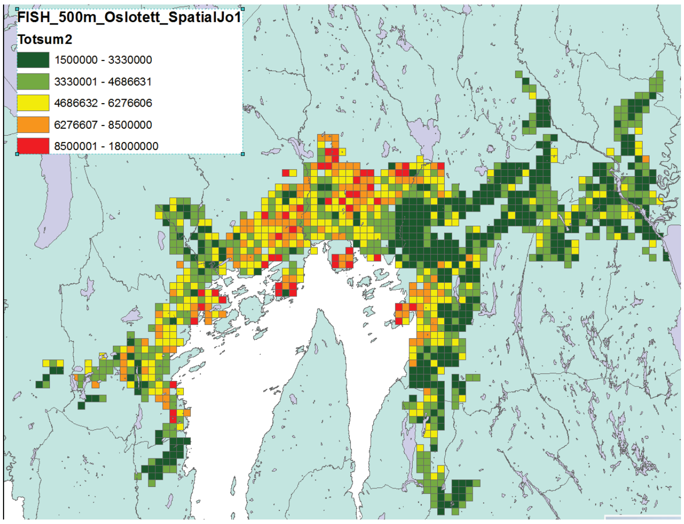
4.2. Examining and determine the best suitable output format

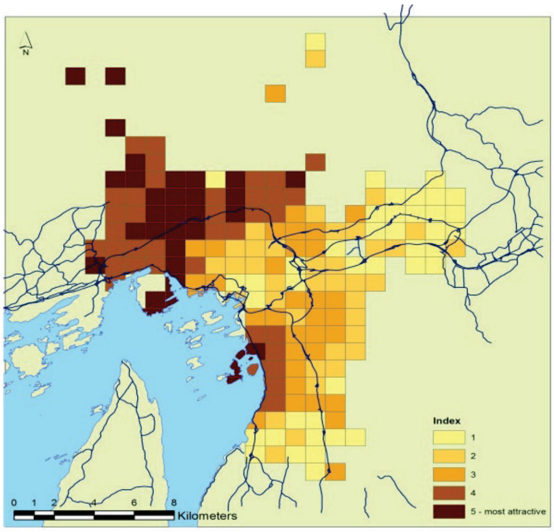
The data gathered is related to the location of properties or buildings. The output could therefore have been on property centroid points or building points, however, as the aim of this work is to identify attractive areas various options for dissemination were discussed.

The two output formats explored were an area layer based on geographical analysis similar to the urban centre zones or a predefined geographical grid dataset. It was found most adequate to choose the geographical grid as output format based on the following reasoning:

1. The data is point based and Statistics Norway has series of dataset that are converted from points to grid data. This make issues as confidentiality easier to handle since the confidentiality commission at Statistics Norway has handled similar data previously.
2. Publishing the attractive urban areas make it also easier for the users of the data, since they are familiar with the grid data as an output format for other statistics from Statistics Norway. They will also be able to easily do comparisons and further analyses based on the output.
3. Geographical grid is a Statistical unit that is handled by the Inspire directive which also make the transfer of the model to other European countries easier.

Predefined geographical grid dataset 500m X 500m



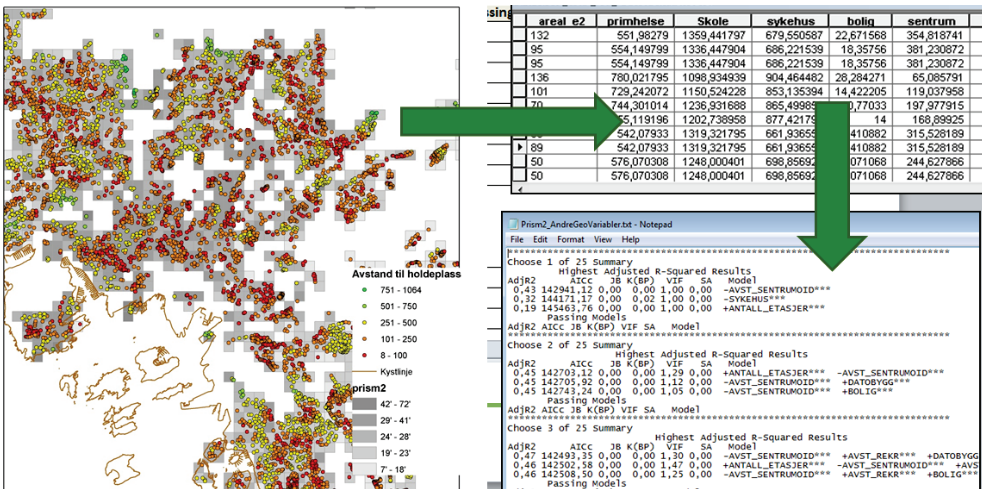


4.3. Exploratory regression analysis in order to obtain insights in the relationships

Based on the real estate data and the geo-derived parameters, we did an exploratory regression analysis (using the ArcGIS tool for this) to obtain insights in the relationships. We structured the data as described with the population being real estate sales. Using sales as the ultimate "ground truth" may be discussed, but this is an easy way to get objective data for attractiveness. In the regression analysis, we used the information provided in conjunction with the sales in addition to the geo-parameters.

Linear Transformation of explanatory variables
The exploratory regression analysis is based on Ordinary Least Square-regression analysis, testing for linear correlation. As the scope and scales of potential explanatory variables vary largely, Linear transformation of variables was tested on all explanatory variables found to be significant contributors, probing whether any linear transform might better correlation. The Dependent variables are unchanged.

Figure 4.1. Regression analysis to gain insight in to which parameters are important



Choice of dependent variable
Seeking to explain housing prices raises also the question whether it is "**Total sales price"** or "**Price per m^2"** that expresses attractivity best. It is possible to make a case for both approaches, that they both say something about attractivity, but different aspects. A potential buyer might tolerate a higher "**Price per m^2"** if being close to education facilities, restaurants and theatres is more important than amount floor space. At a different place in life the same buyer might prefer/need more space, trying with his or her means to optimize on space, in an as attractive location as possible.

Instead of choosing one of these two approaches we have in the project explored both, finding that most variables are only significant in one of the approaches.
A third approach where similar sized dwellings are compared is also explored. We have called it the **Compared same size-**approach.

Figure 4.2. "Total sales prices", Oslo – Mean within 500m X 500m grid cell

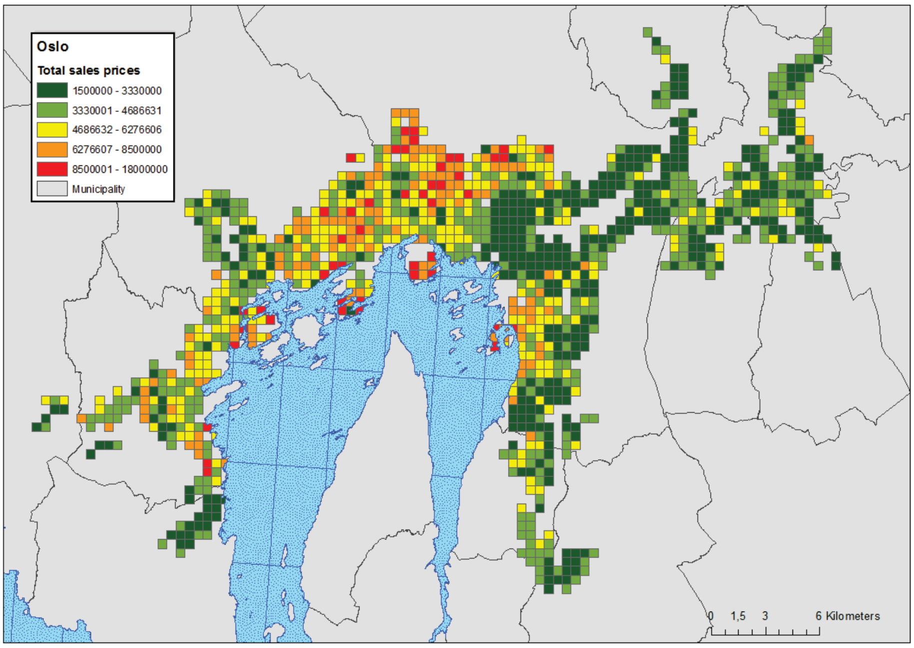
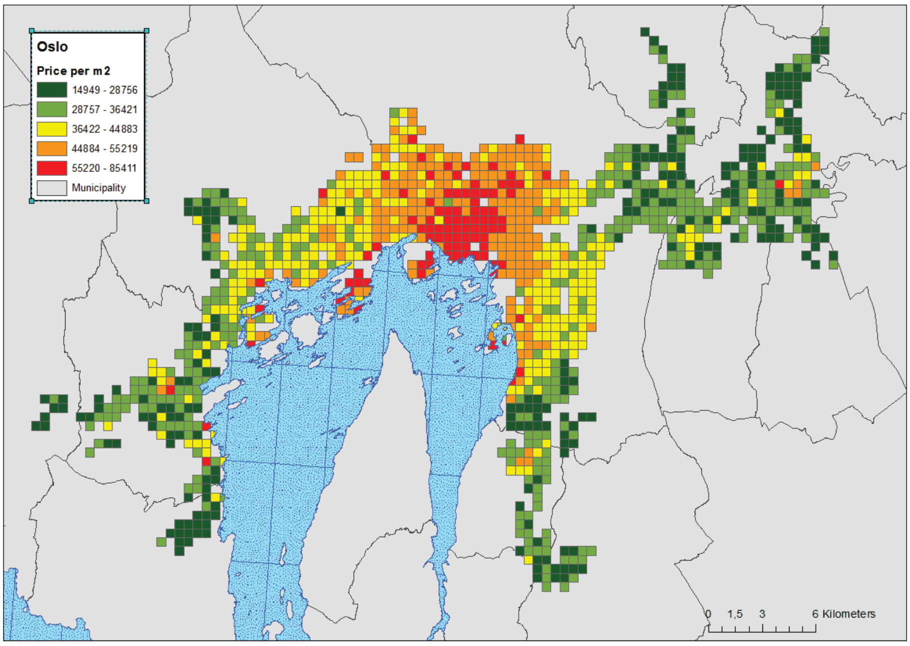


Figure 4.3. Price per m², Oslo - Mean within 500m X 500m grid cell



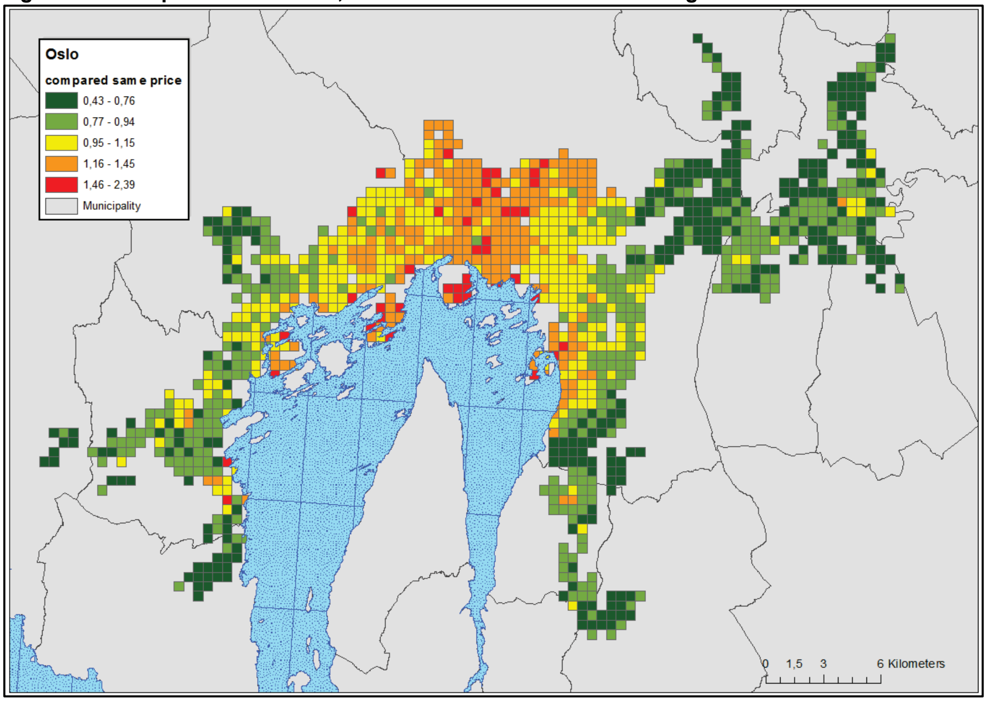
The two illustrations give a visual impression of variation in actual **Total sales prices** and **Price per m^2** in Oslo (2014), based on mean values within a 500m X 500m grid. "Natural Break" is used to divide the values to 5 groups, with quite extreme price differences between most and least expensive areas for both approaches.

There is apparent geographic clustering of house prices throughout the city, especially for **Price per m^2**, gravitating out from the clustering of red cells where Oslo town center lies.

Total sales prices appear in some parts more dispersed, and more clustered in others. The map suggests an east-west divide, commonly also perceived as such in Oslo.

The project methodology seeks to find variables to explain this price variation, using the findings to predict estimates of **Total sales prices** and **Price per m^2**.

Figure 4.4. Compared same sizes, Oslo - Mean within 500m X 500m grid cell



In the **Compared same size-**approach, m^2 floor space is "baked" in to the Dependent variable (that we wish to seek to explain).
The generated values are: *Price per m^2* DIVIDED by *Mean Price per m^2* for all sold similar size dwellings (+- 4 m^2) in the city.

Example:
Kroner per m^2 for properties of $20m^2$ DIVIDED by **Mean Kroner per m^2** for properties $16m^2$-$24m^2$

The average will always be 1.

Square metres floor space is in our project found to be a significant explanatory variable in explaining variation in both "Total sales price" and "Price per m^2". As isolated variable, it can be used to explain 60 per cent of price variation of "Total sales price" (AdjR^2 of 0.60) in Oslo, and 32 per cent of variation in "Price per m^2" (AdjR^2 of 0.32).

In the **Compared same size-**approach, m^2 floor space is "baked" into the Dependent variable, where each sold dwelling "Price per m^2" is compared to "mean Price per m^2" for all sold dwellings with similar floor space (+-4 m^2). Even though our data includes all dwelling sales in Norway for a whole year, we found that including +- $4m^2$ in each comparison was necessary to smooth the data, compensating for natural chance and randomness.

The Compared same size approach can be said to be cleaner in giving more space to non-building-intrinsic explanatory variables. The below shows how this is true for the variable *Education level of population within 250 m*, in itself explaining 47 per cent of the variation in Compared same sizes, in comparison to 35 per cent and 20 per cent for the other approaches.

AdjR^2	**Total sales price**	**Price per m^2**	**Compared same size**
Isolated - *Education level of population within 250 m*	0.20	0.35	0.47

However, when combining all explanatory variables, the highest achievable Adjr^2 for the **Compared same sizes** approach is as low as 0.61 for Oslo, a figure that decreases when the same explanatory variables are tested on Norway's other larger urban settlements. The approach gives some interesting insights on the strength of individual variables, and the fact that size of city effects overall results. Still, the overall AdjR^2 might be said to be too small to be sufficient for meaningful prediction.

The prediction part our project is therefore based solely on the results for the “Total sales price” and “Price per m^2” approaches, with highest achieved AdjR^2 at respectively 0.82 and 0.74 for Oslo. For these approaches, we also only make predictions for cities where highest achieved AdjR^2 => 0.70, which exclude the smaller cities in the project.

The table below specifies the highest achievable AdjR^2 for the three different approaches, with the number of explanatory variables utilized in each model. The explanatory variables utilized in the three approaches are our models for best explanation.

Oslo		
	AdjR^2	**Number of explanatory variables in model**
Total sales price	0.82	8
Price per m^2	0.74	9
Compared same sizes	0.61	7

The ArcGIS Exploratory analysis tool is used. In relation to a chosen dependent variable, the tool firstly tests each explanatory variable isolated, secondly all pairs of variables, thirdly all threesomes of variables, and so on with as many explanatory variables brought into the analysis (upper limit:10). The analysis output specifies the highest achievers (R^2/AdjR^2) at each combination, whether variables in these combinations contribute significantly, in which direction they contribute (+ or -), VIF for the combination, and other measures.

Below is an example of output results for the 3 best combinations of 7 variables in explaining “Price per m^2” in Oslo, where **Passing Models** would specify a combination which passes all criteria set in the tool.

The combination fails on the **Jarque-Bera** test and **Spatial Autocorrelation test**, and is as such not a valid model.

```
Choose 7 of 9 Summary
                                                              Highest Adjusted R-Squared
Results
AdjR²      AICc   JB  K(BP)  VIF   SA   Model
 0,74 475833,93 0,00  0,00 2,82 0,00  +POP_EDUC_L***  -CENTREZ_DIST***  -WATER_DIST***
+FLOOR_SPACE_RECI***  +POP_AGE***  -UNIVERS_DIST***  -BUILDING_AGE***

 0,74 475956,20 0,00  0,00 2,91 0,00  -RESTAURANT_DIST***  +POP_EDUC_L***  -CENTREZ_DIST***
+FLOOR_SPACE_RECI***  +POP_AGE***  -UNIVERS_DIST***  -BUILDING_AGE***

 0,74 476110,96 0,00  0,00 2,90 0,00  -HOSPITAL_DIST***  +POP_EDUC_L***  -CENTREZ_DIST***
+FLOOR_SPACE_RECI***  +POP_AGE***  -UNIVERS_DIST***  -BUILDING_AGE***

       Passing Models
AdjR² AICc JB K(BP) VIF SA   Model
```

After running all combinations, each variables effect is summed up for all combinations:

1. how often a variable is significant
2. how often they contribute in each direction (+ or -)
3. Multicollinearity: VIF and violations on the test
4. Spatial Autocorrelation
5. Passing tests

Below is output summary for 1, 2, 3 and 4 above, for “Price per m^2” in Oslo:

```
Summary of Variable Significance
Variable           % Significant % Negative % Positive
HOSPITAL_DIST           100,00     100,00       0,00
RESTAURANT_DIST         100,00     100,00       0,00
POP_EDUC_L              100,00       0,00     100,00
CENTREZ_DIST            100,00     100,00       0,00
FLOOR_SPACE_RECI        100,00       0,00     100,00
BUILDING_AGE            100,00     100,00       0,00
WATER_DIST               96,88     100,00       0,00
POP_AGE                  96,09       4,30      95,70
UNIVERS_DIST             94,92      83,59      16,41

------------------------------------------------------------------------------

Summary of Multicollinearity
Variable           VIF Violations Covariates
HOSPITAL_DIST      1,91     0      --------
RESTAURANT_DIST    1,52     0      --------
POP_EDUC_L         1,94     0      --------
CENTREZ_DIST       2,97     0      --------
WATER_DIST         1,16     0      --------
FLOOR_SPACE_RECI   1,24     0      --------
POP_AGE            1,14     0      --------
UNIVERS_DIST       2,13     0      --------
BUILDING_AGE       1,22     0      --------

------------------------------------------------------------------------------
Summary of Residual Spatial Autocorrelation (SA)
SA    AdjR²          AICc       JB    K(BP)      VIF   Model
0,000000 0,817885 694356,672574 0,000000 0,000000 3,424741  -RESTAURANT_DIST***  -
CENTREZ_DIST***  -WATER_DIST***  +FLOOR_SPACE_SQR***  +POP_EDUC_L_P5***  +POP_INCOME***
+POP_AGE***  -BUILDING_AGE***
```

For Spatial correlation (Summary point 4), the analysis output specifies further that none of our combinations of variables pass the Spatial autocorrelation test, and that we as such do not have a model that passes all regression analysis tests.

This is also one of the important conclusion of our project. The variables we have at hand are not sufficient to create a model that meets all requirements. We might suspect that access to more intrinsic characteristics on the dwellings might have given variables that would remedy this.

We have proceeded by focusing on the other parameters/criteria

Criteria for final set of explanatory variables in each model, Oslo

We have set the following criteria for final variables in each of the 3 approaches, based on Oslo. Chosen variables meet ALL following criteria 1, 2A, 3B, 3:

1. Have a positive effect of >= 0.01 on total combined AdjR2 for the approach
2. Contribute significantly in explaining variation of the dependent variable, A and B
   - A. Significant more than 95 per cent of times
   - B. Contributes > 85 per cent of times in same direction (+ OR -)
3. No violations of Multicollinearity: VIF < 5

Based on above criteria, following variables are utilised (X) for the three approaches "**Total sales price**", **"Price per m^2"** and "**Compared same size**".

The column Variable short name states the shortened variable names actually used in the datasets.

Data from real estate agencies by dwelling					
Variable type	**Variable**	**Variable short name**	**Total Sales Price**	**Price per m^2**	**Compared same size**
Dwelling	DwellingId				
	Floor space	*Floor_space*	X	X	
	Age of building	*Building_age*	X	X	X
	Number of bedrooms				
Distance to geographic entities/areas	CentreZoneId	*CentreZ_dist*	X	X	X
	Recreational areas				
	Lakes&Rivers & Coastline	*Water_dist*	X	X	X
	Distance to public transport				
	Distance to public rail transport				
	Distance from road with speed limit 60 km/h				
Distance to buildings	Primary Health institutions				
	School (Primary/Secondary)				
	Hospital	*Hospital_dist*		X	X
	Kindergarten				
	University/Higher Education	*Univers_dist*		X	

	Restaurant	*Restaurant_dist*	X	X	X
	buildings built pre 1900				
Intensity-environment	Noise 2011 (day equivalent level in dba)				
	Number of Sun hours				
Population	Household income – before taxes				
	Household income –after taxes	*Pop_Income*	X		X
	Level of education	*Pop_Educ_L*	X	X	X
	Immigration				
	Population with non-western ancestry	*Pop_nonwest*			X
	Age – mean if population	*Pop_age*	X	X	
	Percentage below 18 years old	*Pop_child*			X
Employment	Employees within 5 km				
	Employees within 10 km				

The following chapters 4.4 and 4.5 looks at the exploratory analysis output for each “best” model, respectively for “Total sales price”, “Price per m^2”, and “Compared same sizes”.
Analysis output for Compared same sizes lies in appendix D.

Each chapter starts out with Oslo (which is the city the models are calibrated by), followed by summarized results on how the model fares in all other Norwegian urban settlements > 50 000 inhabitants.

When considering how our Oslo variables perform in other cities, criteria for passing the set criteria are somewhat “relaxed”. Variables should meet ALL following criteria 1, 2A, 3B, 3:

1. Have a positive effect on total combined AdjR^2 for the approach
2. Contribute significantly in explaining variation of the dependent variable, A and B
    - A. Significant more than 80 % of times
    - B. Contributes > 80 % of times in same direction (+ OR -)
3. No violations of Multicollinearity: VIF < 5

4.4. TOTAL SALES PRICE - Best model, data output

4.4.1. TOTAL SALES PRICE – Oslo

The following 8 variables best explain variation of "Total Sales Price" in Oslo, which combined have a $AdjR^2$ of 0.82:

Oslo	$AdjR^2$ (isolated)
RESTAURANT_DIST	0.00
CENTREZ_DIST	0.00
WATER_DIST	0.01
FLOOR_SPACE_SQR	0.60
POP_EDUC_L_P5	0.20
POP_INCOME	0.39
POP_AGE	0.02
BUILDING_AGE	0.02
COMBINED	0.82

Discussion on choice of variables
The three variables that stand out are FLOOR_SPACE_SQR, POP_EDUC_L_P5, POP_INCOME.
An AdjR2 of 0.76 is already reached with just FLOOR_SPACE_SQR and POP_EDUC_L_P5, with 0.78 POP_INCOME is added. 7 variables = 0.82.

Output from the analysis for Oslo (sections on Jarque-Bera and Spatial Autocorrelation removed) is printed below, with summary tables for significance and Multicollinearity. A short discussion of results follows after the output:

```
******************************************************************************
Choose 1 of 8 Summary
Highest Adjusted R-Squared Results
AdjR2      AICc    JB K(BP)  VIF   SA   Model
 0,60 712558,05 0,00  0,00 1,00 0,00  +FLOOR_SPACE_SQR***
 0,39 722346,08 0,00  0,00 1,00 0,00  +POP_INCOME***
 0,20 728602,46 0,00  0,00 1,00 0,00  +POP_EDUC_L_P5***
******************************************************************************
Choose 2 of 8 Summary
Highest Adjusted R-Squared Results
AdjR2      AICc    JB K(BP)  VIF   SA   Model
 0,76 700465,17 0,00  0,00 1,00 0,00  +FLOOR_SPACE_SQR***  +POP_EDUC_L_P5***
 0,72 703917,70 0,00  0,00 1,17 0,00  -CENTREZ_DIST***  +FLOOR_SPACE_SQR***
 0,66 708512,31 0,00  0,00 1,06 0,00  -RESTAURANT_DIST***  +FLOOR_SPACE_SQR***
******************************************************************************
Choose 3 of 8 Summary
Highest Adjusted R-Squared Results
AdjR2      AICc    JB K(BP)  VIF   SA   Model
 0,78 698208,98 0,00  0,00 1,74 0,00  -CENTREZ_DIST***  +FLOOR_SPACE_SQR***  +POP_EDUC_L_P5***
 0,78 698227,31 0,00  0,00 1,61 0,00  -CENTREZ_DIST***  +FLOOR_SPACE_SQR***  +POP_INCOME***
 0,78 699203,57 0,00  0,00 1,23 0,00  -RESTAURANT_DIST***  +FLOOR_SPACE_SQR***  +POP_EDUC_L_P5***
******************************************************************************
Choose 4 of 8 Summary
Highest Adjusted R-Squared Results
AdjR2      AICc    JB K(BP)  VIF   SA   Model
 0,80 696609,64 0,00  0,00 1,82 0,00  -CENTREZ_DIST***  +FLOOR_SPACE_SQR***  +POP_EDUC_L_P5***
+POP_AGE***
```

```
 0,80 696947,34 0,00  0,00 1,62 0,00  -RESTAURANT_DIST***  -CENTREZ_DIST***  +FLOOR_SPACE_SQR***
+POP_INCOME***
 0,80 696975,05 0,00  0,00 1,61 0,00  -CENTREZ_DIST***  +FLOOR_SPACE_SQR***  +POP_INCOME***  -
BUILDING_AGE***
******************************************************************************
Choose 5 of 8 Summary
Highest Adjusted R-Squared Results
AdjR2      AICc   JB K(BP)  VIF   SA   Model
 0,81 695802,67 0,00  0,00 1,85 0,00  -CENTREZ_DIST***  +FLOOR_SPACE_SQR***  +POP_EDUC_L_P5***
+POP_AGE***  -BUILDING_AGE***
 0,81 695886,00 0,00  0,00 1,61 0,00  -CENTREZ_DIST***  +FLOOR_SPACE_SQR***  +POP_INCOME***
+POP_AGE***  -BUILDING_AGE***
 0,81 695917,28 0,00  0,00 3,07 0,00  -CENTREZ_DIST***  +FLOOR_SPACE_SQR***  +POP_EDUC_L_P5***
+POP_INCOME***  +POP_AGE***
******************************************************************************
Choose 6 of 8 Summary
Highest Adjusted R-Squared Results
AdjR2      AICc   JB K(BP)  VIF   SA   Model
 0,81 694851,54 0,00  0,00 3,22 0,00  -CENTREZ_DIST***  +FLOOR_SPACE_SQR***  +POP_EDUC_L_P5***
+POP_INCOME***  +POP_AGE***  -BUILDING_AGE***
 0,81 695237,01 0,00  0,00 1,62 0,00  -RESTAURANT_DIST***  -CENTREZ_DIST***  +FLOOR_SPACE_SQR***
+POP_INCOME***  +POP_AGE***  -BUILDING_AGE***
 0,81 695322,78 0,00  0,00 3,26 0,00  -RESTAURANT_DIST***  -CENTREZ_DIST***  +FLOOR_SPACE_SQR***
+POP_EDUC_L_P5***  +POP_INCOME***  +POP_AGE***
******************************************************************************
Choose 7 of 8 Summary
Highest Adjusted R-Squared Results
AdjR2      AICc   JB K(BP)  VIF   SA   Model
 0,82 694467,76 0,00  0,00 3,36 0,00  -RESTAURANT_DIST***  -CENTREZ_DIST***  +FLOOR_SPACE_SQR***
+POP_EDUC_L_P5***  +POP_INCOME***  +POP_AGE***  -BUILDING_AGE***
 0,82 694668,87 0,00  0,00 3,31 0,00  -CENTREZ_DIST***  -WATER_DIST***  +FLOOR_SPACE_SQR***
+POP_EDUC_L_P5***  +POP_INCOME***  +POP_AGE***  -BUILDING_AGE***
 0,81 695038,88 0,00  0,00 1,62 0,00  -RESTAURANT_DIST***  -CENTREZ_DIST***  -WATER_DIST***
+FLOOR_SPACE_SQR***  +POP_INCOME***  +POP_AGE***  -BUILDING_AGE***
******************************************************************************
Choose 8 of 8 Summary
Highest Adjusted R-Squared Results
AdjR2      AICc   JB K(BP)  VIF   SA   Model
 0,82 694356,67 0,00  0,00 3,42 0,00  -RESTAURANT_DIST***  -CENTREZ_DIST***  -WATER_DIST***
+FLOOR_SPACE_SQR***  +POP_EDUC_L_P5***  +POP_INCOME***  +POP_AGE***  -BUILDING_AGE***
******************************************************************************

------------------------------------------------------------------------------

          Summary of Variable Significance
Variable           % Significant % Negative % Positive
FLOOR_SPACE_SQR      100,00       0,00     100,00
CENTREZ_DIST          98,44      84,38      15,62
POP_AGE               98,44       3,12      96,88
BUILDING_AGE          98,44      98,44       1,56
WATER_DIST            96,88     100,00       0,00
POP_INCOME            96,88       3,12      96,88
RESTAURANT_DIST       96,09      83,59      16,41
POP_EDUC_L_P5         95,31       0,78      99,22

------------------------------------------------------------------------------

       Summary of Multicollinearity
Variable            VIF Violations Covariates
RESTAURANT_DIST     1,51     0      --------
CENTREZ_DIST        2,56     0      --------
WATER_DIST          1,09     0      --------
FLOOR_SPACE_SQR     1,64     0      --------
POP_EDUC_L_P5       3,42     0      --------
POP_INCOME          3,21     0      --------
POP_AGE             1,20     0      --------
BUILDING_AGE        1,21     0      --------

------------------------------------------------------------------------------
```

```
Table Abbreviations
AdjR² Adjusted R-Squared
AICc  Akaike's Information Criterion
JB    Jarque-Bera p-value
K(BP) Koenker (BP) Statistic p-value
VIF   Max Variance Inflation Factor
SA    Global Moran's I p-value
Model Variable sign (+/-)
Model Variable significance (* = 0,10; ** = 0,05; *** = 0,01)
------------------------------------------------------------------------------
```

A ninth variable considered was "Distance to Hospital", which is included in the "Price per m^2"-approach. Added as a ninth variable, increased distance impacts prices negatively in 87.5 per cent of all combinations, a higher percentage than distance to "Restaurant" and distance to "Town centre" (centre zone). However, adding it does not contribute to a higher AdjR^2, at the same time dragging down the + or – contribution percentage of other variables.

4.4.2. TOTAL SALES PRICE - Oslo with other cities

Applying the Oslo-model to other Norwegian cities with a population 50 000
The table below summarizes results for all cities. Blank (-) values express that there is absolutely no correlation to the dependent variable, while .00 indicates correlation (AdjR^2) > 0 < 0.005.
Cities are ordered by size, with a row for *population,* with Norway's second largest city Bergen being the only city in addition to Oslo passing all criteria.

Total sales sums - How much of price variation we are able to explain in Norway's 9 largest cities. AdjR^2 for each variable isolated, and total combined AdjR^2. ***AdjR^2 of 1 = 100 per cent***

AdjR^2									
Urban Settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
Population	*958 378*	*250 420*	*210 874*	*175 068*	*113 534*	*108 636*	*91737*	*60583*	*50917*
RESTAURANT_DIST	.00	.00	-.01	-	.01	.01	-	.00	.02
CENTREZ_DIST	.00	.02	-	-	-	.01	-	-	.14
WATER_DIST	.01	.01	-	.00	.00	-	.01	.03	.02
FLOOR_SPACE_SQR	**.60**	**.62**	**.67**	**.64**	**.53**	**.45**	**.48**	**.45**	**.53**
POP_EDUC_L_P5	**.20**	**.09**	**.04**	**.10**	**.23**	**.14**	**.14**	**.18**	**.08**
POP_INCOME	**.39**	**.23**	**.26**	**.24**	**.32**	**.25**	**.23**	**.22**	**.24**
POP_AGE	.02	.00	-	.00	.00	.00	-	-	.01
BUILDING_AGE	.02	.04	.00	.00	.05	.04	.06	.08	.01
COMBINED	**.82**	**.77**	**.75**	**.79**	**.74**	**.58**	**.57**	**.7**	**.64**

Type 1 (black) = significant variable, contributes consistently in same direction (+ OR –) to price variation
Type 2 (blue) = contributes in same direction (+ OR –) as type 1-figures*, but NOT consistently or as a non-significant variable*
Type 3 (brown) = contributes in opposite direction (+ OR –) as type 1-figures

Short discussion on values throughout all cities.
The combined $AdjR^2$ is clearly highest for Oslo (0.82), with the 3 second largest cities (similar to each other in size) at $AdjR^2$ levels between 0.75 and 0.77. $AdjR^2$ drops further with the smaller cities, to levels below 0.70.

A general conclusion that can be drawn is that the variables in the Oslo-model cannot be used to sufficiently explain and predict "Total sales price" for all Norwegian cites with a population > 50 000. One might argue that prediction based on $AdjR^2$ levels lower than 0.70 might be seen as not meaningful.

The three variables FLOOR_SPACE_SQR, POP_EDUC_L_P5, POP_INCOME continue be to the "weightiest" in explaining variation of "Total sales price" in all other cities.

4.4.3. Calculating coefficients - Ordinary Least Squares regression

The above chosen variables are used as basis in the ArcGIS tool Ordinary Least Squares (OLS) linear regression, gaining coefficients and basis for prediction, which is done in the later prediction part of the project.

Coefficients for dependent variable **Total sales price - Oslo**

Variable	**Coef**
Intercept	-3639487.42215
RESTAURANT_DIST	-164.69926393500
CENTREZ_DIST	-98.65370343370
WATER_DIST	-142.13425651900
FLOOR_SPACE_SQR	596942.51901100000
POP_EDUC_L_P5	138.17765320700
POP_INCOME	1.45942711512
POP_AGE	40625.91287910000
BUILDING_AGE	-7605.85554977000

Chapter 6 show how these coefficients are used at this stage. The chosen variables are combined with Norway's Georeferenced building register, performing prediction of 1. "Total Sales Price" and 2. "Price per m^2" on this basis.

See Chapter 6 for information on calculated coefficients for other cities, and overview of full OLS-reports.

4.5. PRICE PER M^2 - Best model, data output

4.5.1. PRICE PER M^2, Oslo

The following 9 variables best explain variation in “Price per m^2” in Oslo, which all combined have a AdjR^2 of 0.74:

Oslo	AdjR^2 (isolated)
HOSPITAL_DIST	0.30
RESTAURANT_DIST	0.24
POP_EDUC_L	0.35
CENTREZ_DIST	0.43
WATER_DIST	0.04
FLOOR_SPACE_RECI	0.32
POP_AGE	0.01
UNIVERS_DIST	0.19
BUILDING_AGE	0.16
COMBINED	**0.74**

While 3 variables can predict nearly all variation found for “Total sales price” (for Oslo AND other cities), there is a much more spread when explaining “Price per m^2”.

Distance to centre zone (CENTREZ_DIST) is much more important, with an AdjR^2 of 0.43. POP_EDUC_L and FLOOR_SPACE_RECI are also important in “Price per m^2”. (as they are in “Total sales price”), and the 3 combined variables can account for an AdjR^2 of 0.69, just 0.05 short of the total achievable 0.74.

There is a quartet of variables which express distance to city amenities: distance to centre zone (CENTREZ_DIST), distance to restaurant buildings (RESTAURANT_DIST), distance to university buildings (UNIVERS_DIST) and distance to hospitals (HOSPITAL_DIST. Though similar, the Multicollinearity between them is quite low (VIF < 3), so all are included in the model.

POP_INCOME (important for Total sales price) is not found to be significant in explaining variation in “Price per m^2”, and is not included in our model.

Output from the analysis (sections on Jarque-Bera and Spatial Autocorrelation removed) is printed below, with summary tables for significance and Multicollinearity. Discussion of results follows after the output:

```
*****************************************************************************
Choose 1 of 9 Summary
Highest Adjusted R-Squared Results
AdjR²      AICc   JB K(BP)  VIF   SA   Model
 0,43 493879,98 0,00  0,01 1,00 0,00  -CENTREZ_DIST***
 0,35 497075,99 0,00  0,00 1,00 0,00  +POP_EDUC_L***
 0,32 497961,38 0,00  0,00 1,00 0,00  +FLOOR_SPACE_RECI***
*****************************************************************************
Choose 2 of 9 Summary
Highest Adjusted R-Squared Results
AdjR²      AICc   JB K(BP)  VIF   SA   Model
 0,65 482787,44 0,00  0,00 1,00 0,00  +POP_EDUC_L***  +FLOOR_SPACE_RECI***
 0,56 488083,27 0,00  0,00 1,15 0,00  -CENTREZ_DIST***  +FLOOR_SPACE_RECI***
 0,51 490280,37 0,00  0,00 1,38 0,00  +POP_EDUC_L***  -CENTREZ_DIST***
*****************************************************************************
Choose 3 of 9 Summary
Highest Adjusted R-Squared Results
AdjR²      AICc   JB K(BP)  VIF   SA   Model
 0,69 479887,68 0,00  0,00 1,65 0,00  +POP_EDUC_L***  -CENTREZ_DIST***  +FLOOR_SPACE_RECI***
 0,68 480360,22 0,00  0,00 1,14 0,00  +POP_EDUC_L***  +FLOOR_SPACE_RECI***  -UNIVERS_DIST***
 0,67 481026,86 0,00  0,00 1,23 0,00  -RESTAURANT_DIST***  +POP_EDUC_L***  +FLOOR_SPACE_RECI***
*****************************************************************************
Choose 4 of 9 Summary
Highest Adjusted R-Squared Results
AdjR²      AICc   JB K(BP)  VIF   SA   Model
 0,71 478372,92 0,00  0,00 1,73 0,00  +POP_EDUC_L***  -CENTREZ_DIST***  +FLOOR_SPACE_RECI***
+POP_AGE***
 0,71 478588,80 0,00  0,00 1,68 0,00  +POP_EDUC_L***  -CENTREZ_DIST***  +FLOOR_SPACE_RECI***  -
BUILDING_AGE***
 0,70 479068,52 0,00  0,00 1,16 0,00  +POP_EDUC_L***  +FLOOR_SPACE_RECI***  -UNIVERS_DIST***  -
BUILDING_AGE***
*****************************************************************************
Choose 5 of 9 Summary
Highest Adjusted R-Squared Results
AdjR²      AICc   JB K(BP)  VIF   SA   Model
 0,73 476605,42 0,00  0,00 1,75 0,00  +POP_EDUC_L***  -CENTREZ_DIST***  +FLOOR_SPACE_RECI***
+POP_AGE***  -BUILDING_AGE***
 0,72 477406,26 0,00  0,00 1,18 0,00  +POP_EDUC_L***  +FLOOR_SPACE_RECI***  +POP_AGE***  -
UNIVERS_DIST***  -BUILDING_AGE***
 0,72 477814,93 0,00  0,00 1,94 0,00  -RESTAURANT_DIST***  +POP_EDUC_L***  -CENTREZ_DIST***
+FLOOR_SPACE_RECI***  +POP_AGE***
*****************************************************************************
Choose 6 of 9 Summary
Highest Adjusted R-Squared Results
AdjR²      AICc   JB K(BP)  VIF   SA   Model
 0,73 476155,57 0,00  0,00 2,82 0,00  +POP_EDUC_L***  -CENTREZ_DIST***  +FLOOR_SPACE_RECI***
+POP_AGE***  -UNIVERS_DIST***  -BUILDING_AGE***
 0,73 476317,19 0,00  0,00 1,94 0,00  -RESTAURANT_DIST***  +POP_EDUC_L***  -CENTREZ_DIST***
+FLOOR_SPACE_RECI***  +POP_AGE***  -BUILDING_AGE***
 0,73 476391,25 0,00  0,00 1,76 0,00  +POP_EDUC_L***  -CENTREZ_DIST***  -WATER_DIST***
+FLOOR_SPACE_RECI***  +POP_AGE***  -BUILDING_AGE***
*****************************************************************************
Choose 7 of 9 Summary
Highest Adjusted R-Squared Results
AdjR²      AICc   JB K(BP)  VIF   SA   Model
 0,74 475833,93 0,00  0,00 2,82 0,00  +POP_EDUC_L***  -CENTREZ_DIST***  -WATER_DIST***
+FLOOR_SPACE_RECI***  +POP_AGE***  -UNIVERS_DIST***  -BUILDING_AGE***
 0,74 475956,20 0,00  0,00 2,91 0,00  -RESTAURANT_DIST***  +POP_EDUC_L***  -CENTREZ_DIST***
+FLOOR_SPACE_RECI***  +POP_AGE***  -UNIVERS_DIST***  -BUILDING_AGE***
 0,74 476110,96 0,00  0,00 2,90 0,00  -HOSPITAL_DIST***  +POP_EDUC_L***  -CENTREZ_DIST***
+FLOOR_SPACE_RECI***  +POP_AGE***  -UNIVERS_DIST***  -BUILDING_AGE***
*****************************************************************************
Choose 8 of 9 Summary
Highest Adjusted R-Squared Results
AdjR²      AICc   JB K(BP)  VIF   SA   Model
 0,74 475711,71 0,00  0,00 2,91 0,00  -RESTAURANT_DIST***  +POP_EDUC_L***  -CENTREZ_DIST***  -
WATER_DIST***  +FLOOR_SPACE_RECI***  +POP_AGE***  -UNIVERS_DIST***  -BUILDING_AGE***
 0,74 475816,95 0,00  0,00 2,90 0,00  -HOSPITAL_DIST***  +POP_EDUC_L***  -CENTREZ_DIST***  -
WATER_DIST***  +FLOOR_SPACE_RECI***  +POP_AGE***  -UNIVERS_DIST***  -BUILDING_AGE***
 0,74 475926,81 0,00  0,00 2,97 0,00  -HOSPITAL_DIST***  -RESTAURANT_DIST***  +POP_EDUC_L***  -
CENTREZ_DIST***  +FLOOR_SPACE_RECI***  +POP_AGE***  -UNIVERS_DIST***  -BUILDING_AGE***
*****************************************************************************
```

```
Choose 9 of 9 Summary
Highest Adjusted R-Squared Results
AdjR²      AICc   JB K(BP)  VIF   SA   Model
 0,74 475700,41 0,00  0,00 2,97 0,00  -HOSPITAL_DIST***  -RESTAURANT_DIST***  +POP_EDUC_L***  -
CENTREZ_DIST***  -WATER_DIST***  +FLOOR_SPACE_RECI***  +POP_AGE***  -UNIVERS_DIST***  -BUILDING_AGE***
******************************************************************************


 Summary of Variable Significance
Variable           % Significant % Negative % Positive
HOSPITAL_DIST          100,00     100,00       0,00
RESTAURANT_DIST        100,00     100,00       0,00
POP_EDUC_L             100,00       0,00     100,00
CENTREZ_DIST           100,00     100,00       0,00
FLOOR_SPACE_RECI       100,00       0,00     100,00
BUILDING_AGE           100,00     100,00       0,00
WATER_DIST              96,88     100,00       0,00
POP_AGE                 96,09       4,30      95,70
UNIVERS_DIST            94,92      83,59      16,41

------------------------------------------------------------------------------

       Summary of Multicollinearity
Variable            VIF Violations Covariates
HOSPITAL_DIST      1,91     0      --------
RESTAURANT_DIST    1,52     0      --------
POP_EDUC_L         1,94     0      --------
CENTREZ_DIST       2,97     0      --------
WATER_DIST         1,16     0      --------
FLOOR_SPACE_RECI   1,24     0      --------
POP_AGE            1,14     0      --------
UNIVERS_DIST       2,13     0      --------
BUILDING_AGE       1,22     0      --------

------------------------------------------------------------------------------

Table Abbreviations
AdjR² Adjusted R-Squared
AICc  Akaike's Information Criterion
JB    Jarque-Bera p-value
K(BP) Koenker (BP) Statistic p-value
VIF   Max Variance Inflation Factor
SA    Global Moran's I p-value
Model Variable sign (+/-)
Model Variable significance (* = 0,10; ** = 0,05; *** = 0,01)
```

4.5.2. PRICE PER M² - Oslo with other cities

Applying the Oslo-model to other Norwegian cities with a population 50 000
The table below summarizes results for all cities. Blank (-) values express that there is absolutely no correlation to the dependent variable, while .00 indicates correlation (AdjR^2) > 0 < 0.005. The cities are ordered by size, with a row for *population,* with Norway's second largest city Bergen being the only city in addition to Oslo passing all criteria.

Price per m² - How much of price variation we are able to explain in Norway's 9 largest cities. AdjR^2 for each variable isolated, and total combined AdjR2.
AdjR^2 of 1 = 100 per cent

AdjR^2									
Urban settlement	Oslo	Bergen	Stavanger	Tr.heim	Drammen	Fredrikstad	Skien	Kristiansand	Ålesund
Population	*958 378*	*250 420*	*210 874*	*175 068*	*113 534*	*108 636*	*91 737*	*60 583*	*50 917*
HOSPITAL_DIST	.30	.17	.01	.18	.20	.09	.01	.20	.08
RESTAURANT_DIST	.24	.21	.01	.17	.17	.05	.08	.31	.13
POP_EDUC_L	.35	.30	.06	.10	.03	.01	.00	.05	-
CENTREZ_DIST	.43	.30	.04	.35	.14	.03	-	.30	.02
WATER_DIST	.04	.03	.04	.14	.08	.03	-	.03	-
FLOOR_SPACE_RECI	.32	.52	.66	.62	.39	.37	.28	.42	.37
POP_AGE	.01	.00	.00	.01	.15	.14	.08	.19	.07
UNIVERS_DIST	.19	.27	.02	.12	.13	.02	.03	.23	.02
BUILDING_AGE	.16	.10	.13	.12	.11	.06	.05	.09	.00
COMBINED	**.74**	**.73**	**.75**	**.77**	**.67**	**.52**	**.45**	**.70**	**.51**

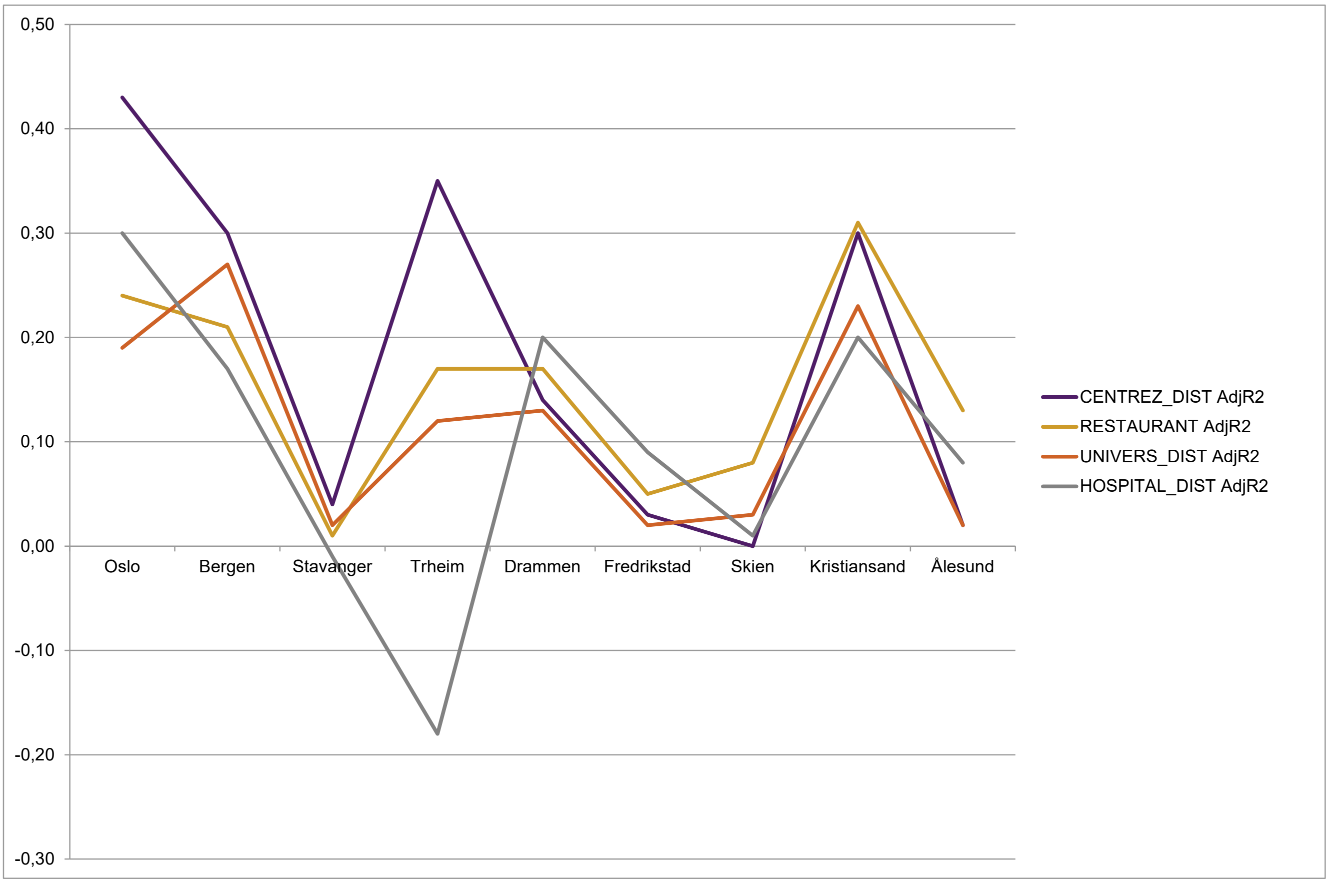
Type 1 (black) = significant variable, contributes consistently in same direction (+ OR –) to price variation
Type 2 (blue) = contributes in same direction (+ OR –) as type 1-figures*, but NOT consistently or as a non-significant variable*
Type 3 (brown) = contributes in opposite direction (+ OR –) as type 1-figures

Short discussion on values throughout all cities.
A general conclusion that can be drawn is that the variables in the Oslo-model cannot be used to sufficiently explain and predict "Price per m²" for all Norwegian cites with a population > 50 000. Any prediction based on AdjR^2 levels lower than 0.70 might be seen as not meaningful, which is the case for all but one (Kristiansand) of cities smaller than Trondheim, with levels falling to 0.45.

Questionable is also individual variable values in the smaller cities. One might argue that one in larger cities could expect more predictable macro level factors (for example attractive to live in town centre, or closeness to water a positive). Several of these factors may understandably fall away at a lower scale (for example no specific need to live in town centre, all amenities are easily accessed at a low cost). Local factors count understandably more, comparison between the cities being more difficult, as our model values behave more erratically.
Macro level factors would therefore weigh relatively more than building intrinsic factors in larger cities, on the city scale. This is true for Oslo and m² floor space, with a considerably smaller AdjR^2 than next biggest cities Bergen, Trondheim and Stavanger. It is also true for distance to town centre for Oslo, Bergen and Trondheim, Norway's main university towns with large student populations. Stavanger is the odd one out, but is a newer large city with growth based on the oil industry, with no "higher education" tradition and fewer students. The "pull" of its town centre counts less.

There are clear similarities for the four "centrality" variables within each of the four largest cities in the approach. If the pull of city centre is high (CENTREZ_DIST), then so it also is for distance to restaurants (RESTAURANT_DIST) and distance to higher education facilities(UNIVERS_DIST). Distance to hospitals (HOSPITAL_DIST) behaves somewhat differently, the correlation is not generally true in all cities.
As would be naturally expected, there are multicollinearity issues between some of these variables, for some of the cities. This is especially true for UNIVERS_DIST and HOSPITAL_DIST, where half of the cities (though not the two largest) have multicollinearity with CENTREZ_DIST.
How AdjR^2 for these four variables are linked to each other become very clear when plotted together in the diagram below:

Figure 4.5. Price per m^2 - isolated AdjR^2 for 4 distance variables. Distances to: city centre, restaurant buildings, higher education buildings and hospitals

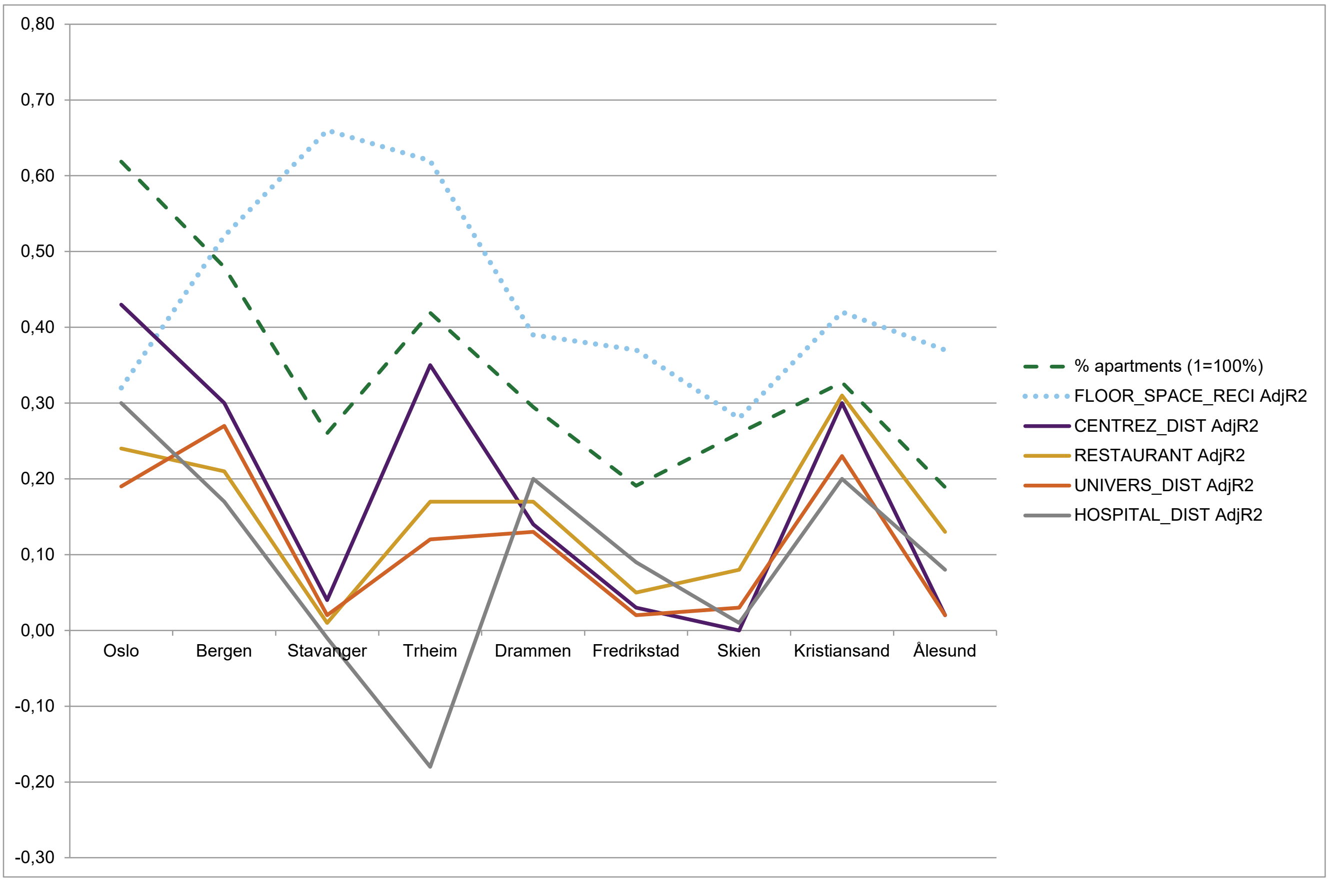


Below we add two more variables into the same diagram. One of them is the percentage of apartments[12] found in each of the urban settlement (dotted green line). There is a clear correlation between percentage apartments in a city and our four variables. We might speculate that this variable in Norway expresses a degree of urbanity, which is then not just a function of city size, effecting a "pull" on "Price per m^2".

We have also added AdjR^2 for *m^2 floor space (*FLOOR_SPACE_RECI, dotted light blue line*)*. For the largest 4-5 cities we see that rising centrality effects (on "Price per m^2") leave less space for *m^2 floor space* as explanatory variable for "Price per m^2":

[12] Not a detached house, semi-detached house or free-standing house.

Figure 4.6. Price per m² - in addition to variables in figure 4.5 : Percentage of apartments found in each city, and AdjR² for variable floor space.



4.5.3. Calculating coefficients - Ordinary Least Squares regression

The above chosen variables are used as basis in the ArcGIS tool Ordinary Least Squares (OLS) linear regression, gaining coefficients and basis for prediction, which is done in the later prediction part of the project.

Coefficients for dependent variable **Total Sales Price - Oslo**

Variable	Coef
Intercept	-3639487.42215
RESTAURANT_DIST	-164.69926393500
CENTREZ_DIST	-98.65370343370
WATER_DIST	-142.13425651900
FLOOR_SPACE_SQR	596942.51901100000
POP_EDUC_L_P5	138.17765320700
POP_INCOME	1.45942711512
POP_AGE	40625.91287910000
BUILDING_AGE	-7605.85554977000

Chapter 6 show how these coefficients are used at this stage. The chosen variables are combined with Norway's georeferenced building register (Cadastre), performing prediction of 1. "Total Sales Price" and 2. "Price per m²" on this basis.

See Chapter 6 for information on calculated coefficients for other cities, and overview of full OLS-reports.

5. Obtaining insights - Variables

The chapter contains a discussion on each variable, and how it performs in each our two main approaches, as indicators for variation in 1. “Total Sales Prices” and 2. “Price per m^2”.

	Data from real estate agencies by dwelling				
Chapter	**Variable type**	**Variable**	**Variable short name**	**Total Sales Price**	**Price per m^2**
5.1	Dwelling	DwellingId			
5.1.1		Floor space	*Floor_space*	X	X
5.1.2		Age of building	*Building_age*	X	X
5.1.3		Number of bedrooms			
5.2/5.2.1	Distance to geographic entities/areas	CentreZoneId	*CentreZ_dist*	X	X
5.2.2		Recreational areas			
5.2.3		Lakes&Rivers & Coastline	*Water_dist*	X	X
5.2.4		Distance to public transport			
5.2.5		Distance to public rail transport			
5.2.6		Distance from road with speed limit 60 km/h			
5.3 /5.3.1	Distance to buildings	Primary Health institutions			
5.3.2		School (Primary/Secondary)			
5.3.3		Hospital	*Hospital_dist*		X
5.3.4		Kindergarten			
5.3.5		University/Higher Education	*Univers_dist*		X
5.3.6		Restaurant	*Restaurant_dist*	X	X
5.3.7		Buildings built pre 1900			
5.4/5.4.1	Intensity-environment	Noise 2011 (day equivalent level in dba)			
5.4.2		Number of Sun hours			
5.5 /5.5.1	Population	Household income – before taxes			
5.5.2		Household income –after taxes	*Pop_Income*	X	
5.5.3		Level of education	*Pop_Educ_L*	X	X
5.5.4		Immigration			
5.5.5		Population with non-western ancestry	*Pop_nonwest*		
5.5.6		Age – mean if population	*Pop_age*	X	X
5.5.7		Percentage below 18 years old	*Pop_child*		
5.6/5.6.1	Employment	Employees within 5 km			
5.6.2		Employees within 10 km			

5.1. Variable type: Dwelling

5.1.1. Floor space was found to be a significant variable in explaining variance in both of our approaches, for all cities.

For "Total sales price" the value for Oslo (0.60) is somewhat lower than the three next largest cities (0.62, 0.67, 0.64), while the difference is much larger for "Price per m^2", with 0.32 for Oslo and 0.52, 0.66, 0.62 for the next three largest cities.

The differences suggest that Oslo is a more socioeconomically divided city than the three next largest.
"Price per m^2" for an apartment of a set size varies more by location in Oslo than in the other cities. Other factors (as education, income) affect more of the price variation in Oslo than for the next largest cities.

	Linear Transformation of FLOOR_SPACE to better correlation
Total Sales Price	**Square root →** FLOOR_SPACE_SQR
Price per m^2	**Reciprocal →** FLOOR_SPACE_RECI

Total Sales Price									
AdjR^2									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
FLOOR_SPACE_SQR	0.60	0.62	0.67	0.64	0.53	0.45	0.48	0.45	0.53

Price per m^2									
AdjR^2									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
FLOOR_SPACE_RECI	0.32	0.52	0.66	0.62	0.39	0.37	0.28	0.42	0.37

5.1.2 Age of building (Adjusted) was in Oslo found to be a significant contributor in explaining variance in both of our approaches. The original variable we tested for was age of building in relation to 2014.
Example: 2013 = 1, 2012 = 2
We found the variable to not contribute significantly. Many buildings built in the pre-war era are however seen as highly attractive (as well as often being in historical town centres), while post-war buildings are often seen as less attractive.

We tried to build this factor into an adjusted age of building variable:
Pre-war adjustment: all buildings built before 1945 are given age value=5
We found the adjusted variable to be generally significant in both approaches, for nearly all cities:

Total Sales Price									
AdjR²									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
BUILDING_AGE	0.02	0.04	0.00	0.00	0.05	0.04	0.06	0.08	0.01

Price per m²									
AdjR²									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
BUILDING_AGE	0.16	0.10	0.13	0.12	0.11	0.06	0.05	0.09	0.00

5.1.3 Number of bedrooms was in Oslo found to have very high multicollinearity with m^2 floor space, but with somewhat lower results. Of the two we chose to use m^2 floor space.

5.2. Variable type: Distance to geographic entities/areas

5.2.1 Distance to CentreZone was found to be a significant variable for Oslo for both approaches. There might be several centre zones within the city, where this variable express distance to the largest (in amount of employees) within the entire urban settlement.

In explaining total sales prices, the variable is valid only for Oslo, and then only at a very low level.

Total Sales Price									
AdjR²									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
CENTREZ_DIST	0.00	0.02	-	-	-	0.01	-	-	0.14

In explaining "Price per m²" the variable is significant throughout nearly all urban settlements, and is a strong indicator of variation in price per m². The "pull" of the city centre on "Price per m²" varies quite extremely between the cities, with a AdjR² of 0.43 for Oslo, 0.30 and 0.35 for Bergen and Trondheim, but as low as 0.04 for 3rd largest city Stavanger.

Price per m²									
AdjR²									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
CENTREZ_DIST	0.43	0.30	0.04	0.35	0.14	0.03	-	0.30	0.02

For "Price per m²", there are clear similarities for a set of "centrality" variables within most cities. If the "pull" of the city centre is high (CENTREZ_DIST), then so is equally true for RESTAURANT_DIST and UNIVERS_DIST, while HOSPITAL_DIST is more erratic.

Chapter 4.5.2 shows and discusses the relationship between these four variables throughout all cities, in explaining "Price per m²". The values are plotted together with m² floor space and a variable showing spread of building types within the cities.

5.2.2 Distance to Recreational areas was for Oslo not found to be a significant contributor in explaining variance in our approaches.

One might assume that living immediately adjacent to a recreational area might increase the attractivity of a dwelling. However, on the city scale our analysis does not find it to be true that distance to recreational areas generally effects "Total sales price" or "Price per m²". A reason might be that they are generally perceived as close enough for everybody. Another possibility is that looking at recreational areas in general is a too broad approach, and that splitting these into different types might give different results.

Our results actually show that higher distance to recreational areas correlates with higher price per m², as city centre areas often are furthest away from recreational areas. Urban pull "pulls" more than recreational areas.

5.2.3 Distance to Lakes&Rivers&Coastline was for Oslo found to be a significant contributor in explaining variance in all approaches. We have tested for only Coastline and also only Lakes&Rivers. We chose to incorporate these in one. We were able to do this without in general losing explanation effect.

The variable can to some degree indicate variation in total sales prices, for some cities. For other cities, it has no effect at all.

Total sales approach									
AdjR²									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
WATER_DIST	0.01	0.01	-	0.00	0.00	-	0.01	0.03	0.02

In explaining Price per m^2 the variable is much more potent, with higher values than for total sales prices. Closeness to water is attractive, but part of the strength of this value must also be ascribed to Norwegian cites largely being coastal towns, with city centres often originating from harbour areas.

Price per m²									
AdjR²									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
WATER_DIST	0.04	0.03	0.04	0.14	0.08	0.03	-	0.03	-

5.2.4 Distance to bus stop and 5.2.5 Distance to public rail transport:

Our variables for public transport were for Oslo not found to be a significant contributor in explaining variance in any of our approaches. Having not found the variables to be significant for Oslo, we also tested these separately for Bergen, Stavanger, Trondheim and Drammen. The tendency for all cities is actually that "Total sales price" increases by distance to public transport.

For "Price per m^2", we find in exploratory analysis the variable to contribute generally 50/50 positively/negatively to price variation. This does not mean that public transport is not important. As is true for our findings on several other services (as primary schools), the results suggest that access to public transport is close enough to often not matter pricewise, in relation to attractivity. A conclusion might be that city planning has been successful in distribution this specific service, access is close enough. Other variables matter more.

We might have seen the variable matter if our city delineation was based on administrative boundaries, encompassing all types of population density, where distance might matter more.

5.2.6 Distance from road with speed limit 60 km/h: The variable was for Oslo not found to be a significant contributor in explaining variance in any of our approaches.

5.3. Variable type: Distance to buildings

As a general comment, it might be expected that distance to services such as primary health institutions or Primary/Secondary schools is shorter in urban settlements than in more sparsely populated areas. The urban settlements used for delineating city borders in this project is based on density of buildings and population, and by definition exclude sparsely populated areas.

Our findings are that several services to the population (as primary schools) are close enough to often not matter pricewise, in relation to attractivity. A conclusion might be that city planning has been successful in distribution this specific service, access is close enough. Other variables matter more.

We might have seen some of these variables matter if the city delineation was based on administrative boundaries, encompassing all types of population density (more sparsely populated areas), where distance might matter more.

5.3.1 Distance to primary Health institutions was in Oslo not found to be a significant contributor in explain variance in our approaches.
This does not mean that the existence of this amenity does not matter. On the city scale it is, however, not significant. A conclusion might be that city planning has been successful in distribution this specific service, access is close enough. Other variables matter more.

5.3.2 Distance to School (Primary/Secondary) was in Oslo not found to be a significant contributor in explaining variance in our approaches.
As above for primary health institutions; this does not mean that the existence of this amenity does not matter. On the city scale it is, however, not significant. A conclusion might be that city planning has been successful in distribution this specific service, access is close enough.

Another variable related to schools is mean exam results of schools. This was a variable we in the project group wished to test for. We were however not able to obtain with the necessary accuracy to make a valid comparison.

5.3.3 Distance to Hospital was in Oslo found to be a significant valid contributor in the "Price per m^2"-approach.

For other cities, we find very high Multicollinearity to CENTREZ_DIST for some, and also two cities where closer proximity to hospitals in general are negative indicator of "Price per m^2" (Stavanger and Trondheim). A conclusion is that hospitals not necessarily are perceived as attractive neighbours, or at least that they not always are placed in attractive areas.

This does not mean that the existence of hospitals in the city does not matter. One would assume the opposite, but that already living within the city is close enough.

Price per m^2									
AdjR^2									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
HOSPITAL_DIST	0.30	0.17	0.01	0.18	0.20	0.09	0.01	0.20	0.08

For "Price per m^2", there are clear similarities for a set of "centrality" variables within most cities. If the "pull" of the city centre is high (CENTREZ_DIST), then so is equally true for RESTAURANT_DIST and UNIVERS_DIST, while HOSPITAL_DIST is more erratic.

Chapter 4.5.2 shows and discusses the relationship between these four variables throughout all cities, in explaining "Price per m^2". The values are plotted together with m^2 floor space and a variable showing spread of building types within the cities.

5.3.4 Distance to Kindergarten was in Oslo not found to be a significant contributor in explain variance in our approaches.

5.3.5 Distance to University/Higher Education was tested and found to be a significant valid contributor in the "Price per m^2"-approach, passing criteria for Oslo and Bergen. For the other larger cities there are Multicollinearity issues, especially with CENTREZ_DIST.

Def: building Type "University/Higher Education", only buildings > 1 500 m^2 floor space which contain an auditorium/ classroom.

For "Price per m^2", there are clear similarities for a set of "centrality" variables within most cities, which UNIVERS_DIST belongs to. As discussed above for Hospitals, chapter 4.5.2 shows and discusses the relationship between these four variables throughout all cities, in explaining "Price per m^2".

Price per m^2									
AdjR^2									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
UNIVERS_DIST	0.19	0.27	0.02	0.12	0.13	0.02	0.03	0.23	0.02

There is clearly correlation between distance to Higher education facilities and house prices, and it is not always the case that this is due to Multicollinearity to CENTREZ_DIST. There does seem to be a tendency for higher education facilities to be located in more affluent areas.

This does not necessarily mean that the location is the reason for their attractiveness. One might suggest that competition between higher education facilities both between cities and within a city make it necessary for them to be located in attractive areas, both from historical and current perspectives.
A historical perspective is locating to town centres due to practical necessity. The 1970's and 80' saw new opportunities and change of preferences, with some education facilities locating to the outskirts of towns, true for example of University of Trondheim-Dragvoll. Current discussions (as for Dragvoll) include a wish to relocate to more central areas, in response to changed preferences of students and employees.

A relevant question is also the general existence of Universities/Higher Education in a city, and the number of students studying there. Oslo, Bergen and Trondheim are historically Norway's university towns, and have the highest number of students. Stavanger is the odd out, and is a newer large city with "boom-town"-type growth based on the oil industry. It does not share the same "higher education" traditions and has far fewer students. The "pull" of higher education facilities AND its town centre counts less.

The following table correlates "pull" of city centre (CENTREZ_DIST) with number of students studying in the main municipalities in our four largest cities.

These are the municipalities in which the urban settlement's main centre zone falls within. Both values are relatively small for Stavanger.

4 largest urban settlements: Number of students in main municipality & isolated AdjR² for distance to city centre

Municipality	Oslo	Bergen	Stavanger	Tr.heim
Number of students	74 169	33 555	12 305	35 474
AdjR² - CITY CENTRE-DISTANCE	.43	.30	.04	.35

5.3.6 Distance to Restaurant was in Oslo found to be a significant valid contributor in the "Price per m²"-approach, passing also criteria for most cities. There is Multicollinearity to CENTREZ_DIST for some cities, but less than for *UNIVERS_DIST and* HOSPITAL_DIST. This is understandable as restaurants generally can be found in all parts of town, the variable picking on the price (attractivity) "pull" from local centres throughout the city. Restaurants may often be co-located with other distributed amenities as Shopping centres or libraries.

Price per m²									
AdjR²									
Urban settlement	Oslo	Bergen	Stavanger	Tr.heim	Drammen	Fredrikstad	Skien	Kristiansand	Ålesund
RESTAURANT_DIST	0.24	0.21	*0.01*	0.17	0.17	0.05	0.08	0.31	0.13

For "Price per m²", there are clear similarities for a set of "centrality" variables within most cities. If the "pull" of the city centre is high (CENTREZ_DIST), then so is equally true for RESTAURANT_DIST and UNIVERS_DIST, while HOSPITAL_DIST is more erratic.

Chapter 4.5.2 shows and discusses the relationship between these four variables throughout all cities, in explaining "Price per m²". The values are plotted together with m² floor space and a variable showing spread of building types within the cities.

RESTAURANT_DIST contributes also in the "Total sales price"-approach, though at very low levels.

Total Sales Price									
AdjR²									
Urban settlement	Oslo	Bergen	Stavanger	Tr.heim	Drammen	Fredrikstad	Skien	Kristiansand	Ålesund
RESTAURANT_DIST	0.00	0.00	*0.01*	-	0.01	0.01	-	0.00	0.02

5.3.7 Distance to "buildings built pre 1900" was in Oslo not found to be a significant contributor in explaining variance in our approaches.

5.4. Variable type: Intensity/environment

5.4.1 Noise-variable was in Oslo not found be a significant contributor in explaining variance in our approaches.

One might assume that living immediately adjacent to high levels of traffic noise might reduce the attractivity of a dwelling. Our analysis does however not find this to be true on the city scale, "Total sales price" or "Price per m^2" does not in general fall by increased noise on the city scale. We have made attempts on adjusting the variable, testing for dwellings with noise values over certain thresholds. We expected this to give some expected correlation, but were not able to achieve this.

We would expect noise to actually matter in a negative fashion, but this might very well be at a very local scale, over certain thresholds. For "Price per m^2" we actually find a correlation between rising price and rising decibel noise, as town centres with higher price per m^2 also have higher noise levels. This is in similar fashion to what we found for recreational areas, city centre areas are often furthest away from recreational areas.

5.4.2 Number of sun hours was in Oslo not found to be a significant contributor in explaining variance in our approaches. It is possibly true that Oslo is not densely enough habituated for real estate developers to build in areas with significantly less sun hours than others. This would be a drawback; other vacant areas would be prioritized.

5.5. Variable type: Population characteristics

In our project, we have chosen to look at population characteristics within 250 meters of each dwelling sales, testing variables that we think might have a correlation to dwelling sales prices. We have probed different sized buffer zones, and found 250 metres to be sufficient for a balance between enough address points to make valid averages, and detail. These variables we probe are household income, education level, immigration, age and children. This gave us data on:

1. Average household income within 250 meters of a dwelling sale
2. Average education level (persons => 26 years hold) within 250 meters of a dwelling sale
3. Percentage immigrants within 250 meters of a dwelling sale
4. Mean age within 250 meters of a dwelling sale
5. Percentage children (< 18 years old) within 250 meters of a dwelling sale

If there is a significant correlation, the relevant variable/variables can be said to be indicators for how dwellings are pushed up or down the demand scale of "supply and demand", location pushing up or down the price people are willing to pay, within a price scope for that specific area.

Finding a significant correlation between housing prices and socioeconomics of a neighbourhood population (as these variables are) suggest that socioeconomics matter, and the strength of that correlation for a city. These differences can be said to point to how more or less "divided" Norway's cities are.

5.5.1 Household income - before tax: We found the following "after tax" version to give the better results.

5.5.2 Household income - after tax is found to be a significant contributor especially to “Total sales price”, with a $AdjR^2$ of 0.39. It consistently contributes highly also in all other cities (>0.2), and meets set criteria of multicollinearity, significance and direction. We have not found it a significant contributor to “Price per m^2”.

Total Sales Price									
$AdjR^2$									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
POP_INCOME	**0.39**	**0.23**	**0.26**	**0.24**	**0.32**	0.25	0.23	0.22	0.24

Looking at Oslo and “Total sales prices”, POP_INCOME isolated has a $AdjR^2$ of 0.39, considerably higher than the other cities. The trend is that income matters, but it matters even more in the largest city Oslo. That income matters in what people are prepared to pay for a dwelling is not surprising. The maps below for Oslo also visualise that there is a geographic divide to mean income levels, which also correlates with total sales sums. Adjacent neighbourhoods are more likely to share the same values than not. There is a geographical divide, where Oslo is the most divided.

The map illustrations below show 500 X 500m grid cells covering the urban settlement of Oslo:
Map 5.1. visualizes mean “Household income” for all persons in each cell.
Map 5.2. visualises mean “Total sales price” for each cell.

Similarities are apparent, and we have shown that household income explains also 39 per cent (AdjR^2 = 0.39) of variation in “Total sales price” in Oslo.

Figure 5.1. Map 1. – Household income

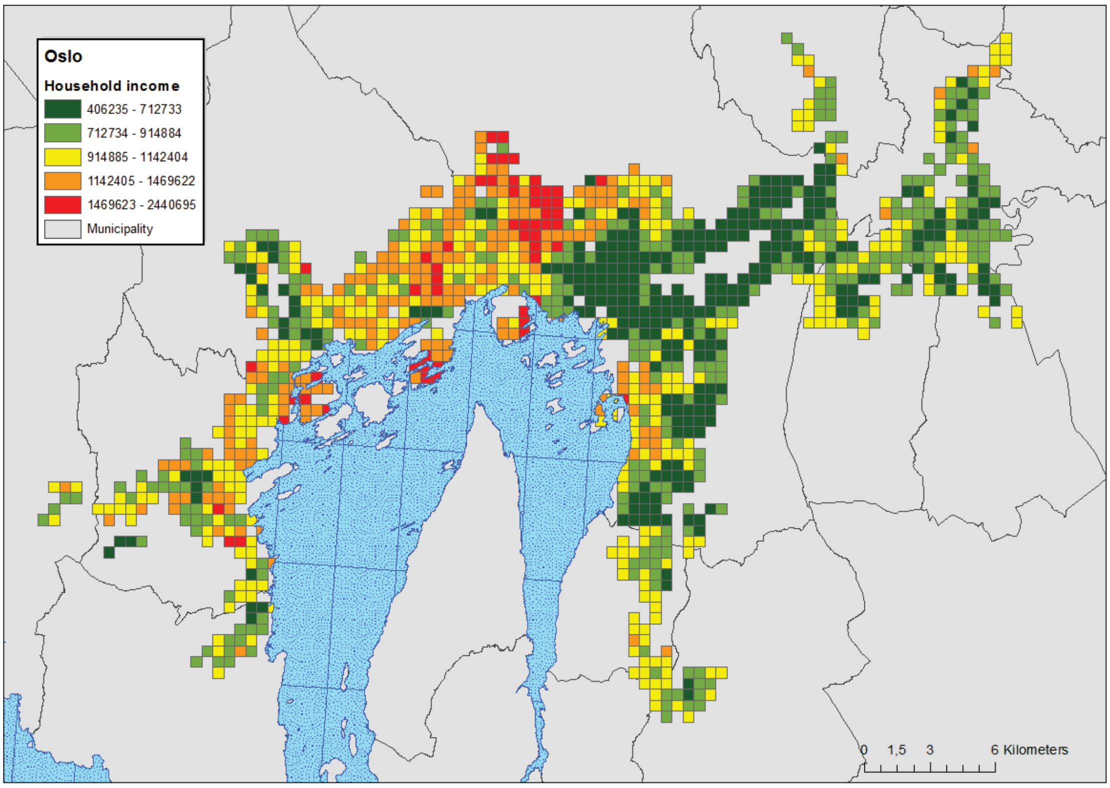


Figure 5.2. Map 2. – Total sales price

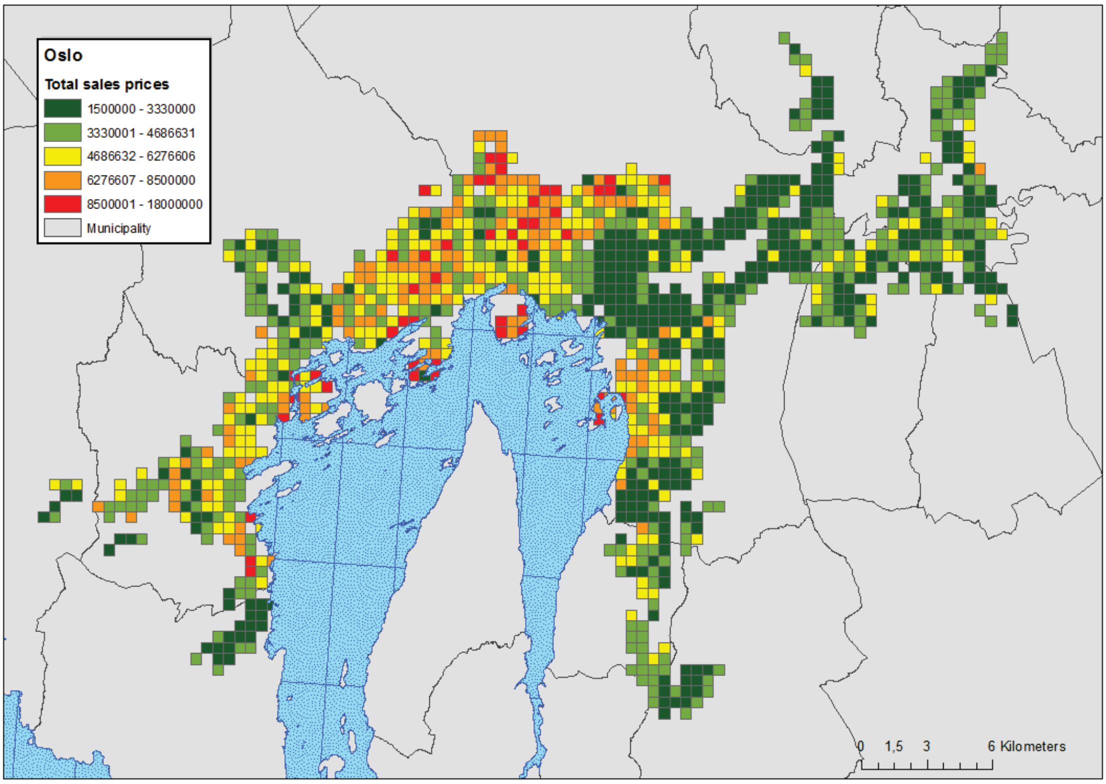
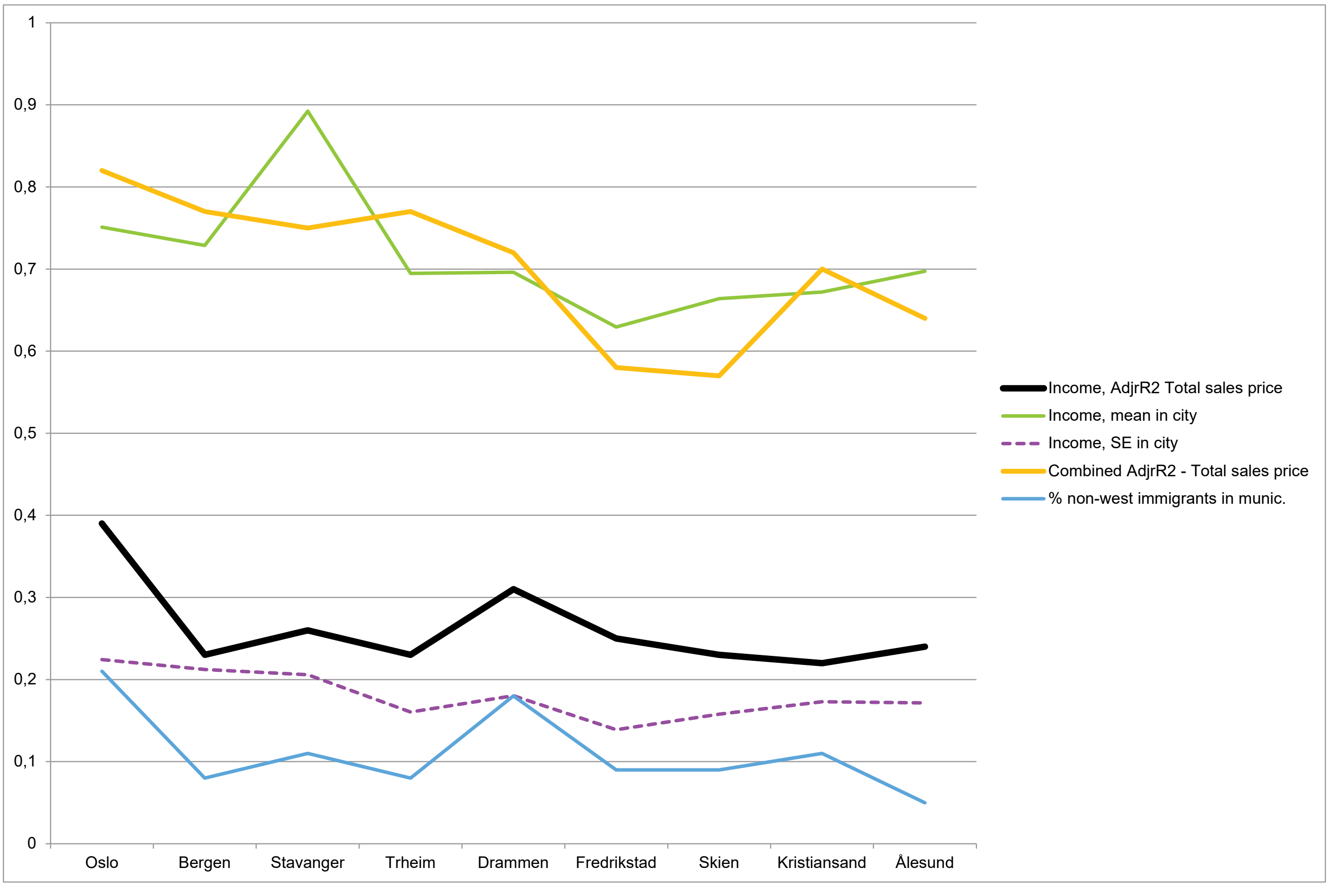


Figure 5.3. Total sales prices: The diagram looks at isolated AdjR² for income, mean income within each city, standard error for the distribution of income within the city and combined AdjR² for the approach. Visualized is also percentage non-western immigrants by municipality.



"Total sales price"-approach: The diagram looks at isolated AdjR^2 for income, mean income within each city, standard error for the distribution of income within the city and combined AdjR^2 for the approach. Visualized is also percentage non-western immigrants by municipality. This is official municipal statistics available from Statistics Norway, and values shown are for the municipalities which each urban settlement's main centre zone falls within.

From the diagram, we can read that mean income is highest in Stavanger, with Oslo second. There is no clear correlation between isolated AdjR^2 for income and mean income level, or its standard error.
There is however an apparent correlation between AdjR^2 for income and percentage non-western immigrants for municipalities. This might suggest that non-western households are highly overrepresented at the lower income levels. Despite this, multicollinearity between variables POP_INCOME and Pop_nonwest is low, with VIF of 1.42.

AdjR^2	AICc	JB	K(BP)	VIF	SA	Model
0,42	724726,87	0,00	0,00	1,32	0,00	+POP_EDUC_L_P5*** +POP_INCOME***
0,40	725518,87	0,00	0,00	1,40	0,00	-Pop_nonwest *** +POP_INCOME***
0,23	731393,36	0,00	0,00	1,70	0,00	-Pop_nonwest*** +POP_EDUC_L_P5***

5.5.3 Level of education is found to be a significant valid contributor in all tested approaches, possibly that of our non-intrinsic variables most useful in explaining price variation.

In the project, we have chosen to only include adults => 26 of age, excluding persons for which Statistics Norway have no education level information.

The data is based on NUS2000 codes, designating an education level from 1 to 8 for all persons. All NUS2000 codes are assigned a corresponding international code (ISCED97). See appendix B and D for details.

Looking at Oslo, level of education isolated has an AdjR^2 of 0.2, 0.35 and 0.47 (compared same size) in all three approaches, considerably higher than the other cities. The variable is also generally a significant strong contributor in all other cities, with second largest city Bergen also holding second highest values. Size matters, but it is not all about size, as 3rd largest city Stavanger has considerably lower values than 4th largest city Trondheim.

An interesting finding is that education levels actually better indicate attractivity than a variable as immigration. One can more accurately describe neighbourhood-characteristics by education levels than immigration levels. A conclusion might be that "choices we make", as taking education, actually is more important than where you come from.

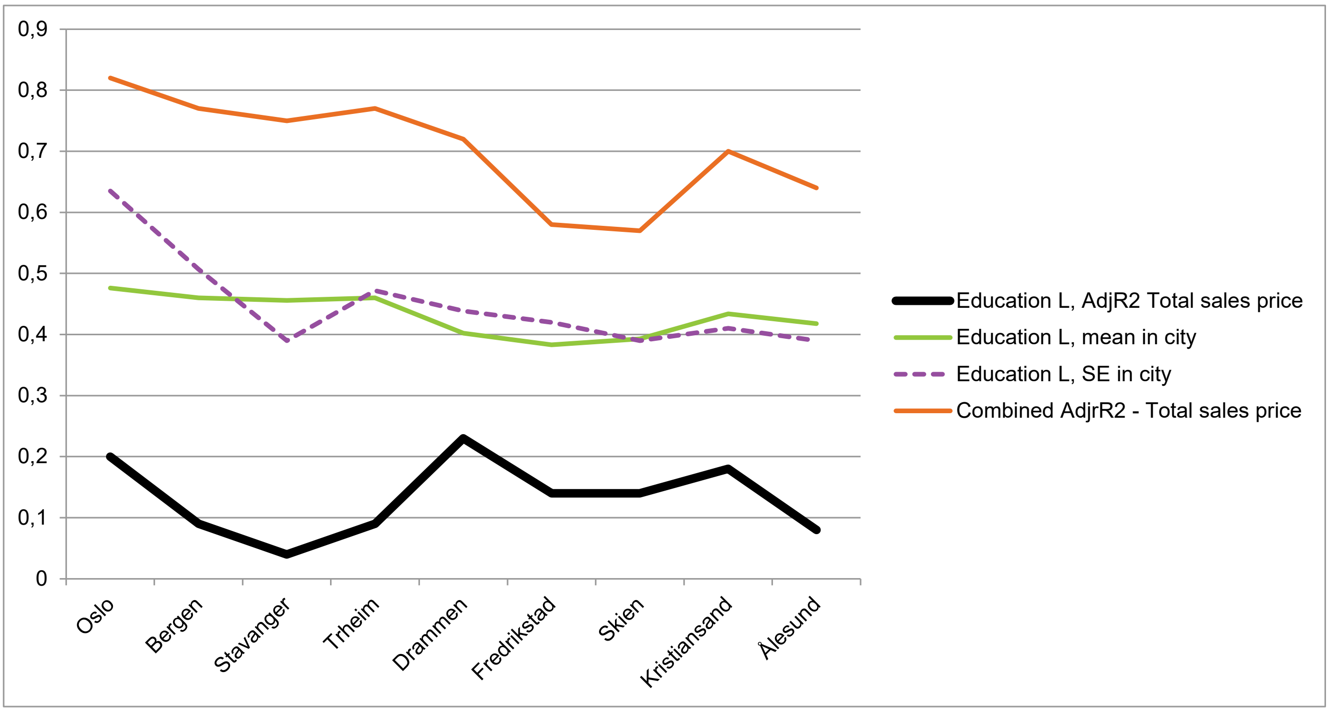
POP_EDUC_L and in the three approaches:

Total Sales Price									
AdjR^2									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
POP_EDUC_L_P5	0.20	0.09	0.04	0.10	*0.23*	*0.14*	*0.14*	*0.18*	*0.08*

Price per m^2									
AdjR^2									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
POP_EDUC_L	0.35	0.30	0.06	*0.10*	0.03	0.01	0.00	0.05	-

Compared same sizes					
AdjR^2					
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**
POP_EDUC_L_P5	0.47	0.23	0.15	0.22	0.15

Figure 5.4. Total sales prices and Education – AdjR² correlated with Education L-mean and standard error

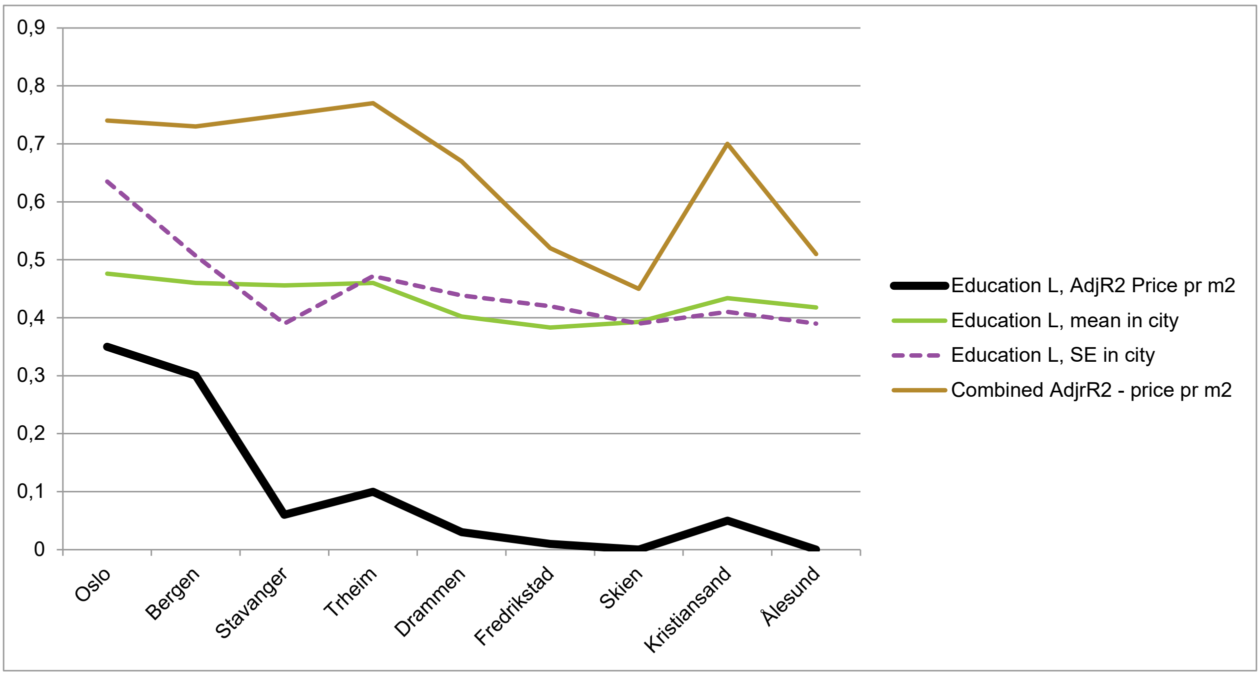


"Total sales price"-approach: The diagram looks at mean education level within each of all cities, the standard error of its distribution, AdjR^2 for isolated Education L, and combined AdjR^2 for the approach.

From the diagram, we can read that the mean education level is highest in Oslo, and falls with size of city. Interesting is that there seems to be a correlation between to which degree education levels vary within each city (standard error) and to which degree the variable Education L can be used to explain variation in total sales prices (Education L, AdjR^2 – Total sales price).

The lower standard error for Stavanger indicates that the population of Stavanger is more homogenously educated than the other cities, giving a smaller AdjR^2 for its Education L.

Figure 5.5. Price per m² and Education – AdjR² correlated with Education L-mean and standard error



"Price per m^2"-approach: Education level-mean and Education level-standard error are the same as above in the diagram for "Total sales price". AdjR^2 for "Total sales price" is swapped for AdjR^2 for "Price per m^2".

The same types of correlation are apparent in this "Price per m^2"-approach as the above "Total sales price"-approach. Isolated Education L, AdjR^2 seems however to be even more effected by standard error than is the case for "Total sales price".

The maps below show 500m X 500m grid cells covering the urban settlement of Oslo:

Figure 5.6. Map 1 visualizes mean Pop_Educ_L for all persons in each cell. There is a clear geographical pattern to the variation in mean education levels throughout Oslo.

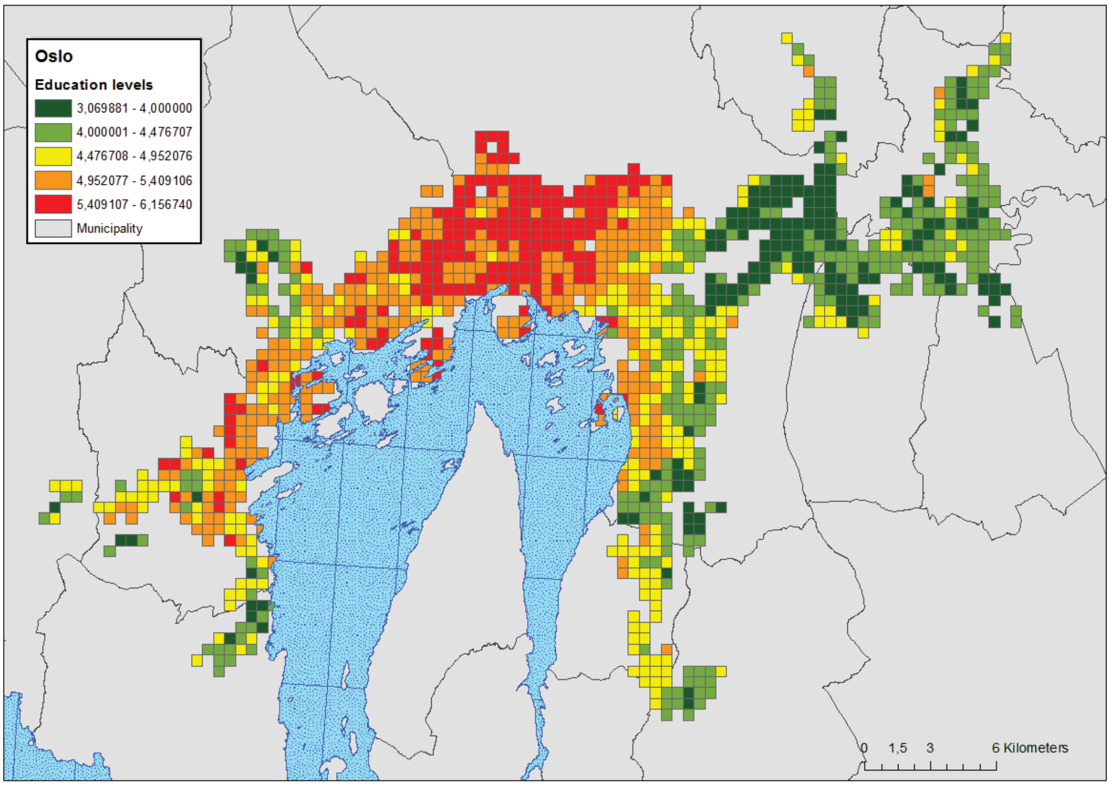


Figure 5.7. Map 2. Visualises mean Price per m². There are apparent similarities to map 1, and we show that Pop_Educ_L can be used to explain 35 per cent (AdjR² = 0.35) of this variation

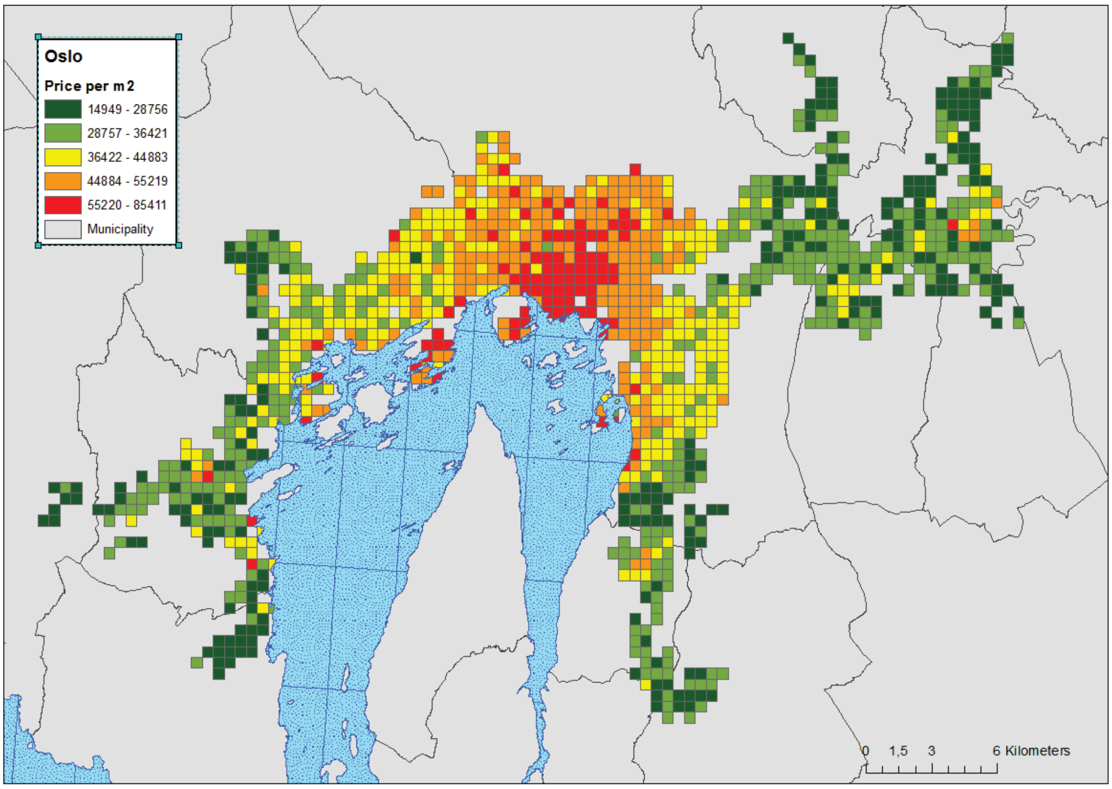


Figure 5.8. Map 3. Visualises mean Total sales prices. Pop_Educ_L explains also 20 per cent (AdjR^2 = 0.20) of this variation

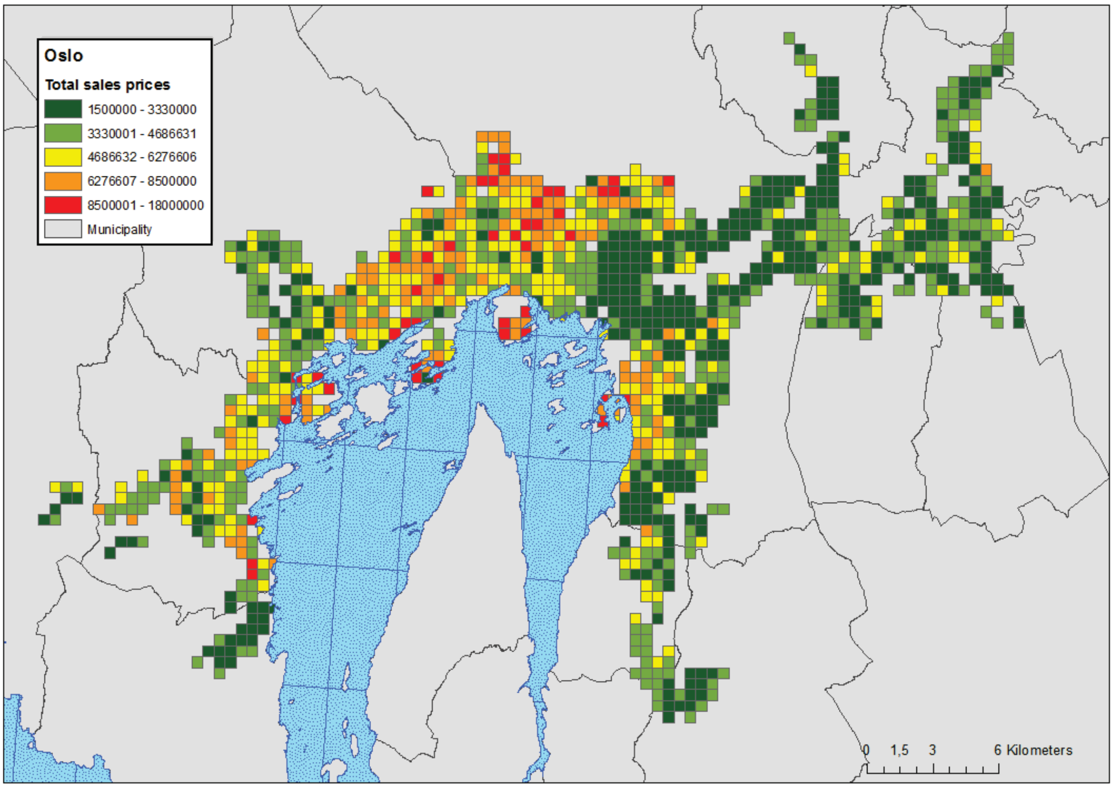
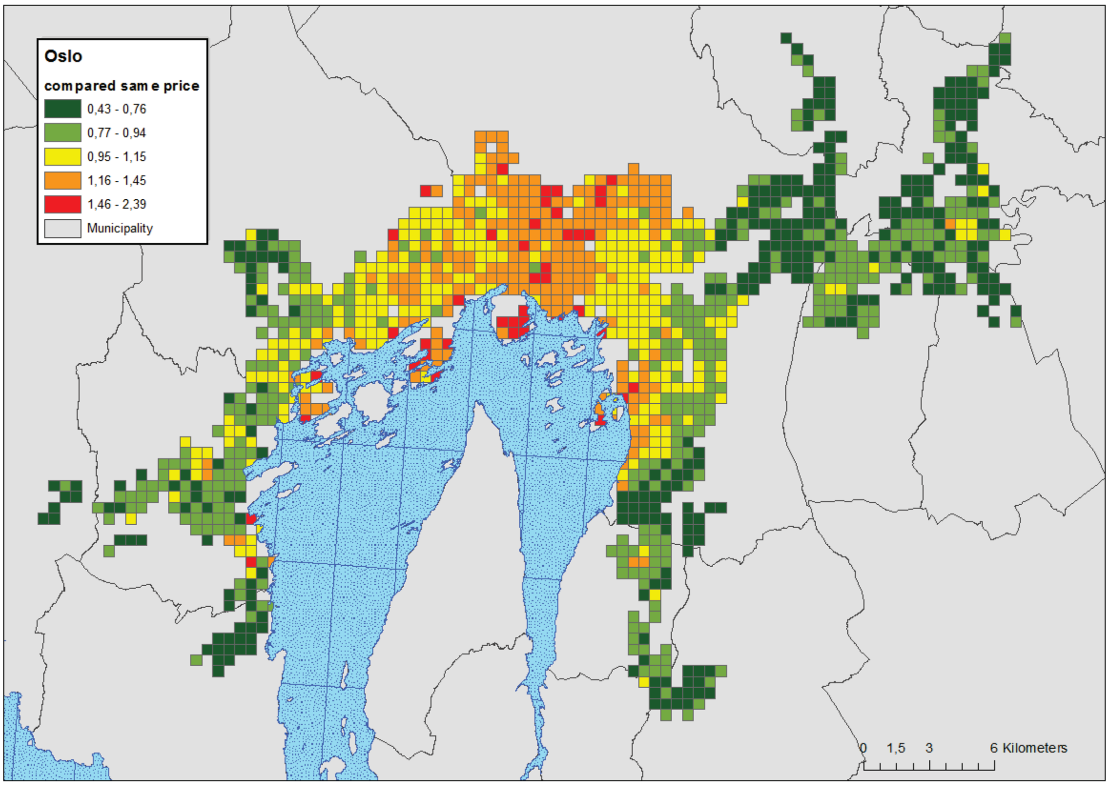


Figure 5.9. Map 4. Visualises mean compared same sizes. Pop_Educ_L explains also 47 per cent (AdjR^2 = 0.47) of this variation



5.5.4 Immigration: The variable immigration was for Oslo found to be less significant contributor than variable "Population with non-western ancestry"

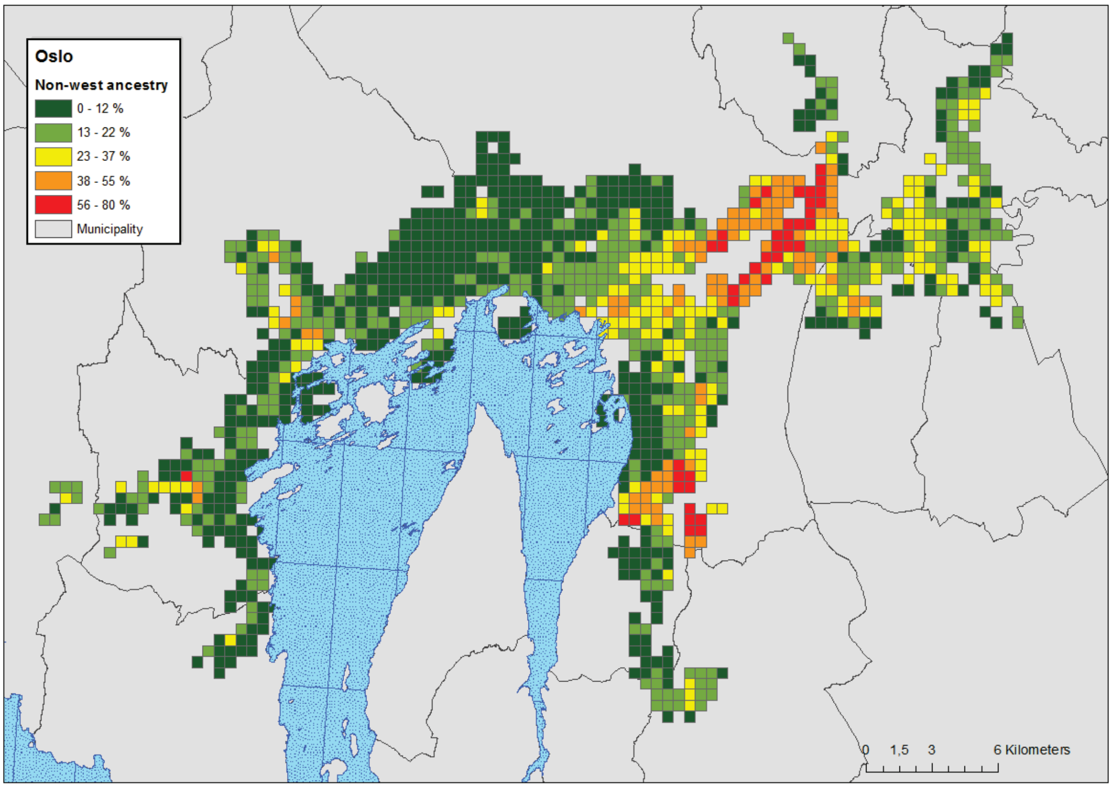
5.5.5 Population with non-western ancestry is for Oslo not found to be a significant contributor for "Total sales price" and "Price per m^2".

Though there in Oslo is correlation between Population with non-western ancestry and these approaches, we did not find increased values to consistently contribute to higher prices. Our findings are that income and education better cover the variation this variable otherwise would cover.

As shown above in the section for Income, there is however a strong suggestion that households of non-western ancestry are highly overrepresented at the lower income levels, and in the "Compared same size"-approach we also find the variable to a be significant contributor, with a $AdjR^2$ of 0.20 for Oslo. It is consistently significant in the five largest cities, but at much lower levels than for Oslo.

Four map illustrations for Oslo below show mean levels of «Population with non-western ancestry», correlated with three maps showing mean Education levels, mean income levels and mean total sales prices:

Figure 5.10. Map 1: Population with non-western ancestry, percentage within 250m of each dwelling sale, mean within grid cell.



Map 1 suggests a geographic divide in Oslo, with clustered higher values to the North-East and South-East. The areas visually correlate to areas with lower total sales prices (Map 2), lower household income (Map 3), and to lower mean education levels (Map 4).

Figure 5.11. Map 2: Total sales prices for all dwelling sales in 2014, mean within grid cell

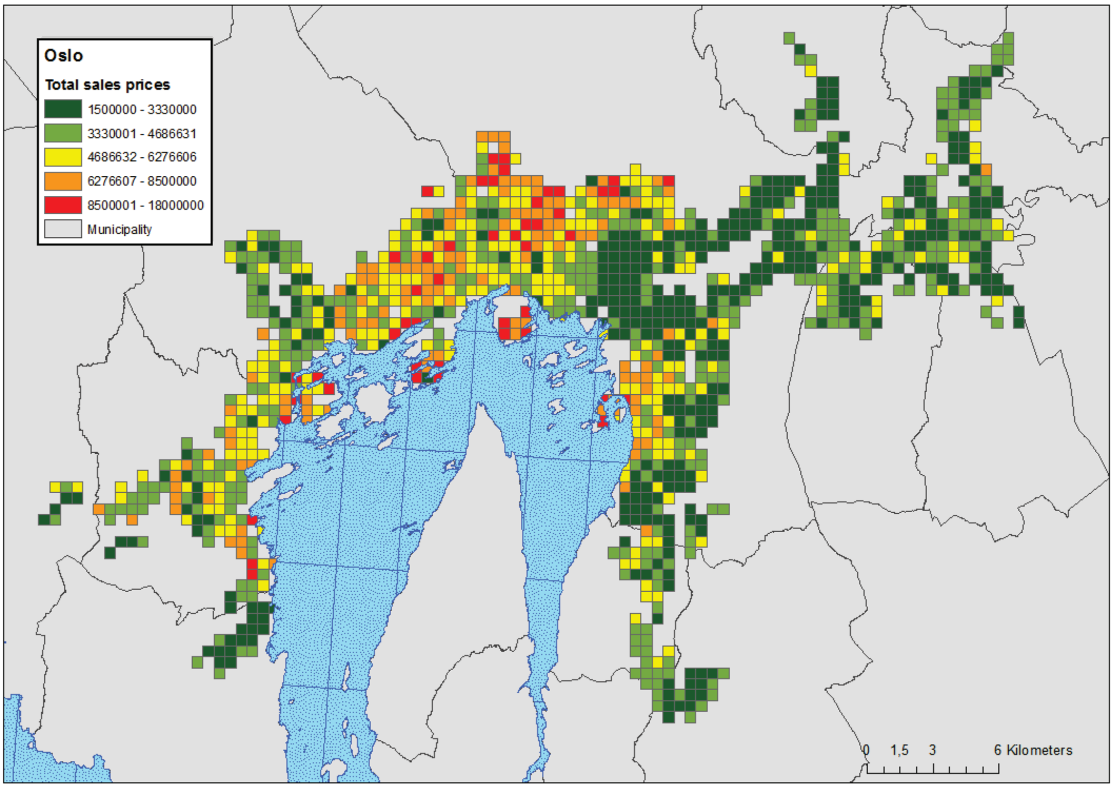


Figure 5.12. Map 3: Mean income (before tax) of households within 250m of each dwelling sale, mean within grid cell.

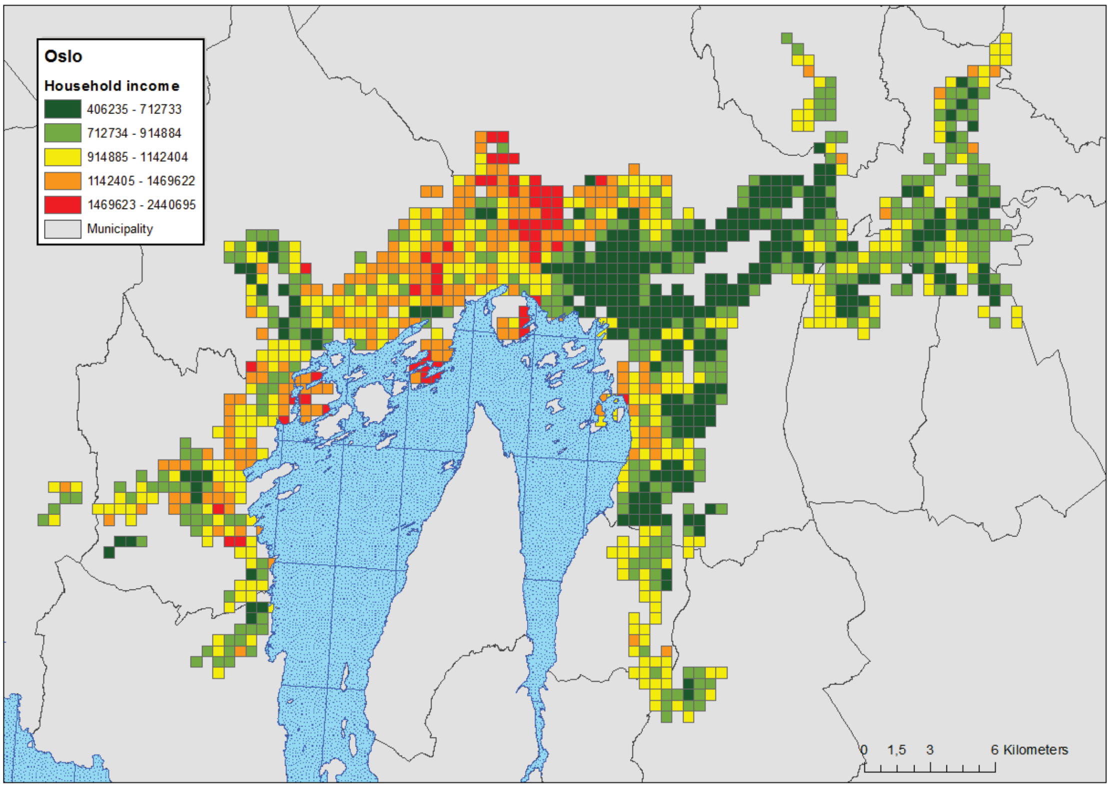
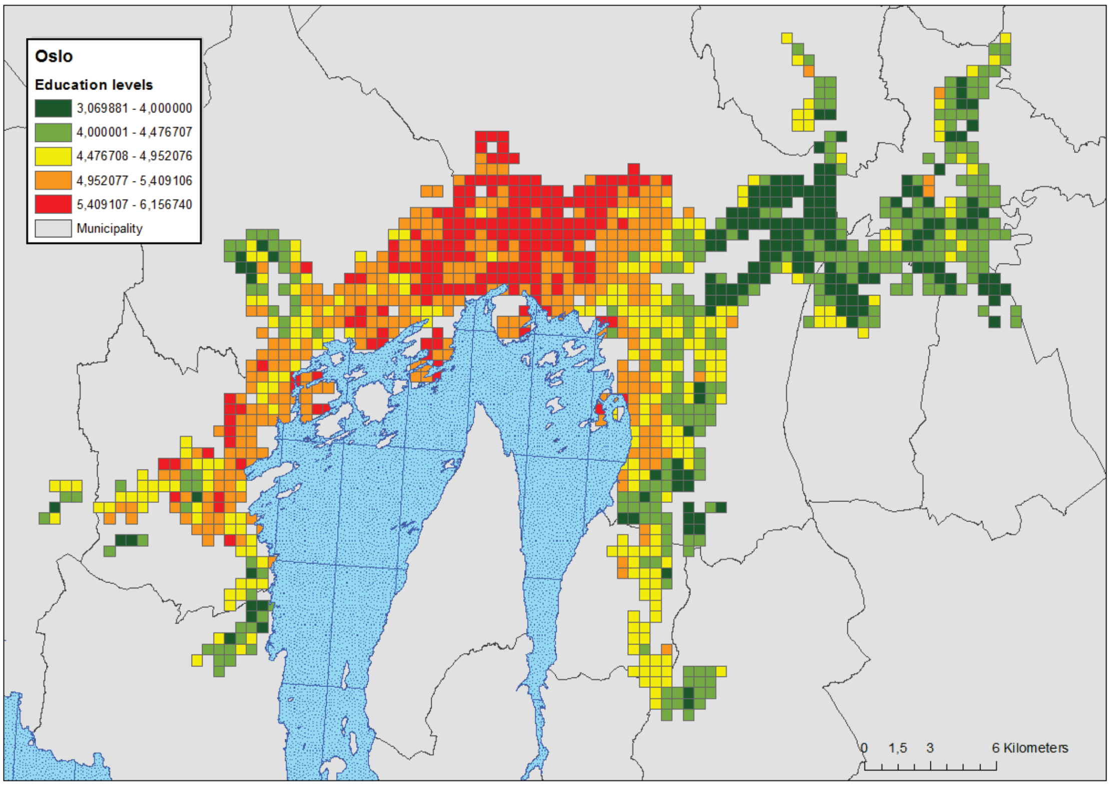


Figure 5.13. Map 4. Mean education level (1-8) within 250m of each dwelling sale, mean within grid cell.



5.5.6 Age – mean of population is in Oslo found to be a significant contributor in both approaches. The values are generally quite low, but for "Price per m²" we see that the variable is a significant contributor for all cities. This suggests that there is a general correlation between age and wealth in Norwegian cities.

Total Sales Price									
AdjR²									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
POP_AGE	0.02	*0.00*	-	0.00	0.00	0.00	-	-	0.01

Price per m²									
AdjR²									
Urban settlement	**Oslo**	**Bergen**	**Stavanger**	**Tr.heim**	**Drammen**	**Fredrikstad**	**Skien**	**Kristiansand**	**Ålesund**
POP_AGE	0.01	0.00	0.00	0.01	0.15	0.14	0.08	0.19	0.07

5.5.7 Percentage below 18 years old performs somewhat similarly to above POP_AGE, but with less explanatory value. In the compared same size-approach the variable was the most the valid of the two.

5.6. Variable type: Employment

“5.6.1 Employees within 5 km” and “5.6.2 Employees within 10 km” were not found to be significant variables in explaining variance in our approaches for Oslo.

Having found the variables not to be significant for Oslo, we also tested these separately for Bergen, Stavanger and Trondheim; finding the same results.

This does not mean that employment is not important, rather that it might be a reason for living in the city itself, not so much where one lives within it. As is true for our findings on several other services (as primary schools), the results suggest that access to employment is close enough to often not matter pricewise, in relation to attractivity.

This would depend on accessibility through a well-functioning public transport system or road system. Our general findings for our public transport variables are that access to these in Norwegian cities is good enough to not matter attractivity wise, coinciding with this conclusion.

We might have seen the variable matter if our city delineation was based on administrative boundaries, encompassing all types of population density, where distance might matter more.

6. Production of attractive urban areas – final attractivity datasets

6.1. Calculate and join chosen explanatory variables to Norway's georeferenced building register

Data from **building register**, by building			
Variable type	**Variable**	**Total Sales Price**	**Price per m^2**
Building *(type Dwelling)*	Building_id		
	Number of dwellings	X	X
	Dwelling area, dwellings	X	X
	Floor space, dwelling	X	X
	Age of building	X	X
Building centroid	X coordinate	X	X
	Y coordinate	X	X
Distance to geographic entities	CentreZone	X	X
	Lakes&Rivers & Coastline	X	X
Distance to Buildings	Hospital		X
	University/Higher Education		X
	Restaurant	X	X
Population	Household income –after taxes	X	
	Level of education	X	X
	Age – mean of population	X	X

Norway's georeferenced building register is used as basis for the prediction part of the project. All chosen explanatory variables above (CentreZone, Lakes&Rivers, etc.) were therefore recalculated in relation to the georeferenced buildings in this dataset.

The dataset is much larger than the georeferenced real estate data used for analysis and creation of our coefficients, as the building register encompasses all buildings in Norway.

A building can hold several dwellings (variable Number of dwellings). Total useful floor space for all dwellings in a building is in variable Dwelling area, dwellings.

In the project, our focus is in dwellings, as this is how are real estate data is organised. Our analysis takes therefore into account number of dwellings in a building, calculating mean m^2 floor space for each dwelling. Variable Dwelling area, dwellings is divided by Number of dwellings to get Floor space, dwelling.

6.2. Calculate predicted Total Sales Price for each building

We calculated predicted **Total sales price** for each building (type dwelling) in the point based building dataset, coefficients separately weighted, by urban settlement.

Coefficients for dependent variable **Total sales price - Oslo**

Variable	**Coef**
Intercept	-3639487.42215
RESTAURANT_DIST	-164.69926393500
CENTREZ_DIST	-98.65370343370
WATER_DIST	-142.13425651900
FLOOR_SPACE_SQR	596942.51901100000
POP_EDUC_L_P5	138.17765320700
POP_INCOME	1.45942711512
POP_AGE	40625.91287910000
BUILDING_AGE	-7605.85554977000

Calculation of "Total sales price" – Oslo

```
(-3639487.42215 + ( [RESTAURANT_DIST] * -164.7) + ([CENTREZ_DIST] *
-98.65) + ([Water_dist] * -142.13) + ([FLOOR_SPACE_SQR] * 596942.5)
+ ([POP_EDUC_L_P5] * 138.2) + ([POP_INCOME] *1.459) +  ([POP_AGE]
*40625.913) +  ([BUILDING_AGE] * -7605.855)) *  [Number of dwellings]
```

Appendix D gives overview of coefficients for the 4 largest cities.

6.3. Calculate predicted Price per m^2 for each building

We calculated predicted **Price per m^2** for each building (type dwelling) in point based building dataset, coefficients separately weighted, by urban settlement.

Coefficients for dependent variable **Price per m^2 - Oslo**

Variable	Coef
Intercept	-13587.30470390000
HOSPITAL_DIST	-0.10373293612
RESTAURANT_DIST	-0.91701610756
POP_EDUC_L	8110.60785656000
CENTREZ_DIST	-0.50031758780
WATER_DIST	-1.65258466752
FLOOR_SPACE_RECI	777209.62615300000
POP_AGE	472.27411045900
UNIVERS_DIST	-0.32628683747
BUILDING_AGE	-89.08159680120

Calculation of **Price per m^2 – Oslo**

```
(-13587.3047039 + ( [HOSPITAL_DIST] * -0.10373293612)  +
([RESTAURANT_DIST] * -0.91701610756) + ([POP_EDUC_L] *
8110.60785656) + ([CENTREZ_DIST] * -0.5003175878) + ([Water_dist] *
-1.65258466752) + ([FLOOR_SPACE_RECI] * 777209.626153) + ([POP_AGE]
*472.274110459) +  ([UNIVERS_DIST] * -0.32628683747) +
([BUILDING_AGE] * -89.08159680120)) * [Number of dwellings]
```

Appendix D gives an overview of coefficients for the 4 largest cities.

6.4. Join to 500m X 500m statistical grid, compute mean values

We joined the buildings point data to a 500m X 500m statistical grid.
Mean “Total sales price” and “Price per m^2” (by dwelling) was calculated for each cell.

6.5. Calculate two grid-based attractivity indexes per urban settlement

For each of the four largest urban settlement (excluding cells with 3 or less dwellings):

- We calculated a **Total sales price - Attractivity index** from medium predicted “Total sales price” in grid cell → index range from 1 to 10, Quantile grouping, where 10 is 10 per cent highest priced dwellings.

- We calculate a **Price per m^2 - Attractivity index** from medium predicted “Price per m^2” in grid cell → index range from 1 to 10, Quantile grouping, where 10 is 10 per cent highest price per m^2 dwellings

Grid-based attractivity index datasets were produced for the four largest urban settlements, as shapefiles.

Attractivity_grid_shp.shape		
Table Feature	Description	Data Type
SSBID	**Grid-Id_500m**	Text
TETTSTEDSN	**Urban settlement name**	Text
Totsum_Attract	**Total sales price - Attractivity index**	Double
PerM2_Attract	**Price per m^2 - Attractivity index**	Double

Projection: ETRS_1989_UTM_Zone_33N

The following pages visualise these datasets as maps:

Figure 6.1. Oslo, Total sales price, predicted

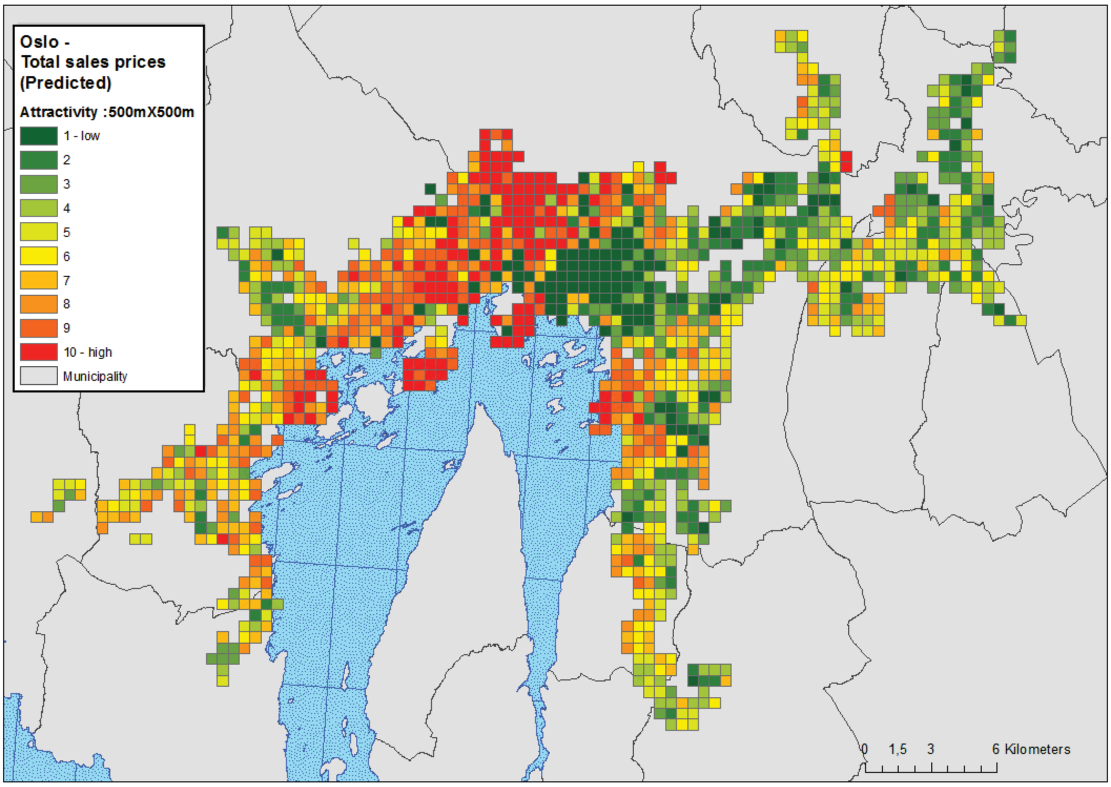


Figure 6.2. Oslo, Price per m², predicted

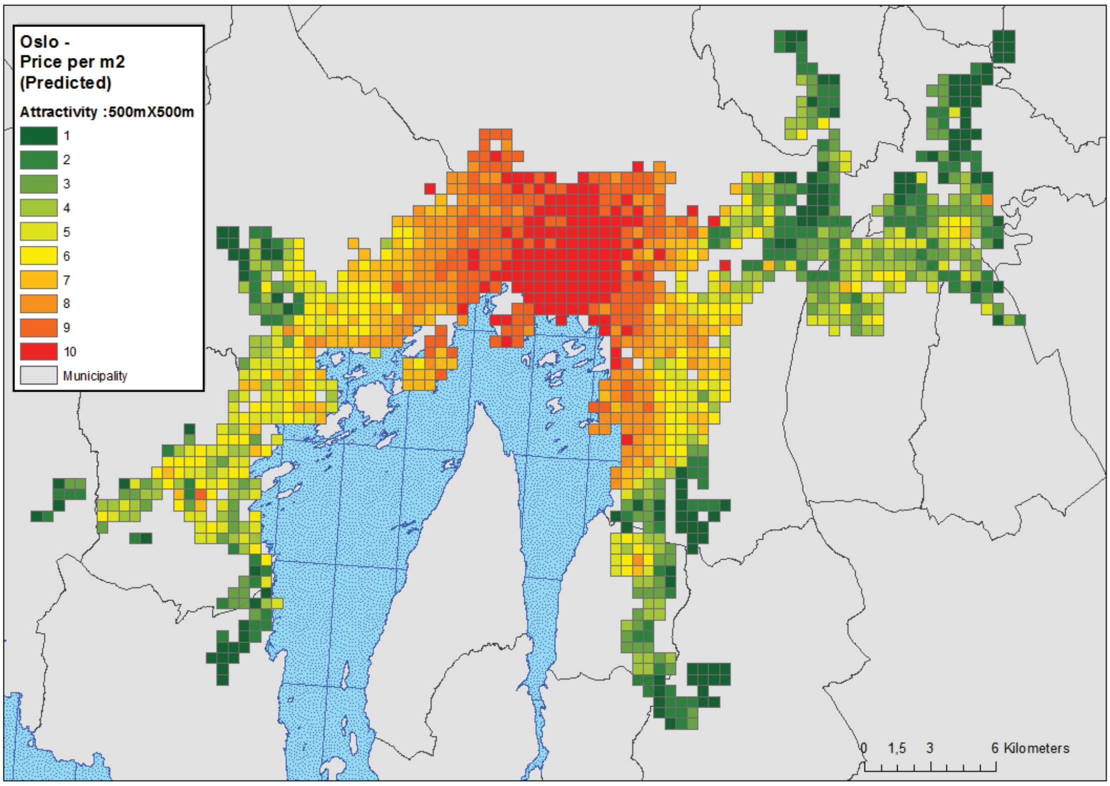


Figure 6.3. Bergen, Total sales price, predicted

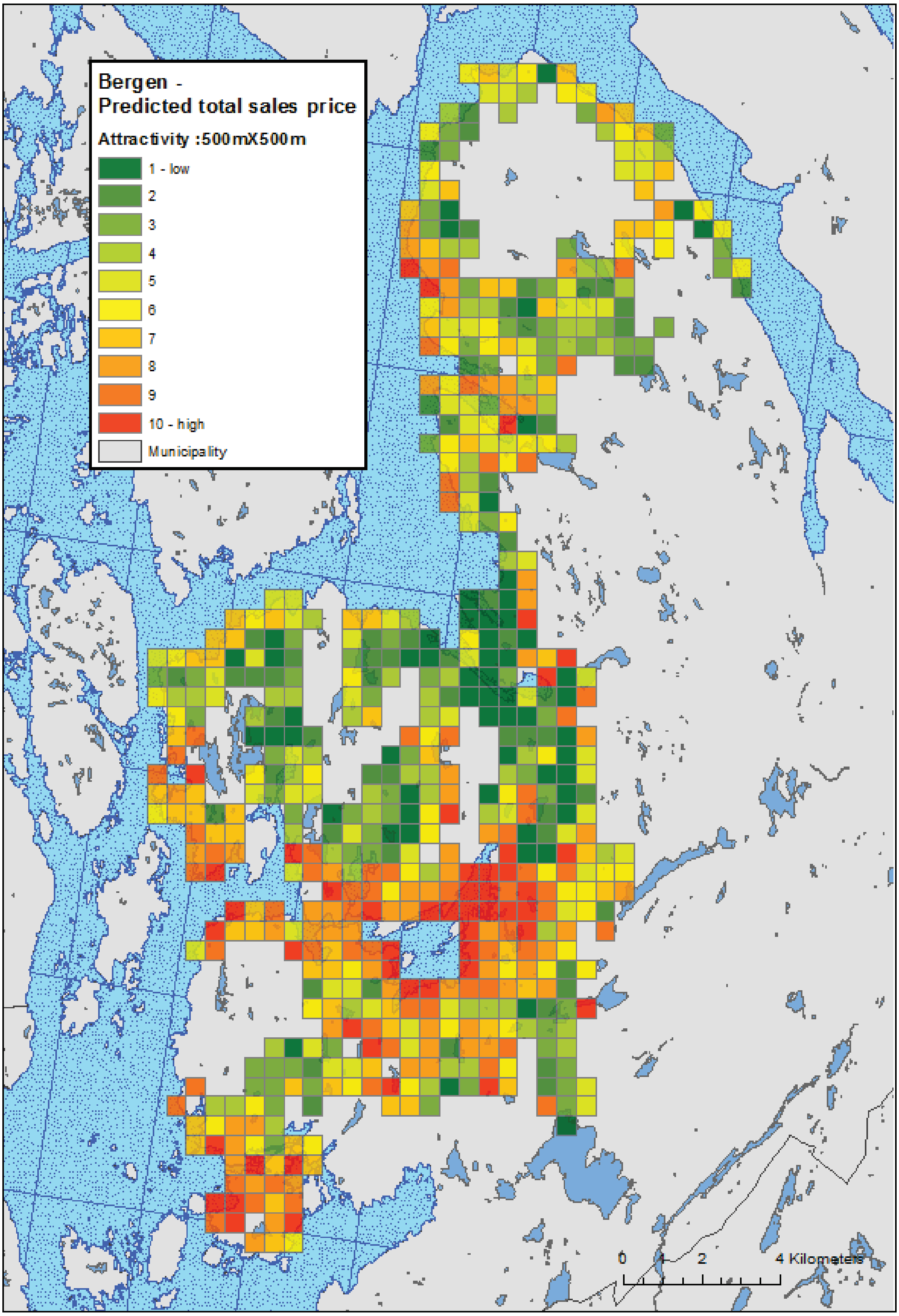


Figure 6.4. Bergen, Price per m², predicted

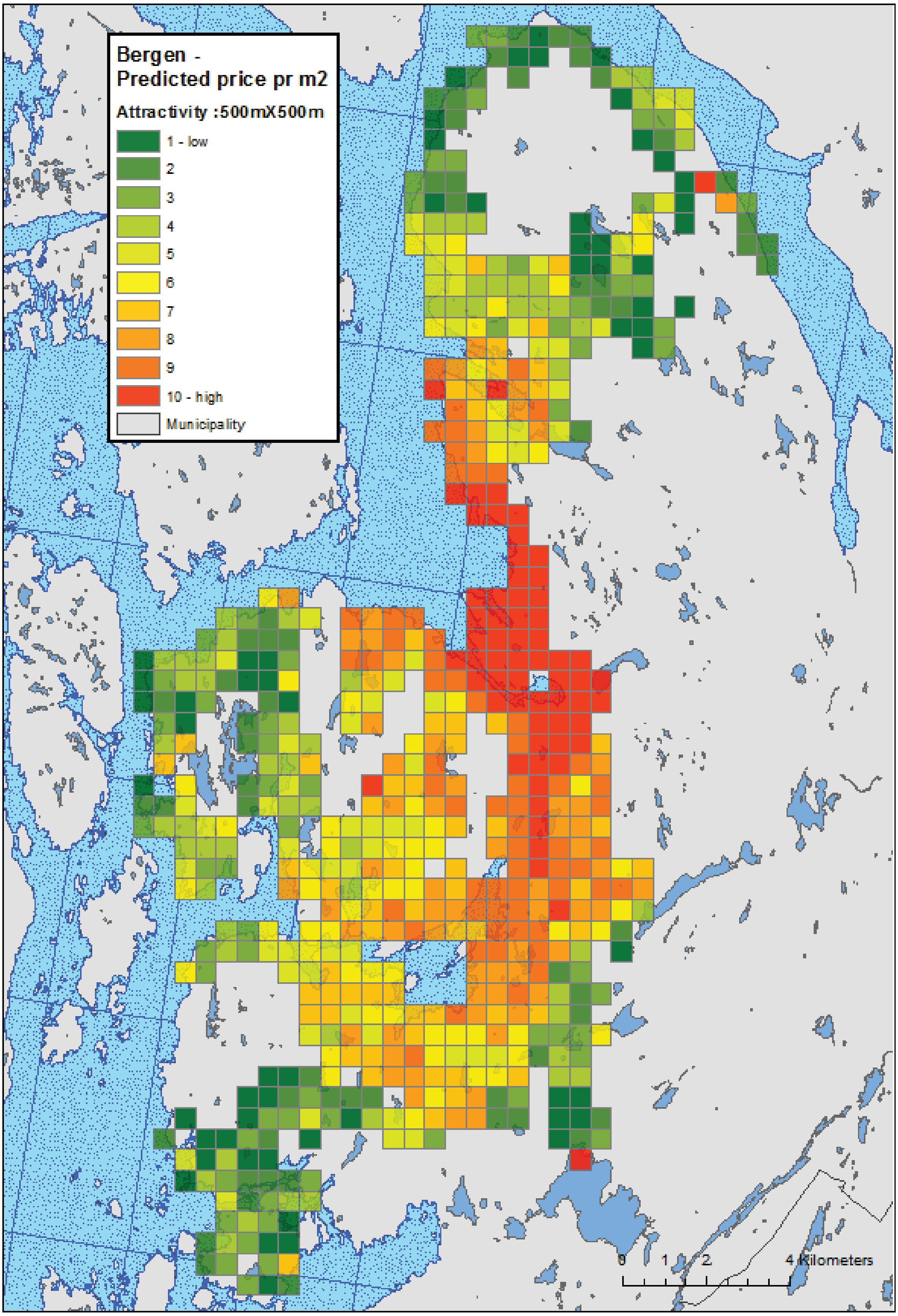


Figure 6.5. Trondheim, Total sales price, predicted

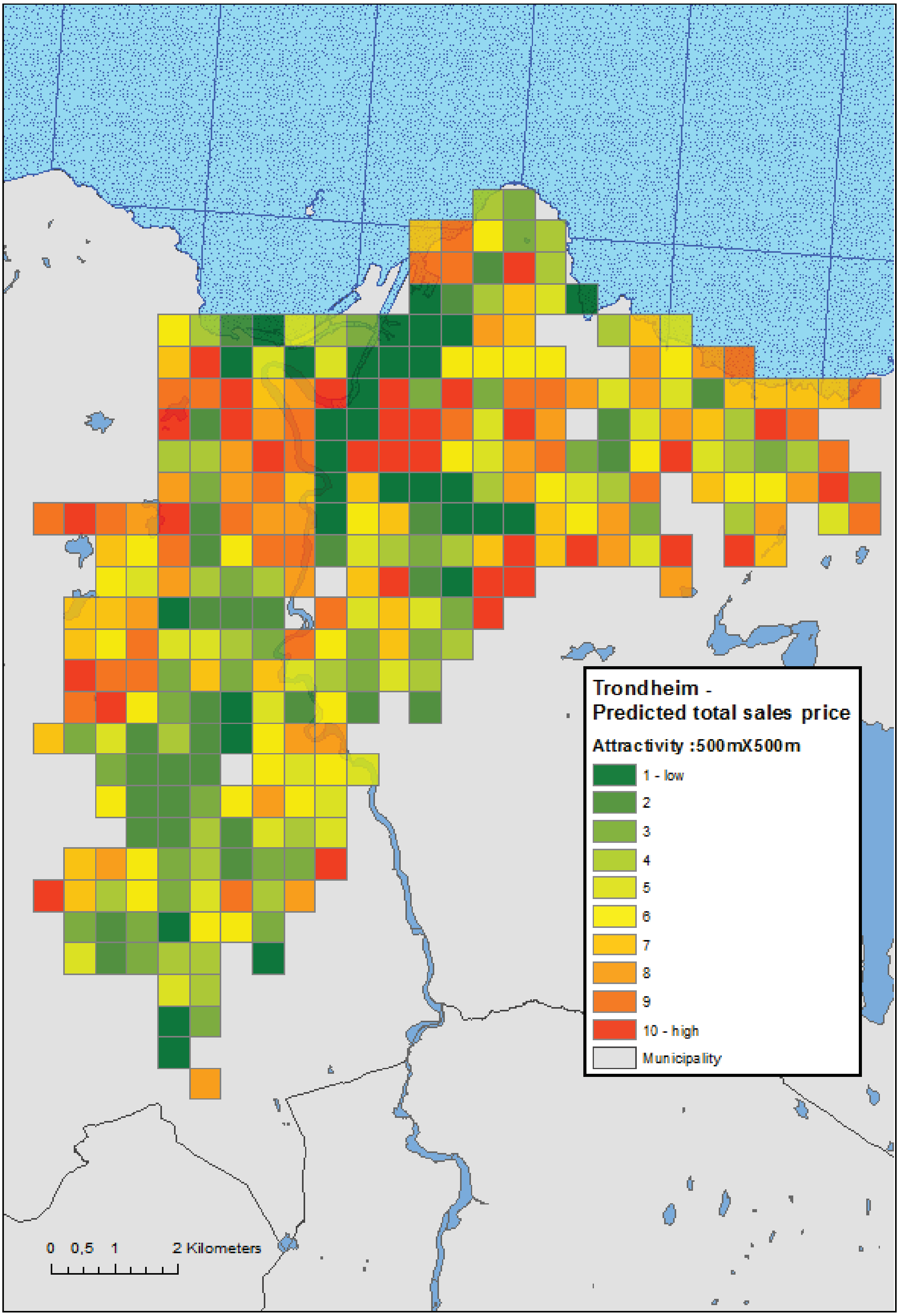


Figure 6.6. Trondheim, Price per m^2, predicted

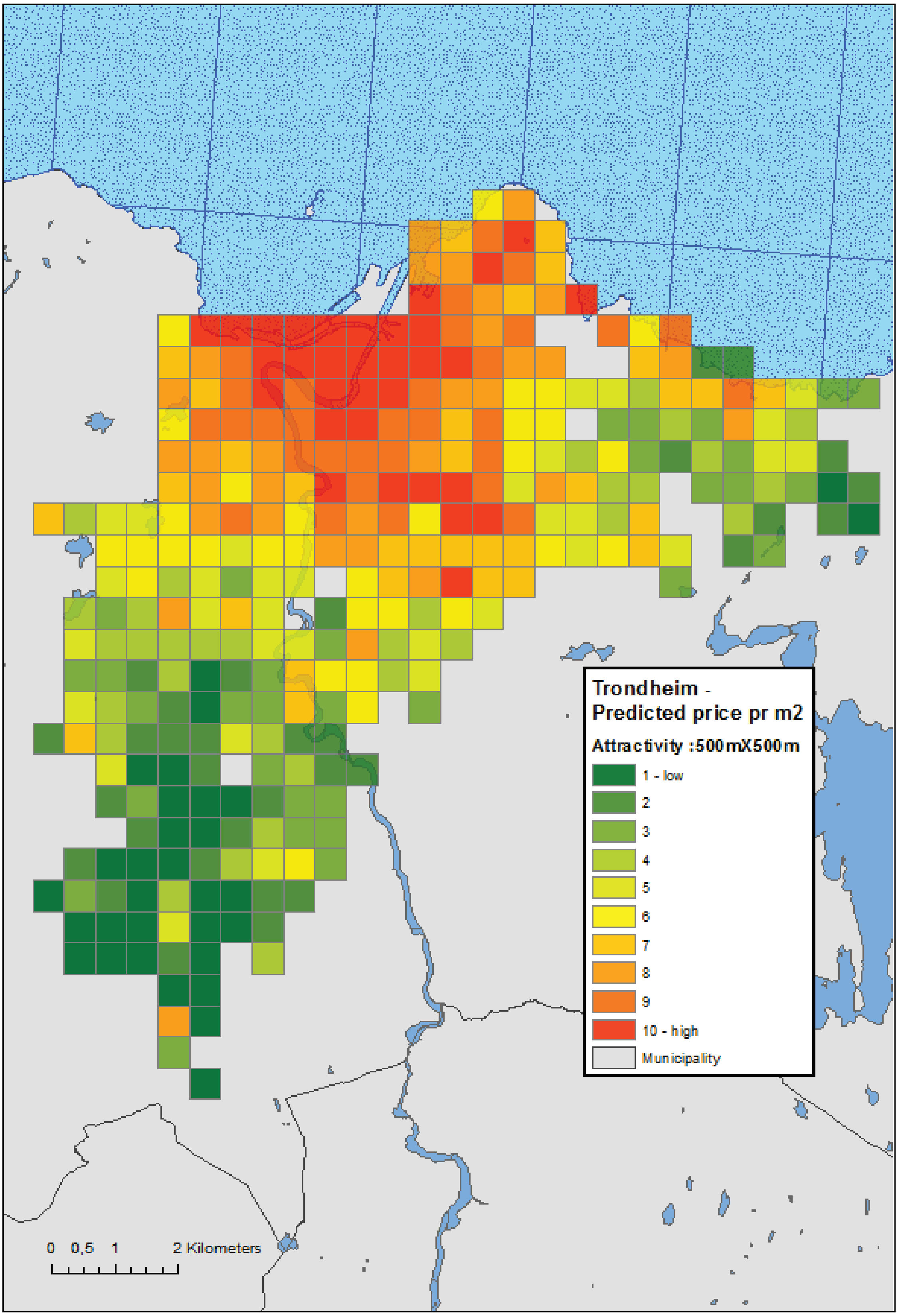


Figure 6.7. Stavanger/Sandnes, Total sales price, predicted

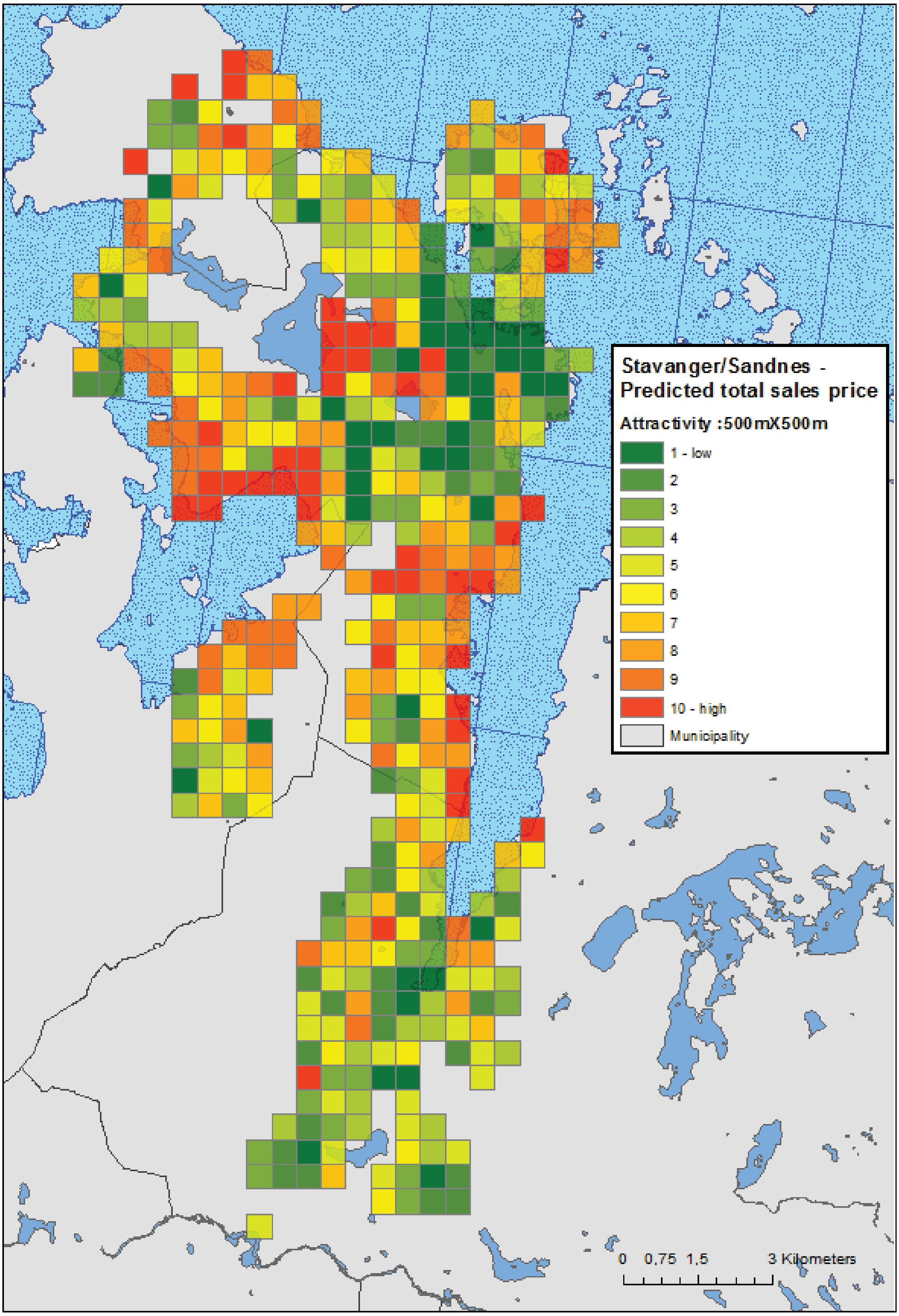
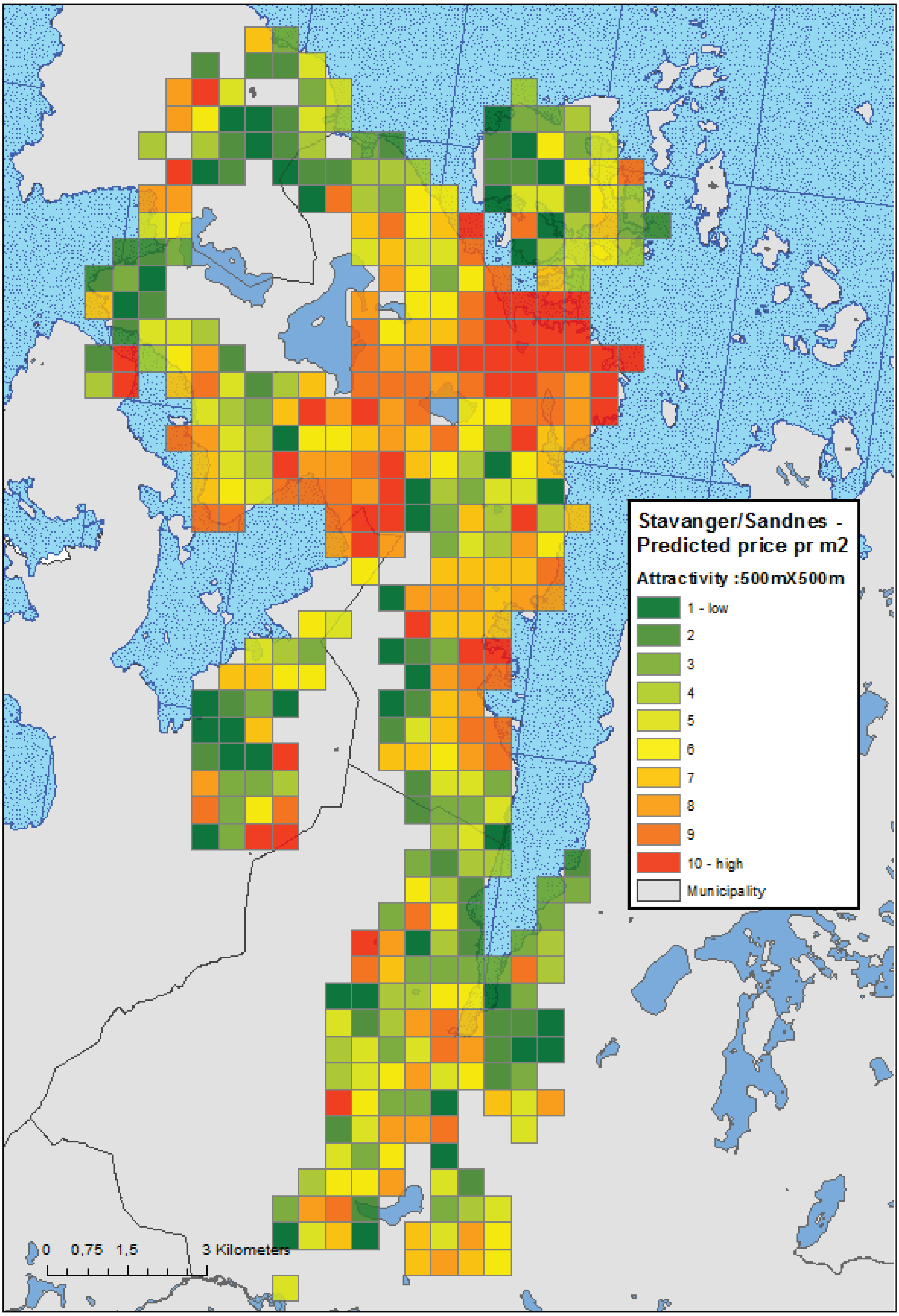


Figure 6.8. Stavanger/Sandnes, Price per m², predicted



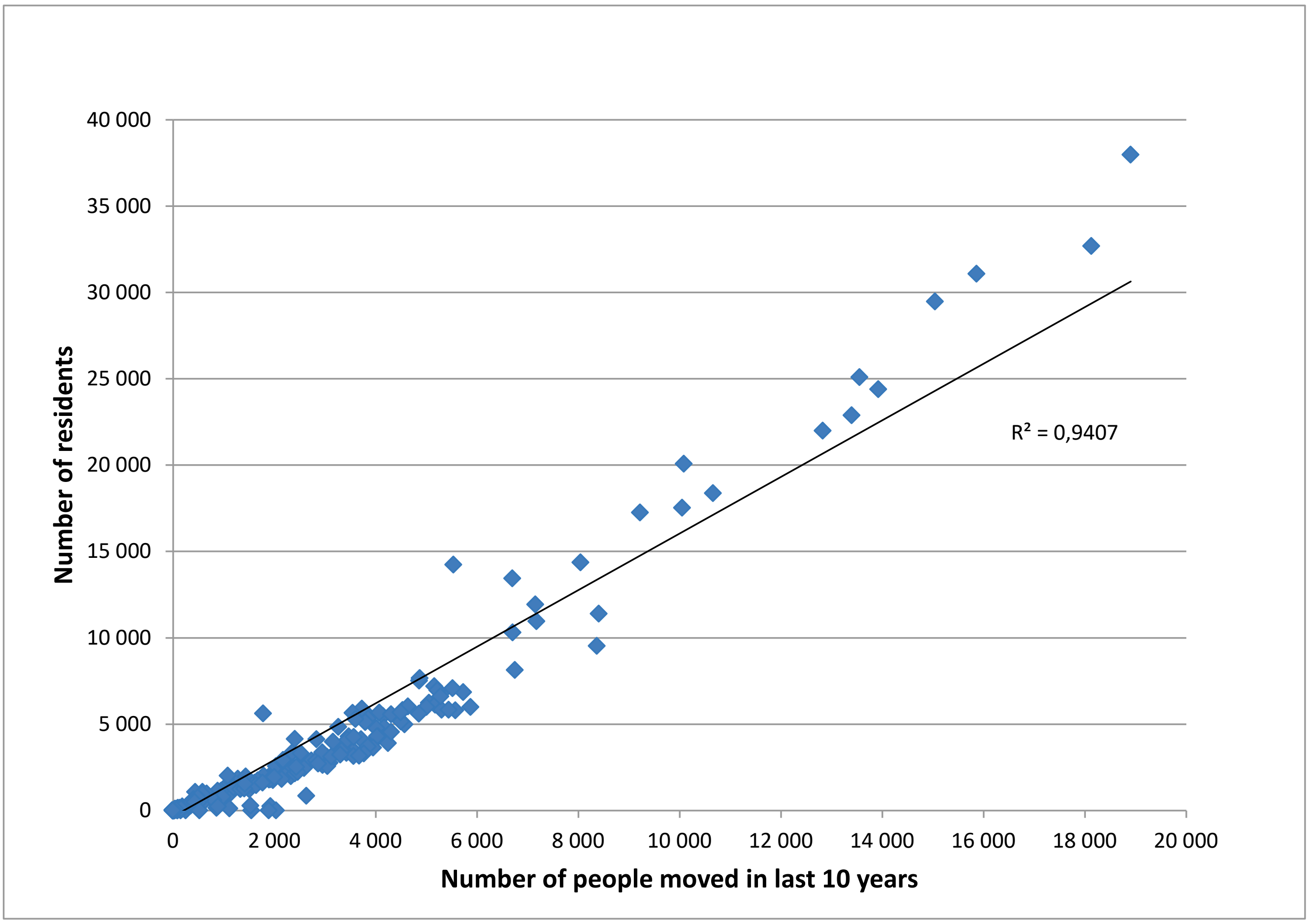
7. Movement of people

One of the parameters thought to be of significance is movement of people within the municipality or within the urban settlement/ city. We have prepared data for all movements within Oslo municipality the last 10 years. Each movement has been georeferenced for “To” and “From” address. We have explored the data and tried to visualise the movements by aggregating movements to km^2 grids. See figure 7.1.

Figure 7.1. Number of people (within the municipality) moving in to km^2 grids

There are, however, many reasons to move from one area to another, not just the goal of living in a more attractive neighbourhood. People of different ages may have different needs concerning dwelling areas; from families with small children, students and young adults to elderly people. Divorce, sickness, or other circumstances may also ultimately lead to changes in dwelling area and not necessarily to more attractive areas. What we found was that the population density closely follows number of in-moving people (see Figure 7.2). As total population is more readily accessible, we did not pursue internal movements within the urban settlement/ city any further.

Figure 7.2. Number of people moved in to the km² grid in relation to resident population



APPENDIX A – Literature overview

Throughout the first year of the project the project participants have met a range of persons with knowledge about quality of life and urban planning. Apart from the "Quality of Life in Cities Perception Survey in 79 European cities" produced by the European Commission, Directorate-General for Regional and Urban Policy, the following literature have been relevant for the results in the project:

1. Spacescape (2015). *The Executive Office and Regional Planning Committee of Stockholm, Värdering av stadskvaliteter i Stockholmsregionen*. http://www.spacescape.se/wp-content/uploads/2015/05/projektrapport_stockholmlan_stadskvalitet.pdf

2. Alexander Ståhle. (2011). Stadsrum och stadsutveckling. Paper presented on Byromskonferansen, https://www.regjeringen.no/globalassets/upload/subnettsteder/framtidens_byer/byrom/byromskonferansen_16062011/stadsrum_och_stadsutveckling_alexander_staahle_16062011.pdf
3. Alexander Ståhle. (2011, May). Stadskvaliteter efterfrågas. *Arkitekten*. https://www.regjeringen.no/globalassets/upload/subnettsteder/framtidens_byer/byrom/stadskvaliteter_efterfragas_alexander_staahle_arkitekten.pdf

Spacescape is a research-driven consulting company in urban planning. Spacescape carried out an Evaluation of urban quality in Stockholm on behalf of The Executive Office and Regional Planning Committee of Stockholm. SpaceScape demonstrated that the housing price in Stockholm is to 90 per cent explained by seven urban qualities plus a socioeconomic index as control variable. Their conclusion is according to the company based on 7000 observations. The qualities of importance are:
Proximity to City
Access to the train, metro, tram stop
Access to the pedestrian street network
Access to urban activities
Access to the green areas
Proximity to water
Block Shape
Socioeconomic Index

4. Carlsen, F. and S. Leknes (2015). For whom are cities good places to live? (Working paper series 1/2015). https://innsidawls.itea.ntnu.no/syndicator-web/public/files/8b479d83-6888-39d9-aefc-b304c6f193b8

The study looks at the relationship between the residency quality variables and population size in a region to distinguish between urban and rural "amenities". Not compared differences between different urban areas.
Communication with Stefan Leknes raised issues about comparisons in between urban areas as Multicollinearity issues having a wide range of town- quality variables. It's generally more of everything in the cities; better shopping, better and more varied cultural offerings, several restaurants and cafés as well as better developed infrastructure etc. Leknes raised the issue how problematic it can be to distinguish in between these quality variables.

5. Barstad, Anders (1997). Store byer, liten velferd? Om segregasjon og ulikhet i norske storbyer, Sosiale og økonomiske studier 97. Statistisk sentralbyrå.
6. Barstad, Anders (2003). Levekår i storbyene: Noen bydeler er særlig utsatte, Samfunnsspeilet, 2003/2. Statistisk sentralbyrå.

Barstad is carrying out research on living conditions and social participation at Statistics Norway. The two articles above are handling the issues of welfare and quality of life in the bigger cities of Norway. These issues are relevant for the work of identifying items that are important for attractiveness in urban areas.

7. Albouy, David (2015). What Are Cities Worth? Land Rents, Local Productivity, and the Total *The Review of Economics and Statistics,* July 2016, 98(3): 477–487. doi: 0.1162/REST_a_00550
8. David Albouy et al. (2015) Urban Population and Amenities: The Neoclassical Model of Location. http://davidalbouy.net/quantities.pdf

Albouy is trying to explain variation within cities regarding population, density, and land supply based on three amenity types: quality-of-life, productivity in tradable, and productivity in non-tradables.

9. Poulhès, Mathilde. (2016). *A Room with a View or Rear Window? Hedonic prices of the Parisian real-estate*. Paper presented at EFGS Conference Paris, November 17th, 2016. http://www.efgs.info/wp-content/uploads/conferences/efgs/2016/S8-5_presentationV1_MathildePoulhes_EFGS2016.pdf

Poulhès is looking in to variation of real-estate prices in Paris, and asks the question; "What are the prices of extrinsic housing characteristics?"
Real-estate prices are not only due to intrinsic characteristics, but also localization, neighbourhood, quality, job accessibility, amenities, etc.
Her study utilizes hedonic model; explaining individual choices and prices by the differentiation of goods.

APPENDIX B

1. Identifying comparable data

Statistics Norway is the central body for preparation and dissemination of official statistics in Norway stipulated in the Statistics Act. This gives Statistics Norway a unique situation to access a wide range of data. Statistics Norway is also administrating own registers.

1.1. Statistical registers at Statistics Norway

In the 1950s and 60s official statistics began to some degree to be based on administrative registers, making it possible to use the same data for different statistical purposes and reducing the burden of data collection. The production of official statistics in Norway has increasingly relied on official registers and other administrative data over the last decades, and this has become a characteristic of the official statistics in the Nordic countries. Today Statistics Norway uses about 60 such registers.

1.2. Data used for identifying attractive urban areas

1.2.1. The Cadastre (Ground properties, addresses, buildings and dwellings)

http://kartverket.no/en/Land-Registry-and-Cadestre/

The Cadastre is Norway's official property register. The system is comprised of properties, property boundaries, addresses and buildings. The different parts are interconnected. The Norwegian Mapping Authority is central Cadastre authority and is responsible for managing the Norwegian Cadastre and associated regulations. Municipalities are local Cadastre authorities and are responsible for updating the cadastre.

In this project buildings with information on geographic location, building type and base area of great importance.

1.2.2. The Central Population Register (CPR)

https://www.ssb.no/en/befolkning

The CPR is continuously updated, but in this study the data obtained is for 1st January 2014. Statistics Norway is linking the population data to the address data in the Cadastre in order to geolocate each individual. Nationally 99.7 per cent of the residents are geolocated via the Cadastre. As for addresses, there are established procedures for connecting residents to buildings.

1.2.3. Immigration and immigrants

http://www.ssb.no/en/innvandring-og-innvandrere

Statistics Norway publishes immigration statistics on regal residents.

In this project, we look at the immigration variables for all residents, geolocating each resident to its address by use of the Central Population Register (CPR).

1.2.4. Education

http://www.ssb.no/en/utdanning/

Statistics Norway publishes statistics on Education.

In this project, we look at Education level of individuals, geolocating each individual by use of the Central Population Register (CPR).

1.2.5. Income and consumption

http://www.ssb.no/en/inntekt-og-forbruk
Statistics Norway publishes statistics on income and consumption.
In this project, we look at household income, geolocating each household to its address by use of the Central Population Register (CPR).

1.2.6. Urban settlements

http://www.ssb.no/en/befolkning/statistikker/beftett/
http://www.geonorge.no > *Tettsteder*

1. A hub of buildings shall be registered as an urban settlement if it is inhabited by at least 200 persons (60 - 70 dwellings).
2. The distance between the buildings shall normally not exceed 50 metres. Deviations are allowed for areas that cannot/are not to be occupied, for example parks, sports facilities, industrial areas or natural barriers such as rivers or arable land. Also included are agglomerations that naturally belong to the urban settlement with up to a distance of 400 metres from the centre of the urban settlement.

Urban settlements are geographical areas with dynamic boundaries. Thus, the number of urban settlements and their boundaries will change over time, depending on construction activity and changes of resident population.
The delimitation of the urban settlements is independent of administrative boundaries.
In the project we seek to create attractiveness datasets for urban settlements, instead of municipalities. Urban settlements can be said to better portray the actual city as functional entity, with for example the urban settlement of Oslo spreading into 9 different municipalities.

1.2.7. National topographical map for Norway 1:50 000

www.geonorge.no > *N50 Kartdata*

The Norwegian Mapping Authority is responsible for updating and administering the national map databases N50-N5000 Map Data
The N50 Map database is in vector format, covers all of mainland Norway, and is adjusted for use at scales between 1:25 000 to 1:100 000 metres. The product correlates to the Norwegian paper map series 1:50 000.
The N50 Map database is updated continuously, and is distributed weekly.

1.2.8. Data from real estate agencies

The population is all dwellings sold on the open market. Statistics Norway receives data from the website FINN.no, through cooperation with the Norwegian Association of Real Estate Agents (NEF) and the Association of Real Estate Undertakings (EFF). FINN.no cooperates with most important real estate agencies in Norway. From the second quarter of 2005, the house price index also includes data from Notar AS. From this point, all main real estate brokers are included in the survey. FINN.no and Notar AS report their sales monthly.
The data obtained from FINN.no and Notar for our project holds sales of dwellings in Norway for the entirety of 2014.They includes all types of dwellings, freeholder and in housing cooperatives, and give information on single dwelling sales, such as price, square metres floor space and address information.

In the project, we have geolocated each dwellings sale to the centroid of the property it lies upon, by geolocating to the Property registry in the Cadastre. A *different option we tested and chose not to follow was to geolocate to the Building registry in the Cadastre, giving more precise building coordinates and more intrinsic information on the building. This however precluded geolocation of most housing cooperative dwellings.*

1.2.9. Urban centre zones

https://www.ssb.no/en/natur-og-miljo/statistikker/arealsentrum/
www.geonorge.no > *Sentrumssoner*
Statistics Norway delimits the urban centre zones annually and makes them freely available for download. The centre zones as defined as follows:

1. A centre zone is an area composed of one or more centre cores with a surrounding with a zone of 100 meters around.

2. A central core is an area of more than three different categories of main businesses groups with activities related to centre zones. This implies that in addition to retail business, public administration or health-/social services or other businesses related to social and personal service must be represented. The distance between these businesses cannot be more than 50 meters.

3. There must be at least 50 employees in the centre zone.

1.2.10. Recreational areas

https://www.ssb.no/en/natur-og-miljo/statistikker/arealrek
www.geonorge.no > *Rekreasjonsareal*

There is no nationwide mapped information about either play and recreation area or areas for recreational walk. In this statistical work, it is therefore chosen to identify areas that may have potential as a recreational area and area for recreational walk.

Basically, we have chosen to distinguish between areas for recreational walk and play / recreation area only by land size. In the calculation of access to recreation areas, we also include areas for recreational walk as defined here. That is all recreation areas larger than 5 acres, including over 200 acres. We have not looked at whether areas are regulated in terms of municipal land use planning (zoning, site plan or municipal) or prepared in the form of playground equipment, walking paths and trails and more. The terms "recreation area and areas for recreational walk" must not be confused with "recreational area" or "outdoor areas" used in the planning context.

The following areas are included in play- and recreation areas and areas for recreational walking:

- Forest, open solid ground, wetlands, bare rock, gravel and boulder fields, parks and sports fields cf. Statistics standard classification of areas for statistical purposes.
- Lakes and ponds that are less than 1 acre are also included.

Sports fields that are not normally available for public recreational activities are not included.

1.2.11. Noise

https://www.ssb.no/natur-og-miljo > *Stoy (Støy)*
Statistics Norway calculate noise levels and noise annoyance index for the most important noise sources in order to monitor a national environmental target. The method is based on existing noise mapping efforts by the different authorities responsible for the infrastructure causing noise. In addition, Statistics Norway makes simplified noise calculations for all relevant dwellings not covered by the existing maps. As a result, all dwellings (addresses with residents) have an assigned noise level. The data set for road traffic noise is used in this study.

1.2.12. Statistical grids

https://www.ssb.no/natur-og-miljo/geodata
https://www.ssb.no/natur-og-miljo > *Statistiske rutenett*
Statistics Norway has defined grid for the use of official statistics (see Documents 2009/9). The grid has a grid - ID as the coupling key towards grid statistics. In the study population (DSF), housing (cadastre) and building data (cadastre) connected to a grid with grid size 500 m^2.

1.2.13. National route database from Norwegian Travel AS (NRI)

Norwegian Travel AS (NRI) is responsible for developing and maintaining a national route database containing all public transport in Norway, as well as direct flights between Norway and abroad. This occurs when receiving, processing, testing and quality control of route data sent from- or on behalf of all operators licensed for public transport in Norway. The contents of the database can be used free of charge by operators wishing to convey information to the travelling public.

1.2.14. Digital Terrain Model

www.geonorge.no > *DTM 10*
The Digital terrain model with heights in a grid of 10 x 10 metres. The terrain model is a grid model with resolution (grid size) 10 x 10 metres. It is also possible to download files at a 20 x 20 metre resolution. Files with 20 metre resolution expires at the end of 2013. The grid is generated from a so-called hybrid DTM structure with the program SCOP. Height information underlying terrain model contours and highlights from various other datasets.

1.2.15. Migration

https://www.ssb.no/en/befolkning
Data on migration is regularly prepared in conjunction with migration statistics. We have prepared the data further by producing grid statistics on the population moves within the municipality. This has been done both as net moves in to each grid and as arrows indicating the moves. We thought this would indicate which areas are more attractive than others. However, there is a close relationship between number of moves and total population. We were able to separate the moves in to age groups, and certain patterns may be observed, but this parameter was not utilized in the final testing for Oslo because of the above described drawbacks.

1.2.16. Employment

http://www.ssb.no/en/arbeid-og-lonn/
Statistics Norway publishes statistics on employment, also linking this data to Statistical grids.
In this project, we look at number of employees within certain distances, thereby getting an estimate on accessible possible jobs.

APPENDIX C

1. Structuring and georeferencing data

In order to compare the various data all were referred back to their most precise location. All three parts of the cadastre have been utilized: Property centroids from the Property register, Address centroids from the Address register, and Building centroids from the Building register. They have generally been used as follows:

1. For our *dwellings sales dataset,* the most precise obtainable point location is **property centroid,** giving a dataset where the main variables we chose to use are as following:
   - 2 unique variables: Total sales sum, m^2 floor space
   - 1 deduced variable: Price per m^2
   - Location (Property-, farm-, House section-id) → property centroid, x y
2. For population data - "who lives where?", the **Address centroid** is the best obtainable location.
3. For variables on buildings the most precise obtainable location is the **Building centroid**

This chapter describes the various data variables that were brought along from the source data and how this was structured in a geodatabase. Chapter 3.4 shows how the different variables are added in to the **real estate agencies by dwelling-dataset,** prepared for analysis

1.1. Point data

The key three components for our point data is the Cadastre: Property register, Building register and Address register in, each holding X,Y-coordinates representing their entities.
This data was combined with population data by addresses and housing price data by dwelling. Making use of database keys within the cadastre it was possible to make relations in between the following point datasets:

Properties in Cadastre - Centroid Point		
Tabell egenskap	Table Feature	Data Type
KOMMUNENR	MunicipalityId	Long Integer
GATENR_GAARDSNR	Street-Id_Property-Id	Long Integer
HUSNR_BRUKSNR	House-id and farm-id	Long Integer
BOKSTAV_FESTENR	House section-Id	Long Integer
X_KOORDINAT	X coordinate	Double
Y_KOORDINAT	Y coordinate	Double

Buildings in Cadastre – Point		
Tabell egenskap	Table Feature	Data Type
KOMMUNENR	MunicipalityId	Long Integer
BYGNINGSNR	BuildingId	Long Integer
BYGNINGSTYPE	Buildingtype	Long Integer
ANTALL_ETASJER	Number of floors	Long Integer
ANTALL_BOLIGER	Number of dwellings	Long Integer
BRUKSAREAL_TOTALT	Dwelling area, total	Double
BRUKSAREAL_BOLIG	Dwelling area, dwelling	Double
KOMMUNENR	MunicipalityId	
GATENR_GAARDSNR	Street-Id_Property-Id	
HUSNR_BRUKSNR	House-id and farm-id	
BOKSTAV_FESTENR	House section-Id	
Boa	Area measurements	Text
p_rom		Double
Bta		Double
Bra		Double
Grflate		Text
Tattbrkd	Date – building taken in use	Text
Areal_e2	Building ground surface, total	Text
Byggstat	Building status	Text
Naering	Business	Text
X_KOORDINAT	X coordinate	Double
Y_KOORDINAT	Y coordinate	Double

Address (with Population) in Cadastre – Point		
Tabell egenskap	Table Feature	Data Type
GATENR_GAARDSNR	Street-Id_Property-Id	Long Integer
HUSNR_BRUKSNR	House-id and farm-id	Long Integer
BOKSTAV_FESTENR	House section-Id	Long Integer
adr0_6	No of persons, age < 6	Long Integer
adr6_16	No of persons, age 6-16	Long Integer
Adr16_20	No of persons, age 16-20	Long Integer
Adr20_40	No of persons, age 20-40	Long Integer
Adr40_67	No of persons, age 40-67	Long Integer
Adr67_	No of persons, age > 67	Long Integer
Alle	No of persons, total	Long Integer
Flytte5	No of persons, moved last 5 years	Long Integer
X_KOORDINAT	X coordinate	Double
Y_KOORDINAT	Y coordinate	Double

Data from real estate agencies by dwelling		
Tabell egenskap	Table Feature	Data Type
BRUKSENHETID	DwellingId	Double
KOMMUNENR	MunicipalityId	Long Integer
GATENR_GAARDSNR	Street-Id_Property-Id	Long Integer
HUSNR_BRUKSNR	House-id and farm-id	Long Integer

BOKSTAV_FESTENR	House section-Id	Long Integer
Boa	Area measurements	Text
p_rom		Double
Bta		Double
Bra		Double
ant_soverom	No bedrooms	Double
Byggeaar	Age of building *	Double
Laanetakst	Price indication and price by date of sale	Double
Pris		Double
Totalsum		Double
pris_m^2	Price per m^2	Long Integer

* Age of building (Adjusted)
- Variable in register is construction year.
Used variable is age of building in relation to 2014. Example: 2013 = 1, 2012 = 2
Pre-WW2 adjustment: all buildings built before 1945 are set to an age value=5

Georeferenced to **Address-point**

The Central Population Register (CPR)		
Tabell egenskap	Table Feature	Data Type
GATENR_GAARDSNR	Street-Id_Property-Id	Long Integer
HUSNR_BRUKSNR	House-id and farm-id	Long Integer
BOKSTAV_FESTENR	House section-Id	Long Integer
Personnummer	Personal identity number	Long Integer

1.2. Area data

The following area datasets were linked or generated in relation to **Real estate dwelling - Point** data, adding variables to the dataset. Geographical analysis: overlay and distance calculations.

Geographical grid 1km – area		
Tabell egenskap	Table Feature	Data Type
rute_1000m	Grid-Id_1000m	Double

Geographical grid 250m – area		
Tabell egenskap	Table Feature	Data Type
rute_250m	Grid-Id_250m	Double

Geographical grid 500m – area		
Tabell egenskap	Table Feature	Data Type
rute_500m	Grid-Id_500m	Double

Urban settlements – area		
Tabell egenskap	Table Feature	Data Type
Tettstedsnr	Urban settlement-id	Double

Distance to **Centre zone – area**		
Tabell egenskap	Table Feature	Data Type
Sentrum	CentreZoneId	Long Integer

Distance to **Recreational areas – area**		
Tabell egenskap	Table Feature	Data Type
Rekr	Recreational areas	Long Integer

Distance to **N50 fresh water – area**		
Tabell egenskap	Table Feature	Data Type
Innsjø_elvbekk	Lakes&Rivers	Long Integer

Distance to **Coastline – area**		
Tabell egenskap	Table Feature	Data Type
Kystlinje	Coastline	Long Integer

1.3. Analyses datasets

In order to get variables based on literature and own findings within the project some datasets were generated by analyses

1.3.1. Analyses datasets - other geographic entities

Values for all variables are generated as Distance to the **real estate dwelling - Point** data.

Distance to **other geographical entities**		
Tabell egenskap	Table Feature	Data Type
Holdeplass	Distance to public transport	Double
Bane	Distance to public rail transport	Double
avst_veg60	Distance from road with speed limit 60 km/h	Double

1.3.2. Analyses datasets - buildings

Data and Building types reside in **Buildings register in Cadastre.**
Values for all variables are generated in relation to the **real estate dwelling - Point** data.

Georeferenced to **Building-point**

Distance to **building types**–		
Tabell egenskap	Table Feature	Data Type
BYGNINGSNR	BuildingId	Long Integer
Primhelse	Distance to primary Health institutions	Double
Skole	Distance to school	Double
Sykehus	Distance to hospital	Double
Barnehage	Distance to kindergaten	Double
Universitet/Høyskole	University/Higher Education*	Double
Restaurant	Restaurant	Double

*Building Type “University/Higher Education”, with only buildings > 1 500 m^2 floor space AND contain auditorium/ classroom.

Georeferenced to **Building-point**

Number of large surface buildings in 1k grid cell built pre-1900 – analyses dataset		
Tabell egenskap	Table Feature	Data Type
BYGNINGSNR	BuildingId	Long Integer
gamlestorehus1900	Built buildings pre 1900	Long Integer

1.3.3. Analyses datasets - Population

These data were combined firstly using **Personal identity number** as database key, secondly with **Address** to point-georeference the data**.** Values for all variables are generated in relation to each **real estate dwelling – Point**

Georeferenced to **Address-point**

Income - medium for all households within 250 metres of real estate dwelling – Point		
Tabell egenskap	Table Feature	Data Type
Personnummer	Personal identity number	Long Integer
Husholdningsinntekt – før skatt	household income – before taxes	Long Integer
Husholdningsinntekt – etter skatt	household income –after taxes	Long Integer

Georeferenced to **Address-point**

Education - medium for all persons aged 25 or more within 250 metres of real estate dwelling – Point		
Tabell egenskap	Table Feature	Data Type
Personnummer	Personal identity number	Long Integer
Utdanningsnivå	Level of education*	Float

Level of education*
The data uses NUS2000 codes, designating an education level from 1 to 8 for all persons.
All NUS2000 codes are assigned a corresponding international code (ISCED97) for international reporting purposes. Detailed list of NUS2000 codes with corresponding international codes (ISCED97):
https://www.ssb.no/en/utdanning/statistikker/utniv/aar/2016-06-20?fane=om#content

In the project, we have chosen to only include adults => 26 of age, excluding persons for which Statistics Norway have no education level information. This is true for many immigrants, and approximately 750 000 persons in Norway lack information of education level in our registers.

Georeferenced to **Address-point**

Immigration - Percentage of total population within 250 metres of real estate dwelling – Point		
Tabell egenskap	Table Feature	Data Type
Personnummer	Personal identity number	Long Integer
Innvandrere	Immigrants*	Long Integer
Befolkning med ikke-vestlig bakgrunn	Population with non-western ancestry *	Long Integer

Immigrants*
As of 1 January 2016, around 848 000 persons resident in Norway were either immigrants (699 000) or born in Norway to two immigrant parents (150 000). The variable encompasses these two groups, which combined make up 16 per cent of the population of Norway.

Population with non-western ancestry*
The variable is firstly based on the population variable “Country background of grandparents”, where all Norwegian residents are assigned a 3-digit country-code, identifying the country background of their grandparents. A main country code is designated, based on these four country codes.

Depending on this country-code, each person is designated to one of two possible groups, which we for ease here call “West” and “Non-West”. This is a Statistics Norway official grouping of countries for statistical purposes, and is divided as follow:
“West”: EU/EEA, USA, Canada, Australia and New Zealand
“Non-West”: Asia, Africa, Latin America, Oceania excl. Australia and New Zealand, and Europe outside EU/EEA

Georeferenced to **Address-point**

Other Population variables: in relation to population within 250 metres of real estate dwelling – Point		
Tabell egenskap	Table Feature	Data Type
Personnummer	Personal identity number	Long Integer
Andel barn	Percentage Children (<18 years old)	Double
Snitt alder	Medium age	Double

Georeferenced to centroid-point for **Geographical grid 500m**

Employees - Number of jobs (by employees) found within 5 km and 10km distance - “as the crow flies”		
Tabell egenskap	Table Feature	Data Type
Emp_tot5km	Employees within 5 km	Long Integer
Emp_tot10km	Employees within 10 km	Long Integer

1.3.4. Analyses datasets - intensity

Values for all variables are generated in relation to the **real estate dwelling - Point** data.

Noise		
Tabell egenskap	Table Feature	Data Type
Stoyexp (2011 døgnekvivalentnivå - dba)	Noise 2011 (day equivalent level in dba)	Long Integer

Sun hours – generated from DTM 10 – Grid dataset		
Tabell egenskap	Table Feature	Data Type
Soltimer	No. of Sun hours	Double

APPENDIX D

1. COMPARED SAME SIZES - Data Output

1.1. Oslo

Oslo	AdjR^2 (isolated)
RESTAURANT_DIST	0.18
AND_REGANN	0.20
CENTREZ_DIST	0.30
WATER_DIST	0.05
POP_EDUC_L_P5	0.47
POP_INCOME	0.08
AND_UNDER18	0.16
BUILDING_AGE	0.14
HOSPITAL_DIST	0.23
COMBINED	0.61

The overall combined AdjR^2 for **Compared same size** in Oslo is 0.61, clearly lower than 0.74 and 0.82 for the two other approaches.

Square meters floor space is “baked” into the dependent variable in this approach, a variable strongly contributing to the overall result in the two other approaches.
An apparent advantage of the **Compared same size** approach is therefore that it might open up space for other explanatory variables.
The model in itself holds just one building intrinsic variable (BUILDING_AGE), and our analysis results show that all “best 3”-combinations up to 6 variables actually do not include the variable.

The apparent very high value is POP_EDUC_L_P5, with an AdjR^2 of 0.47, 77% of the total achievable AdjR^2 of 0.61. For our main approaches POP_EDUC_L scored 0.21 and 0.35, with higher total AdjR^2's.
A new variable in this approach (not in the two others) is AND_REGANN, which here pass the set criteria for including it.

```
*****************************************************************************
Choose 1 of 9 Summary
      Highest Adjusted R-Squared Results
AdjR²      AICc    JB K(BP)  VIF   SA   Model
 0,47 -11853,04 0,00  0,00 1,00 0,00  +POP_EDUC_L_P5***
 0,30  -5502,20 0,00  0,00 1,00 0,00  -CENTREZ_DIST***
 0,23  -2997,96 0,00  0,00 1,00 0,00  -HOSPITAL_DIST***
*****************************************************************************
Choose 2 of 9 Summary
              Highest Adjusted R-Squared Results
AdjR²      AICc    JB K(BP)  VIF   SA   Model
 0,53 -14500,34 0,00  0,00 1,06 0,00  +POP_EDUC_L_P5***  -AND_UNDER18***
 0,53 -14492,77 0,00  0,00 1,34 0,00  -CENTREZ_DIST***  +POP_EDUC_L_P5***
 0,51 -13691,86 0,00  0,00 1,13 0,00  -RESTAURANT_DIST***  +POP_EDUC_L_P5***
*****************************************************************************
Choose 3 of 9 Summary
                         Highest Adjusted R-Squared Results
AdjR²      AICc    JB K(BP)  VIF   SA   Model
 0,56 -16178,33 0,00  0,00 1,66 0,00  -CENTREZ_DIST***  +POP_INCOME***  -AND_UNDER18***
 0,55 -15458,56 0,00  0,00 1,39 0,00  -CENTREZ_DIST***  +POP_EDUC_L_P5***  -
BUILDING_AGE***
 0,55 -15403,29 0,00  0,00 1,83 0,00  -CENTREZ_DIST***  +POP_EDUC_L_P5***  -
AND_UNDER18***
*****************************************************************************
Choose 4 of 9 Summary
                                   Highest Adjusted R-Squared Results
AdjR²      AICc    JB K(BP)  VIF   SA   Model
 0,58 -17462,12 0,00  0,00 3,32 0,00  -CENTREZ_DIST***  +POP_EDUC_L_P5***
+POP_INCOME***  -AND_UNDER18***
```

```
 0,58 -17036,36 0,00  0,00 1,84 0,00  -CENTREZ_DIST***  +POP_INCOME***  -AND_UNDER18***
-BUILDING_AGE***
 0,57 -16725,45 0,00  0,00 1,88 0,00  -RESTAURANT_DIST***  -CENTREZ_DIST***
+POP_INCOME***  -AND_UNDER18***
******************************************************************************
Choose 5 of 9 Summary
                                           Highest Adjusted R-Squared Results
AdjR2      AICc   JB K(BP)  VIF   SA   Model
 0,60 -18095,16 0,00  0,00 3,39 0,00  -CENTREZ_DIST***  +POP_EDUC_L_P5***
+POP_INCOME***  -AND_UNDER18***  -BUILDING_AGE***
 0,59 -17845,78 0,00  0,00 3,37 0,00  -RESTAURANT_DIST***  -CENTREZ_DIST***
+POP_EDUC_L_P5***  +POP_INCOME***  -AND_UNDER18***
 0,59 -17649,18 0,00  0,00 2,42 0,00  -AND_REGANN***  -CENTREZ_DIST***  +POP_INCOME***
-AND_UNDER18***  -BUILDING_AGE***
******************************************************************************
Choose 6 of 9 Summary
                                                   Highest Adjusted R-Squared Results
AdjR2      AICc   JB K(BP)  VIF   SA   Model
 0,60 -18415,32 0,00  0,00 3,43 0,00  -RESTAURANT_DIST***  -CENTREZ_DIST***
+POP_EDUC_L_P5***  +POP_INCOME***  -AND_UNDER18***  -BUILDING_AGE***
 0,60 -18253,65 0,00  0,00 3,43 0,00  -CENTREZ_DIST***  -WATER_DIST***
+POP_EDUC_L_P5***  +POP_INCOME***  -AND_UNDER18***  -BUILDING_AGE***
 0,60 -18206,11 0,00  0,00 4,42 0,00  -AND_REGANN***  -CENTREZ_DIST***
+POP_EDUC_L_P5***  +POP_INCOME***  -AND_UNDER18***  -BUILDING_AGE***
******************************************************************************
Choose 7 of 9 Summary
                                                           Highest Adjusted R-Squared
Results
AdjR2      AICc   JB K(BP)  VIF   SA   Model
 0,60 -18552,83 0,00  0,00 4,49 0,00  -RESTAURANT_DIST***  -AND_REGANN***  -
CENTREZ_DIST***  +POP_EDUC_L_P5***  +POP_INCOME***  -AND_UNDER18***  -BUILDING_AGE***
 0,60 -18525,41 0,00  0,00 3,47 0,00  -RESTAURANT_DIST***  -CENTREZ_DIST***  -
WATER_DIST***  +POP_EDUC_L_P5***  +POP_INCOME***  -AND_UNDER18***  -BUILDING_AGE***
 0,60 -18455,72 0,00  0,00 3,79 0,00  -HOSPITAL_DIST***  -RESTAURANT_DIST***  -
CENTREZ_DIST***  +POP_EDUC_L_P5***  +POP_INCOME***  -AND_UNDER18***  -BUILDING_AGE***
******************************************************************************
Choose 8 of 9 Summary
                                                                   Highest Adjusted R-
Squared Results
AdjR2      AICc   JB K(BP)  VIF   SA   Model
 0,61 -18646,64 0,00  0,00 4,50 0,00  -RESTAURANT_DIST***  -AND_REGANN***  -
CENTREZ_DIST***  -WATER_DIST***  +POP_EDUC_L_P5***  +POP_INCOME***  -AND_UNDER18***  -
BUILDING_AGE***
 0,60 -18604,94 0,00  0,00 4,94 0,00  -HOSPITAL_DIST***  -RESTAURANT_DIST***  -
AND_REGANN***  -CENTREZ_DIST***  +POP_EDUC_L_P5***  +POP_INCOME***  -AND_UNDER18***  -
BUILDING_AGE***
 0,60 -18556,10 0,00  0,00 3,81 0,00  -HOSPITAL_DIST***  -RESTAURANT_DIST***  -
CENTREZ_DIST***  -WATER_DIST***  +POP_EDUC_L_P5***  +POP_INCOME***  -AND_UNDER18***  -
BUILDING_AGE***

------------------------------------------------------------------------------
        Summary of Variable Significance
Variable       % Significant % Negative % Positive
RESTAURANT_DIST            100,00     100,00       0,00
CENTREZ_DIST   100,00     100,00       0,00
POP_EDUC_L_P5             100,00       0,00     100,00
BUILDING_AGE       100,00     100,00       0,00
WATER_DIST          99,22     100,00       0,00
AND_REGANN            98,04      92,55       7,45
AND_UNDER18           97,25      93,73       6,27
POP_INCOME        96,86       5,49      94,51
HOSPITAL_DIST             95,29      97,65       2,35

------------------------------------------------------------------------------
     Summary of Multicollinearity
Variable        VIF Violations Covariates
HOSPITAL_DIST     1,94     0      --------
RESTAURANT_DIST     1,55     0      --------
AND_REGANN     2,58     0      --------
CENTREZ_DIST     2,97     0      --------
WATER_DIST   1,13     0      --------
POP_EDUC_L_P5      4,94     0      --------
POP_INCOME 3,30     0      --------
AND_UNDER18    2,70     0      --------
BUILDING_AGE 1,28     0      --------

------------------------------------------------------------------------------
Table Abbreviations
AdjR2 Adjusted R-Squared
AICc  Akaike's Information Criterion
JB    Jarque-Bera p-value
K(BP) Koenker (BP) Statistic p-value
VIF   Max Variance Inflation Factor
SA    Global Moran's I p-value
Model Variable sign (+/-)
Model Variable significance (* = 0,10; ** = 0,05; *** = 0,01)

------------------------------------------------------------------------------
```

1.2. *Oslo* with other cities

Applying the Oslo-model to 5 largest Norwegian cities

The combined AdjR^2 is clearly highest for Oslo (0.61), with the 4 second biggest cities at AdjR^2 levels between 0.27 and 0.44.
POP_EDUC_L_P5 is relatively high in all cities, at over 50% of total achieved AdjR^2.

The table below summarizes results for all 5 largest cities. The results in this approach are nor used for further prediction, as total AdjR^2 is relatively low.

Compares same sizes - How much of price variation we are able to explain in Norway's 5 largest cities. AdjR^2 for each variable isolated, and total combined AdjR^2.

AdjR^2					
	Oslo	**Bergen**	**Stavanger**	**Trondheim**	**Drammen**
Population	*958 378*	*250 420*	*210 874*	*175 068*	*113 534*
RESTAURANT_DIST	.18	.14	-	.13	.05
AND_REGANN	.20	.04	.02	.06	.10
CENTREZ_DIST	.30	.10	.02	.21	.07
WATER_DIST	.05	.03	.03	.12	.06
POP_EDUC_L_P5	.47	.23	.15	.22	.15
POP_INCOME	.08	.02	.09	.02	.06
AND_UNDER18	.16	.05	.00	.09	.12
BUILDING_AGE	.14	.14	.07	.12	.14
HOSPITAL_DIST	.23	.08	.00	.12	.09
COMBINED	**.61**	.39	**.27**	**.38**	**.44**

APPENDIX E: Coefficients

Coefficients for dependent variable **Price per m² - Oslo**

Variable	Coef
Intercept	-13587.30470390000
HOSPITAL_DIST	-0.10373293612
RESTAURANT_DIST	-0.91701610756
POP_EDUC_L	8110.60785656000
CENTREZ_DIST	-0.50031758780
WATER_DIST	-1.65258466752
FLOOR_SPACE_RECI	777209.62615300000
POP_AGE	472.27411045900
UNIVERS_DIST	-0.32628683747
BUILDING_AGE	-89.08159680120

Coefficients for dependent variable **Total sales price - Oslo**

Variable	**Coef**
Intercept	-3639487.42215
RESTAURANT_DIST	-164.69926393500
CENTREZ_DIST	-98.65370343370
WATER_DIST	-142.13425651900
FLOOR_SPACE_SQR	596942.51901100000
POP_EDUC_L_P5	138.17765320700
POP_INCOME	1.45942711512
POP_AGE	40625.91287910000
BUILDING_AGE	-7605.85554977000

Coefficients for dependent variable **Price per m^2 - Bergen**

Variable	Coef
Intercept	-7713.58131803000
HOSPITAL_DIST	0.06648987709
RESTAURANT_DIST	-0.43653439651
POP_EDUC_L	5798.33561125000
CENTREZ_DIST	-0.46144645688
WATER_DIST	-3.00203064421
FLOOR_SPACE_RECI	893694.08116500000
POP_AGE	349.76230673000
UNIVERS_DIST	-0.43818095301
BUILDING_AGE	-116.99735552800

Coefficients for dependent variable **Total sales price - Bergen**

Variable	Coef
Intercept	-2344096.24620000000
RESTAURANT_DIST	-105.71453592200
CENTREZ_DIST	-55.49647668670
WATER_DIST	-230.77653000000
FLOOR_SPACE_SQR	440048.89105900000
POP_EDUC_L_P5	185.28117497600
POP_INCOME	0.48695168530
POP_AGE	34859.32533760000
BUILDING_AGE	-9560.17326391000

Coefficients for dependent variable **Price per m^2 - Stavanger**

Variable	Coef
Intercept	-17306.71004590000
HOSPITAL_DIST	0.84801618671
RESTAURANT_DIST	-0.15224580203
POP_EDUC_L	6354.51775363000
CENTREZ_DIST	0.13756698264
WATER_DIST	-0.55010147558
FLOOR_SPACE_RECI	1333052.85512000000
POP_AGE	326.29757795400
UNIVERS_DIST	-0.41802193674
BUILDING_AGE	-97.32012872050

Coefficients for dependent variable **Total sales price - Stavanger**

Variable	Coef
Intercept	-2792520.12452000000
RESTAURANT_DIST	19.47462218690
CENTREZ_DIST	-16.81135079100
WATER_DIST	-43.48560456110
FLOOR_SPACE_SQR	422530.96142500000
POP_EDUC_L_P5	214.44035050900
POP_INCOME	0.92321492042
POP_AGE	33636.25927110000
BUILDING_AGE	-11557.00054300000

Coefficients for dependent variable **Price per m^2 – Trondheim**

Variable	Coef
Intercept	5727.03100837000
HOSPITAL_DIST	0.29368263028
RESTAURANT_DIST	-1.30488436087
POP_EDUC_L	2926.37426736000
CENTREZ_DIST	-1.00799106829
WATER_DIST	-0,59001798890
FLOOR_SPACE_RECI	926982.17856800000
POP_AGE	303.73978051800
UNIVERS_DIST	0.13546251148
BUILDING_AGE	-94.08378334330

Coefficients for dependent variable **Total sales price – Trondheim**

Variable	Coef
Intercept	-1507734.01292000000
RESTAURANT_DIST	-97.90814885930
CENTREZ_DIST	-75.68383720320
WATER_DIST	-39.77955240060
FLOOR_SPACE_SQR	379916.61355900000
POP_EDUC_L_P5	133.90884179400
POP_INCOME	0.40038716949
POP_AGE	26436.06410850000
BUILDING_AGE	-7370.85458140000

Statistics Norway

Postal address:
PO Box 8131 Dept
NO-0033 Oslo

Office address:
Akersveien 26, Oslo
Oterveien 23, Kongsvinger

E-mail: ssb@ssb.no
Internet: www.ssb.no
Telephone: + 47 62 88 50 00

ISBN 978-82-537-9595-9 (electronic)

Statistisk sentralbyrå
Statistics Norway

Design: Siri Boquist