STAT 112

FINAL PROJECT REPORT

Analysis of Factors Affecting Danish Residential Property Sales: A Data-Driven Approach with Visualized Insights

Prepared by:

Metehan Baydar, Cemile Bilgir, Arca Büyükçatalbaş, Abdulsamet Demir, Arif Emre Kesik, İlayda Ilgıt Köroğlu and Yiğit Efe Özdemir.

Instructor: Prof. Dr. Ceylan Yozgatlıgil

January 2025

1. Abstract

This report analyzes the development of house prices in Denmark since 1885 and the economic, social and historical factors that have influenced this development. The report examines how house prices are affected by factors such as garden availability, energy efficiency, property type and distance from the city center, regional location. To better understand these relationships, a comprehensive dataset was compiled, and visualizations were created. After data pre-processing, cleaning and organizing, exploratory data analysis (EDA) techniques were used to analyze market trends and interactions between key variables. It also assesses the impact on the housing market of important historical events such as the Second World War and modern construction techniques. In conclusion, the multifaceted and time-varying structure of Denmark's housing market has been shaped by social, environmental and historical dynamics.

2. Introduction

The Danish residential market is an important factor affecting the country's economy. Residential prices, regional variations and market trends are some of the main factors affecting the residential market. This report aims to investigate in-depth the factors affecting house prices and includes a set of research questions. The inferences drawn from the graphs prepared for these questions allow for the development of policies on the residential housing market. In addition, this report provides guidance for those with specific investment objectives.

3. Data tidying and Cleaning Steps

The data was first examined to find out what kind of errors there might be, and when these errors were edited, they were checked again to fix them and obtain the final version. The process proceeds as follows:

The csv file of the data set was uploaded to Google Colab and converted to DataFrame with the help of the pandas .read_csv function.

The data were opened with the df.info() function to check that the data types are correct, most of them were seen as data type object.

Then it was observed with the df.head() function that there were special characters in the names of some columns and that uppercase and lowercase letters were not taken into account. These need to be corrected. It was noticed that the same value was represented more than once with df.tail() and these need to be cleaned.

df.iloc() function was used to check if there was any separation or header issue and no error was found.

Then, df.head() function was used to check if each column had duplication or one variable. No problem has been observed. Therefore, the use of stacking or unstacking operations was not required.

To start organizing the names, first all headers in the DataFrame were renamed with df.rename() function, all letters in lower case, thus they are all the same format.

After the headers were named, the value counts (frequencies) of the categorical columns were checked. To simplify the process, the for loop was used with the df.value_counts() function.

Also, the central tendency and dispersion values of the numerical columns were examined using the df.describe() function. For was used to simplify this process.

Following this, the Not Avilable(NA) values in the data set were examined with the df.isnull() function. A total of 943 null values were found.

After that, the mentioned certain columns' valeus are organized to have only certain unique values. This is done by the str.strip() and str.replace() functions. The "@" sign was replaced with a in 4 columns, and the "%" sign was replaced with nothing in 3 columns.

After that, the categorical columns were case-adjusted using the str.title() and str.lower()

functions. Many of them were changed to title format.

After all this, the check was done again, and some columns' strings were seen as NA.

To fix this problem, the type of categorical columns was changed to "object" and NA strings were changed to formats where they can be read as NA with Python's replace ("strings", pd.NA) function. After the change, the data type was converted back to string. In addition, the "none" data in the heating type was changed to "no heating" because in fact those data are perceived as NA since they appear to be none, but it is stated that there is no heating system there.

Then, since some of the data was initially interpreted as objects by Python, the data type of the numeric columns was changed to the appropriate ones with the df.astype() function.

And after that, a new price_dkk column was calculated by multiplying the old price_dkk column by 0.14 to change the currency from Danish Krone to US Dollar.

In the next step, the NA values were worked on again.

First, NA percentages were calculated on a row basis and those with NA values more than 60% in their columns were filtered to avoid excessive data loss.

Secondly, NA values in categorical columns were replaced with mode because median or mean values cannot be used in categorical values.

In the next step, NA values in numerical columns were replaced with median because outliers were still in the dataset.

Then, decimal points of some columns were rounded to two since there were data with more than ten decimal places using the df.round() function.

After that, outliers in numerical columns were replaced with means.

To do this, the first quartile (Q_1) and the third quartile (Q_3) were first calculated with the df.quantile() function and then these were used to calculate Interquartile Range(IQR).

Whiskers were found by subtracting 1.5 IQR from Q_1 and adding 1.5 IQR to Q_3. In the next step, rows with outliers were detected and the mean value of those columns was calculated.

After the calculation, outliers were replaced with the mean value using the df.loc() function.

Finally, the frequencies and unique values of the categorical columns, the central tendency and distribution values of the numerical columns were checked again and how they changed after data cleaning was observed.

4. Exploratory Data Analysis

4.1 Data Description

The dataset contains information about Danish residential properties and economic factors over multiple observations. It is designed to study housing market trends, property attributes, and their relationships with economic indicators.

4.2 Variable Description

1. Region (Categorical): Geographical region where the property is located.

- Categories: Capital, Zealand, Southern Denmark, Central Jutland, North Jutland

2. Property Type (Categorical): Type of residential property.

- Categories: Apartment, Detached House, Terraced House, Farmhouse

3. Built Year (Integer): The year the property was constructed

- Range: 1900–2024

4. Size sqm (Integer): Size of the property in square meters.

- Range: 50–300

5. Rooms (Integer): Number of rooms on the property.

- Range: 1–10

6. Condition (Categorical): The general state of the property.

- Categories: New, Good, Needs Renovation

7. Heating Type (Categorical): Type of heating system installed on the property.

- Categories: Central, Electric, Gas, None

8. Energy Label (Categorical): Energy efficiency rating of the property.

- Categories: A, B, C, D, E, F, G

9. Price DKK (Integer): Sale price of the property in Danish Krone.

- Range: 500,000–10,000,000

10. Sale Year (Integer)

Description: Year the property was sold.

Range: 1992–2024

11. Garden Availability (Categorical): Indicates whether the property has a garden.

- Categories: Yes, No

12. Distance to City (km) (Float): Distance of the property to the nearest city in kilometers.

- Range: 0.5–50.0

13. School Rating (Integer): Rating of nearby schools (scale 1 to 5).

- Range: 1–5

14. Parking Available (Categorical): Indicates whether parking is available on the property.

- Categories: Yes, No

15. Nomination Interest Rate (%) (Float): Danish nominal interest rate (annualized, not converted to quarterly).

- Range: 0.1–10.0

16. DK Annual Inflation Rate (%) (Float): Danish annual inflation rate (annualized, not converted to quarterly).

- Range: 0.1–5.0

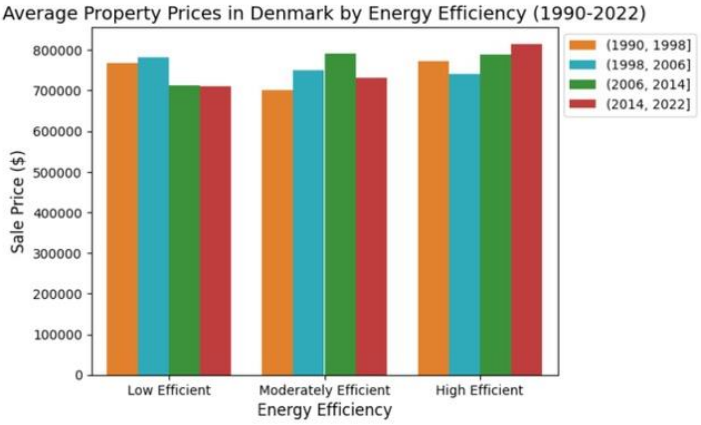
17. Yield on Mortgage Credit Bonds (Float): Yield on 30-year mortgage credit bonds (without spread).

- Range: 0.5–7.0

4.3 Research Questions

4.3.1 What is the relationship between property prices in Denmark from 1990 to 2022 and energy efficiency of properties?

(İlayda Ilgıt Köroğlu)



Visualization 1: Average Property Prices in Denmark by Energy Efficiency (1990-2022)

According to the Danish Energy Agency (2023), energy productions from renewable energy have been increasing and non-renewable energy products are decreasing. It is expected that the increase in efficiency standards and the increase in demand for sustainable living spaces will lead to significant changes in the housing market. It is also important to keep in mind that the government incentives from the government agreement in 2012, which Petersen (2014) mentioned in the report published by the Denmark government, started around that year. Examining this relationship will show how energy efficiency affects property prices over time. To examine this relationship, it was thought that it would be more accurate to examine it with a bar chart, since we have both categorical and numerical data. Both examine different categories separately and the change over time becomes more readable and easier. In order to prevent data clutter in energy efficiency and because it is more understandable to read the data by labeling it rather than reading it with letters, A and B efficiency ratings were divided into three as "Highly Efficient", C and D efficiency ratings "Moderately Efficient", and E and F efficiency ratings "Low Efficient". At the same time, the years are divided into 4 intervals as 1990-1998 (first), 1998-2006 (second), 2006-2014 (third), 2014-2022 (last) to make it easier to read and understand.

First, When the price of low-efficiency properties is looked at, it starts at approximately 750 thousand U.S dollars (USD), in the first-year interval and increases until the second-year interval. This indicates that such properties are still acceptable at first. However, it has dropped to an average of 700 thousand USD in the third and last year interval.

This decline suggests a gradual reduction in the demand for properties with low energy efficiency, which, consequently, leads to a decrease in their value. The government's policies encouraging energy efficiency and customers' turning to environmentally friendly properties may have led to this price decrease because according to the data from the Danish Energy Agency, renewable energy sources are increasingly being used and while importance is being given to sustainability, it can be said that energy efficiency is also very likely to be given importance.

When we look at Moderately Efficient in the graph, it is seen that these energy efficient properties have an average price of 700 thousand USD in the first-year interval, and this increases to an average of approximately 800 thousand USD by the third-year interval. The reason for this much price increase is the period when environmental awareness increased and incentives for energy efficiency became increasingly stronger. In last year's interval, there was a decrease in the average price value. The interpretation that can be made here is that the market value of properties with moderate efficiency decreased with the increase in demand for more efficient properties. This is likely due to the tendency of homeowners to switch to more efficient properties.

When the price of low-efficiency properties is looked at, it starts at approximately 750 thousand USD, in the first-year interval and increases until the second-year interval. This indicates that such properties are still acceptable at first. However, it has dropped to an average of 0.7 million USD in the third and last year interval. This decline suggests a gradual reduction in the demand for properties with low energy efficiency, which, consequently, leads to a decrease in their value. The government's

policies encouraging energy efficiency and customers' turning to environmentally friendly properties may have led to this price decrease because according to the data from the Danish Energy Agency, renewable energy sources are increasingly being used and while importance is being given to sustainability, it can be said that energy efficiency is also very likely to be given importance.

When looking at the price values of high efficiency properties, although the average was approximately 750 thousand million USD in the first-year interval, a decrease is seen in the second-year interval, but in the third and last year interval, the price values are seen to increase and exceed an average of 800 thousand USD. The increase in the price is due to the increase in environmental awareness and more demand for energy efficiency. It is possible that it reflects the increase in government support and the interest of customers in environmentally friendly properties. During this period, high efficiency properties gained more value with advantages such as low energy costs and sustainability.

As a result, the changes in price of properties with the energy efficiency label in Denmark are due to environmental and economic factors. High efficiency properties, due to their environmental friendliness and energy saving lifestyle, have seen and are expected to see a greater increase in their prices compared to other types of efficiency. Moderately efficient properties are also seen as the second choice after high efficiency properties according to low efficiency, depending on purchasing power. In general, low efficiency properties lose value while high efficiency properties gain value.

4.3.2 What are the regional trends in the distribution of property conditions (Good, Needs Renovation, New) in Denmark, and how do they relate to average property prices? (Arca Büyükçatalbaş)

Visualization2: Average Property Price by Region and Condition (with Percentages)

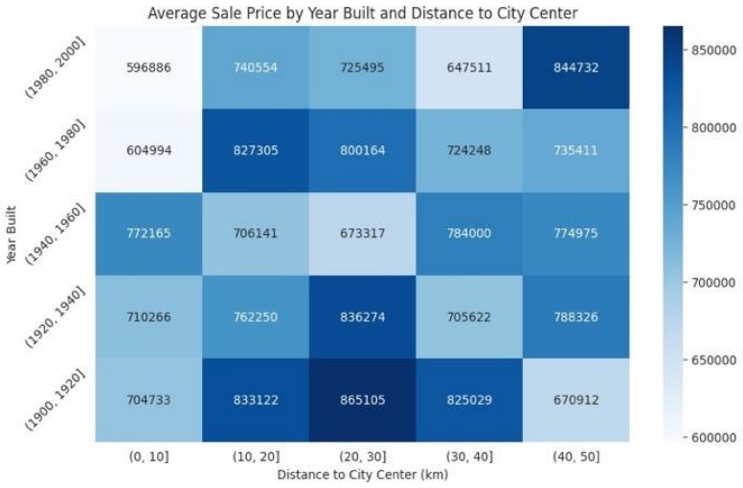
The visualization offers an in-depth look at average property prices across various regions in Denmark, categorized by property condition—specifically, Good, Needs Renovation, and New. A stacked bar chart is chosen for this visualization because it effectively displays both the absolute and relative proportions of each property condition within each region, allowing for easy comparison across categories and regions. This format is particularly well-suited for illustrating trends, as it highlights not only the total value of average prices but also how different property conditions contribute to those totals.

Each area shows distinct trends in property conditions, revealing differences in housing stock and market dynamics. In the Capital region, for instance, there's an even distribution: 34.6% of properties are rated as "Good," while 32.5% fall into the "Needs Renovation" category, and 32.9% are classified as "New." Central Jutland mirrors this balance with 34.0% of properties in "Good" condition, 32.9% needing renovation, and 33.1% labeled as new constructions. Zealand stands out with the highest share of "New" properties at 35.2%, indicating significant recent investments, whereas "Needs Renovation" and "Good" properties are quite evenly divided at 32.4% and 32.5%, respectively.

In contrast, North Jutland has the largest percentage of "Good" properties at 37.6%, but the fewest new constructions at 28.6%, suggesting a market rich in well-preserved older homes. Southern Denmark also shows a balanced breakdown with 33.3% of properties classified as "Good," 32.4% needing renovations, and 33.7% being new developments.

Overall, properties classified as "Needs Renovation" consistently account for about one-third of the housing market, with North Jutland showing the lowest at 28.6% and Southern Denmark the highest at 32.4%. The percentage of "New" properties varies from 28.6% in North Jutland to 35.2% in Zealand, with regions like Zealand and the Capital showcasing newer and potentially pricier developments. Conversely, "Good" properties remain a strong part of the market in most regions, ranging from 32.5% in Zealand to 37.6% in North Jutland, reflecting steady demand for well-kept homes. This information sheds light on the diverse housing landscape in Denmark, where varying property conditions and prices mirror broader trends in development and market stability.

4.3.3 How does the distance to the city center and built years affect the prices? (Metehan Baydar)



Visualization3: Average Sale Price by Year Built and Distance to City Center

This research question examines how the year of construction of real estate and its distance from the city center affect real estate prices. It is observed that houses that are 20-30 km away from the city center attracted more attention between the years 1900-1920, and similar effects on prices are observed when approaching and moving away from the city center between these years. In the years 1920-1940, as in the years 1900-1920, houses that are 20-30 km away from the city center attract more attention, then houses that are 40-50, 10-20, 0-10, and finally 30-40 km away attract more attention. Between the years 1940-1960, unlike previous years, houses that were 30-40 km, 40-50 km, and 0-10 km away attracted great attention, then houses that were 10-20 and 20-30 km away, respectively, and a decrease in real estate prices is observed compared to previous years. It was observed that between the years 1960-1980, houses 10-30 km away attracted the most interest, then houses 30-50 km away, and houses closest to the city center attracted the least interest. Finally, between the years 1980-2000, it was observed that houses farthest from the city center attracted the most interest, and houses 10-20, 20-30, 30-40, and 0-10 km away attracted the most interest, respectively.

It can be deduced that the price of real estate is directly proportional to the interest they receive. There are many variables on real estate prices such as whether the infrastructure is modern or antique, but their effects are not examined and observed in this research question. In summary, it is observed that there is no direct proportion between people's interest in the city center over the years. The reason for using heat map visualization is to show three variables clearly and beautifully.

4.3.4 How do changes in the desirability of different property types over time affect their sale prices? (Cemile Bilgir)

To provide an answer to this question, it is necessary to examine the preference and price-based comparison of property types over time.

Examining the price-based comparison of the 4 different property types (farmhouse, detached house, apartment, terraced house) in the dataset with a general distribution visualization allows for accurate inferences. It is very efficient to use a box plot to evaluate statistical information such as minimum value, maximum value, quartiles, and median in a single graph and according to four different property types.

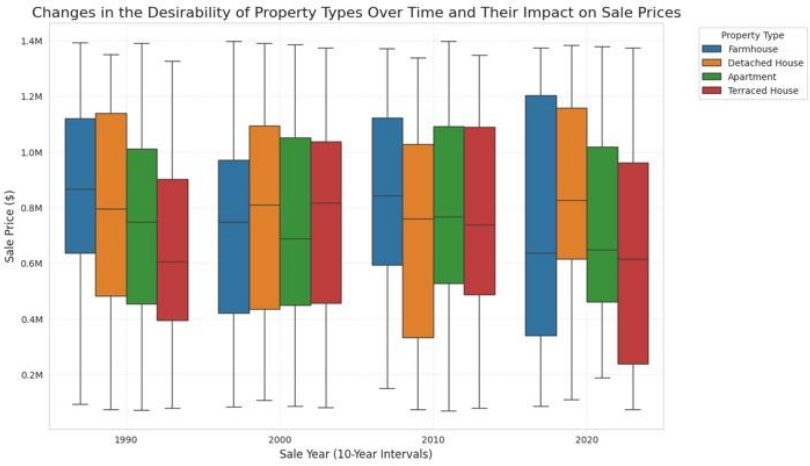
When the box plot (Visualization 4) in the figure is analyzed, it is seen that although the farmhouse had a high selling price compared to other types in the beginning (1990s), its selling price decreased significantly in the following years (1990s to 2020s). This may be due to the increasing urbanization of the world, the lack of preference for farmhouses, and the resulting decline in the price of farmhouses in line with the supply-demand balance. On the other hand, when we look at the interquartile range for farmhouses, we see that it has widened considerably in recent years (2020s), from which we can conclude that there is a lot of variation between prices. Also, compared to other types of real estate, the farmhouse has the highest price in recent years. In recent years, with the increase of farmhouse projects with luxury and more modern amenities, the gap between the prices of such luxury housing and ordinary farmhouses has grown significantly. The high cost of luxury projects has contributed to the widening of this difference, and it can be assumed that this has also caused the interquartile range to increase.

For detached houses, the plot shows that there is generally an average selling price for these houses. There have been only minor changes over the years. Even if the price range has increased or decreased over time, it has not increased or decreased significantly. As a result of these inferences, it can be assumed that detached houses have not lost their popularity over the years and therefore there has not been a major change in their selling prices.

In the analysis of apartment prices, we observe that the price range has remained almost stable over the years. On the other hand, even if the average price experienced a slight increase and fluctuation between the 1990s and 2010s, it decreased considerably in the 2020s. It can be argued that the reason for this is that urbanization increased a lot in the 2000s, but with the preference to return to nature over time, people do not prefer cities and therefore do not prefer apartments. This situation can also be considered as a factor in the decline in prices.

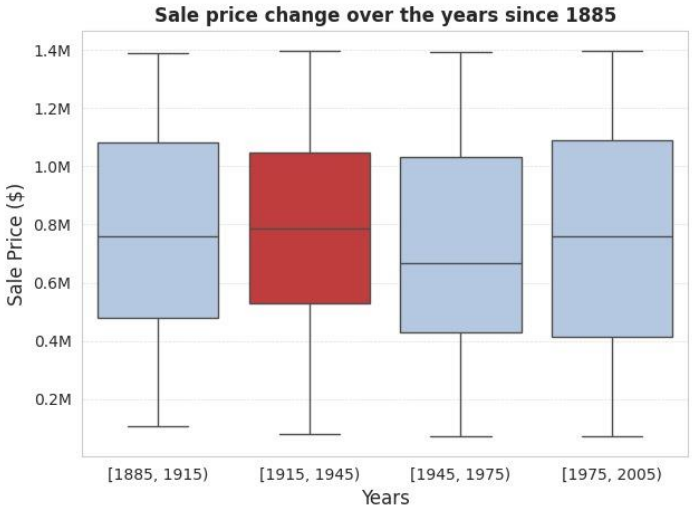
When we examine terraced houses, we infer that the price range of houses has increased in recent years. This increase may be due to urban transformation projects in the country. These projects aim to demolish old buildings and replace them with new buildings. Since many terraced houses are old, this reduces the number of available terraced houses, increases demand, and leads to price diversification. At the same time, the fact that terraced houses had the highest average sales in the 2000s compared to other years may be due to the widespread urbanization in those years, as was the case for apartment buildings. In the average price valuation, terraced houses had a lower average price at the beginning (the 1990s), maybe because people preferred larger houses with gardens at that time. Again, the average price increases in the 2010s due to urbanization and decreases in the 2020s due to migration from urban to rural areas in recent years.

To sum up, the preference of different property types over time directly affects their sale prices. Over the years, factors such as urbanization, and economic and demographic changes have influenced the demand for property types and thus their prices.



Visualization4: Changes in the Desirability of Property Types Over Time and Their Impact on Sale Prices

4.3.5 How have house prices in Denmark evolved since 1885, and what economic, social, and historical factors have influenced these changes? (Abdulsamet Demir)



Visualization5: Sale price change over the years since 1885

By exploring this question, one can get an idea of how housing prices in Denmark have evolved since 1885 and how various reasons, like societal changes or significant historical events, influenced them. By examining these trends, one can gain a deeper insight into housing market dynamics and historical events.

A simple box plot was used to answer this question. The boxplot has the year values from 1885 to 2005 on the x-axis and price values as U.S. Dollars on the y-axis. On the second box, a different color was used since it has a different behavior than others.

The first thing one can see is prices mostly follow a decreasing trend over the years except for the years between 1915 and 1945. Also, the median values do not follow any pattern in the year groups and the range is similar for all year groups.

The first group 1885-1915 shows Denmark's early industrial era. The price of houses looks competitive due to their historical charm and maybe even their durability. This suggests older, well-preserved homes can find customers who appreciate architectural heritage.

The second group, 1915-1945, tells a different story. The price values between 1915 and 1945 have a slightly bigger median and quantile 1 value. These years have seen both the world wars and the destruction they have caused. The houses that survived this period probably had higher construction standards and better quality; thus, they sold for a higher price than in other periods.

The years 1945 and 1975 show that after World War II, Denmark experienced an economic recovery and urban development period which led to better housing standards. The house prices are consistent, indicating practical post-war designs.

In the last period 1975-2005, house prices have a wider range, indicating the variety of modern techniques and materials. Also, the reason behind high prices in this period is probably the energy efficiency and contemporary designs of those constructions and not their historic value.

This analysis shows how factors such as historical value, post-war urbanization, and modern construction methods influence house pricing. Furthermore, this analysis indicates the price of a house is not only shaped by its age but also by its quality, design, and relevance to the needs of the customers. Examining these trends gives valuable insights into how the housing market in Denmark has evolved around historical and economic contexts.

4.3.6 How Do Average Property Prices Vary Over Time Between Properties with and Without Gardens? (Arif Emre Kesik)

To answer this question, the change in average prices in dollars over the years should be analyzed in two separate line plots for houses with and without a garden. The use of line plots is preferred since time-dependent changes will be analyzed and conclusions will be drawn. On the same y-axis, the left-hand side plot shows the price change of houses with gardens and the right-hand side plot shows the price change of houses without gardens over the years. To obtain a more readable and understandable display, sales years are grouped and analyzed at five-

year intervals. Also, the y-axis starts at $660,000 to get a more consistent view by indexing the minimum average price. The sum of the increases and decreases between 1992 and 2021 shows an increase in the price of houses with a garden and a decrease in the price of houses without a garden.

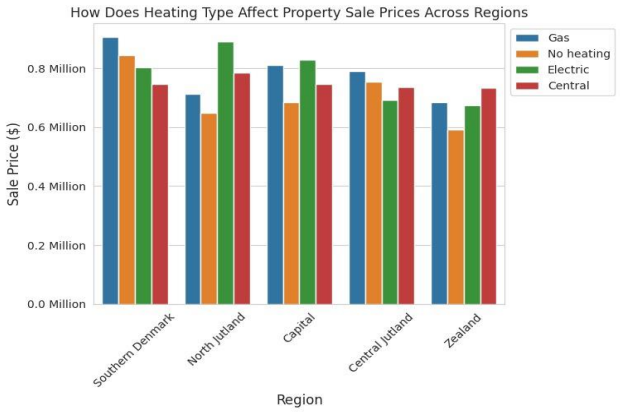
Between 1997 and 2011, there is a total increase in the price of houses with a garden. After a significant increase in the first 10-year interval, there was a slight decrease in the second 10-year interval, while the last 10 years showed a slower increase compared to the first 10 years. Conversely houses without a garden have shown a steady decrease and the rate of decrease has shown a steady increase.

In the last two 10-year intervals, the situation experienced in the houses without a garden above is seen for the houses with a garden and shows a continuous decrease in these dates, and it can also be seen that the rate of decrease is higher in the last 10-year interval than in the previous 10 years. In houses without a garden, it is seen that the average price increased significantly in the first 10 years and decreased again in the last 10-year period.

As a result, the average price of houses with gardens has increased with fluctuations, while houses without gardens have decreased. In some year intervals, both have increased and decreased, while in some year intervals, one has increased while the other has increased. This shows that the demand and price of a house does not only depend on whether it has a garden or not, but that the price change depends on other factors.

Visualization6: Average Prices Over the Years for Properties with/without Garden

4.3.7 How Does Heating Type Affect Property Sale Prices Across Regions? (Yiğit Efe Özdemir)



Visualization7: How Does Heating Type Affect Property Sale Prices Across Regions

A clustered bar chart was used for this research question because it is an effective way to compare heating types based on the location of the houses.

In Southern Denmark, houses without heating are found to be the second most expensive. This could be because the warmer weather in the south makes heating less necessary, increasing the preference for these houses. Also, the gas heating type is found to be the most expensive and the central heating type is found to be the cheapest.

In North Jutland, houses with electric heating are observed to be the most expensive, while houses without heating are the cheapest. This might be because the colder weather in the north causes houses without heating to be less preferred, thereby lowering their prices.

In the Capital Region, differences in prices among heating types are relatively small. However, houses with electric heating are the most expensive and no heating type is the cheapest. This might show that electric heating is the primary choice in this region.

In Central Jutland, the highest average prices are seen in houses with gas heating, and the lowest average prices are seen in houses with electric heating. This might indicate that gas heating is most preferred in this region,

In Zealand, no significant difference in sale prices among heating types is observed. However, houses with central heating are noted to have the highest average prices, while houses without heating have the lowest. This might show that central heating is the preferred heating type in this region.

Additionally, when average prices are compared across all regions, it is observed that Southern Denmark and North Jutland have the most expensive houses, while Zealand has the least expensive ones. This situation might be due to variations in economic activity, demand, and housing supply across these regions.

5.Conclusion

Housing prices in Denmark are shaped by a dynamic interplay of factors, including garden presence, energy efficiency, property conditions, and proximity to city centers. Over time, houses with gardens have been appreciated, while those without them have declined. Energy-efficient properties, especially high-efficiency ones, have seen rising demand, driven by environmental awareness and government policies. Regional differences in property conditions and heating preferences reflect local economic and climatic factors. Historical events, such as World Wars and post-war recovery, have influenced price trends, while modern construction techniques and urbanization have shaped recent patterns. These findings highlight the multifaceted nature of Denmark's housing market, shaped by societal, environmental, and historical forces.

References

1. Danish Energy Agency, *Energy in Denmark 2021 (Internet edition)*. Danish Energy Agency, 2023. [Online]. Available: https://www.ens.dk/en/analyses-and-statistics/annual-and-monthly-statistics. [Accessed: Jan. 17, 2025].

2. R. H. Petersen, "Strategy for energy renovation of buildings: The route to energy-efficient buildings in tomorrow's Denmark," Danish Government, 2014. [Online]. Available: https://energy.ec.europa.eu/document/download/2e31c276-3562-4f85-a8bc-dd84c7a4c6bd_en?filename=2014_article4_en_denmark.pdf. [Accessed: Jan. 17, 2025].