

DAT565/DIT407 Assignment 2

Vaibhav Talari
talari@chalmers.se

Stefán Ólafur Ingimarsson
stefanla@chalmers.se

2024-09-05

This report is submitted by group 25 for **assignment two** in *Introduction to Data Science & AI*.

1 Problem 1: Scraping House Price

We used the BeautifulSoup library in order to parse all the 40 HTML files and output the *agg_result.csv* file which contains all the extracted information in readable .csv format.

2 Problem 2: Analyzing 2022 House Sales

- The five-number summary is as follows:
 - Minimum: SEK 1,650,000.
 - Maximum: SEK 10,500,000.
 - Median: SEK 5,000,000.
 - First Quartile: SEK 4,012,500.
 - Third Quartile: SEK 5,795,000.
- Refer to figure 1 for the histogram showing the closing prices of houses in the year 2022. The number of bins is calculated as the square root over the total number of data points [1]. We use 13 bins.
- Refer to figure 2 for the scatter plot showing the relation between closing prices (in log scale) against boarea in meter square.
- Refer to figure 3 for the scatter plot showing the relation between closing prices (in log scale) against boarea in meter square with an added dimension of rooms.
- *Discuss shortly what the figures tell about the prices and the relationship between the different variables.*

After plotting the data, the figures showed that the closing price increased as the living area increased. There were exceptions, as some of the properties had a higher closing price than other properties that had the same amount of living area. An example of this would be Långdammsvägen 2,

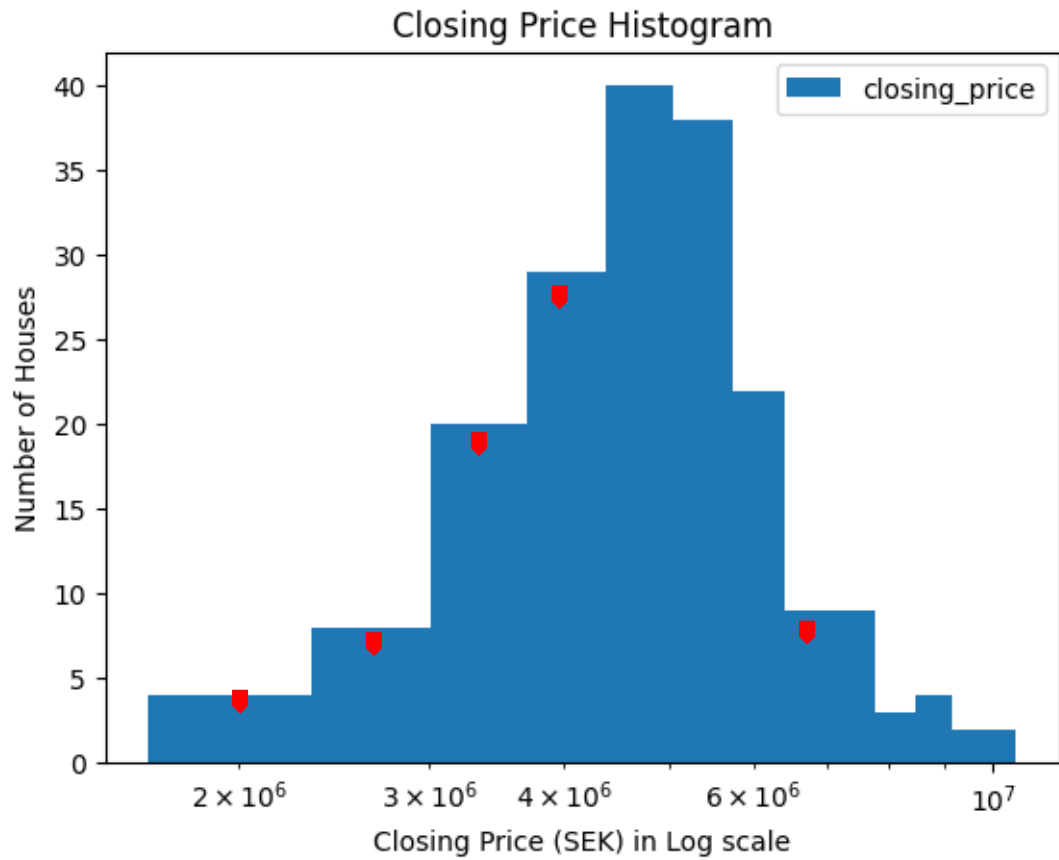


Figure 1: Histogram of Closing Prices of Houses in the Year 2022

Marstrandsön, Kungälv kommun. This property only had a living area of 28.0 m^2 , yet it had a closing price of 7,125,000 SEK. Marstrandsön being one of the more highly desirable area in Kungälv due to it being an historic and exclusive area[2], among other factors. This most likely drives up the price. Since the demand is so high, people are willing to pay more money for less living space. The plots indicated that there were a few other properties that were sold for a larger amount of money despite not having much living area.

Overall, the price increased as the living area increased, with exceptions for exclusive properties such as the one previously mentioned.

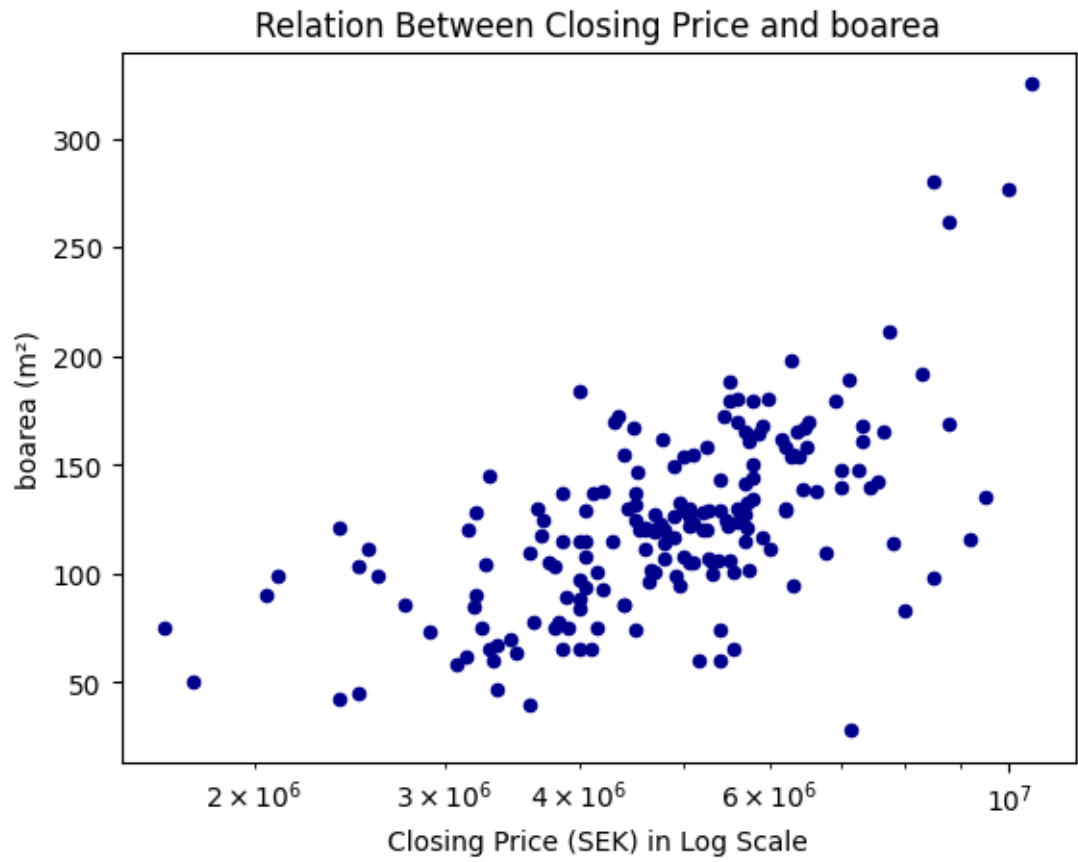


Figure 2: Scatter Plot of Closing Prices against boarea (meter square) in the Year 2022

References

- [1] Wikipedia contributors. *Histogram*. Accessed: 2024-11-16. 2024. URL: https://en.wikipedia.org/wiki/Histogram#Number_of_bins_and_width.
- [2] unknown. "About Marstrand". In: (2019). URL: <https://www.marstrand.se/en/about-marstrand/>.

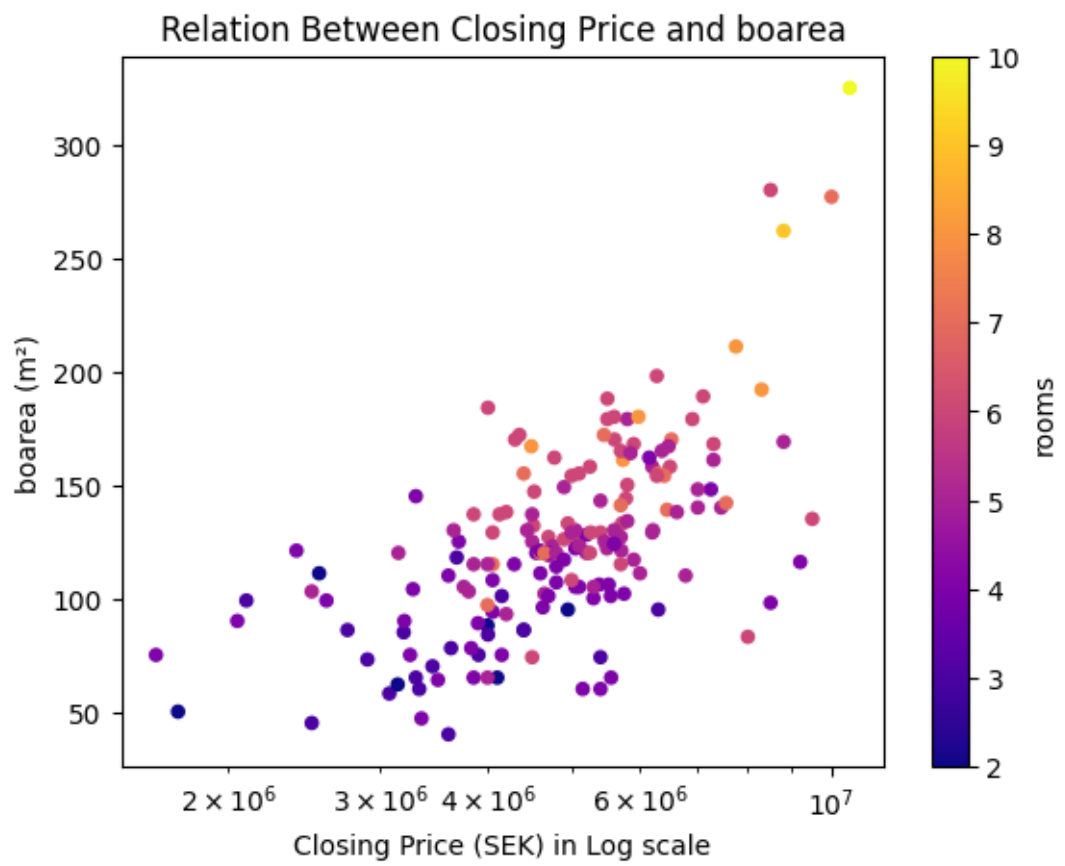


Figure 3: Scatter Plot of Closing Prices against boarea (meter square) in the Year 2022 with changing colours by room