

## Research motivation (10%)

*The research question/business problem is clearly articulated and important. The choice for the research method (e.g., regression analysis) is motivated well. The way of deployment (e.g., PDF report, dashboard, ...) is useful and accessible to potential knowledge users, and clearly communicates the conclusions of the analysis. The workflow is of potential use to other students and the larger scientific community.*

The dataset is created by Team 2 of the course Online Data Collection and Management (oDCM), on behalf of Tilburg University. There was no financial funding involved in the creation of this dataset.

**Research motivation:** The housing shortage in the Netherlands in the last few years has led to a serious housing crisis experienced today. Especially young people are encountering major difficulties in finding the right place for an affordable price. Also during the last political party elections, the housing crisis was a very much debated topic. For this reason, Team 2 realized the significance of creating a dataset that can give both house seekers and house sellers more insights into the housing market in the Netherlands. Mainly, the aim of creating this dataset is to make a house's selling duration more predictable. For both parties this can lead to a more efficient housing market: house sellers can better predict for what period their house will be on sale, whilst house seekers get insights into the selling rate in a particular city. On top of historical selling prices and selling durations, the dataset includes all other information that was initially provided on the house product page. This means that information about a house's roof type, number of floors, energy label or even location of garden is also included. This can give house seekers a general idea of what a house with their preferences could cost.

**Research question:** How does the selling price of a house affect the selling duration?

**Research method:** A regression analysis was performed, using the following variables:

Dependent variable = selling duration

Independent variable = selling price

Control variable = city

Performing a regression analysis helps to predict the relationship between the independent and dependent variable. In this case: the effect of the selling price of the house on the selling duration. This is shown through the following equation:

$$\text{Selling duration} = 6.38420686945145 + 0.00012867853603877 * \text{Price}$$

The regression coefficient (0.00012867853603877) depicts the change in selling duration for one unit change in the selling price. This is a very small, yet positive effect and can also be seen in the following scatter plot:

(link to scatter plot)

The resulting p-value is

[illegible]

which shows that the effect is significant, because the value is lower than our (SIGNIFICANCE LEVEL).

Additionally, a Shiny app is produced with two sliders for the variables construction year and selling price. The app can be opened by clicking on the following URL: <https://wouter1997.shinyapps.io/averagesellingduration/>. Construction year is a variable that was left out of the final regression analysis because it showed an incredibly weak relationship. However, it is included in the Shiny app because it allows for the selection of a subset of houses. For example, the dataset of Eindhoven that was created on March 7, 2021 consists of 2,187(?) entities. When people use the app and are interested in a particular price, they can choose themselves whether they want to incorporate all listed houses or limit to only recently built houses for example.

**Conclusion:** The regression analysis, scatter plot and Shiny App show the small, yet positive relationship between selling price and selling durations of houses on the market. This is of significance to both house sellers and house seekers, in terms of higher predictability of the Dutch housing market.