

Data sources

In this section the data used in this report will be introduced.

Boliga

Data was collected from Boliga.dk which is one of the largest, independent web-portal for real-estate sales in Denmark. It has 1.1 million active views each month (boligagruppen.dk). The data scraped from Boliga provides unique insight into each real-estate posted on their website, such as pricing, m², municipality, coordinates, and access to the BBR - the Danish Building and Housing Register, which contains information about all Danish properties.

Additional features

As we want to predict the pricing of real-estates in the current market, we want to add additional features, which potentially could increase the predicting power of our model.

hvorlangterder.dk is a find-closest demo developed by Viamap, a digital map provider. This API contains information about distances to distances from an address to conveniences such as hospitals, schools and shopping. Obtaining these distances for all of the listed housing and merging the two datasets, gives additional arguments to geographical-control. The data from hvorlangterder.dk can allow more micro-oriented differences within a given municipality. Which otherwise would not have been captured.

Socioeconomic data on municipality level is collected from **statistikbanken.dk** & **statistik.politi.dk**. These factors include measures for reported crime, socioeconomic index, mean income and level of completed education, etc. These measures are thought to influence the pricing of real estate in a given municipality.