Introduction to Social Data Science – Project Description

1. Group number?

Group 5

2. Names of students in group?

Moritz Bilstein (wmj863), Daniel Thorsten Bruns (mws487)

3. What is your research question?

What determines housing prices in the greater Copenhagen area and how does the performance of a machine learning approach compare to the models in the literature?

4. What kind of data are you planning on using? How will you get access to these data?

We utilize an API to access housing information from the website https://www.boligsiden.dk. The final dataset contains various housing characteristics as well as location data. We combine this with neighborhood characteristics such as public school ratings that have been established as important determinants by the literature.

5. What will your data analysis be like? Will you use machine learning? How?

We visualize the areal differences in housing prices for different housing types in heat maps, where we use Open Street Maps as our primary source. Our model uses a linear regression approach with regularization to predict housing prices.

6. Have you already identified other papers within this area that you can use in a literature review? If so, name a few and explain what they do in one sentence only.

There is extensive literature on the prediction of housing prices using a variety of different methods including machine learning algorithms. We will relate our paper to the most relevant research to our empirical strategy.

7. How do you 'contribute' to the literature?

We apply a prediction model to the local context of Denmark using highly up-to-date data.