Predicting Housing Prices in Denmark
A Machine Learning Approach

Group 23: He SHI, Mingjie XING, Qi ZHANG
Department of Economics, University of Copenhagen Summer 2023

August 15, 2023

1. What is your research question?

   This project is intended to build a predictor of residential housing prices in cities in Denmark with a machine learning hedonic model based on interior features such as house area and location, and exterior features such as income level of the neighbourhood, weather condition and cities' economic development level. The project mainly covers major cities in Denmark, e.g., Copenhagen and Aarhus. The frequency of the data will be monthly. The trained model based on 2023 data will be tested on 2022 and 2021 data.

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

   - House Prices Data:

     We are going to scrape the individual level house prices from boligsiden, a major real estate broker in Denmark covering housing prices as monthly data of 36 months scale and detailed features of houses in different regions with an open data access. We will scrape the data from `https://www.boligsiden.dk/`. Structural features in data includes area, location, age and owner income. Web-scraping of the data is allowed for academic use.

   - Income Data:

     Pay level distribution by region can be found on STATISTICS DEN-

MARK under the Pay Level subsection of Income and Earnings. We are not going to carry out time series analysis due to the constraint of available price data.

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

   We are going to leverage a

   We are going to apply the polynomial regression on the data to obtain the coefficient of deciding factors on house price in different cities in Denmark and in different temporal sphere.

   We will handle the under-fitting and over-fitting problem when applying the machine learning model to get a more precise version of the house price prediction by leveraging LASSO/ Ridge regression.

4. 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 are several papers recently on top journals in real estate economics predicting housing prices with machine learning models. **?** applies non-parametric machine learning model with k-fold cross-validation to examine the U.S. commercial real estate appraisal to shrink the deviations between market values and subsequent transaction prices. **?** augments the traditional hedonic model with Gradient Boosting Machine algorithm to better predict the housing prices in Beijing.

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

   The main contribution of this project includes: 1) creating a comprehensive dataset on Denmark residential real estate market; 2)