Scraping data from Boliga.dk

Boliga.dk is an independent Danish property portal for real estate in Denmark. Boliga.dk is not a real estate agency (although it does provide the “Boliga Self Sale” ("Boliga Selvsalg") service, which allows property owners to list and sell their homes directly, without involving traditional real estate agents) but aggregates data from various sources, including both real estate agencies and private sellers, to provide comprehensive property listings and market information. Priserne hentes fra tinglysningen! While the specific list of real estate agencies that Boliga.dk collects data from is not publicly disclosed, the platform aims to provide a wide and diverse range of property listings from across Denmark.

Boliga.dk also offers a "Sold Properties" service that provides users with access to historical sales data and information about properties that have been sold in Denmark. These data go back to 1992.

Boliga.dk was used for gathering data about sold properties in Denmark in the period from 1992 to 2022 (both included). It was decided to focus on the categories villa (“villa”), townhouse (“rækkehus”), and condominium (“ejerlejlighed”). Thus, excluding the categories vacation house (“fritidshus”) and country estate (“landejendom”). These exclusions were made based on the assumption, that 1) vacation houses are often bought by people living in other municipalities making it impossible to base house prices on municipality based features such as income, and 2) that amount of land used for agricultural or farming purposes impacts the price of country estates in a way that would make for a skewed comparison on square meter prices with traditional houses.

Denmark consists of 98 municipalities, and since Boliga.dk allows searching sold listings based on these geographical features, 98 individual web scrapes were performed. Each scrape consisting of all sold listings from 1992 to 2022 in the aforementioned categories. For each municipality the amount of sold listings varied from 7613 (Albertslund) to 297.624 (Copenhagen) making it a total of XXXX sold listings for the 98 municipalities combined.

As Boliga.dk only shows 50 listings per page, a loop was created that scrolled through every page in the search results. The Python Pandas library was used for harvesting the HTML tables on each page, append them to a Pandas DataFrame, and then save the result in a CSV-file.