Pandas objects can be combined in different ways according to the nature of features in the dataset. Relational database style operations are based on linking keys together, thereby maintaining the relationship between the combined datasets (McKinney, 2018, s. 231).

The data collection and scraping process provided a total of 15 Datasets from Boliga, Hvorlangterder, DAWA, the Danish Police, Social og Indenrigsministeriet and Danmarks Statistik. This section will describe how each dataset was merged, using pandas relational database styled *merge* and *join* operations.

**Boliga.dk**

An evaluation of the Boliga datasets features and their ability to support further data collection led to the utilization of the following features:

* [Longitude, Latitude]: Geographical placements of the properties
* Municipality: a numeric code for municipalities in Denmark

The Boliga dataset acts as a master dataset and was joined or merged upon throughout the process, and acts as the “left” of all operations. The Longitude and Latitude features serve as specific coordinates for valued properties but are in few cases repetitive regarding different apartments from the same complex. The municipality code was used for translation purposes between the master dataset and other datasets.

**DAWA**

The DAWA dataset was scraped with an input of a distinct municipality code from the master dataset, returning a dataframe of municipality names for each code. This dataframe was merged onto the Boliga dataframe as a many-to-one merge with municipality code as the merge keys. The scraped municipality names from DAWA was used as the merge key for all further merges of municipal data.

**Danmarks Statistik, Danish Police & Social og Indenrigsministeriet**

Of the municipality-based data collected, the merging of these to our master dataframe was subsequently performed identically. Every dataset contained a column for unique municipality names and values for the given feature of the dataset. A many-to-one left merge was performed with the master dataframe, using municipality names. The result being a master dataframe containing municipal specific features for properties.

The Danish Police datasets contained totals for different categories of crimes reported within municipalities. These datasets were outer joined with an index set to contain municipality names. Afterwards, the values were added to each other, providing a total of reported crimes within each municipality. The police statistics webpage does not specify whether different types of reported crimes can relate to a single case, but we assessed that the number of crimes reported provides a meaningful feature either way. The dataset containing total reported crimes per municipality, was merged with the master dataset in the same fashion as previous municipality-based features.

**Hvorlangterder.dk**

The hvorlangterder scrape provided location specific features, taking an input of Latitude and Longitude coordinates. The scraping function, which was created for returning location-based features, created a column for the row specific values. Therefore, a merge operation was not necessary, but could alternatively have been managed with a one-to-one inner-merge id. As few coordinates regarding apartment properties are repetitive, these are not suitable for merging upon.

# Bibliography

McKinney, W. (2018). Data Wrangling: Join, Combine and Reshape. In W. McKinney, *Python for Data Analysis* (pp. 231-246). Sebastopol: O'Reilly Media Inc.