# Applied Data Analysis Capstone Project
# Real Estate in Copenhagen

Gesine Wanke

July 7, 2020

## 1 Introduction/Business Problem

The following project regards the question of finding a place to live in the city of Copenhagen. The analysis is made for customers searching for a place to buy in Copenhagen either as a home or an investment. It will analyse the housing prices of Copenhagens city districts, as well as the average annual house price increase of the last 5 years as coropleth maps. Additionally, it will use the Fourscquare location data to analyse the neighborhoods around the public transport system in Copenhagen according to their venues. This analysis will make it easy for costumers to define an area of the search for their future home or real estate investment.

## 2 Data

The Data for the analysis is available from serveral sources. The .geojson files for the districts of Copenhagen are retrieved from opendata.dk (`https://www.opendata.dk/city-of-copenhagen/bydele`) and are downloaded within the python script. These are used to define the borders of the coropeth map for each city district.
The housing prices and housing price history are available for each ity district of Copenhagen from Boliga's hompeage(`https://www.boliga.dk/boligpriser`), but can not scraped directly. The relevant data is copied to a .csv-file to be read into a pandas data frame from the git-hub repository.
For the analysis of the neighborhoods in the districts and their venues the metro-system of Copenhagen is used. The metro stations are extracted from wikipedia (`https://en.wikipedia.org/wiki/List_of_Copenhagen_Metro_stations`). The links to each wikipedia entry are used to extract the location data for the metro stations. The Fourthsqare data base is used to extract a list of venues around 2km within a radius of each metro station. This distance is chosen as biking and public transport are very common means of transportation for Copenhageners.