

Applied Data Analysis Capstone Project

Real Estate in Copenhagen

Gesine Wanke

July 7, 2020

1 Introduction/Business Problem

The following project regards the problem of finding a place to live in the city of Copenhagen. The place to live should be in an attractive neighborhood, affordable for the customer and the value of the property should increase over time.

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 Copenhagen's city districts, as well as the average annual house price increase of the last 5 years as coropleth maps. Additionally, it will use the Foursquare location data to analyse the neighborhoods around the public transport system in Copenhagen according to their venues. This analysis will make it easy for customers 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 several sources.

2.1 Copenhagen district borders

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. The file contains the borders as latitude and longitude for each district. These are used to define the borders of the coropleth map.

2.2 Copenhagen real estate prices

The housing prices and housing price history are available for each city district of Copenhagen from Boliga's homepage (<https://www.boliga.dk/boligpriser>). The extracted data is the price per square meter of an apartment over the last 5 years. The data can not be scraped directly but the relevant data is copied to a .csv-file to be read into a pandas data frame from the git-hub repository. The file contains the name of the district, as well as the price per square meter.

2.3 Copenhagen neighborhoods

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 a wikipedia-list(https://en.wikipedia.org/wiki/List_of_Copenhagen_Metro_stations). From the list the station name, as well as the link to the wikipedia entry of each respective station is extracted. The links to each wikipedia entry are used to

extract the location data for the metro stations from the pages. The Fourthsquare 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.