

# Coursera Capstone Project

*Where to move in Mannheim, Germany*

By: Tobias Höfchen

April 2021



Source : <https://www.gruenderkueche.de/gruenden-in-mannheim/attachment/mannheim-wasserturm-panorama-baden-wuerttemberg-deutschland/>

## Table of contents

|                            |   |
|----------------------------|---|
| Abstract .....             | 3 |
| Introduction .....         | 3 |
| Data .....                 | 3 |
| Methodology .....          | 3 |
| • Webscraping.....         | 3 |
| • Data wrangling.....      | 3 |
| • API calls.....           | 3 |
| • K-Means Clustering ..... | 3 |
| • Data visualization.....  | 4 |

## Abstract

If you move to another city, you always have to decide to which part of the city you want to move. In this project i will recommend districts in mannheim based on a preceeded data analysis. This analysis includes e.g. cost of living per district, distance to public transport and distance to supermarkets. The analysis will be performed by using different data science techniques and additional auxiliary techniques like webscraping.

## Introduction to the problem

When you plan to move to another city you have to carry out a lot of decisions. One with outstanding importance is the decision where to rent/buy a flat/house. This decision has a major impact on your experience of the new city and your overall feeling. Just imagine you live far away from the next supermarket and public transport is far away plus you haven't got a car.

This project tries to help people to avoid this situation by analyzing the different districts of mannheim. Therefore, factors like the accessibility of places you need for all day living (e.g. supermarkets) and the overall cost of housing in the different districts will be taken into account.

## Data

The most important aspect of every data science project is data. If the data quality is bad no algorithm can produce good output. Hence, in this project the following data is used to get a good overview over the local circumstances:

- List of districts of Mannheim. Only the districts from this list will be analyzed. The data will be extracted by webscraping from [https://de.wikipedia.org/wiki/Liste\\_der\\_Stadtbezirke\\_und\\_Stadtteile\\_von\\_Mannheim](https://de.wikipedia.org/wiki/Liste_der_Stadtbezirke_und_Stadtteile_von_Mannheim)
- Geospatial data (Latitude, Longitude coordinates). These are needed to visualize the results and to assign the venue data to the districts.
- Venue data. This data will be fetched via foursquare.com API.
- Rental prices per squaremeter will be fetched from <https://www.wohnungsboerse.net/mietspiegel-Mannheim/586> by webscraping the page.

## Methodology

The mentioned data will be analyzed by applying different data science techniques. They can be separated into the following classes:

- **Webscraping:** Webscraping is basically extracting information from a given html by downloading the sourcecode of the page. In this project especially the package requests and beautifulsoup will be used.
- **Data wrangling:** This techniques simply describes aggregating and cleaning the data.
- **API calls:** Some websites offer APIs to interact with the underlying database.
- **K-Means Clustering:** This algorithm basically tries to find groups with similarities in given data by calculating their distance.

- **Data visualization:** Every result has to be visualized in order to be easier understood. Therefore, packages like matplotlib or folium will be used.