Finding Where to Live in a New City:

A Case Study for the city of Hamburg, Germany

Mihael Machado de Souza

1. INTRODUCTION

Finding a place to live is always a challenge. When looking for a new home, to purchase or rent, you must consider whether the rent price is okay, the quality of the house/apartment, how much renovating it might need, whether it is already furnished or not. You also need to take into account the neighborhood the apartment is in and what type of amenities it offers. This can also highly depend on one's own family structure. If young children are involved, does it have good schools nearby? If you don't have your own vehicle, does it have good access to public transport? What about supermarkets? Or restaurants? How noisy is the neighborhood? These are just some of the questions we have to think about when choosing the next place to call home.

Now, on top of all of that, assume you're also moving to a city you now nothing about. You got this great opportunity, but it is in a whole different state/country. You have no family or connections there, so you can't just ask for suggestions. Google is useful, but you first have to know what to search for. As an example, Hamburg has 104 quarters. Where to start? General searches, such as "Where to live in Hamburg?", will find hundreds of pages worth of information. But without some firsthand knowledge, how to separate the good advice from the bad? Or to know if whoever wrote the advice has similar priorities to yours?

In those circumstances, a data driven approach can help you jumpstart your search the right way, e.g. using a recommender system. On an individual level, if you can cross-reference what types of businesses or amenities are available in each neighborhood with their average rent prices, how safe they are and how far they are from your work place, for example, you can direct your searches towards a handful of preferred locations as opposed to the whole city. And you would know, in advance, that those neighborhoods/quarters are suited for your own personal preferences. On a business level, realtors or real state agencies could develop short questionnaires to probe a client's preferences and use that to get tailored recommendations for each potential client. City administration, itself, could add such questionnaires to their own websites and produce these recommendations for anyone looking to move into the city.

2. DATA

2.1 Data Sources

To build our content-based recommender system, we first have to gather the data that we will use to calculate the weights for our content matrix. For this proof of concept, we'll leverage data from three sources: Foursquare (https://pt.foursquare.com/), ImmobilienScout24 (http://immobilienscout24.de/) and BingMaps (https://www.bing.com/maps). Our quarters from Hamburg, with their MultiPolygon bounding boxes, are obtained from the geojson file made available in https://github.com/mihaelsouza/neighborhoods. We define the central coordinate for each quarter as the average latitude and longitude within each bounding polygon.

With Foursquare, we will gather data from all venues available within 500 meters of a quarters' central coordinate. To build our recommender system, we will then categorize these venues in some broad categories. For example, if we compared our application to a movie recommender system, these would be the genres associated with each film. For Foursquare, they already provide a top-level category system that encompass their set of unique venue labels (https://developer.foursquare.com/docs/build-with-foursquare/categories/). As a first step, we will use these 10 categories. They are:

- Arts & Entertainment → Group venues like Museums, Theaters and Stadiums.
- College & University → Group different university and higher education buildings.
- Event → Location of general seasonal events, e.g. Christmas Markets and Parades.
- Food → Groups Restaurants, Diners and Coffee Shops, for example.
- Nightlife Spot → Groups Bars, Pubs and Night Clubs.
- Outdoors & Recreation → Green areas (e.g. Parks) and Athletics venues (e.g. Gym).
- Professional & Other Places → Business Services and Government Buildings.
- Residence → Private Homes, Housing Developments and Assisted Living.
- Shop & Service → General service providers (e.g. ATM) and stores (e.g. Malls).
- Travel & Transport → General travel (e.g. Hotels), rentals (e.g. Cars) and public transport.

Aside from venue information, we also need rent prices across all neighborhoods. On a more general manner, this could be obtained from state reports. We would like, however, this information to be as current as possible. In that case, we will search for all available homes for rent in the ImmobilienScout24 website. Their search result page contains the listing's address, price (in € per month) and size (in m²). We will scrape this page and gather data from all listings, and then normalize the price based on the listing's size to get a price per square meter for each neighborhood.

Finally, we will use BingMaps' API to calculate the travel distance between each quarter central coordinate and the address provided by the potential client. This address would be, for example, of his/her new work address. We will use this information to calculate a distance index between each quarter and the place of work and consider that into our recommender system as well. If no address is given, we can simply set the index to zero and not take it into account when creating our recommendation.

Together, information from the three data sources mentioned above will form the content matrix for our recommender system. Then, we will create a set of questions that addresses each of the categories defined to gather the client's perspective/preferences according to each category. This will establish our user profile that, in combination with the content matrix, will allow us to obtain neighborhood/quarter recommendation tailored for each client.