Neighborhood selection for relocation

1. Introduction/Background:

A 2019 study by the US Census Bureau tracking the migration within USA showed that approximately 32 million moved to a different apartment or home in 2018. This corresponds to the relocation of approximately 10% of the US population. People move for a multitude of reasons - the top three reasons being bigger/cheaper house, better neighborhood and better job prospects. The latter of these - the job-related relocation industry, is worth approximately $25 billion itself! Relocation expense is a sizeable chunk of the job offer and it is estimated that on an average, firms spend approximately $16 million every year on employee relocation.

While the process of choosing the next apartment or house can feel quite daunting given the different variables at play, it becomes more taxing when it's a job-related relocation. The employers provide the new hires ~2 weeks to accept a job offer and then an additional 4 weeks to report for the new role. To manage the logistics of moving out from a known location to another neighborhood or state in such a short time span can become daunting. Not having enough time to scope out the new location, most of the employees rely on online research for choosing the right neighborhood. As most of this information is available online through different sources, scraping the web for information and using data science techniques to process the data can speed up neighborhood research. This would help the employee use the available time effectively and reduce the strain of the relocation process.

Business Problem:

While neighborhood selection is usually a personal choice, this assignment focusses on the following two factors:

- Safety - as measured by the crime statistics per 1000 people of the city and
- Facilities such as schools, grocery stores, gyms, parks etc.

This project aims to classify neighborhoods based on these factors. To highlight the diversity of neighborhoods, Minnesota state is chosen as it has a mix of big and small cities.

2. Data Sources:

Information about the list of cities and counties in Minnesota and their population is obtained from Wikipedia (here). Crime statistics information was harder to get as the most up to date databases required a paid account. An article listing one such report was used to scrape the crime stats information (here). As crime statistics were not available for all the cities in Minnesota, the data was curtailed to include only those cities with a complete set of information.

The spatial location data (latitude, longitude) was obtained using the Geocoder package and added to the table. Neighborhood facilities within a 1-mile radius were searched for each city using Foursquare API and neighborhoods were classified according to their categories.

This combined dataset was used for clustering neighborhoods by K-means classification method.