# Finding a 4 Year Home

Joseph Kilroy

August 2018

Coursera Applied Data Science Capstone

## 1. Introduction

### 1.1 Background

Every year students go off to college where they will hopefully spend the next four years of their life. While this is the case for most students some spend one semester at an institution and return home because they were not happy. There are many reasons for this but I common reason is that they did not like the area. It could be anything from a location being too rural, too crowded, or not enough to do. I spent some time as a college counselor and would teach a course called college and career planning. During this course I would have students look at four categories while choosing schools. These were Admission Requirements, Academics, Cost, and Campus Life. I include Campus Life because this will be the next four years of their life. Academics should be the main focus of their life but if they are unhappy it will make it hard to stay motivated and succeed.

### 1.2 Problem

Students will spend four years getting a degree. Ideally, they will spend time researching the university and the area around it to make sure it is a good fit. Most colleges have plenty of on campus activities and venues but sometimes students need to get away from campus. Maybe it is going to a nice pizza joint with classmates after a big exam or taking the person you meet in Biology class to the movies. Some students like to work out at different gyms or do CrossFit. These could all be challenging if you pick a campus that does not have venues nearby that you are interested in. The goal of this project is to match students with a college that has similar venues as where they live. I will use Waukesha Wisconsin as the student's home base and use the four-year universities in Wisconsin as the schools of choice.

### 1.3 Interest

There are many people who might be interested in a project like this. Students, Teachers, and College Counselors would all be interested in using this project to help students find a university. People looking to move for a new job but like where they currently live could use this to find neighborhoods like where they currently live. Anyone who is looking at moving locations could use this to find a place to live that matches what they enjoy.

## 2. Data

### 2.1 Data Sources Prep

To begin I needed to collect some data about each school. I used the following Wikipedia site to get a list of all post-secondary schools in the state of Wisconsin. https://en.wikipedia.org/wiki/List_of_colleges_and_universities_in_Wisconsin. From this site I created a table with name and type of each school in the state of Wisconsin. Because I am only focusing on four-year universities I went through the list and deleted any college that was a tech school, specialty school, or a two-year college. I then went to each school website to get the address of the school. Once I had the address, I entered it into google to get the longitude and latitude of each school. Here is a copy of the CSV with the first 5 schools.

| School | Address | Location | Type | Latitude | Longitude |
|---|---|---|---|---|---|
| Alverno College | 3400 S 43rd St | Milwaukee | Master's university | 42.983 | -87.967 |
| Beloit College | 700 College St | Beloit | Baccalaureate college | 42.503 | -89.031 |
| Cardinal Stritch University | 6801 N Yates Rd | Milwaukee | Doctoral/research university | 43.142 | -87.906 |
| Carroll University | 100 N East Ave | Waukesha | Master's university | 43.024 | -88.221 |
| Carthage College | 2001 Alford Park Dr | Kenosha | Baccalaureate college | 42.622 | -89.822 |

The Latitude and Longitude is to get the venues from Foursquare. The other columns are to give students more information on the university they were matched with. Then I gathered the same data for Waukesha Wisconsin.

### 2.2 Data Program

In this stage of the project I used the latitude and longitude for each school to make a call to the foursquare and get a list of all venues within 3200 meters of the coordinates. I choose this distance because I figured it is a reasonable distance for someone without a car to travel. This data will be used to do a K-means cluster with Waukesha Wisconsin to determine the schools that are in the same cluster as Waukesha. Then I will create a data frame with limited venues to see if we get different results. I am doing the second test because I do not want to be matched based off venues that are not visited. For example, someone who does not own a pet would not want to be matched based off the number of pet stores.