# **Location for a new Donut Shop in Madison, WI**

#### **Jesse Griffiths**

## **Introduction/Business Problem**

The right location is crucial to the success of any venue. Here we'll determine the best location for an aspiring donut shop owner to open shop in Madison, WI. To do this we'll segment the city into distinct regions and examine the prevalence of different venue types in those regions. Then, we'll find a region that is similar to the region with the highest rate of donut shops/bakeries, but doesn't have a high rate itself.

The idea is that venues that do well in one region should also do well in a similar region. We then want to make sure that the donut shop is not in a region that already has many because we want to avoid competition as much as possible.

## **Data**

We'll be using Foursquare location data for Madison, WI. We'll use the coordinates of each location to calculate the geographic distances between clusters of venues, and segment the city into regions accordingly. For example, this is how we would differentiate between downtown and campus. Then, we'll examine venue types contained in each region to determine the similarity between regions. For example, two regions that have a lot of coffee shops and Indian restaurants would likely be more similar than a region that doesn't have either.

## **Methodology**

Here's the general outline of what we'll be doing with the data.

1. Segment Madison, WI into regions based on venue locations
2. Determine frequency of venue types for each region
3. Determine similarity of regions based on venue types to find the best region for a new donut shop

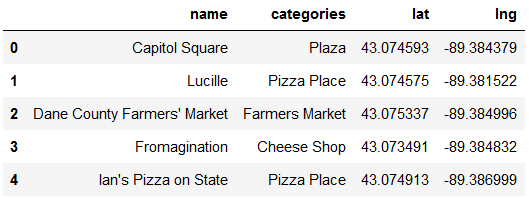
### 

### 

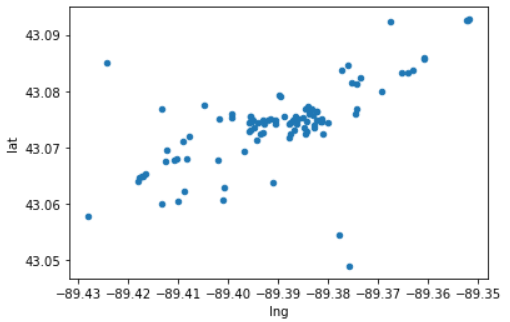
### **1. Segment Madison, WI into regions based on venue locations**

We'll use Foursquare data to find clusters of venues to define our regions. We'll attempt k-means clustering. If that does not segment the city in a useful way, we'll define the cluster centers manually.

We made a request for venue data in an 8000 yard radius of the Capitol building in Madison, WI, which roughly defines the city. Here is a sample of that data once it was cleaned up and converted from json to a pandas dataframe:

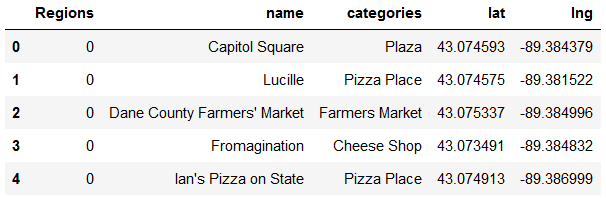


We then plotted each venue’s latitude and longitude on a scatter plot to make sure that the data returned was nicely spread out. Below you can see that plot, which shows the shape of Madison fairly well:

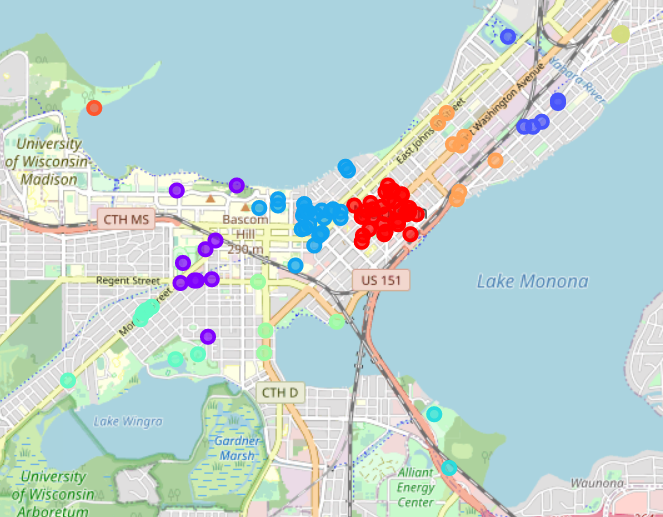


In the above scatterplot we can see two clusters near the center. The one on the left is the college campus, and the one on the right is downtown. When I run k-means to define the regions, these will need to be distinct. That will be the pass/fail condition for using k-means for this purpose.

We ran k-means clustering for 10 clusters on this data to group the venues by proximity. Here is a sample of the resulting data:



And here are the clusters mapped:

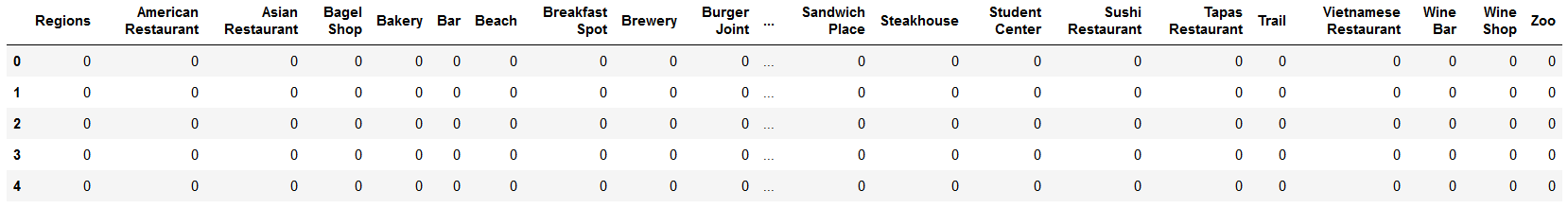


The map above shows that k-means clustering was able to split downtown from campus, which was what I was worried about before. Thankfully, we don't have to go in and manually define regions.

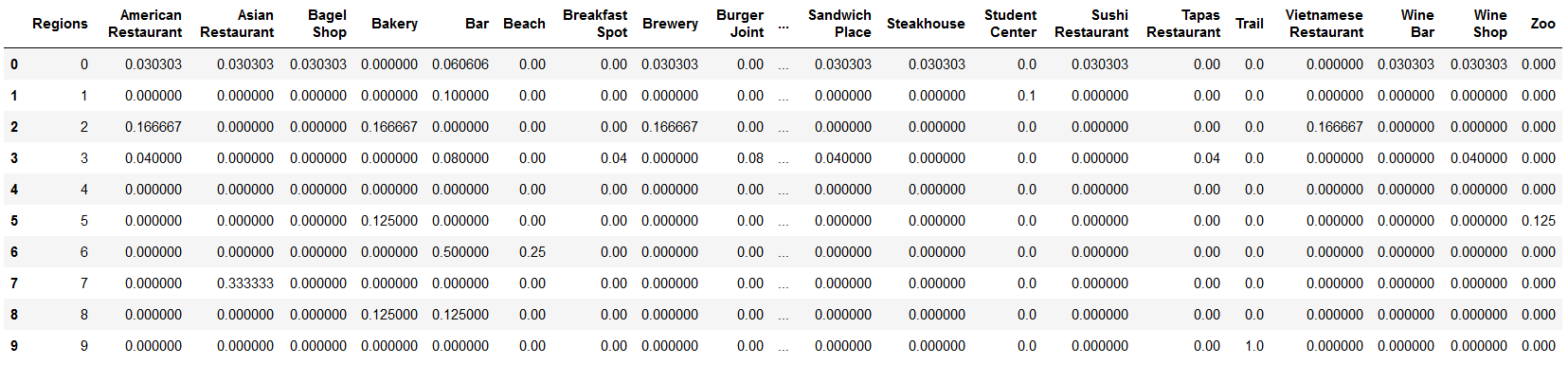
### **2. Determine frequency of venue types for each region**

Here we'll use the same Foursquare data to calculate the frequency of each venue type in a region.

I used onehot encoding to turn the venue types into binary values. Here’s a sample of what that looks like:



In the above table each row is a venue with its region along with its type by a 1 in the corresponding column. This was useful for calculating frequencies of venue types in a region, shown below.



In the above table each row is a region, and each column shows the frequency of a venue type within that region.

### **3. Determine similarity of regions based on venue types to find the best region for a new donut shop**

Here we'll use a nearest-neighbors approach to determine the similarity of different regions, using Euclidean distance to determine similarity.

It turns out that only one region, Region 1, has a venue categorized as a ‘Donut Shop’, so this will be the region to compare the other regions to.

## 

## **Results**

Here we'll calculate the difference from each region to Region 1 to see which is the most similar. After looping through each region and calculating the difference to Region 1, the region with the smallest difference to Region 1 was Region 3. This tells us that the ideal location for a new donut shop is Region 3, because it is the most similar to the only region that has a donut shop.

## **Discussion**

Region 1 is the west side of the college campus, and Region 3 is the east side. Because of this, it makes sense for business types in one region to exist in the other.

The donut shop in Region 1 is on a busy road that connects downtown Madison to the suburbs. It's also worth noting that the suburbs of Madison are the home of a large tech company with 10,000 young employees, who often live downtown. The high volume of commuters through Region 1 could explain the success of the donut shop, with people coming in to pick up breakfast for themselves and coworkers.

Region 3 connects the campus to the downtown area. It has high foot-traffic with a lot of places to stop in and grab quick snacks (ice cream shops, cheese shops, etc.) A donut shop would fit the profile of the region perfectly.

## **Conclusion**

It's useful to keep symmetries in mind when solving problems such as these. In this example, looking at one side of campus gives us business ideas for the other side. This shows that value in keeping the defining characteristics of a city in mind when planning where to open a business. In general, a city's campuses, lake fronts, downtown areas, etc. can be broken down into sub-regions that can be compared. What's working in one could possibly work in another.