IBM Data Science

Capstone Presentation

Background

- The United States is the 4th largest country by area with a rich secondary school system

- It has over 7,000 registered secondary schools

- For young adults the environment around the university is just as important as the university education.

- The availability of cheap restaurants, foreign food, museums, bars and night life factors into many individual's choice of a university.

Problem Statement

- It is difficult for low income students to travel and assess the ambience of distant secondary schools
- It's possible to use existing data and machine learning to:
  - Identify a nearby university to assess the ambiance for a university farther away
  - Identify additional universities that share similar ambiance to a previously visited university
  - Identify universities to not consider based off the similar ambiance to a previously visited university
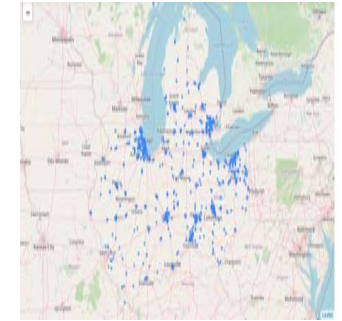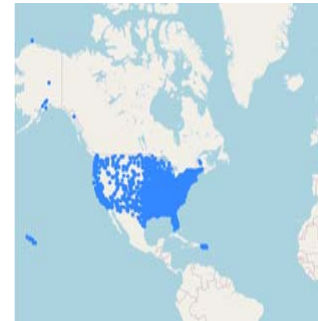
Data

- Data from Foursquare and the National Center for Education Statistics combined to create university ambience groupings
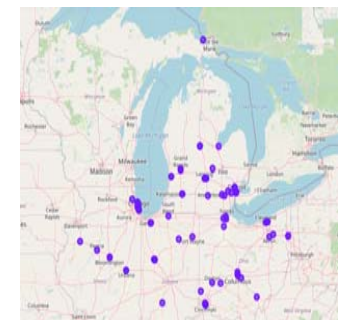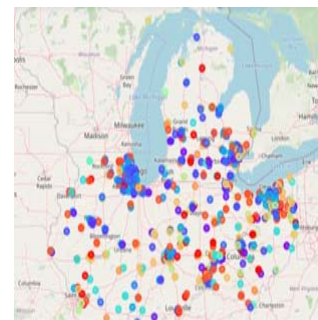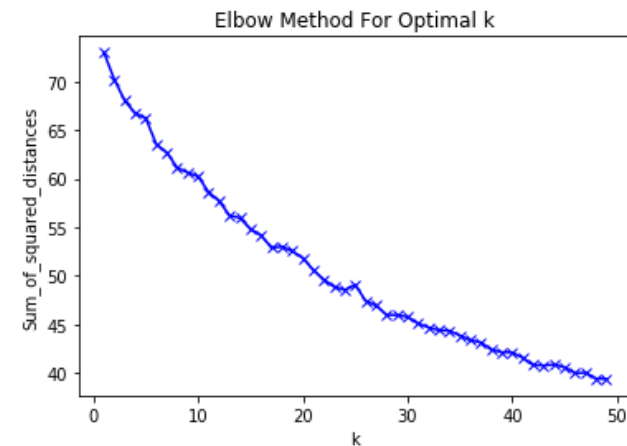- Nearby venues are sorted and ranked for each university

	School	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	ATA College - Cincinnati	Clothing Store	Gas Station	Mexican Restaurant	Supplement Shop	American Restaurant	Sandwich Place	Gym / Fitness Center	Pet Store	Restaurant	Discount Store
1	Abcott Institute	Fast Food Restaurant	Sandwich Place	Cosmetics Shop	Grocery Store	Coffee Shop	Gym / Fitness Center	BBQ Joint	Bank	Bar	Gas Station
2	Adler University	Hotel	Theater	Coffee Shop	Middle Eastern Restaurant	Gastropub	Salad Place	Donut Shop	Arts & Crafts Store	Sandwich Place	Fountain
3	Adrian College	Pizza Place	Café	Hotel	Discount Store	Optical Shop	Sandwich Place	Video Store	Fast Food Restaurant	Gym / Fitness Center	Gas Station
4	Adult and Community Education-Hudson	Discount Store	History Museum	Public Art	Food Court	Fast Food Restaurant	Soccer Stadium	Caribbean Restaurant	Gas Station	Chinese Restaurant	Park
...	...	...	...	...	...	...	...	...	...	...	...
886	Wright State University-Main Campus	Mexican Restaurant	Coffee Shop	Café	Sandwich Place	Pizza Place	Steakhouse	Kebab Restaurant	Burger Joint	Breakfast Spot	Smoke Shop
887	Xavier University	Pizza Place	Ice Cream Shop	Discount Store	Bar	High School	Home Service	Bowling Alley	College Basketball Court	Sandwich Place	Breakfast Spot

Methodology

1. Gather data from the National Center for Education Statistics
   - Filter universities to the states of Indiana, Ohio, Michigan, and Illinois (done due to limitations on Foursquare API calls for anything larger)

2. Request venue data from the Foursquare API for each university in the data set
   - Store up to 30 recommended venues per university

3. For each university, sort the venues by most popular to least popular

4. Run a K-Means algorithm for all the universities, clustering based off the 10 ten venues
   - K-Means was chosen due to its unsupervised nature and speed of computation
   - Iterate through K values from 1 to 100

5. Use the elbow method to obtain a reasonable K

6. Map the different clusters and display the list of universities in each cluster

K-Means

- K-Means chosen due to speed of algorithm and ability to run unsupervised
- K-Value optimized by elbow method

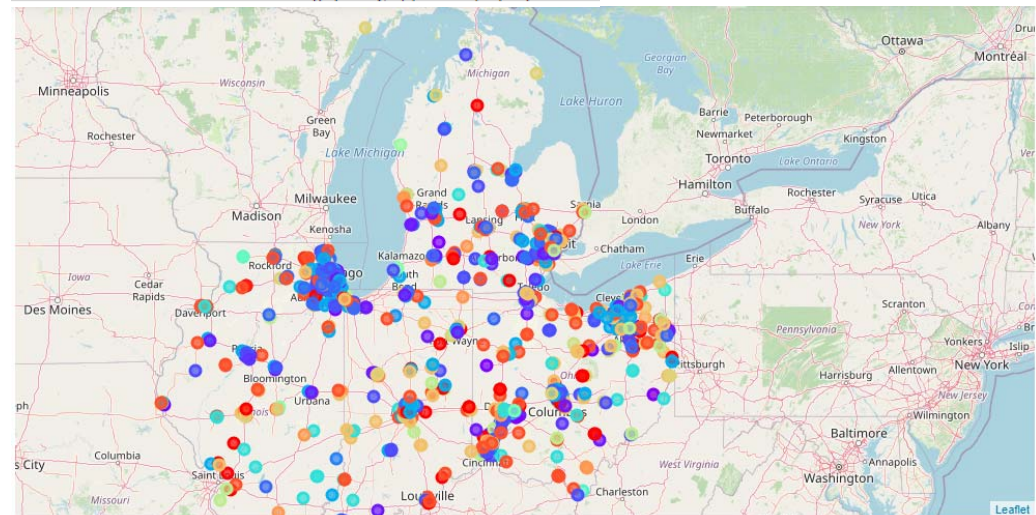Elbow Method For Optimal k

Results

- Thirty school clusters
- Viable dataset for finding similar schools based off of ambiance


```
*****Universities in cluster 0 *****
                                              NAME
                     Southwestern Illinois College
                                 Blackburn College
                          Chicago State University
       City Colleges of Chicago-Kennedy-King College
                             John A Logan College
                Lewis and Clark Community College
                             Prairie State College
          Southern Illinois University-Edwardsville
                                   Hanover College
           PJ's College of Cosmetology- Jeffersonville
                             Huntington University
                 Ivy Tech Community College-Kokomo
                Ivy Tech Community College-Richmond
                            Indiana University-East
                        Ravenscroft Beauty College
               PJ's College of Cosmetology- Greenfield
                   PJ's College of Cosmetology- Muncie
```

Recommendations

- Increase analysis to entire U.S. secondary school system

- Include crime, income, population density into analysis

- Create graphic interface forfuture students to use