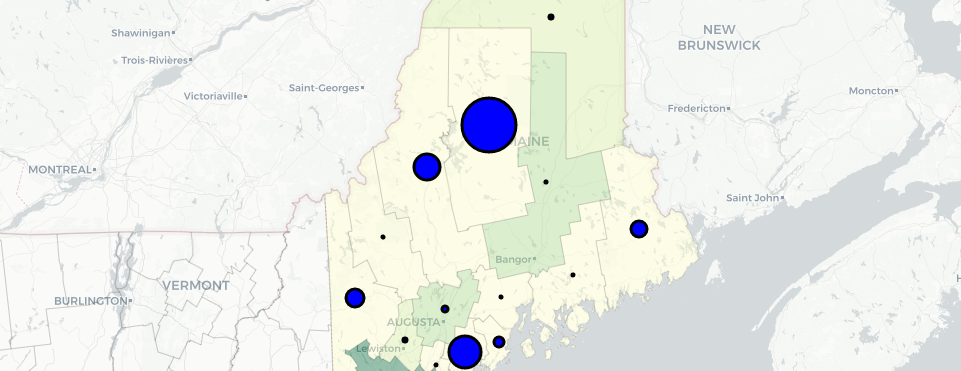
IBM Data Science Professional — Capstone Project

Population Deltas for the State of Maine and Potential Impact on Local Businesses

Link to the project notebook on githup: <https://github.com/kristianjackson/Coursera_Capstone/blob/master/IBM%20Data%20Science%20Professional%20Capstone%20Project.ipynb>



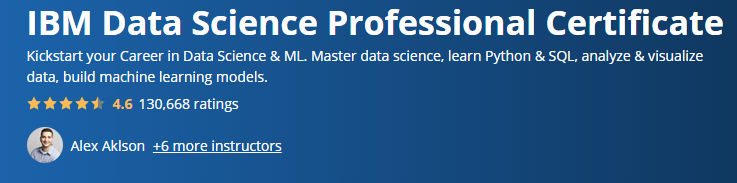
Welcome to the state of Maine

**Credentials!**

I work as a senior technology consultant for the federal government, and beyond that, I have been interested in technology since my first Atari 2600! For most of my career, I worked in the private sector and relatively stable positions. That meant that I never put much stock in certifications. I viewed them as money-sinks, a vehicle that people would use instead of experience. After all, those that can do…right?

As a consultant I am not looking to stay in one position, I am looking to new challenges all the time. That means that people need to be able to figure out my level of knowledge rapidly and with a degree of confidence that makes them feel comfortable bringing me in to fill a gap in their resources. Credentials and certifications help that along.

I decided to embark on the path of the IBM Data Science Professional through [Coursera](https://www.coursera.org/professional-certificates/ibm-data-science).



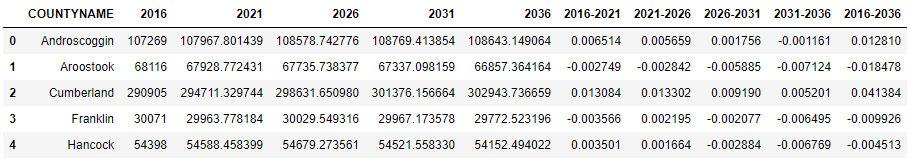
Coursera/IBM Data Science Professional

It’s a series of 9 courses that culminate in a capstone project of your choosing. The coursework involves becoming familiar with IBM’s product offering, Watson, and their branded notebook environment, and also working with Python, gathering data, the basic ML models, and finally the capstone.

**Capstone Project**

Most of the coursework is fairly straight forward, you review a concept, do some activities to reinforce the ideas, and move on. For the capstone, you’re set loose. Pick a topic, gather the data, do the analysis, draft a report, and present it either via slide deck or as I have decided to do, via a blog post.

I started with an idea that I wanted to do something related to the state of Maine because that is where I grew up and my family mostly lives there. So I started gathering data. What I came across was population change projections with observed data from 2016 and projected out to 2036.



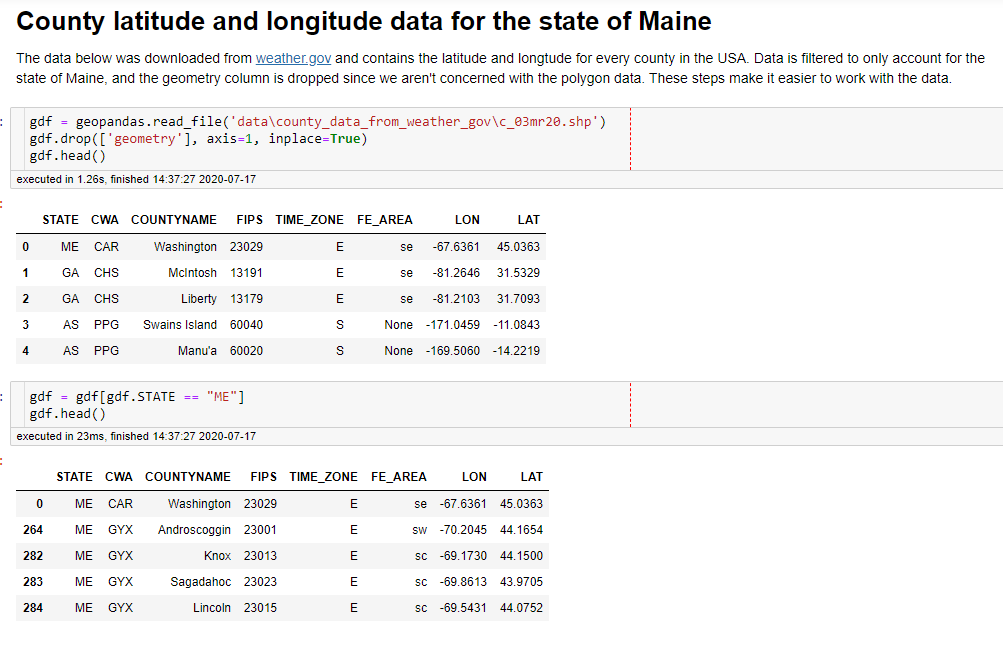
A sample of the population projection data

Initially, I planned on seeing if I could incorporate some COVID-19 data. Maine does populate county-based COVID-19 data but when I inspected it I decided that it was not interesting, especially not with any intersection with the population data.

In another part of the coursework, we had gathered data from [Foursquare](https://foursquare.com/) via their [Developer API](https://developer.foursquare.com/places). Now we were thinking about something interesting. I decided that my analysis was going to examine the population of the state of Maine, projected changes, and the business/venue data from Foursquare in an attempt to evaluate what areas are more likely to see businesses succeed based on people willing to buy products.

I have made the [notebook](https://github.com/kristianjackson/Coursera_Capstone/blob/master/IBM%20Data%20Science%20Professional%20Capstone%20Project.ipynb) available via my [GitHub](https://www.github.com/) account so I am not going to discuss some of the mundane set up portions of the work. I am going to highlight the major steps.

As discussed, I already had the population delta projects, but I needed a way to map that data. The data set did not contain latitude/longitude data so I needed to get that from somewhere. After a quick search, I found the data at [weather.gov](https://www.weather.gov/gis/Counties). Then I needed to clean the data and merge it into my pandas data frame:



Data preparation and cleaning

After the data clean up we are left with two data frames: a data frame that has the population projections and a data frame that has the latitude, longitude, FIPS code, and other assorted data that we aren’t interested in. To make the rest of the exploration easier we can merge the data sets keyed on the COUNTYNAME parameter:



Merging the two data frames into one

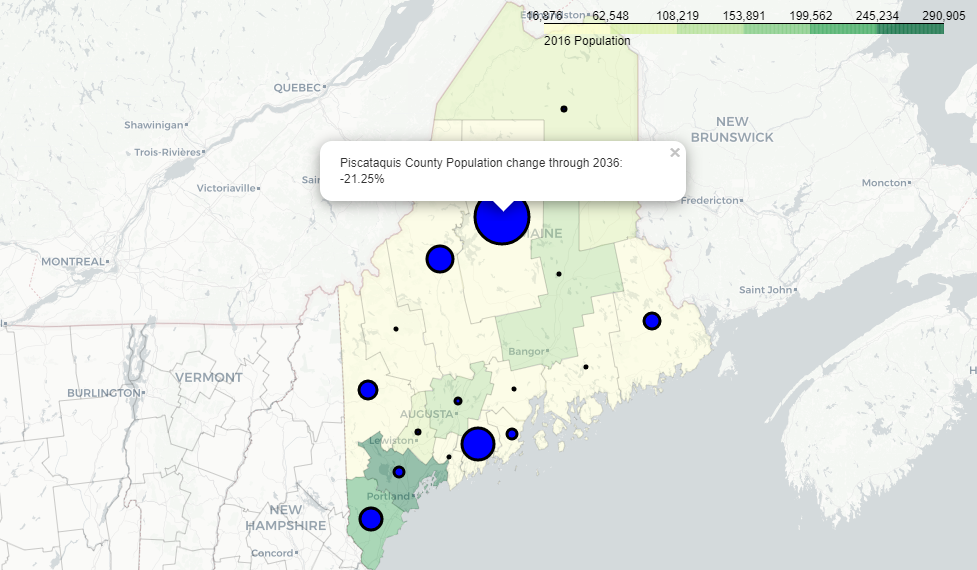
**Mapping the state**

Now that we have some data lets make a basic map of the state:



Code to make a map of Maine

The code starts by creating a map object, centered on the state of Maine, that we can apply different layers too. First, we apply a chloropleth of the population, as observed in 2016 by county. The lighter the color the less the population. This will give us a visual touchpoint for the rest of the data, a starting point. The second layer that we create overlays a circle on the county. The radius of the circle is set by the percent of the population difference projected for 2036. By referencing the circles we can see which areas are projected to increase or decrease in population. The actual percentage is indicated when we click the circle.



The state of Maine as a chloropleth

With this analysis, we see that most of the population is concentrated in the bottom portion of the state. Since I can’t represent a negative radius size, I used the absolute values of the percent change. As indicated in the popup, Piscataquis County is expecting a negative 21.25% change in the population, and they aren’t very populated, to begin with.

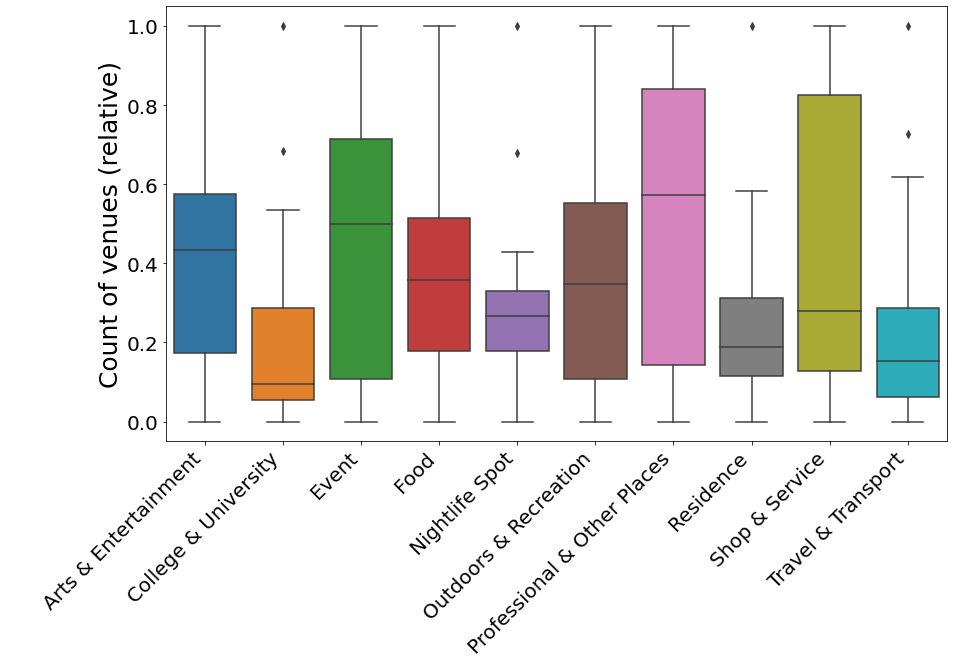
**Foursquare and Venue Data**

Looking at the projected population tells a story about who is going where, but what are they going to do when they get there? Foursquare attempts to help us with that question by making data available via a freely available [developer API](https://developer.foursquare.com/places). Gathering venue data can be used to show how developed a location is.



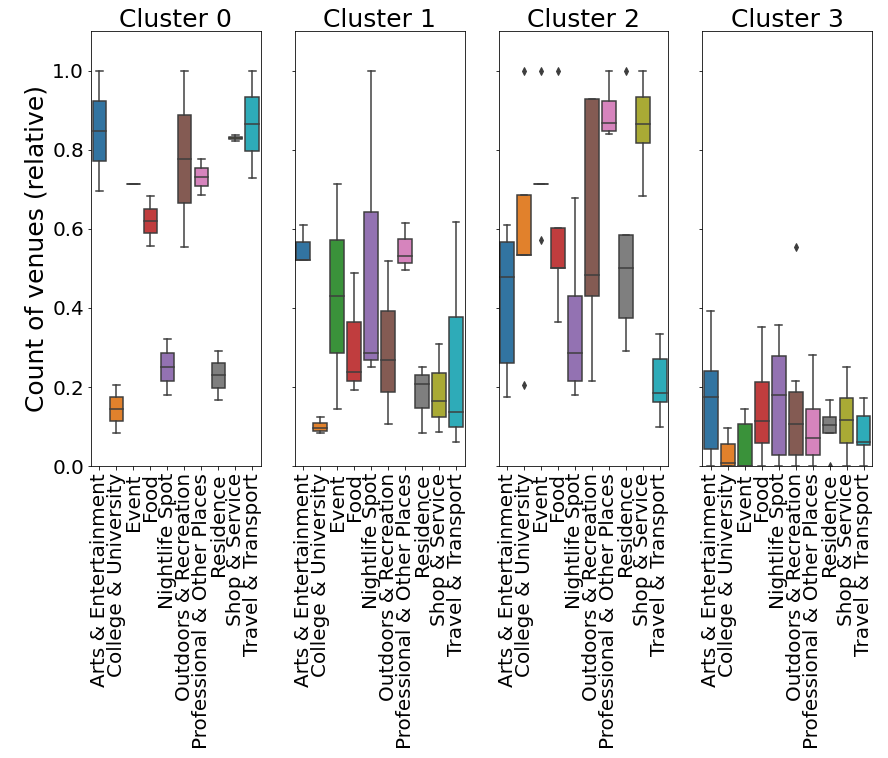
Code for getting category data from Foursquare

By starting with a list of categories the data can be binned and examined. After getting the categories from the API we can call the explore API to get venue data located near each of the counties. To better understand the results let’s look at the box plots:



Box plots of our venue data across all the state of Maine

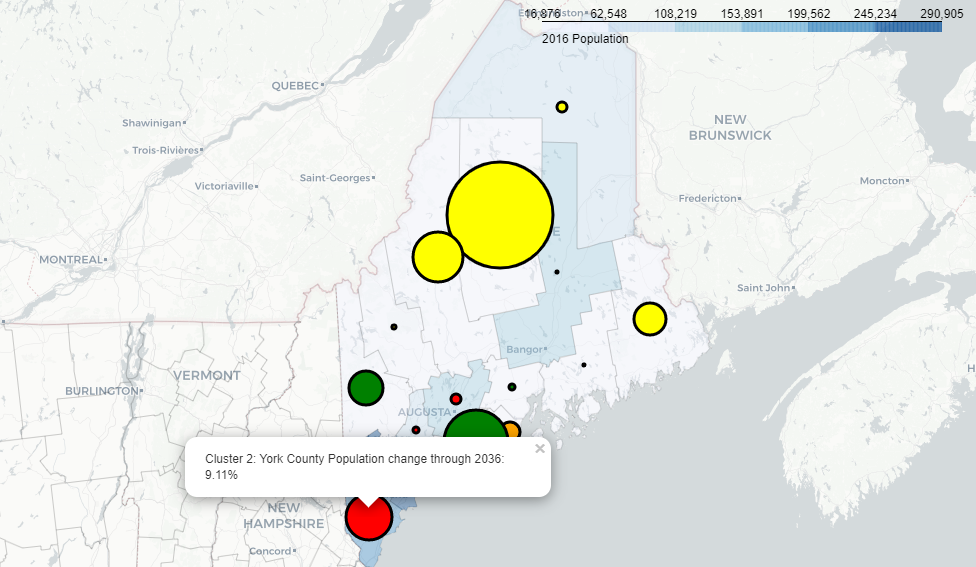
All of the major categories are represented, but how can we apply this to our problem of population change and the potential effect on the economy? For that can apply the K Means clustering method to see if there are any obvious groupings:



Clusters of data for the state of Maine

After exploring the number of clusters make sense, 4 provides us a clear picture of the state. Cluster 4 is underdeveloped as there are few venues. Cluster 0 is highly developed in certain areas: Arts and Entertainment, Food, Outdoor Activities, and Travel stand out as being important to the areas. Clusters 1 appears to be a developing area based on the expanse of types of venues. Cluster 2 is lacking in certain areas but also more densely developed than Cluster 0.

Discussing this data is exciting, but visualization of it would tell a better story. One method of visualization would be to make the population change indicators on the map change color based on the cluster that the county belongs to.



Clusters added as colors

As the coloring indicates, there are 4 clusters represented on the map. Cluster 2 includes York county. York county is heavily seasonal by nature with tourists coming to check out Maine’s seafood, trees, and beaches. The popup indicates that the population of York county is growing. and Reviewing the cluster box plots shows that Cluster 2 is established with multiple types of venues represented.

The last bit of this is deciding if we can answer the question we started with: can we determine if population change will impact the economy of the state? The last bit of this data is anecdotal. Generally speaking, the southern part of the state is the most populous portion. The clusters seem to indicate that they are also the most developed portion of the state. While this data doesn’t prove that a change in the population could affect the economy it does show that the development is correlated with the population. Therefore we could conclude that a change in the population could indeed affect the future development of businesses in these areas.

In the areas where Maine is more sparsely populated there is already a lack of development. If we extrapolate that data we could say that further reduction in population is not likely to affect the economy, but if there were a way to increase the population if holds that the economy might also develop better.

**Conclusions**

We started with the premise that population change might impact the economy of the state of Maine. After mapping the population starting point, overlaying the population projections, and examining the venue data from Foursquare, combined with personal experience and insight we have concluded that population and development are at least correlated. We cannot draw a causal conclusion since this is based on forecast data and significantly beyond the scope of what we analyze.

If Maine wants to increase the development of businesses, it should also investigate programs or measures to increase the population growth in those lower populated areas.