Data Science Capstone Report

IBM Data Science Professional Certificate

Introduction

Boundaries have always been historically important in the development of population centers. The border between New York and New Jersey plays a huge role on the development of the New York City Metro as well as the border of Kansas and Missouri in the Kansas City Metro. The problem with state boundaries is that the different laws, codes, and taxes can cause unequal urban development and can lead to lopsided development in a metropolitan area. An area with a high sales tax maybe geared predominantly to residential while an area with a low sales tax and high income tax will be heavily commercial. There are also other hidden implications state boundaries going through metropolitan areas can have and this project will look at that.

In the project two metro cities will be compared for their similarity. One is the Sioux City, Iowa metroplex, and the other being the Sioux Falls, South Dakota Metroplex. The Sioux City Iowa metro is spread over three states: Iowa, Nebraska, and South Dakota. The Sioux Falls area is spread soley over South Dakota. For some more background, Iowa has fairly high income tax while low sales tax and no grocery tax. Nebraska is fairly low across sales, property, and income taxes. South Dakota has a grocery tax, however, has no income tax and has no corporate income tax. Their laws and regulations have a very business friendly slant, and as such, attract high tech fields as well as medical and professional fields. Iowa and Nebraska have higher overall taxes but focus more on infrastructure and education while South Dakota puts a slightly less of a focus.

The main focus and problem to be looked at, however, is how do states lines affect a metropolitan area. Do they help or hurt development? Is development really lopsided? Really

looking at these two metropotanilans areas and looking at the difference between Sioux Falls which has no boundaries, and Sioux City, which has three state boundaries will be the focus.

Data

To obtain the data, I first looked at a Zip Code Map, and manual pulled the ZipCodes of the Neighborhoods I wanted. I then went on a the U.S. Census website and found the Lat Lon's for the zip codes. I then mannual appended the income data to the csv file for each zip code. This allowed me to do further analysis and also was serve as a reference point in the discussion. Geospatial Data was obtained through Foursquare API and by putting in the various Lat-Lons and finding venues close by those Lat-Lons

Methodologies.

First, the csv files were read into the Jupyter notebook using. Using Watson Studios, this was simply done by clicking the add data button. Once these files were read, I imported the various libraries and files that would be needed. I then set up my four square credentials and then used a for statement to obtain input the necessary information into the URL for Foursquare and to output into columns the result. I then did a count() command to count the total number of venues for each neighborhood. Afterwards, I used one-hot encoding as my first machine learning method. This was a great method to use as I was to get a number between 0 and 1 and then able to rate the top ten venues in each neighborhood. I then sorted out the the numbers from the one hot encoding and got the top 10 venues for each neighborhood. I then used K-cluster to try and find similar clusters. This proved somewhat successful, but really, looking the data nothing could be gleamed

Results

	NAME	ZIPCODE	LAT	LON	INCOME	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Downtown Sioux City	51101	42.49156	-96.39943	17,649	1	Bar	Mexican Restaurant	Pizza Place	Coffee Shop
1	North Sioux City	51105	42.51332	-96.37071	33,390	0	Bar	Pizza Place	Mexican Restaurant	Coffee Shop
2	Briar Cliffs	51104	42.53499	-96.40241	59,096	1	Bar	Coffee Shop	Mexican Restaurant	Pizza Place
3	Riverside	51103	42.52066	-96.44160	38,181	1	Bar	Coffee Shop	Mexican Restaurant	Pizza Place
4	Singing Hills	51106	42.46039	-96.32263	54,415	4	Bar	Sandwich Place	Mexican Restaurant	Pizza Place
5	Sergant Bluff	51111	42.41448	-96.38793	56,318	3	Fast Food Restaurant	Pizza Place	Sandwich Place	Mexican Restaurant
6	South Sioux Nebraska	68776	42.47802	-96.46125	51,389	2	Bar	Gas Station	Coffee Shop	Mexican Restaurant
7	Dakota Dunes SD	57049	42.52041	-96.50418	73,567	4	Bar	Gas Station	Pizza Place	American Restaurant

	NAME	ZIP	LAT	LON	INCOME	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Downtown Sioux Falls	57103	43.54379	-96.69381	56,319	2	Mexican Restaurant	Bar	American Restaurant	Sandwich Place
1	Central Sioux Falls	57105	43.52211	-96.73395	57,295	0	Coffee Shop	Fast Food Restaurant	Mexican Restaurant	Café
2	North Sioux Falls	57104	43.59735	-96.70126	36,850	0	Mexican Restaurant	Sandwich Place	American Restaurant	Fast Food Restaurant
3	Northwest Sioux Falls	57107	43.60164	-96.82520	61,213	0	Hotel	American Restaurant	Construction & Landscaping	Sandwich Place
4	Southwest Sioux Falls	57106	43.50241	-96.83039	61,620	4	Coffee Shop	Fast Food Restaurant	Burger Joint	American Restaurant
5	South Sioux falls	57108	43.47331	-96.68963	94,094	3	Park	Fast Food Restaurant	Sandwich Place	Coffee Shop
6	East Sioux Falls	57110	43.54969	-96.63113	68,409	2	Park	Mexican Restaurant	Sandwich Place	Bar
7	Tea	57064	43.46329	-96.87032	81,845	1	Gas Station	Water Park	Ice Cream Shop	Pet Store

Discussion

Looking at the data the clearly is a difference between the Tri State Sioux City Metro and the single state Sioux Falls Metro. With a grocery tax in South Dakota, Dakota Dunes mainly has gas and restaurants with Iowa having most of the grocery and retail. South Dakota has incentives on their medical industry so seeing incentives and this can be seen looking at the Optical Shop venue in the Dakota Dunes list. Bars Dominate the Sioux City Metro while Parks are the seemingly dominant feature in Sioux Falls Sioux Falls has a greater diversity of venues by far. Income distribution is far more uniform in Sioux Falls than in Sioux City

When looking at all the data and even looking at clustering, it is easy to see that there have several different areas tends to segment the metro area in a somewhat harmful way. Lack of community unity from an urban planning standpoint is evident in the Sioux City metro. The Sioux Falls metro is much more uniform and has more evident central vision. It is clear from the data people live in Nebraska and South Dakota and work in Iowa. This increased the need for transportation which can have other unintended consequences. Segmentation, as shown from the data leads to less diversity community venues and leads to a less uniform metro area.

Conclusion

Having state boundaries running through a metropolitan area can take away from urban planning cohesion and lead to large groups of segmentation which have net negative effect on the community. Urban planners and politicians alike to need to figure ways out to make special incentives to prevent the creation of bedroom community and work only communities which have a negative effect on the metro area as a whole. A proposal in the Kansas City area will make a special tax incentive to help lessen the effects and differences between state lines . The data suggests that this is needed in areas where political boundaries cross population centers