M. Pierce

Data Science Capstone

Introduction:

Choosing where to live can be a difficult decision. There are many factors that go into choosing a location to live. The audience for this project is young professionals, with no children, starting their careers in the city of Milwaukee, Wisconsin. This project focuses on Milwaukee County and the cities within. Each city in the county is unique and different and has different offerings. This project will explore locations, clusters, population, and venues. Once cleaned and visualized the data can be used for an individual to make decisions where to start looking to live.

Data:

The data location, population, and density came from https://simplemaps.com/data/us-cities. This data set included State, county, city, lat/long, density, and time zone information. This was used to highlight and visualize the locations in relationship to the City of Milwaukee and create k-means clusters. Foursquare data was used to gather information on all venues near the cities/clusters. The venues were sorted by frequency and top 10 lists were created for each city in each cluster.

Methodology:

The process started by segmenting the national city data. The data included all cities, counties and states. Since this project was only concerned with Milwaukee County, Wisconsin the data needed top be cleaned and segmented.

Cleaning and segmenting the DATA: Column names, and data other than Wisconsin.

In [4]:
```
df.rename(columns={'lat':'Latitude', 'lng':'Longitude','state_id': 'State', 'county_name':'County', 'county_fips':'Zip', 'cit
df.head()
```

Out[4]:

	City	State	state_name	Zip	County	Latitude	Longitude	population	Density	timezone
0	South Creek	WA	Washington	53053	Pierce	46.9994	-122.3921	2500	125.0	America/Los_Angeles
1	Roslyn	WA	Washington	53037	Kittitas	47.2507	-121.0989	947	84.0	America/Los_Angeles
2	Sprague	WA	Washington	53043	Lincoln	47.3048	-117.9713	441	163.0	America/Los_Angeles
3	Gig Harbor	WA	Washington	53053	Pierce	47.3352	-122.5968	9507	622.0	America/Los_Angeles
4	Lake Cassidy	WA	Washington	53061	Snohomish	48.0639	-122.0920	3591	131.0	America/Los_Angeles

Once the columns were standardized for this exercise, I dropped unneeded data.

Drop unneeded columns

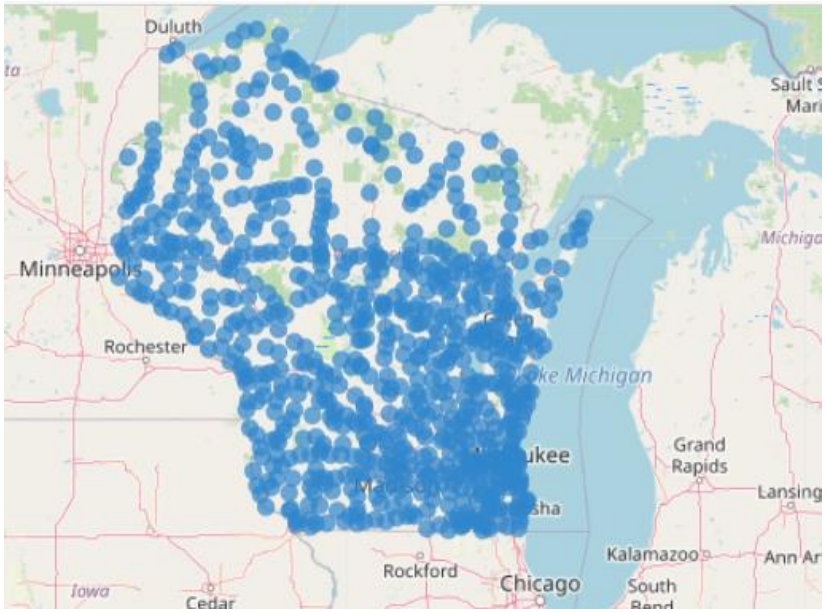
In [5]:
```
df.drop(columns={'state_name', 'population', 'timezone'}, inplace=True)
df.head()
```

Out[5]:

	City	State	Zip	County	Latitude	Longitude	Density
0	South Creek	WA	53053	Pierce	46.9994	-122.3921	125.0
1	Roslyn	WA	53037	Kittitas	47.2507	-121.0989	84.0
2	Sprague	WA	53043	Lincoln	47.3048	-117.9713	163.0
3	Gig Harbor	WA	53053	Pierce	47.3352	-122.5968	622.0
4	Lake Cassidy	WA	53061	Snohomish	48.0639	-122.0920	131.0

M. Pierce

After removing unnecessary columns, using folium, I displayed only cities and counties in Wisconsin. This was still too much data to work with.

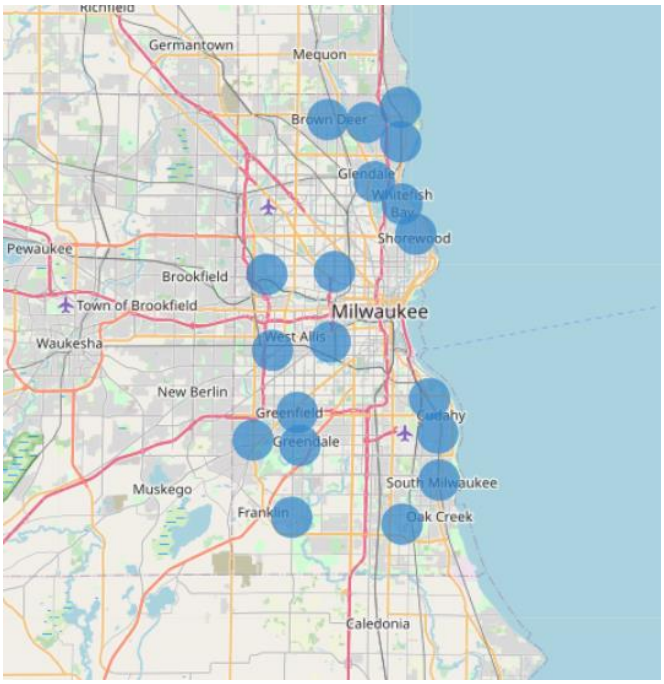


I created a new data frame that included only cities in Milwaukee County.

```
In [10]:  Mil_data = WI_data[WI_data.County == 'Milwaukee']
          Mil_data.shape
          Mil_data
```

Out[10]:

	City	State	Zip	County	Latitude	Longitude	Density
731	Greendale	WI	55079	Milwaukee	42.9370	-88.0018	982.0
766	Oak Creek	WI	55079	Milwaukee	42.8803	-87.9009	493.0
825	Wauwatosa	WI	55079	Milwaukee	43.0615	-88.0347	1409.0
863	Whitefish Bay	WI	55079	Milwaukee	43.1132	-87.9003	2545.0
903	River Hills	WI	55079	Milwaukee	43.1717	-87.9353	119.0
920	Franklin	WI	55079	Milwaukee	42.8854	-88.0104	403.0
940	Fox Point	WI	55079	Milwaukee	43.1581	-87.9013	900.0
958	South Milwaukee	WI	55079	Milwaukee	42.9120	-87.8627	1678.0
1011	West Allis	WI	55079	Milwaukee	43.0068	-88.0296	2033.0
1031	Brown Deer	WI	55079	Milwaukee	43.1743	-87.9750	1050.0
1122	West Milwaukee	WI	55079	Milwaukee	43.0124	-87.9710	1434.0
1150	Milwaukee	WI	55079	Milwaukee	43.0642	-87.9673	2389.0
1157	St. Francis	WI	55079	Milwaukee	42.9716	-87.8729	1418.0
1214	Hales Corners	WI	55079	Milwaukee	42.9410	-88.0491	924.0
1221	Shorewood	WI	55079	Milwaukee	43.0914	-87.8864	3239.0
1266	Greenfield	WI	55079	Milwaukee	42.9619	-88.0051	1233.0
1275	Cudahy	WI	55079	Milwaukee	42.9467	-87.8641	1479.0
1402	Glendale	WI	55079	Milwaukee	43.1288	-87.9277	852.0
1424	Bayside	WI	55079	Milwaukee	43.1827	-87.9017	707.0



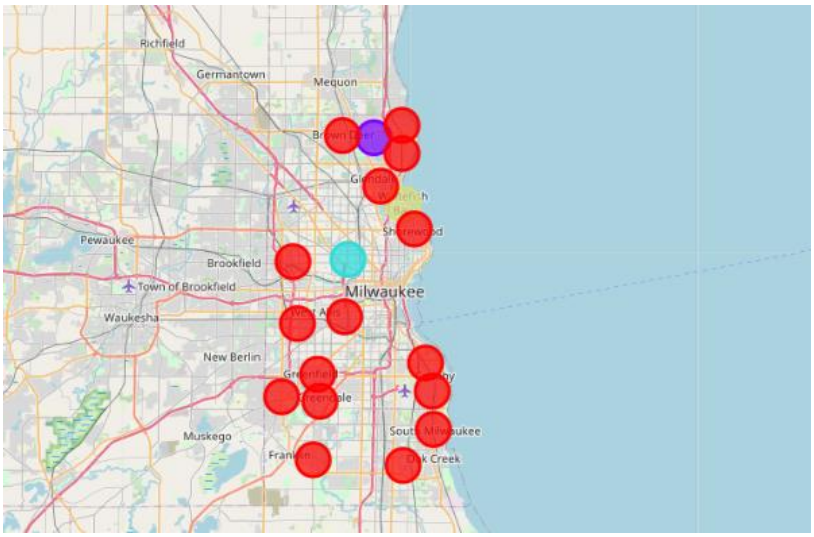
Once prepared, the data can now be enriched by using the foursquare API. The following is a top 10 venues for each city in the county.

Out[44]:

	City	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	Bayside	Garden Center	Home Service	Gym	Park	Flower Shop	Cosmetics Shop	Dance Studio	Deli / Bodega	Diner	Discount Store
1	Brown Deer	Mobile Phone Shop	Hotel	Liquor Store	Gym / Fitness Center	Sandwich Place	American Restaurant	Pool	Café	Electronics Store	Donut Shop
2	Cudahy	Bar	Sports Bar	Thrift / Vintage Store	Dive Bar	Furniture / Home Store	Optical Shop	Discount Store	Gas Station	Park	Flower Shop
3	Fox Point	Pool	Organic Grocery	Flower Shop	Convenience Store	Cosmetics Shop	Dance Studio	Deli / Bodega	Diner	Discount Store	Dive Bar
4	Franklin	Baseball Field	Reservoir	Park	Yoga Studio	Flower Shop	Cosmetics Shop	Dance Studio	Deli / Bodega	Diner	Discount Store
5	Glendale	Pub	Gas Station	Gym / Fitness Center	Dance Studio	Rental Car Location	Park	Baseball Field	Garden Center	Bar	Dog Run
6	Greendale	Park	Yoga Studio	Sports Bar	Arts & Crafts Store	Athletics & Sports	Bakery	Baseball Field	Coffee Shop	Ice Cream Shop	Pizza Place
7	Greenfield	Furniture / Home Store	Fast Food Restaurant	Italian Restaurant	Mobile Phone Shop	Cosmetics Shop	Salon / Barbershop	Shoe Store	Spa	Sushi Restaurant	Sandwich Place
8	Hales Corners	Sporting Goods Shop	Chinese Restaurant	Fast Food Restaurant	Shoe Store	Cosmetics Shop	Mexican Restaurant	Coffee Shop	Pet Store	Pharmacy	Pizza Place
9	Milwaukee	Cosmetics Shop	Bar	Convenience Store	Grocery Store	Pizza Place	Coffee Shop	Liquor Store	Cocktail Bar	Hot Dog Joint	American Restaurant
10	Oak Creek	Trail	Playground	Lounge	Dive Bar	Fast Food Restaurant	Farmers Market	Electronics Store	Donut Shop	Dog Run	Discount Store
11	River Hills	Sculpture Garden	Yoga Studio	Food Truck	Cosmetics Shop	Dance Studio	Deli / Bodega	Diner	Discount Store	Dive Bar	Dog Run
12	Shorewood	Pizza Place	Bakery	Pharmacy	Bar	Gym / Fitness Center	Coffee Shop	Salon / Barbershop	Grocery Store	Mobile Phone Shop	Spa
13	South Milwaukee	Bar	Ice Cream Shop	Coffee Shop	Farmers Market	Mexican Restaurant	Flower Shop	Martial Arts Dojo	Lounge	Bowling Alley	Chinese Restaurant
14	St. Francis	Park	Plaza	Indie Theater	Basketball Court	Bar	Steakhouse	Brewery	Dog Run	Trail	Food Truck
15	Wauwatosa	Coffee Shop	Furniture / Home Store	Pool	Park	Outdoors & Recreation	Spa	Bank	Sporting Goods Shop	Pharmacy	Golf Course
16	West Allis	Bar	Mobile Phone Shop	Soccer Field	Convenience Store	Pub	Chinese Restaurant	Cajun / Creole Restaurant	Café	Sandwich Place	Dive Bar
17	West Milwaukee	American Restaurant	Sandwich Place	Fast Food Restaurant	Pizza Place	Mobile Phone Shop	Grocery Store	Cosmetics Shop	Discount Store	Pub	Big Box Store
18	Whitefish Bay	Café	Gift Shop	Athletics & Sports	Park	Cosmetics Shop	New American Restaurant	Plaza	Pharmacy	Pet Store	Construction & Landscaping

Results:

Once the foursquare data was gathered and encoded, k-means clustering was used to group the cities into 4 clusters.



Each cluster was then output to include the top 10 venues to aid the individual in choosing a location.

Cluster 4

```
In [52]: mil_merged.loc[mil_merged['Cluster Labels'] == 3, mil_merged.columns[[0] + list(range(5, mil_merged.shape[1]))]]
```

Out[52]:

	City	Longitude	Density	Cluster Labels	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
863	Whitefish Bay	-87.9003	2545.0	3	Café	Gift Shop	Athletics & Sports	Park	Cosmetics Shop	New American Restaurant	Plaza	Pharmacy	Pet Store	Construction & Landscaping

```
In [ ]:
```

Discussion:

When the clusters are examined in tabular form, types of venues that are important to the individual can be identified. This is not the only thing that needs to be considered, however. The visualization of the unique clusters on the maps also aid the individual in choosing their ideal location.

Once narrowed down further and a location based on cluster is chosen, more data can be included, such as crime data, housing cost, traffic data etc. This is essentially the first step in choosing a place in Milwaukee to relocate.

Conclusion:

Large data sets can be used to solve complex problems. This simple exercise uses two separate sources and types of data to aid in a personal decision. The set of data could be increased to provide more information and clarity for the individual.