Finding the Best Neighborhood for an NFL Stadium in Portland, OR

Applied Data Science Capstone
David Johnson

Introduction and Business Problem

The National Football League is a 32 team professional sports league that plays American football in the United States. The NFL is very popular, with an average live attendance of 68,400 spectators per game. This average revenue for an NFL team is around $450 million, with approximately two thirds coming from television deals. The remaining third, approximately $150 million per year, comes from local business at the stadium, such as tickets, parking, and concessions.

Portland, Oregon, with a population of around 2.3 million, is one of the largest cities in the United States without an NFL team. There is a potential business opportunity to build a football stadium in Portland, with the hopes of eventually gaining an NFL team. But where should the stadium be built?

This report uses Data Science methods to answer the question of where the stadium should be built. By evaluating which neighborhood in Portland is most similar to the neighborhoods around existing NFL stadiums in other cities we will recommend a construction location to the stakeholders.

Data

Data sources for this project will include:

- List of neighborhoods in Portland, Oregon https://en.wikipedia.org/wiki/Neighborhoods_of_Portland,_Oregon
- List of current NFL stadiums https://en.wikipedia.org/wiki/List_of_current_National_Football_League_stadiums
- Foursquare data for venues https://foursquare.com/developers/apps
- OpenStreetMap https://www.openstreetmap.org/#map=4/38.01/-95.84

The list of neighborhoods in Portland, Oregon and the list of current NFL stadiums will be retrieved from WIkipedia. Wikipeida is a convenient online encyclopedia with a permissive license for use of data.

OpenStreetMap is used to convert street addresses into latitude and longitude. The data was accessed via the nominatim API for both Portland Neighborhoods and NFL stadiums.

Foursquare is a location service that provides information about venues (stores, restaurants, etc) near a given location. Foursquare will be used to investigate venues in different Portland neighborhoods, as well as near the NFL stadiums.

Methodology

Python 3.6 was the programming language used to address this data science problem. Python is a powerful open source programming language, and IBM cloud notebooks were a convenient tool to work and publish results.

First, the “Beautiful Soup” library was used to scrape wikipedia pages for the list of Portland Neighborhoods and the list of current NFL stadiums.

Table 1: List of Neighborhoods

	Name
0	Arlington Heights
1	Forest Park
2	Goose Hollow
3	Hillside
4	Linnton
5	Northwest District
6	Northwest Heights
7	Northwest Industrial
8	Old Town Chinatown
9	Pearl District
10	Portland Downtown

Table 2: List of Stadiums

	Name
0	Arrowhead Stadium
1	AT&T Stadium
2	Bank of America Stadium
3	CenturyLink Field
6	FedExField
7	FirstEnergy Stadium
8	Ford Field
9	Gillette Stadium
10	Hard Rock Stadium

Next, we use the OpenStreetMap API to determine the Latitude and Longitude of each Portland neighborhood, and each NFL stadium.

Table 3: Neighborhoods with Lat/Long

	Name	Latitude	Longitude
0	Arlington Heights	45.519496	-122.710667
1	Forest Park	45.561376	-122.758458
2	Goose Hollow	45.517749	-122.692819
3	Hillside	45.527439	-122.713120
4	Linnton	45.600330	-122.786779
5	Northwest District	45.533013	-122.698845
6	Northwest Heights	45.540806	-122.774354
7	Northwest Industrial	23.598351	58.273050
8	Old Town Chinatown	45.524934	-122.673516
9	Pearl District	45.529044	-122.681598
10	Portland Downtown	45.515274	-122.680025

Table 4: Stadiums with Lat/Long

	Name	Latitude	Longitude
0	Arrowhead Stadium	39.048940	-94.483003
1	AT&T Stadium	32.747842	-97.092844
2	Bank of America Stadium	35.225736	-80.853882
3	CenturyLink Field	47.595346	-122.331644
6	FedExField	38.907687	-76.864487
7	FirstEnergy Stadium	41.506056	-81.699712
8	Ford Field	42.339957	-83.045617
9	Gillette Stadium	42.091253	-71.264465
10	Hard Rock Stadium	25.957920	-80.238838

Next, we use the Foursquare API to determine what venues are near each NFL stadium. These values are then grouped together.

Table 5: Venues Near Each Stadium

	Neighborhood	ATM	Accessories Store	Airport Terminal	American Restaurant	Amphitheater	Antique Shop	Aquarium	Arcade	Art Gallery
0	AT&T Stadium	0	0	0	0	0	0	0	0	0
1	Arrowhead Stadium	0	0	0	0	0	0	0	0	1
2	Bank of America Stadium	0	0	0	0	0	0	0	1	0
3	CenturyLink Field	0	0	0	2	0	1	0	0	1
4	FedExField	0	0	0	2	0	0	0	0	0
5	FirstEnergy Stadium	0	0	0	0	0	0	0	0	0
6	Ford Field	0	0	0	4	0	0	0	0	0
7	Gillette Stadium	0	2	0	1	0	0	0	0	0
8	Hard Rock Stadium	0	0	1	0	0	0	0	0	0
9	Heinz Field	0	0	0	2	0	0	0	0	0

We use Kmeans with k=1 on this dataset. This is an unusual application of kmeans, which is normally used to separate data into multiple clusters. Our desire is to find the “center” of a

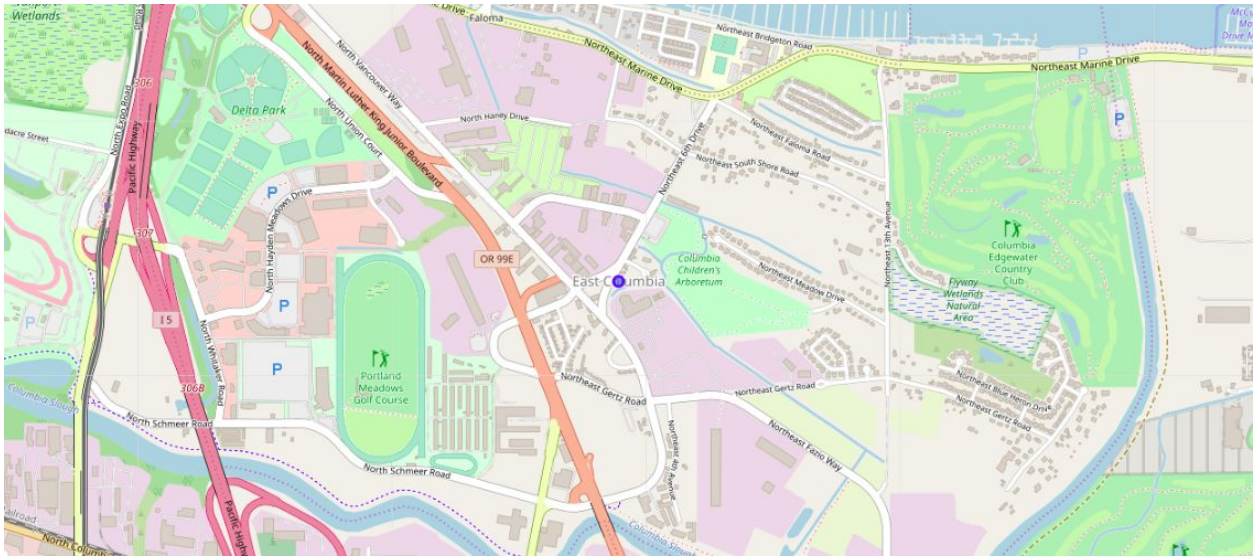
known group, i.e. neighborhoods with NFL stadiums. Once this “center” is determined we can see which Portland neighborhood is most similar. The result is:

	Neighborhood	ATM	Accessories Store	Airport Terminal	American Restaurant	Amphitheater	Antique Shop	Aquarium	Arcade	Art Gallery	...
0	Stadium	0.034483	0.068966	0.034483	1.034483	0.068966	0.034483	0.068966	0.034483	0.172414	...

We import a euclidean distance function from scikit learn, and iterate through the grouped venues in each neighborhood to find which neighborhood is the smallest distance from the nominal center.

Results

Based on this analysis, the East Columbia neighborhood of Portland is most similar to the average neighborhood in the vicinity of an NFL stadium. This makes sense when visualizing the location using Folium. Other NFL stadiums are built in the vicinity of parking and other athletic facilities. We can see that the nearby country club and golf course are similar venues.



Discussion

The methodology described produced an intuitive result. Investigating the East Columbia neighborhood shows it to have similar characteristics to the area near other NFL stadiums. Further investigation might vary the radius of nearby venues used for the Foursquare query.

Using 500 meters produced a relatively low total number of nearby venues. The “center” of the stadium neighborhood had a value of <1 for each category of venue. Expanding the radius may influence the most similar neighborhood, with a future result being more densely populated.

Conclusion

Using Data Science methods with the Python programming language produced a satisfactory answer to a complicated question. Which Portland neighborhood is best to build an NFL stadium? The result that the East Columbia neighborhood is best can be empirically justified by the fact that it is the most similar to the vicinity of existing NFL stadiums when measured by number and category of nearby venues.