Introduction/Business Problem

Where to build a new affordable housing development in Chicago?

Background

Chicago may be named one of Pricewaterhouse Cooper's 'Cities of Opportunities' but it is also a city which is highly segregated and suffering from stark income inequality that has been dubbed "a combination of Manhattan smashed against Detroit." (Semuels, 2018) This is especially obvious and divisive when it comes to housing, with some families homeless, a deadly prospect given the freezing temperatures of cities around Lake Michigan. Fortunately, the city government and the Chicago Housing Authority (CHA) have been working to provide rental housing priced below market rates and other housing subsidies to support the poor and disenfranchised. (CBTA, 2018)

Problem

Even though there are 350,000 low-income households, there are only 230,000 units available to these households. (DuPaul, 2018) This shows that there is still a strong need for housing assistance in Chicago. While issuing housing vouchers may help to solve the problem in the short-term, there is a need for the city government and the CHA to build new housing to ensure that there is enough housing stock in the long-term.

The question then is where these new affordable housing developments should be built. Ideally, these developments should be in areas where new residents are close to their **support networks** (presumably other less well-to-do residents who are living in current affordable housing developments) and to **amenities** that support the residents' lives, such as public transport and parks.

As such, data that might contribute to determining the best location to build these include the current amount of affordable housing stock in each community area. This could be used as a proxy for proximity to friends and family for these new residents, which would be an important consideration for them in choosing which affordable housing development to rent. Other relevant data would be a table of common venues in areas closest to current affordable housing developments.

Interest

This would be of most interest to the city government which has to decide the size and location of new affordable housing developments. After all, the city government would want to choose a location which would best support the lives of their residents. Other stakeholders who might be interested are providers of necessities and services to these residents, such as budget grocery chain stores that accept food stamps.

Data

Data Source

The main file used is a CSV file containing thousands of affordable units that are supported by City of Chicago. The list is updated periodically when construction is completed for new projects or when the compliance period for older projects expire, typically after 30 years. It does not include every City-assisted affordable housing unit that may be available for rent, nor does it include the hundreds of thousands of naturally occurring affordable housing units located throughout Chicago without City subsidies. It is available for download from data.gov. This file is required to identify the locations of the existing affordable housing properties and to uncover the common venues close to them using Foursquare location data.

Data Cleaning and Feature Selection

There is a total of 389 affordable housing properties (rows) grouped into 60 community areas and 19 columns, with each column providing information about the affordable property. The main columns used for this project are as follows:

- Property Name
- Community Area that the Property belongs to
- Number of Units in each Property
- Longitude and Latitude of each Property

The data is relatively clean, with all no missing values and all present values in a standardised format. Therefore, no additional data cleaning measure is required.

Methodology

Visual Representation of Housing Units in Chicago
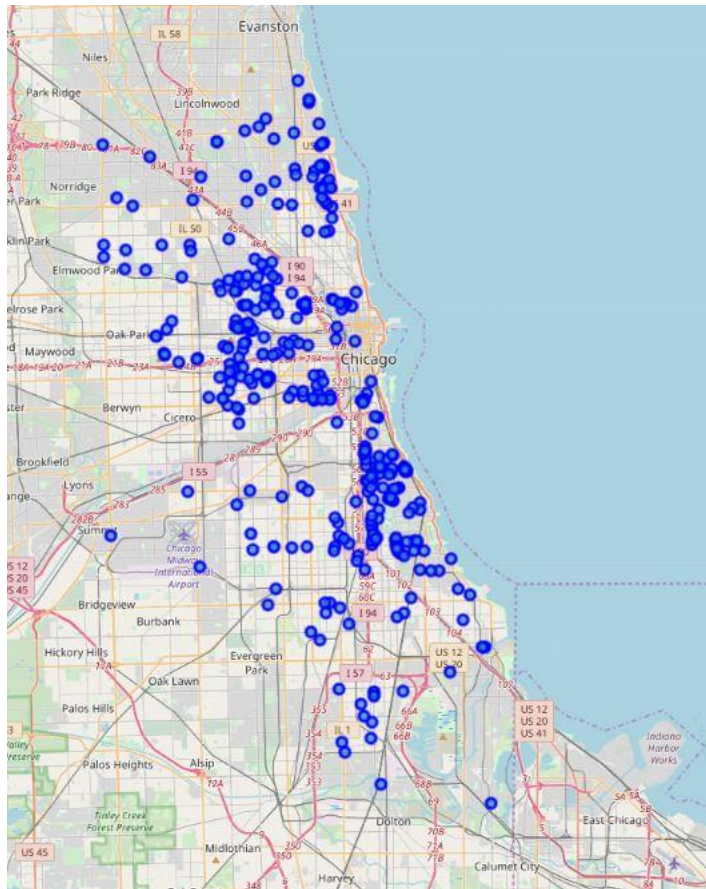

Choice of Community Area

As there are a total of 62 community areas currently with affordable housing developments in Chicago, the first step would be to decide which community area to build another affordable housing development. The criterion used will be the existing total number of units in each community area. It is presumed that the more affordable housing units there already are in a housing unit, the more likely new residents would have a support network nearby.

Figure 1: Cross Section of the horizontal bar graph containing the number of units in each community area

Grand Boulevard is chosen as it has the greatest number of affordable housing units in Chicago (2201 units) compared to the mean of 350 units.

Choice of Property Areas within Grand Boulevard
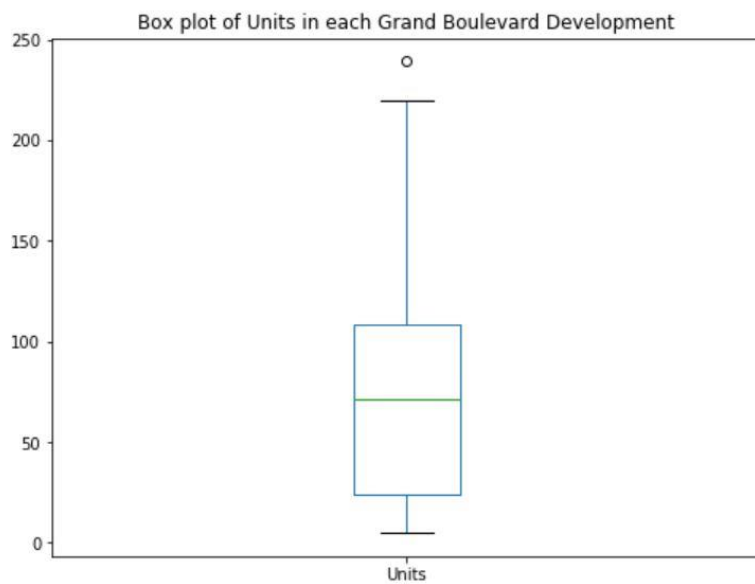
	Units
count	27.000000
mean	81.518519
std	63.778083
min	5.000000
25%	24.000000
50%	71.000000
75%	108.500000
max	239.000000

Based on the box plot and the figures in the statistic description table above, only properties above the 1st interquartile range of 24 units will be selected to ensure that the results from subsequent analysis are significant.

Identify the top 5 locations around each property

Thereafter, Foursquare data will be used to identify the top 5 venues around each property. An example is shown below for each property, and a summary for each property respectively.

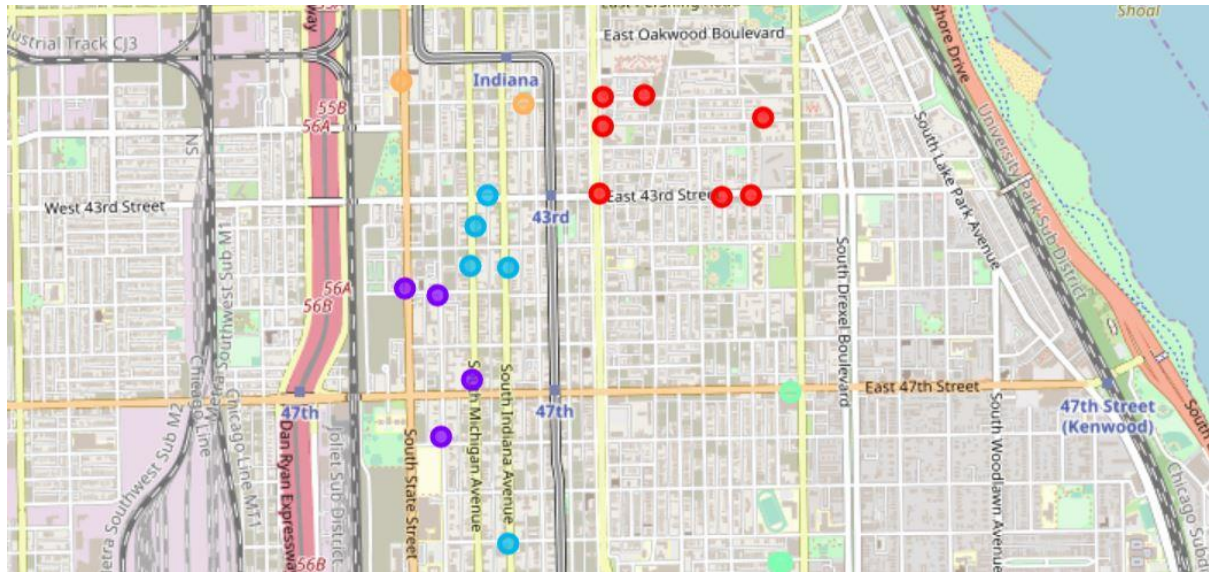|	Property Name	Property Latitude	Property Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Geneva Gables	41.814061	-87.623112	Alice's	41.816453	-87.623775	BBQ Joint
1	Geneva Gables	41.814061	-87.623112	Bronzeville Pizzeria & Grille	41.816615	-87.623032	Fast Food Restaurant
2	Geneva Gables	41.814061	-87.623112	J&J Fish & Chicken	41.816490	-87.620921	Restaurant
3	Geneva Gables	41.814061	-87.623112	Windy City Cigar Lounge	41.816172	-87.620029	Smoke Shop
4	Geneva Gables	41.814061	-87.623112	The Game	41.816226	-87.620010	Sports Bar

	Property Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bronzeville Family Apts.	Grocery Store	Train Station	Coffee Shop	Clothing Store	Sushi Restaurant
1	Bronzeville Senior Apts.	Park	Grocery Store	Coffee Shop	Gas Station	Discount Store
2	Cornerstone Apts.	Baseball Field	Cosmetics Shop	Bus Station	Stadium	Fountain
3	Evans Langley	Baseball Field	Cosmetics Shop	Bus Station	Stadium	Fountain
4	Geneva Gables	Restaurant	Train Station	Sports Bar	BBQ Joint	Smoke Shop

Cluster the data

Subsequently, the data will be clustered through k-means clustering. Clustering is selected because it would be interesting to see how the various properties in Grand Boulevard are grouped based on the proximity of their locations and most common venues. Specifically, k-means clustering is selected because it is computationally faster. Centroids are selected at random in this instance. 5 clusters were chosen as they were deemed to have the most relevant split compared to 3 and 4 clusters.

Results

Visual Representation of the Clusters in Grand Boulevard

Interpretation for each cluster

Cluster 0 – Relaxed Living Cluster

	Property Name	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	Bronzeville Senior Apts.	41.820382	-87.614382	0	Park	Grocery Store	Coffee Shop	Gas Station	Discount Store
6	Hearts United II-The Quincy (scattered sites)	41.816666	-87.609060	0	Coffee Shop	Park	BBQ Joint	Bookstore	Playground
9	Hearts United I-The Langston (scattered sites)	41.819576	-87.608483	0	Coffee Shop	Park	Vineyard	Boutique	History Museum
11	Paul G. Stewart Phase III Tower	41.819229	-87.616452	0	Coffee Shop	Restaurant	History Museum	Clothing Store	Sports Bar
12	Park Boulevard Tower/Grand Renaissance Apts.	41.816750	-87.616649	0	Coffee Shop	Clothing Store	Art Gallery	Grocery Store	Sports Bar
15	Hearts United III-The Leontyne (scattered sites)	41.816602	-87.610556	0	Coffee Shop	Park	BBQ Joint	Boutique	Clothing Store
17	Bronzeville Family Apts.	41.820350	-87.616514	0	Grocery Store	Train Station	Coffee Shop	Clothing Store	Sushi Restaurant

This cluster is in the northeast section of Grand Boulevard and all the properties are to the east of (and close to the) 43rd and Indiana subway stations. The most common venue for this cluster is coffee shops followed by parks and grocery stores which support a more relaxed and wholesome lifestyle.

Cluster 1 – Unhealthy Food Cluster

	Property Name	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3	Legends South - Savoy Square	41.813195	-87.626337	1	Park	BBQ Joint	Fast Food Restaurant	Café	Fried Chicken Joint
7	Progressive Square	41.807699	-87.624560	1	Park	Train Station	Fried Chicken Joint	Liquor Store	Bakery
13	Margaret Ford Manor Independent Living	41.812980	-87.624707	1	Park	BBQ Joint	Fast Food Restaurant	Liquor Store	Bakery
16	Rosenwald Courts Apts.	41.809826	-87.622999	1	American Restaurant	Plaza	Train Station	Liquor Store	Fried Chicken Joint

This cluster is located to the west of Grand Boulevard. The most common venue for this cluster is parks but it is also surrounded by places offering relatively unhealthy food such as fast food, liquor and fried chicken.

Cluster 2 – Tourism Cluster

	Property Name	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Geneva Gables	41.814061	-87.623112	2	Restaurant	Train Station	Sports Bar	BBQ Joint	Smoke Shop
1	Indiana Manor Town Homes	41.813971	-87.621192	2	Restaurant	Art Gallery	Sports Bar	Train Station	BBQ Joint
18	Washington Park SRO	41.803755	-87.621217	2	Restaurant	Sandwich Place	Chinese Restaurant	Convenience Store	Fast Food Restaurant
19	Legends South - Mahalia Place	41.816663	-87.622265	2	Gym	Performing Arts Venue	BBQ Joint	Boutique	Discount Store
20	Legends South - Gwendolyn Place	41.815550	-87.622857	2	Food Truck	Smoke Shop	Park	Bike Rental / Bike Share	Bed & Breakfast

This cluster is located mostly to the west of 43rd station with one location located further south. The most common venue is restaurant and there are attractions such as art galleries and performing arts centres which would attract both local and foreign tourists.

Cluster 3 – Recreation Cluster

	Property Name	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	Cornerstone Apts.	41.803042	-87.607589	3	Baseball Field	Cosmetics Shop	Bus Station	Stadium	Fountain
4	Evans Langley	41.803040	-87.607755	3	Baseball Field	Cosmetics Shop	Bus Station	Stadium	Fountain
10	Shops & Lofts at 47	41.809414	-87.607086	3	Fast Food Restaurant	Convenience Store	Cosmetics Shop	Discount Store	Seafood Restaurant

This cluster is located at the south of Grand Boulevard. The most common venues are related to sports such as baseball fields and stadiums. Make-up is also a prominent feature, with cosmetic shops included as the top 3 venue close to all properties.

Cluster 4 – Town Centre Cluster

	Property Name	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
8	Legends South - Coleman Place	41.820075	-87.620423	4	Southern / Soul Food Restaurant	Flower Shop	BBQ Joint	Convenience Store	Motel
14	Legends South - Hansberry Square	41.820911	-87.626526	4	Motel	Intersection	Train Station	Flower Shop	Southern / Soul Food Restaurant

This cluster is located north of Grand Boulevard and relatively closely to Indiana station. A motel is included as one of the most common locations, along with an intersection and train station, suggesting that this cluster is the hub of the town.

Discussion

Choice of Next Affordable Housing Development: Relaxed Living Cluster 0

The next affordable housing development would best be built in an area close to **Relaxed Living Cluster 0**. This is because the most common venues around it support the most wholesome lifestyle compared to the other clusters. There are many coffeeshops and a park nearby that would serve as spaces for social cohesion. The availability of a grocery store means that residents would have convenient access to daily necessities and can prepare their own food which is healthier and cheaper than eating out all the time. The latter is especially important since the residents, as they stay in affordable housing, are not well-to-do. There are also museums, galleries and bookshops that serve as affordable and healthy entertainment to residents.

On the other hand, the other clusters do not feature venues which are as conducive for healthy living as Cluster 0. Cluster 1 contains mostly unhealthy food options. Cluster 2 has smoke shops that might encourage cigarette smoking. Cluster 3 has recreation options, but they are not as comprehensive as Cluster 0. Cluster 4 is the town centre, which may be convenient for travellers but apparently do not have amenities such as grocery stores that would support healthy living.

Limitations

There are limitations to this project. First, the data set is not representative of **ALL** affordable housing in Chicago and is only a sub-set of what the City of Chicago offers. Thus, there may be other community areas which may be more suitable than Grand Boulevard that might have been missed out in this analysis. This is further supported by the fact that there are only 62 community areas featured in this data set compared to the 77 community areas in the whole of Chicago.

Next, there are many other relevant factors that this project did not take into consideration. For instance, per capita income and crime rate would have been relevant and should be included as it would be preferable to build affordable housing in areas which are relatively middle-class and low crime so that residents can lively safely and comfortably. Most glaringly, the proportion and preference of each racial group living in each community area was not considered. Hence, the choice to locate a new housing development in a community area that is already densely populated with affordable housing may worsen the racial segregation that is endemic in Chicago.

Finally, this project did not consider if there were empty plots of land available in the areas under consideration. If the area is already fully built up, then the city government would either have to find another site or buy over the existing site from owners which may not be feasible given the cost involved.

Conclusion

A raw model with potential to be a viable tool

In conclusion, this project has shown that it is possible to identify areas to build new affordable housing developments with data science. First, the problem was identified. Thereafter, a suitable dataset was found, assumptions about what would constitute an ideal location were made and data science tools such as clustering were used. The results were subsequently visualised with graphs and maps though they are insufficient to deliver the insight expected of data analysts. The final and most crucial step was to interpret the graphs and maps to arrive at insights and reasoning that would uncover and support the decision to be made. This was demonstrated in this project with a firm recommendation of location, which shows that the project has managed to accomplish its objective.

However, there are limitations facing this model as elucidated in the Discussion section above. Should these limitations be overcome, this model will be even more robust and greater reliance can be placed on it to arrive at the optimal location for building new housing developments. The city government would then be pleased to have this tool.

References

CTBA. (2018). A brief overview of Chicago's affordable housing programs. Retrieved from https://budgetblog.ctbaonline.org/a-brief-overview-of-chicagos-affordable-housing-programs-95413a2fb184

DuPaul. (2018). 2018 State of Rental Housing in Cook County. Retrieved from https://www.housingstudies.org/releases/2018-state-rental-housing-cook-county/

Semuels, A. (2018). Chicago's Awful Divide. Retrieved from https://www.theatlantic.com/business/archive/2018/03/chicago-segregation-poverty/556649/