



CHICAGO NEIGHBORHOOD ANALYSIS PROJECT

The Development Company

Date: July 14, 2019

Demetrius Moorer
Dmjc2004@gmail.com

Contents

1. Introduction/Business Problem	2
2. Data Acquisition & Cleaning.....	2
2.1 Data Sources	2
2.2 Data Cleaning	3
3. Methodology.....	3
4. Results.....	4
5. Discussion.....	13
6. Conclusion.....	17
APPENDIX A: Chicago Neighborhood Clusters Extract.....	18

1. Introduction/Business Problem

A new Chicago-area Developer, The Development Company (TDC), is looking to determine which up-and-coming neighborhoods are likely to experience enough price appreciation to support new-construction residential housing.

Demand for new housing units in Chicago is high as new residents are moving in from across the country, looking for spaces with modern amenities & modern finishes. The supply of these types of units is low, leading to bidding wars and trendy neighborhoods on the Northside of the city experiencing rapid price appreciation. TDC would like to work in the historically underserved South and West areas of the city, but they need to understand the underlying market dynamics in these areas.

TDC's business plan is looking for neighborhoods where their finished products can sell for at least \$200 per square foot and within 120 days of being listed for sale.

Note: For purposes of clarity, Chicago Community Areas will be referred to as Neighborhoods.

2. Data Acquisition & Cleaning

2.1 Data Sources

To help TDC understand the underlying market dynamics in the Chicago South and West side neighborhoods, the following data will be collected:

Data	Link to Data	Use Case to Solve Problem
Chicago Community Area Snapshot Dataset from June 2019	https://datahub.cmap.illinois.gov/dataset/community-data-snapshots-raw-data	This data set from the Chicago Metropolitan Agency for Planning (CMAP) is a snapshot of demographics for each of Chicago's 77 neighborhoods. This will allow for obtaining a complete listing of all the neighborhoods in Chicago to conduct analysis activities (i.e. Clustering, Geocoding) at the Neighborhood level.
Google Geocoding API	https://developers.google.com/maps/documentation/geocoding/start	This data source will allow for identifying the center of each Chicago neighborhood for map plotting and retrieval of nearby venues via FourSquare.
FourSquare Explore Venues API	https://developer.foursquare.com/docs/api	This data source will allow for obtaining data on venues in Chicago neighborhoods that are closest to the center of the neighborhood area for a given distance.

		This will be used to determine what basic features each neighborhood has that may be of interest to potential buyers of TDCs products.
Redfin Data Center - Days on Market & Price Per Square Foot 2012 - 2019	https://www.redfin.com/blog/data-center/	This data source will allow for obtaining insights into real estate market activity occurring in the Chicago neighborhoods.

2.2 Data Cleaning

CMAP Chicago Community Area Snapshot

The Original CMAP Dataset contained 231 columns of data for each Chicago neighborhood. Using the Data Dictionary from the CMAP Data Hub, MS Excel was used to drop columns deemed least applicable to the analysis of inherent market value. 108 columns were retained as they were the type of data points that would most contribute to inherent market value (i.e. demographics, proportion of owner-occupants to renters, proportion of vacant land to overall land use).

3. Methodology

Phase One: Population Review

CMAP Chicago Community Area Snapshot

To start the analysis, the final CMAP Community Area Snapshot dataset was run through the KMeans clustering algorithm to cluster the Community Areas. The KMeans algorithm was configured to have 11 clusters. This activity allows for grouping of Chicago neighborhoods that can be investigated further as potential opportunities to start development activities.

Google Geocoding API

The neighborhood name from the final CMAP Community Area Snapshot dataset was used to retrieve location (latitude and longitude) data from Google for all the neighborhoods. A new CCA location dataset was created as a result to create a map and plot the clusters to visualize cluster distribution.

Exploratory Data Analysis

For the selected cluster, visualizations based on CMAP dataset population and median home value data points were generated to understand the neighborhoods in the selected cluster.

Phase Two: Neighborhood Deep Dive

FourSquare Venue Data

For the neighborhoods in the cluster selected at the end of Phase One, the CCA location dataset was parsed to retrieve location data for each neighborhood.

The neighborhoods' location data was used to call the FourSquare Explore API and retrieve venue data for each neighborhood to determine if any of the neighborhoods have venues that potential buyers might be particularly interested in.

The results were captured in a Top Ten Features table to facilitate comparative analysis of neighborhood features.

Phase Three: Neighborhood Real Estate Analysis

Last, for the neighborhoods in the selected cluster, Price Per Square Foot and Days on Market data was retrieved from Redfin's Data Center for Single-Family Attached, Single-Family Detached and Condos.

One assumption made is that the Days on Market extracts for each property type would provide insights into how fast a newly built TDC product can sell while Price Per Square Foot data would provide insights into whether the neighborhood can support the cost of new construction plus profit.

Last, TDC used Box Plots to review Price Per Square Foot and Days on Market data for each property to determine Maximum Price Per Square Foot, Median Days on Market, and detect potential outliers. Based on the review of this data, the top 3 neighborhoods for each property type were selected and Price Per Square Foot line charts (less outliers) were generated to determine if the price appreciation was trending upward or downward for the selected neighborhoods.

4. Results

Population Review Results

After running the KMeans Algorithm, 11 clusters were returned by the algorithm. See Appendix A for the extract. The following table has the cluster results and the Cluster Profile based on the median home value (MED_HV) and the Population fields.

Cluster #	Cluster Color	Neighborhoods	Cluster Profile
0	Red	Ashburn Auburn Gresham Brighton Park Chatham Chicago Lawn East Side Englewood Gage Park Greater Grand Crossing New City North Lawndale Roseland South Chicago Washington Heights West Englewood West Pullman	South and Southwest side neighborhoods with low median home values with declining population.

1	Purple	Lincoln Square Logan Square Near West Side	Northwest and West side neighborhoods with high median home values and increasing population.
2	Dark Blue	Armour Square Douglas Hermosa Hyde Park Kenwood Lower West Side Montclare Mount Greenwood Oakland	South, Southwest, and Northwest side neighborhoods with moderately high median home values and increasing population.
3	Medium Blue	Archer Heights Avalon Park Calumet Heights Clearing East Garfield Park Garfield Ridge Grand Boulevard McKinley Park Morgan Park O'Hare West Elsdon West Lawn Woodlawn	Northwest, West, South, and Southwest side neighborhoods with moderately low median home values and increasing population.
4	Sky Blue	Avondale Beverly Bridgeport Jefferson Park North Park Norwood Park	Northwest and South side neighborhoods with high median home values and increasing population.
5	Aqua Green	North Center	North side neighborhood with very high median home values and increasing population.
6	Light Green	Lake View Lincoln Park Near North Side West Town	North and West side neighborhoods with high median home values and increasing population.
7	Medium Green	Austin Belmont Cragin Humboldt Park Rogers Park South Lawndale South Shore	North, West, and South side neighborhoods with low median home values and declining population.

8	Dark Yellow	Albany Park Dunning Edgewater Irving Park Portage Park Uptown West Ridge	North and Northwest side neighborhoods with moderately high median home values and increasing population.
9	Orange	Edison Park Forest Glen Near South Side The Loop	Northwest, Central, and South side neighborhoods with high median home values and increasing population.
10	Dark Orange	Burnside Fuller Park Hegewisch Pullman Riverdale South Deering Washington Park West Garfield Park	South and West side neighborhoods with very low median home values and declining population.

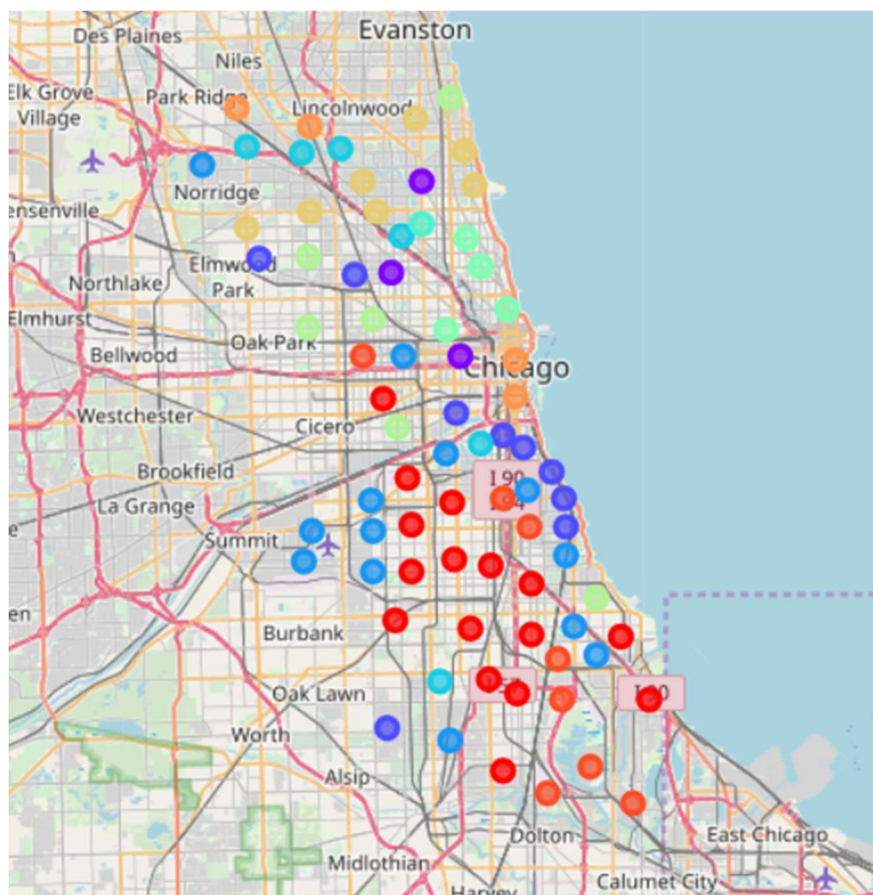


Figure 1: Map of Chicago Neighborhood Clusters

Exploratory Data Analysis

Cluster #2 was selected given that it's cluster profile best matched TDC's business objectives stated in the Introduction.

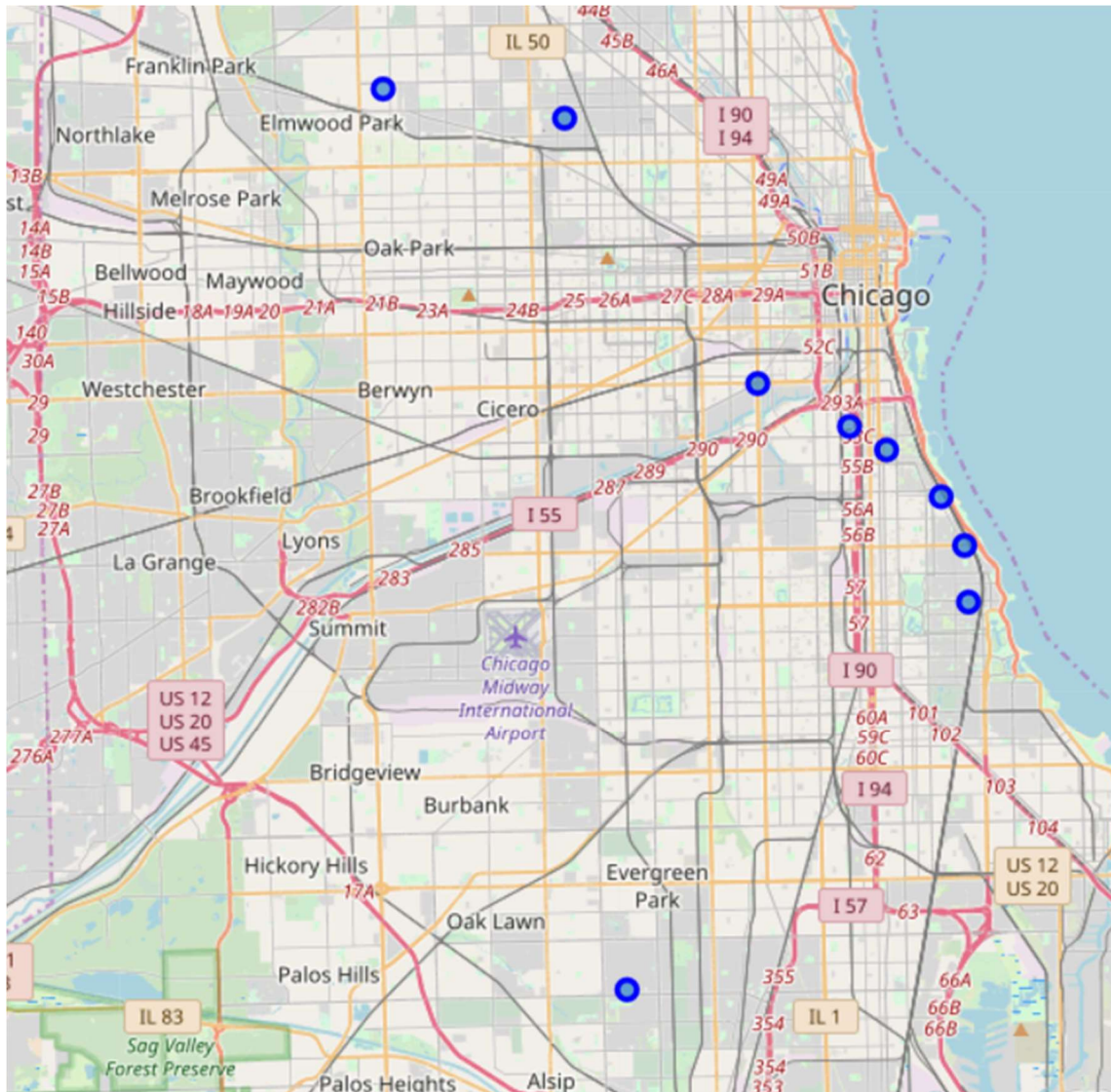


Figure 2: Cluster 2 Chicago Neighborhoods

Using the Median Home Value and Population fields from the CMAP Data Extract, the following visuals were created.

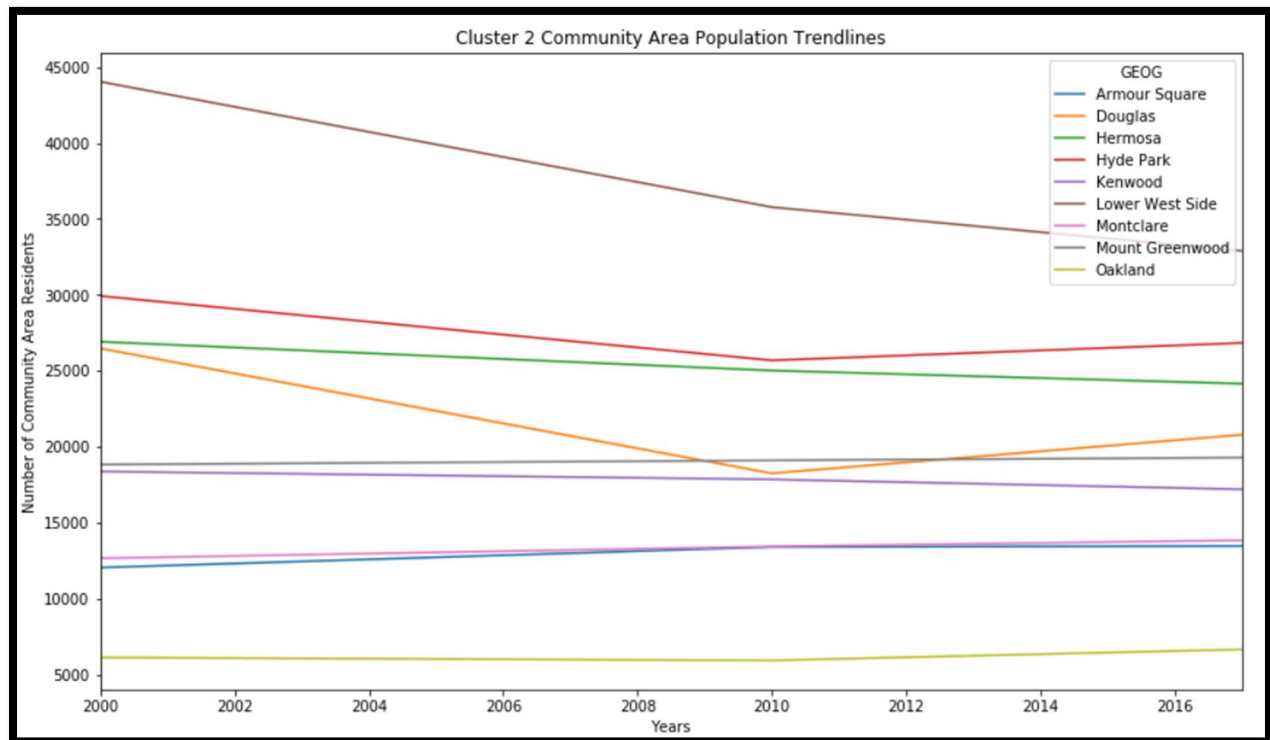


Figure 3: Population Trend Lines for Chicago Neighborhoods in Cluster 2

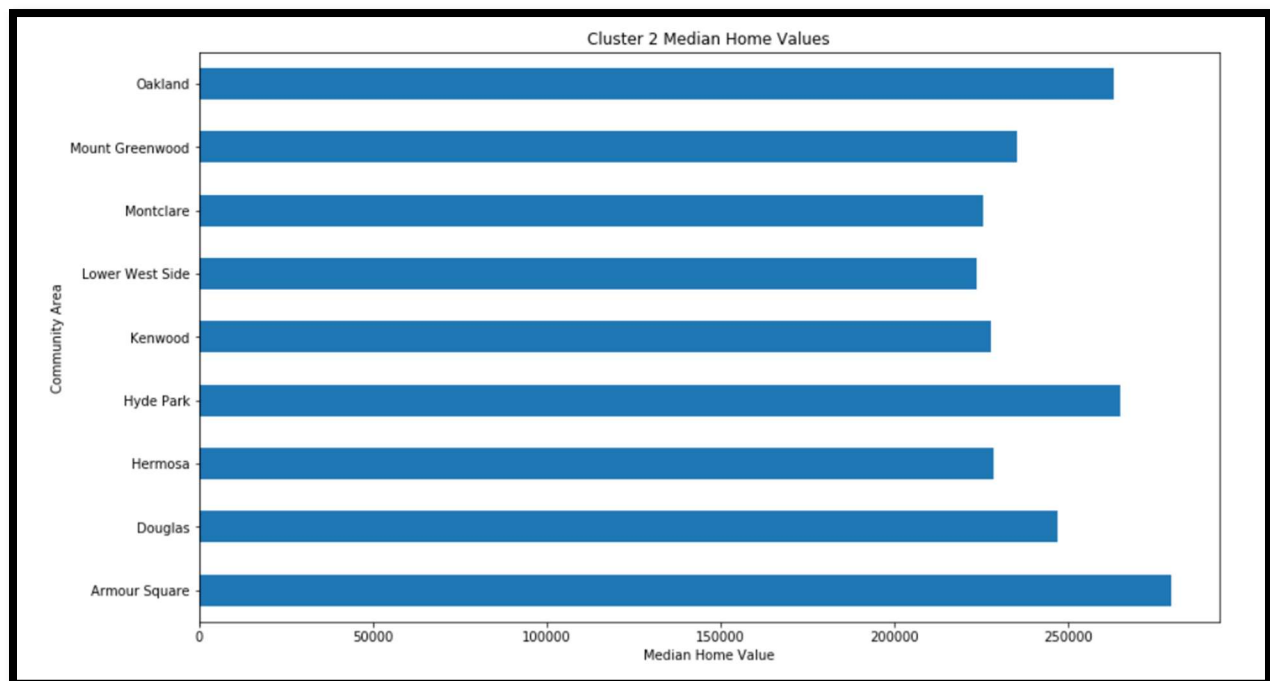


Figure 4: Median Home Values for Chicago Neighborhoods in Cluster 2

Overall, these visuals support the profile assigned to the cluster earlier in this section. This cluster has neighborhoods that are trending towards steady increase in population going forward and the median home values are in the mid \$200,000 range.

Neighborhood Deep Dive

Taking a deep dive into the neighborhoods of Cluster 2, here are the results of the FourSquare analysis of the Top Ten venues types for each neighborhood:

	Neighborhood	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	Armour Square	Chinese Restaurant	Sports Bar	Italian Restaurant	Sandwich Place	Breakfast Spot	Cosmetics Shop	Asian Restaurant	Gas Station	Hot Dog Joint	Fried Chicken Joint
1	Douglas	Fast Food Restaurant	Fried Chicken Joint	Restaurant	Cosmetics Shop	Sandwich Place	Eye Doctor	Wings Joint	Historic Site	Ice Cream Shop	Currency Exchange
2	Hermosa	Mexican Restaurant	Park	Greek Restaurant	Historic Site	Latin American Restaurant	BBQ Joint	Arcade	Gay Bar	Grocery Store	French Restaurant
3	Hyde Park	Bookstore	Sandwich Place	Café	Train Station	Park	Bike Rental / Bike Share	Coffee Shop	Mexican Restaurant	Bus Station	Middle Eastern Restaurant
4	Kenwood	Bus Station	Beach	Gym / Fitness Center	Scenic Lookout	Boutique	Breakfast Spot	Chinese Restaurant	Clothing Store	Cosmetics Shop	Donut Shop
5	Lower West Side	Mexican Restaurant	Plaza	Boat or Ferry	Music Venue	Convenience Store	Pizza Place	Business Service	Brewery	Snack Place	Supermarket
6	Montclare	Pizza Place	Flea Market	Fast Food Restaurant	Mexican Restaurant	Dry Cleaner	Convenience Store	Cosmetics Shop	Currency Exchange	Deli / Bodega	Diner
7	Mount Greenwood	Park	Gym	Sports Bar	Performing Arts Venue	Deli / Bodega	Dry Cleaner	Pizza Place	Convenience Store	Cosmetics Shop	Currency Exchange
8	Oakland	Beach	Park	Yoga Studio	Eye Doctor	Cosmetics Shop	Currency Exchange	Deli / Bodega	Diner	Dive Bar	Donut Shop

Figure 5: Table of Top Ten Venue Types for Chicago Neighborhoods in Cluster 2

Per the results, Oakland, Hyde Park, and Kenwood have amenities that can be attractive to new neighborhood residents (i.e. Beach, Gym, Yoga Studio, Cafe, Park).

Neighborhood Real Estate Analysis

Last, let's review the visuals from the Redfin Data for the Cluster 2 Neighborhoods and start with Single-Family Detached homes.

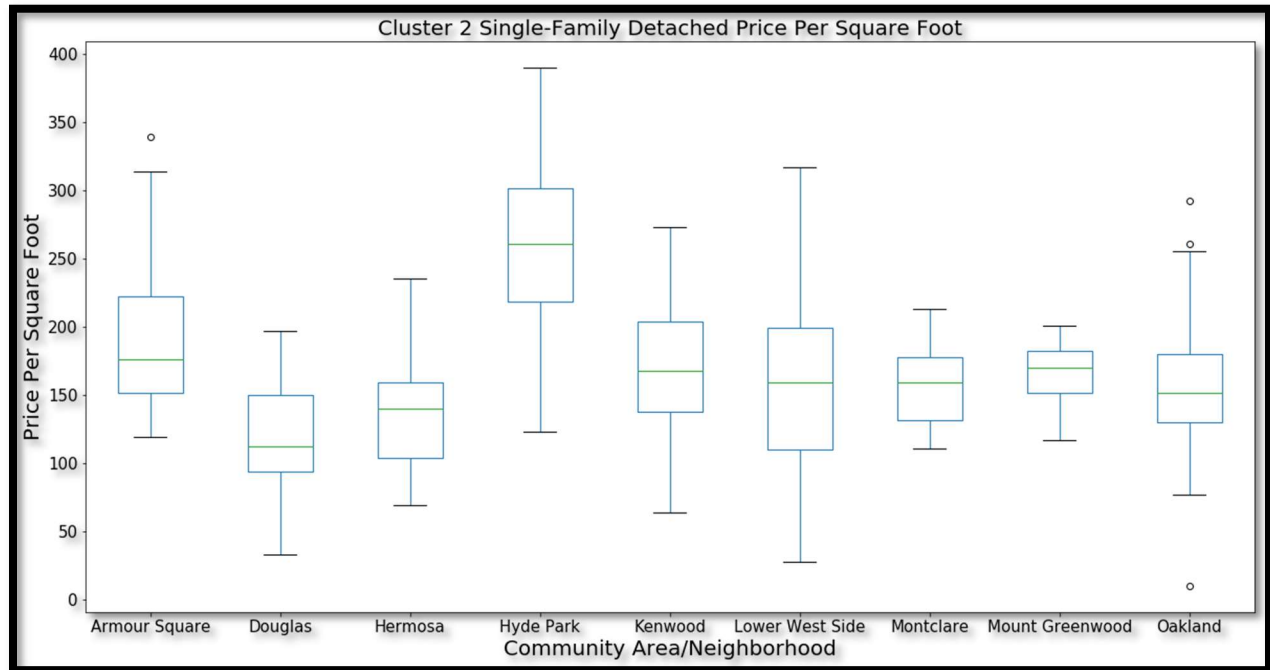


Figure 6: Single-Family Detached Price Per Square Foot Box Plot Diagram for Chicago Neighborhoods in Cluster 2

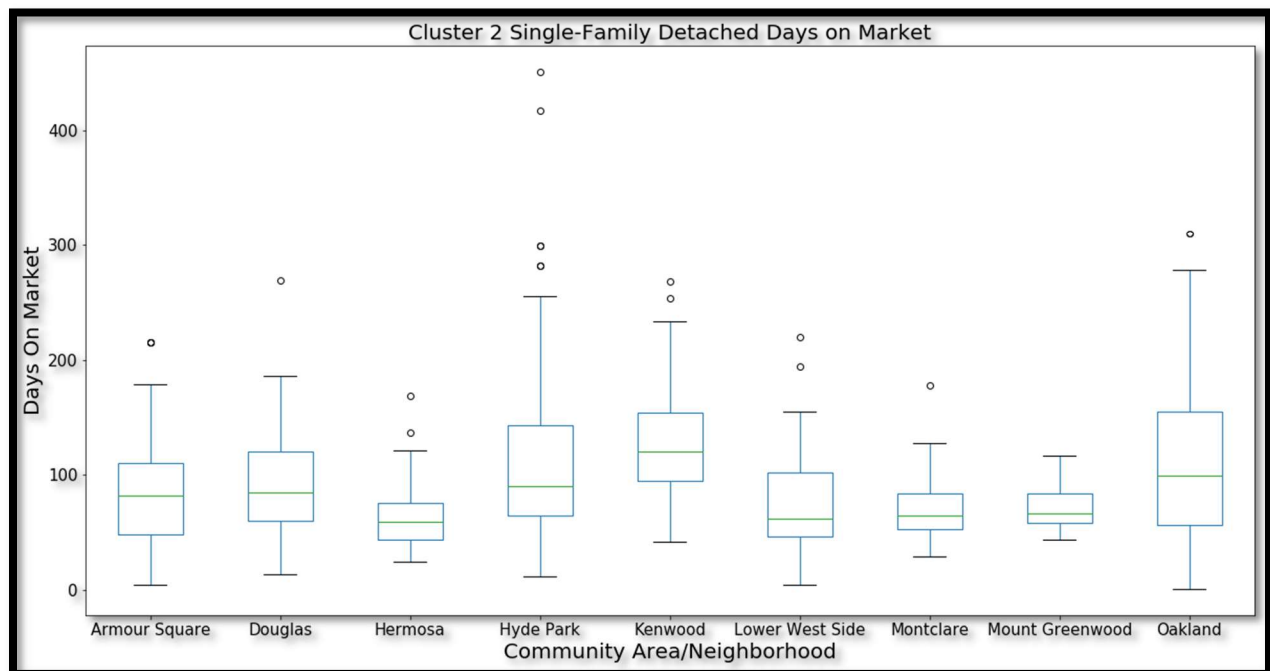


Figure 7: Single-Family Detached Days on Market Box Plot Diagram for Chicago Neighborhoods in Cluster 2

Single-Family Detached homes in the cluster have been selling with median Price Per Square Foot of \$120 – \$260. If a typical Single-Family Detached home is assumed to be around 3000 Square Feet (1000 sq. ft. per floor), this would put the final purchase price between \$360,000 and \$780,000.

The median Days on Market for Single-Family Detached houses in the cluster are mostly coming in under 100 days. Assuming the numbers include time for buyers to obtain financing (approximately 45 days), the data suggests that homes are selling within 30-45 days of being listed.

Moving on to Single-Family Attached homes (a.k.a. Townhouses), below are the visual from the Redfin data.

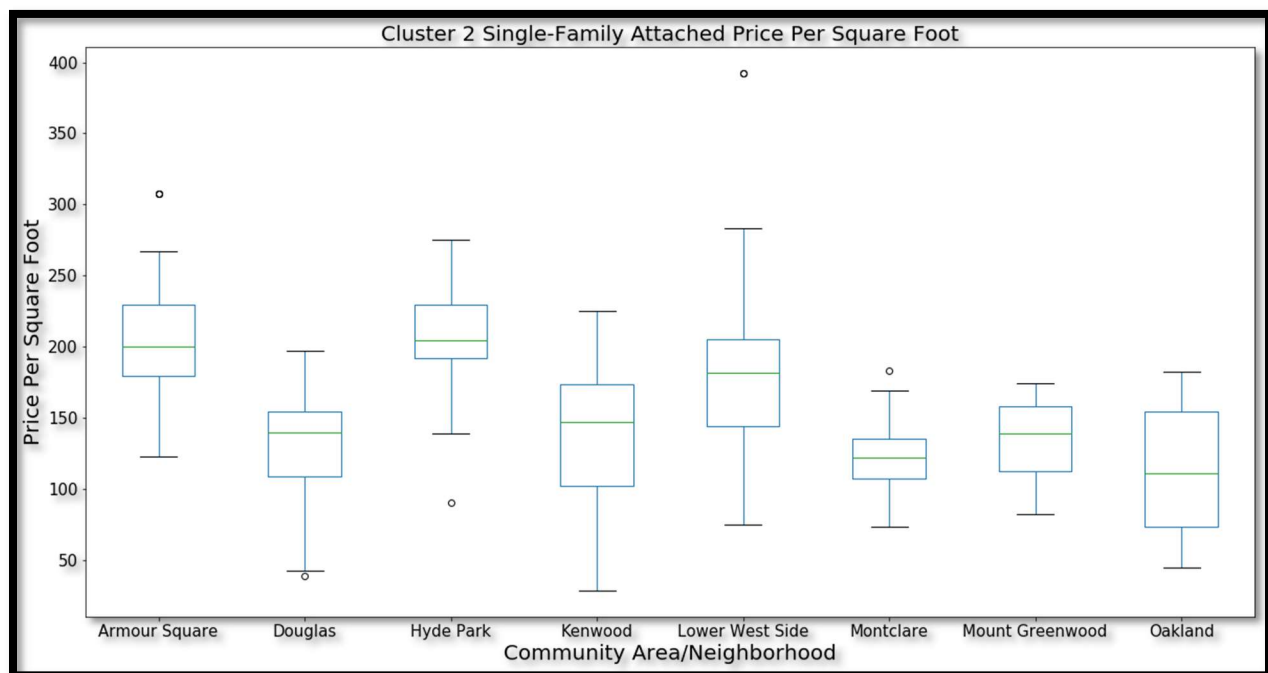


Figure 8: Single-Family Attached Price Per Square Foot Box Plot Diagram for Chicago Neighborhoods in Cluster 2

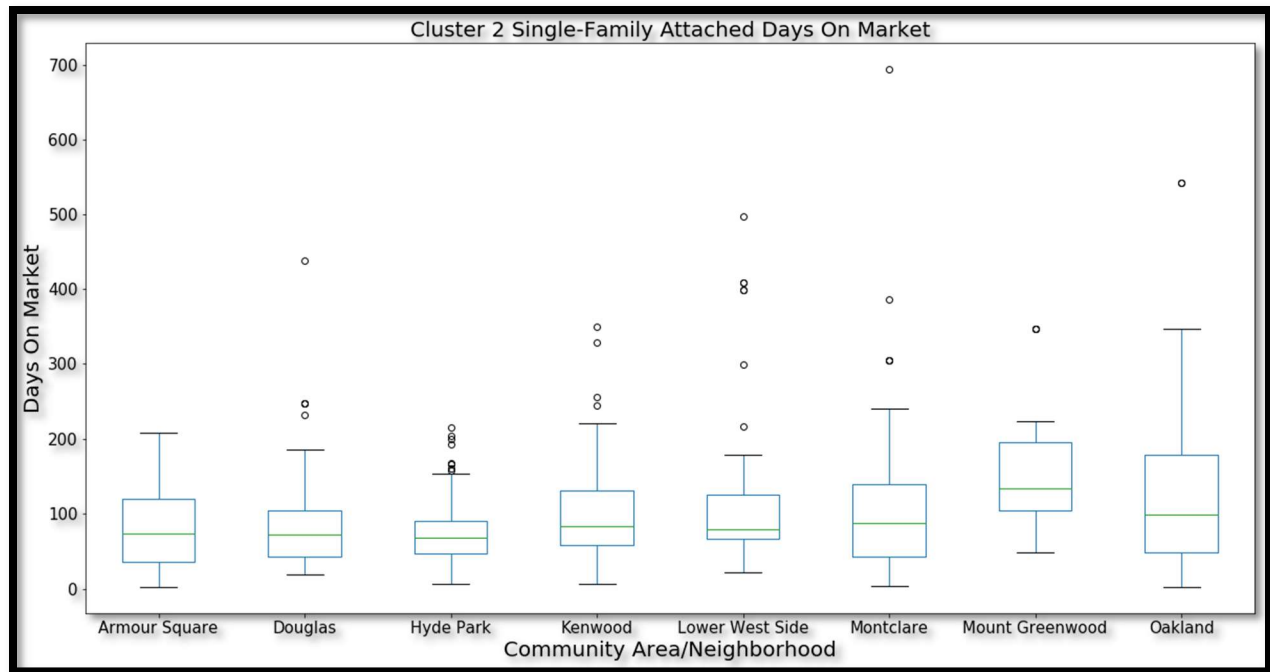


Figure 9: Single-Family Attached Days on Market Box Plot Diagram for Chicago Neighborhoods in Cluster 2

Single-Family Attached homes in the cluster have been selling with median Price Per Square Foot of \$140 – \$210. If a typical Single-Family Attached home is assumed to be around 2400 Square Feet (800 sq. ft. per floor), this would put the final purchase price between \$336,000 and \$504,000.

The median Days on Market for Single-Family Attached houses in the cluster are mostly coming in between 90 - 140 days. Assuming the numbers include time for buyers to obtain financing (approximately 45 days), the data suggests that homes are selling within 45-95 days of being listed.

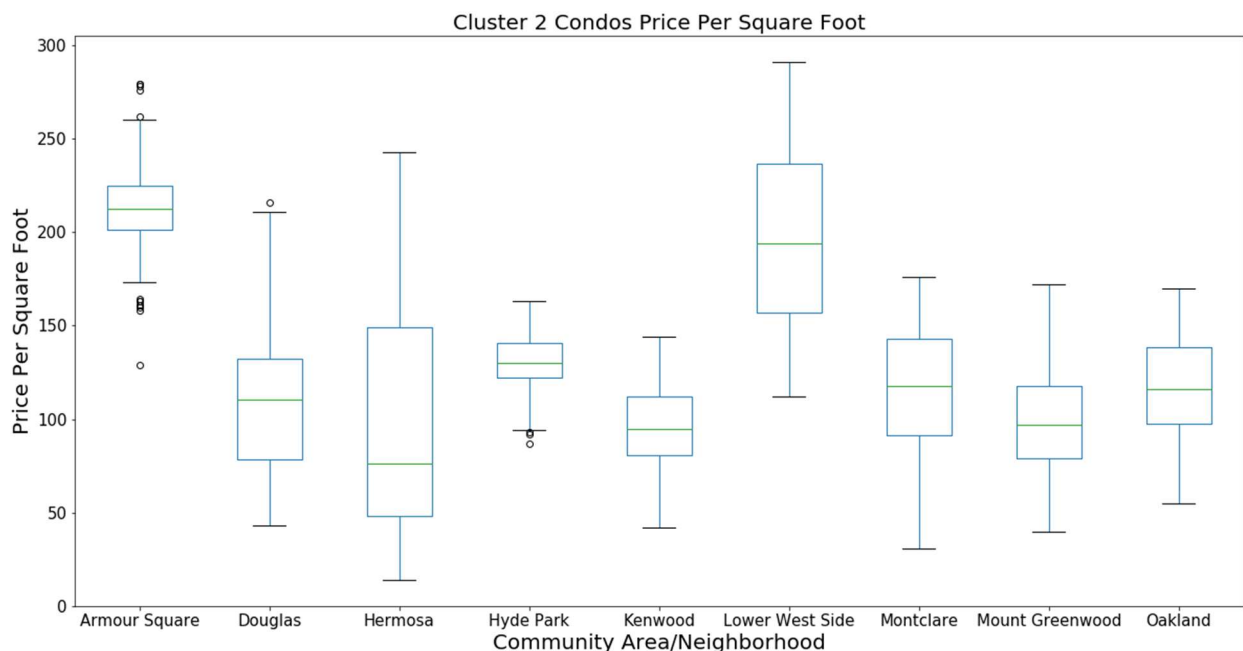


Figure 10: Condos Price Per Square Foot Box Plot Diagram for Chicago Neighborhoods in Cluster 2

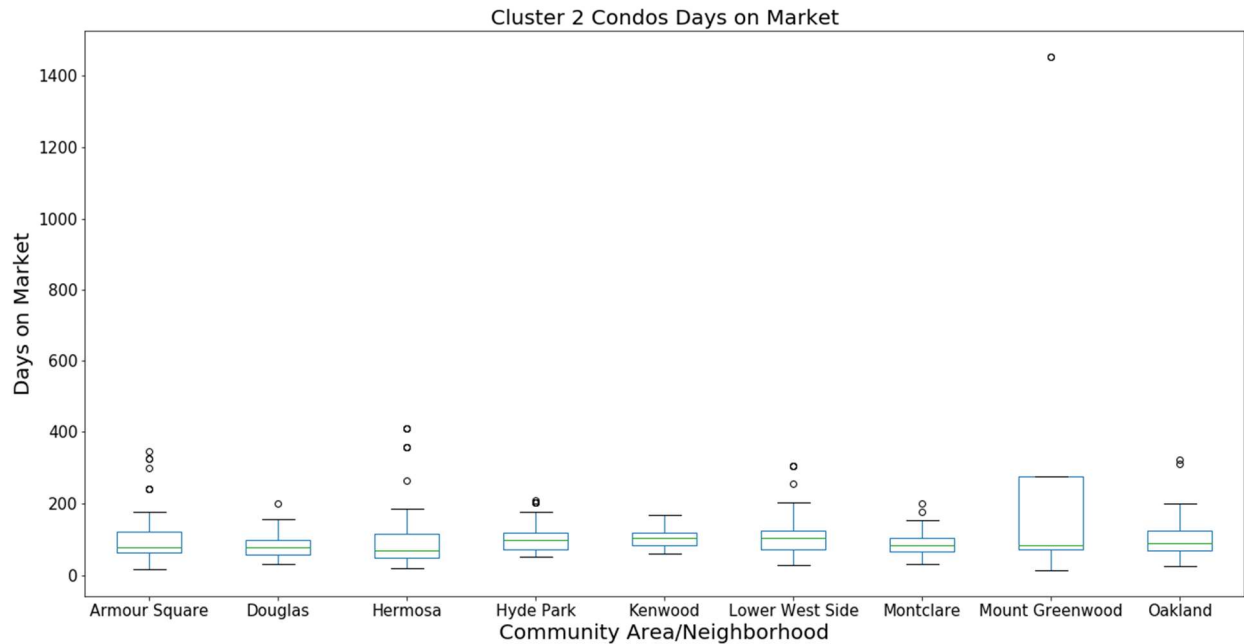


Figure 11: Condos Days on Market Box Plot Diagram for Chicago Neighborhoods in Cluster 2

Condos in the cluster have been selling with median Price Per Square Foot of \$75 – \$220. If a typical 2-bedroom, 2-bath Condo is assumed to be around 1200 Square Feet, this would put the final purchase price between \$90,000 and \$264,000.

The median Days on Market for Condos in the cluster are mostly coming in between 100 - 130 days. Assuming the numbers include time for buyers to obtain financing (approximately 45 days), the data suggests that homes are selling within 55-85 days of being listed.

5. Discussion

Part 1: Population Review Discussion

Out of the 11 clusters that were created by the KMeans algorithm, the cluster that best matched TDC's business plan is Cluster 2 due to its good combination of moderately high home values and the population level is trending upward.

One observation of note is Cluster 5, which contains the northside neighborhood of North Center. In most instances, Lincoln Park and Lakeview are deemed to be the most expensive neighborhoods. However, the clustering analysis identified that North Center has a median home value that is \$100,000 more than those in Cluster 6, which contains Lincoln Park and Lakeview.

Part 1: Exploratory Data Analysis Discussion for Cluster 2

One observation to be called out is that while the population trend line for the Lower West Side neighborhood, which includes popular up-and-coming sub-neighborhood Pilsen, shows a downward trend, this doesn't fully explain what is happening in the area. Per further review of the Chicago Development Community's new network, Curbed Chicago, this area is experiencing gentrification where older dense buildings (i.e. Multi-Family Residential) are being replaced by Single-Family Detached

homes, which could make it appear that the neighborhood is losing population and on a downward trend.

To not contribute to the negative perception of gentrification, it is strongly recommended that seeking to do projects on vacant land first before tearing down pre-existing buildings that may have historical significance.

A second observation regarding the Median Home Values chart is the median home values for the Armour Square neighborhood, at almost \$300,000. It is not clear if the proximity to the CTA Red Line or Guaranteed Rate Field (home of the Chicago White Sox) is what gives this area its median home value. It is recommended to prioritize Armour Square for first projects if the current cost of land acquisition is reasonable.

Part 2: Neighborhood Deep Dive

The Oakland neighborhood has several features that can be the start of a new amenities space on the southside beyond McCormick Square.

The Beach, Parks, and Yoga Studios in Oakland are features that can draw in residents with disposable income and allow for the establishment of new services that are seen in abundance on the North side (i.e. Sidewalk Cafes, Coffee Shops) and require densely populated areas to thrive.

Oakland is directly east of the Douglas neighborhood, which has a large amount of vacant land due to a large amount of homes that had been torn down due to significant deferred maintenance or because it had been previously Public Housing.

Vacant land in the parts of Douglas that border Oakland are recommended to buy because it may be less expensive. Projects in that area can act as “near-prime” housing that can start moving neighborhood rebuilding activities westward.

Part 3: Neighborhood Real Estate Analysis

Because Real Estate Developers are fundamentally putting brand-new housing units into the markets they operate in, it is recommended that the maximum price per square foot be looked at for each neighborhood. Additionally, to ensure that any newly constructed housing units do not eat into profits, looking at median, and possibly minimum for days on market is recommended. Developers can begin marketing activities during Construction phase, reducing time on market once construction is complete.

For each property type, recommendations based on estimates are provided below. The prioritization accounts for likelihood of being able to acquire property at reasonable costs as the key input into the development projects. Additionally, priority one neighborhoods reflect alignment to the business priorities stated in the Introduction.

Single-Family Detached:

Figures 6 & 7 show that Montclare, Douglas, and Hermosa best fit the above criteria for construction of new Single-Family Detached homes. To determine if these are the right neighborhoods to consider, a trend line chart was constructed to see Price Per Square Foot.

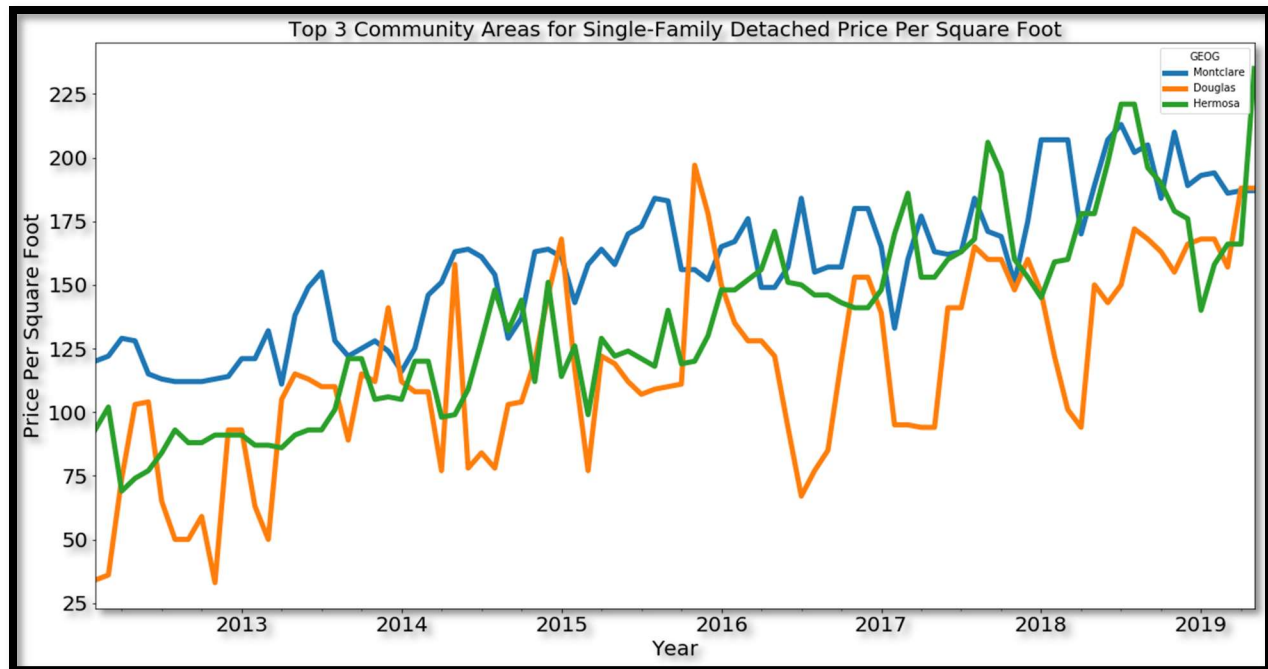


Figure 12: Single-Family Detached Price Per Square Foot for Top 3 Neighborhoods in Cluster 2

Based on a 3000 sq. ft. home, as discussed in the Results section, the following table was constructed:

Priority	Neighborhood	Estimated Sales Price	Estimated Days on Market
1	Hermosa	\$681,000 @ \$227/sq. ft.	150
2	Douglas	\$562,500 @ \$187.5/sq. ft.	150
3	Montclare	\$562,500 @ \$187.5/sq. ft.	200

Single-Family Attached:

Figures 8 & 9 show that Armour Square, Hyde Park, and Lower West Side best fit the above criteria for construction of new Single-Family Attached homes. To determine if these are the right neighborhoods to consider, a trend line chart was constructed to see Price Per Square Foot.

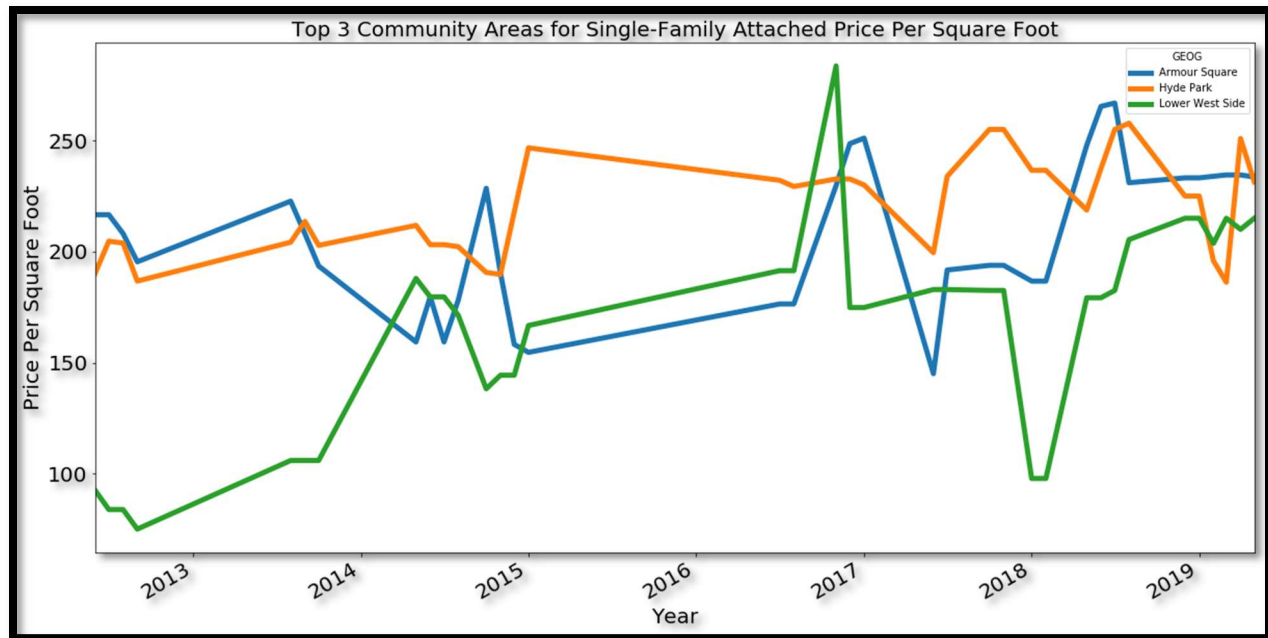


Figure 13: Single-Family Attached Price Per Square Foot for Top 3 Neighborhoods in Cluster 2

Based on a 2400 sq. ft. home, as discussed in the Results section, the following table was constructed:

Priority	Neighborhood	Estimated Sales Price	Estimated Days on Market
1	Armour Square	\$576,000 @ \$240/sq. ft.	150
2	Hyde Park	\$576,000 @ \$240/sq. ft.	150
3	Lower West Side	\$528,000 @ \$220/sq. ft.	150

Condos:

Figures 10 & 11 show that Armour Square, Douglas, and Lower West Side best fit the above criteria for construction of new Condos. To determine if these are the right neighborhoods to consider, a trend line chart was constructed to see Price Per Square Foot.

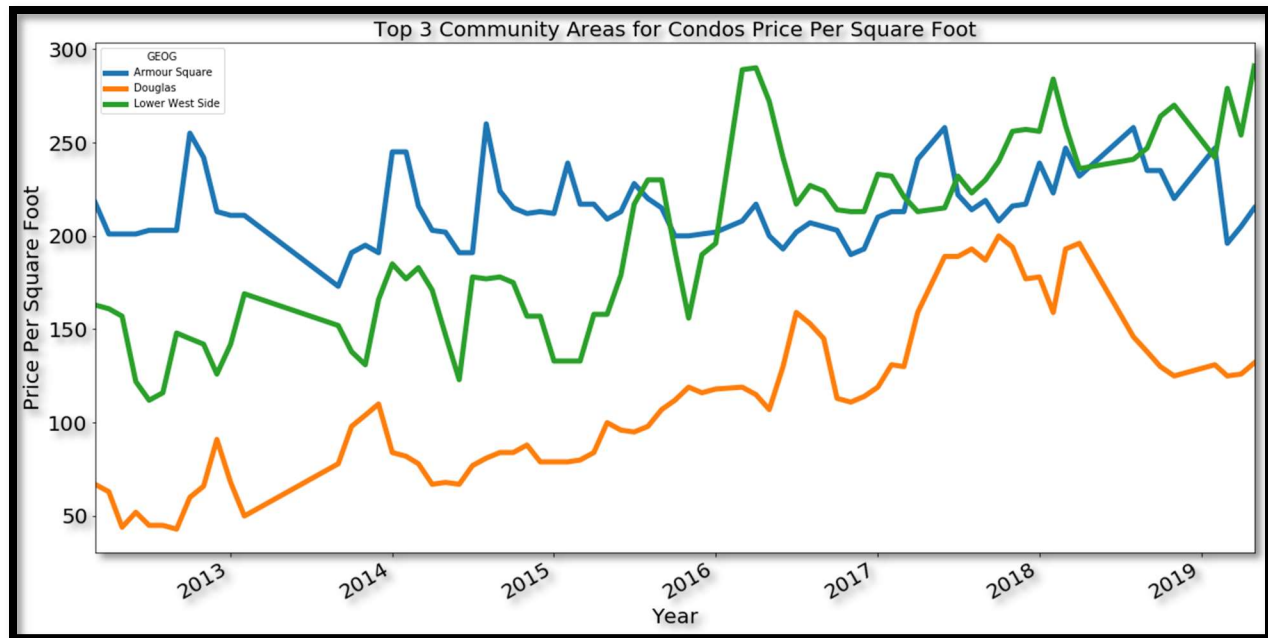


Figure 14: Condos Price Per Square Foot for Top 3 Neighborhoods in Cluster 2

Based on a 1200 sq. ft. home, as discussed in the Results section, the following table was constructed:

Priority	Neighborhood	Estimated Sales Price	Estimated Days on Market
1	Armour Square	\$252,000 @ \$210/sq. ft.	120
2	Lower West Side	\$336,000 @ \$280/sq. ft.	120
3	Douglas	\$168,000 @ \$140/sq. ft.	90

6. Conclusion

Closing Thoughts

Overall, the underlying market dynamics explored in this report show that early investment in near-prime South and West-side neighborhoods can yield good profit potential while delivering community benefits.

In the specific case of Douglas, with booming neighbors such as Oakland and Armour Square, investments in re-densifying Douglas can be a good investment for the developers (i.e. increasing profit potential) while the neighborhood residents get increased access to more and higher-quality services.

Future Projects

Logistic Regression models taking in market data for the Cluster 2 neighborhoods and predicts estimated sales price of finished products.

APPENDIX A: Chicago Neighborhood Clusters Extract



cca_clusters.csv