



## Capstone Project

# 1. Redfin Web Scraper

A scraping tool to extract Redfin web data for the Zip Codes of interest.

## **Introduction:**

To fortify our Market Research and better price our homes, in addition to the new housing competitor data, we also require resale housing (MLS) data. There are various data aggregators available to source this data from and one such is Redfin – a premier MLS listing search engine. Unfortunately, there is no API to tap in and extract data periodically from Redfin.

This is where we would need to create a Redfin Web Scraper to gather resale data by providing Zip codes as the search points.

## **Goal:**

Create a web scraper to scrape Redfin webpages and extract the meaningful information. We want to obtain key data points such as (See Appendix section for more details):

- Property Details like Price, #Beds, #Baths and Sq Ft.
- Text
- Listed by and Last Updated
- Home Facts
- Price Insights
- Property Details
- Schools
- Public Facts
- Transportation
- Affordability Index and perform Price Threshold Testing. Currently this is obtained manually, and we wish to automate this.

**Routine:** Monthly. Additionally, include Period Date with the extracted data.

## **Outcome:**

A dataset which includes key information from Redfin.

## **Measure of success:**

- Automated data extraction.
- An end-to-end data extraction pipeline that yields the desired data routinely.

**Expected output file format:** Any format suitable for consumption such as JSON, CSV or Parquet.

**Version Control:** GitHub

## 2. Cost of Living Estimator:

A tool to estimate cost of living for various Zip codes.

### **Introduction:**

This tool will help us obtain the cost-of-living estimation for a Zip code of interest as cost-of-living influences various lifestyle choices we make, and this can help analyzing how cost of living affects the housing price in an area.

### **Goal:**

Build Cost of Living estimator.

**Reference:** <https://nycdatascience.com/blog/student-works/cost-of-living-estimator/>

## 3. Application Development:

An interactive visualization application to display MLS Data.

### **Introduction:**

The current Pricing Analytics (PA) application is designed as a quick reference tool to view competitors around a Stanley Martin project and benchmark pricing with the competitors.

We would like to build further upon this and leverage the extracted MLS data. This can be achieved either by integrating the MLS data to the current app or create a new framework.

This project includes creating the interactive dashboards to display the MLS data.

### **Goal:**

Create an application that leverages the extracted MLS data and run visualizations.

This includes following features/visualizations:

1. Map (OpenStreetMap) to visualize distance between the SMH products and the listed resale houses.
2. Scatter Plot: Avg Base Price vs Avg Sq Feet.
3. Hypothesis testing.
4. Z-score calculation.
5. Box-Whisker Plot.
6. Correlation Heatmap.

### **Outcome:**

A deployable application to display housing resale data.

## 4. Price Prediction and Analysis

### Introduction:

After we have acquired the necessary data from Redfin, we would want to suggest or predict pricing of our products. Currently we price our products based on some market research and this data-oriented approach will accelerate our pricing endeavors.

To further enrich the dataset, we desire to bring in Federal Reserve Economic Data (FRED - Unemployment Rate, Market Hotness, Economic Indicators, Producer Price Index and House Price Indexes), Market Hotness Data and Building Permits data.

These data points can be used to perform analysis such as:

- How FRED indicators affect the housing prices?

### Goal:

- Enrich data, extract meaningful features, and predict housing prices.
- Perform text analysis on the extracted text.

### Datasets:

1. Redfin MLS Data.
2. Zonda (Active New Housing Competitor) Data.
3. SMH Internal Data.
4. Market Research (FRED).
5. Housing Indicators (FRED).
6. Demographic Information.
7. Text Analysis

### Data Preparation:

- Merge the datasets.
- Perform the necessary feature engineering to obtain the desired features.

### Target Variable:

- Housing Price

### Outcome:

- A Machine Learning model to predicts the housing prices.
- Sentiment Analysis of the obtained text.

## APPENDIX

1518 Scandia Cir, Reston, VA 20190

**\$500,000**

Buy with Redfin: **\$498,224** >

**3**

Beds

**2.5**

Baths

**1,680**

Sq Ft

### Text:

CONTEMPORARY TOWN HOUSE IN SOUGHT AFTER BATANA WOODS IN RESTON. THIS HOUSE FEATURES HARDWOOD FLOORS ON MAIN AND SECOND LEVEL, NEW APPLIANCES WITH GRANITE IN KITCHEN, DINNING ROOM OFF KITCHEN LIVING ROOM LEADING TO PORCH WITH NEW SLIDING DOORS., NEW HVAC SYSTEM NEW PLUMBING, ROOF IS 9 YEARS OLD. CROWN MOLDINGS IN DINNING ROOM THIS TOWN HOUSE IS IN EXCELLENT MOVE IN CONDITION. NEW WASHER AND DRYER WILL BE INSTALLED. IN UTILITY ROOM FRONT LANDSCAPING UPDATED

### Listed by and Last Updated:

Listed by Walt Sobie • Long & Foster Real Estate, Inc.

Redfin last checked: [5 minutes ago](#) | Last updated Nov 8, 2021 • Source: BRIGHT MLS

### Home Facts:

#### Home Facts

Status	Active	Time on Redfin	57 days
Property Type	Townhouse	HOA Dues	\$150/month
Year Built	1973	Style	Contemporary
Community	RESTON SEC 19 BLK 2 BENT	Lot Size	1,587 Sq. Ft.
MLS#	VAFX2019762		

Figure 1: House Facts (MLS # is an important data point here)

### Price Insights:

#### Price Insights

List Price	\$500,000	Est. Mo. Payment	\$2,397
Redfin Estimate	\$497,696	Price/Sq.Ft.	\$298
Buyer's Brokerage Commission	2.5%		

## Property Details:

### Property Details for 1518 Scandia Cir

Interior Features	
<b>Bedroom Information</b> <ul style="list-style-type: none"><li># of Bedrooms On 1st Upper Level: 3</li></ul> <b>Bathroom Information</b> <ul style="list-style-type: none"><li># of Bathrooms (Full) On 1st Upper Level: 1</li><li># of Bathrooms (Half) On Main Level: 1</li><li># of Bathrooms (Full) On 1st Lower Level: 1</li><li># of Bathrooms (Full) On All Upper Levels: 1</li><li># of Bathrooms (Full) On All Lower Levels: 1</li><li># of Bathrooms (Half): 1</li><li># of Bathrooms (Full): 2</li></ul> <b>Interior Information</b> <ul style="list-style-type: none"><li>Appliances: Built-In Microwave, Built-In Range, Dishwasher, Disposal, Dryer - Electric, Exhaust Fan, Oven/Range - Gas, Refrigerator, Washer, Water Heater</li><li>Door Features: Insulated, Sliding Glass, Storm</li><li>Flooring Type: Ceramic Tile, Hardwood</li></ul>	<ul style="list-style-type: none"><li>Not Furnished</li><li>Interior Features: Built-Ins, Ceiling Fan(s), Crown Moldings, Family Room Off Kitchen, Floor Plan - Open, Formal/Separate Dining Room, Kitchen - Galley, Window Treatments, Wood Floors</li><li>Living Area Sq.Ft. Source: Assessor</li><li>Wall &amp; Ceiling Types: 9+ Ceilings, Dry Wall</li></ul> <b>Room Information</b> <ul style="list-style-type: none"><li>Living Room, Dining Room, Bedroom 2, Bedroom 3, Kitchen, Family Room, Bedroom 1, Storage Room, Bathroom 1</li><li>Laundry Type: Lower Floor, Washer In Unit</li></ul> <b>Basement Information</b> <ul style="list-style-type: none"><li>Fully Finished</li></ul>
Parking / Garage, Homeowners Association, School / Neighborhood, Utilities	
<b>Parking Information</b> <ul style="list-style-type: none"><li># of Assigned Spaces: 2</li><li>Assigned Parking Space Number(s): 18</li><li>Assigned</li><li>On Street</li></ul> <b>Homeowners Association Information</b> <ul style="list-style-type: none"><li>HOA/Condo/Coop Fee Includes: Common Area Maintenance, Fiber Optics at Dwelling, Lawn Maintenance, Management, Pool(s)</li><li>HOA/Condo/Coop Amenities: Bike Trail, Common Grounds, Community Center, Pool - Outdoor</li><li>HOA Fee: \$150</li><li>HOA Fee Frequency: Monthly</li></ul> <b>School Information</b> <ul style="list-style-type: none"><li>School District Name: FAIRFAX COUNTY PUBLIC SCHOOLS</li><li>School District Source: Listing Agent</li></ul>	<b>Utilities Information</b> <ul style="list-style-type: none"><li>Fiber Optic</li><li>Central Air</li><li>Cooling Type: Central A/C</li><li>Cooling Fuel: Electric</li><li>220 Volts</li><li>Forced Air</li><li>Heating Fuel: Natural Gas</li></ul> <ul style="list-style-type: none"><li>Hot Water: 60+ Gallon Tank, Natural Gas</li><li>Sewer Septic: Public Sewer</li><li>Natural Gas Available, Sewer Available</li><li>Water Source: Public</li><li>Municipal Trash: Yes</li></ul>
Taxes / Assessments	
<b>Tax Information</b> <ul style="list-style-type: none"><li>Agricultural Tax Due: No</li><li>County Tax: \$5,255</li><li>Tax Assessed Value: \$460,910</li><li>Tax Year: 2021</li></ul>	<ul style="list-style-type: none"><li>Tax Annual Amount: \$5,626</li></ul> <b>Assessments Information</b> <ul style="list-style-type: none"><li>Year Assessed: 2021</li></ul>
Lease / Rent Details, Location Details, Listing Information	
<b>Lease Information</b> <ul style="list-style-type: none"><li>Lease Considered: No</li></ul> <b>Location Information</b> <ul style="list-style-type: none"><li>Outside City Limits</li><li>Directions: FROM ROUTE 7, SOUTH ON 606, TURN SOUTH ON WOELE AVE, TURN LEFT ONTO NORTH SHORE DRIVE, TURN LEFT ONTO SCANDIA CIRCLE.</li></ul>	<ul style="list-style-type: none"><li>Location Type: Middle Of Block</li><li>Road Responsibility: City/County</li></ul> <b>Buyer's Brokerage Compensation ⓘ</b> <ul style="list-style-type: none"><li>2.5%</li></ul>
Property / Lot Details	
<b>Property Information</b> <ul style="list-style-type: none"><li>Not Federal Flood Zone</li><li>Total Below Grade Sq. Ft.: 0</li><li>Accessibility Features: 36"+ wide Halls, Doors - Swing In</li><li>Irrigation Water Rights</li></ul>	<ul style="list-style-type: none"><li>Structure Type: Interior Townhouse/Rowhouse</li><li>Window Features: Double Hung, Insulated, Screens, Sliding, Storm</li><li>Building Not Winterized</li></ul> <b>Exterior Information</b>
<ul style="list-style-type: none"><li>Irrigation Water Rights</li><li>Horse: No</li><li>Improvement Assessed Value: \$350,910</li><li>Ownership Interest: Fee Simple</li><li>Property Condition: Excellent</li><li>Year Built Source: Assessor</li><li>Year Major Reno/Remodel: 2019</li></ul> <b>Building Information</b> <ul style="list-style-type: none"><li># of Levels (Total): 5</li><li>Construction Materials: Batts Insulation, Blown-In Insulation</li><li>Foundation Details: Concrete Perimeter, Slab</li><li>Main Entrance Orientation: South</li><li>Other Structures: Above Grade, Below Grade</li><li>Outdoor Living Structures: Porch(es)</li><li>Pets Allowed: Yes</li><li>Pet Restrictions: Size/Weight Restriction</li><li>Roof: Architectural Shingle</li></ul>	<b>Exterior Information</b> <ul style="list-style-type: none"><li>Exterior Lighting, Sidewalks, Street Lights</li></ul> <b>Pool Information</b> <ul style="list-style-type: none"><li>Community Pool</li></ul> <b>Lot Information</b> <ul style="list-style-type: none"><li>Tidal Water: No</li></ul> <b>Land Information</b> <ul style="list-style-type: none"><li>Land Assessed Value: \$110,000</li><li>Soil Types: Clay</li></ul> <b>Above Grade Information</b> <ul style="list-style-type: none"><li>Finished Sq.Ft.: 1,680</li><li>Finished Sq.Ft. Source: Assessor</li></ul> <b>Below Grade Information</b> <ul style="list-style-type: none"><li>Unfinished Sq.Ft. Source: Assessor</li><li>Finished Sq.Ft. Source: Assessor</li></ul>

Details provided by BRIGHT MLS and may not match the public record. [Learn more.](#)

## Schools:

Updated: 12/01/2021

## Schools

This home is within the [Fairfax County Public Schools](#).

**Reston's enrollment policy is not based solely on geography.** Please check the school district website to see all schools serving this home.

### GreatSchools Rating

NR	<b>Chantilly High School Academy</b> Public • 9 to 12 • Choice school	- Students	N/A	7.0 mi Distance
4/10	<b>Forest Edge Elementary</b> Public • Preschool to 6 • Serves this home	541 Students	★★★★★ 27 reviews	0.5 mi Distance
NR	<b>Marshall High School Academy</b> Public • 9 to 12 • Choice school	- Students	N/A	7.4 mi Distance
4/10	<b>Hughes Middle</b> Public • 7 to 8 • Serves this home	1050 Students	★★★★★ 7 reviews	1.9 mi Distance
4/10	<b>South Lakes High</b> Public • 9 to 12 • Serves this home	2512 Students	★★★★★ 14 reviews	2.1 mi Distance
NR	<b>Falls Church High School Academy</b> Public • 9 to 12 • Choice school	- Students	N/A	9.6 mi Distance

School data provided by [GreatSchools](#). School service boundaries are intended to be used as reference only. To verify enrollment eligibility for a property, contact the school directly.

## Public Facts:




### Public Facts for 1518 Scandia Cir

Beds	3	Lot Size	1,587 Sq. Ft.
Baths	3.5	Style	Townhouse
Finished Sq. Ft.	1,680	Year Built	1973
Unfinished Sq. Ft.	—	Year Renovated	—
Total Sq. Ft.	1,680	County	Fairfax County
Stories	2	APN	0181 05020031

Home facts updated by county records on Oct 4, 2021.

## Transportation:

### Transportation in 20190

 <b>15</b> /100 Car-Dependent Walk Score®	 <b>39</b> /100 Some Transit Transit Score®	 <b>22</b> /100 Somewhat Bikeable Bike Score®
---	---	---

This area is **car dependent** — almost all errands require a car. **Transit is available**, with a few nearby public transportation options. There is a **minimal amount of infrastructure for biking**.

## Relative Affordability Report - Elm's Glen

ALL SMH Products, TTM ending May 2021

Total Sales	Avg Closing Price
998	\$239,543

Avg Attached Closing Price	Avg Detached Closing Price
\$178,398	\$252,274

Avg Resale Closing Price	Avg New Con Closing Price
\$238,705	\$249,850

The metrics above do not contain Stanley Martin Homes Subject Site Data

### Price Segments

Lower Price Closing Prices

< \$210500

Low - Mid Price Closing Prices

\$210500 - \$267988.9

Mid - High Price Closing Prices

\$267988.9 - \$304620

Highest Price Closing Prices

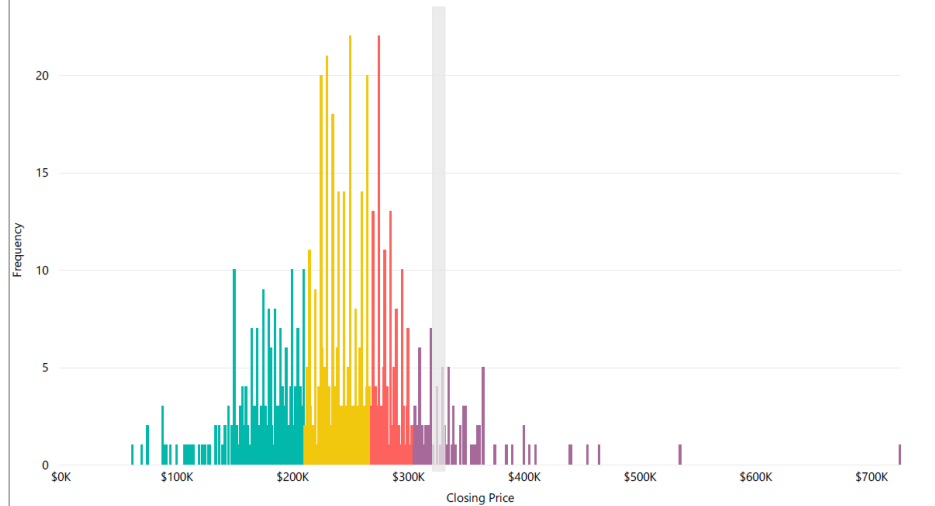
> \$304620

### Avg SMH Projected Price

Avg Projected Closing Price

\$322,803

### Frequency of Closing Price for All Market Closings (SFH, TH, CO) - New and Resale



\*\* SMH projected closing price represented by the shaded areas on the Frequency Graph. \*\*

Lower Price: Lower 30% of closing prices. || Low-Mid Price: Lower middle 40% of closing prices. || Mid-High Price: Upper middle 20% of closing prices. || Highest Price: Top 10% of closing prices



# Relative Affordability Report - Elm's Glen

ALL Detached Products, TTM ending May 2021

Total Sales  
802

Avg Closing Price  
\$252,274

Total Resale Sales  
763

Avg Resale Closing Price  
\$251,496

Total New Con Sales  
39

Avg New Con Closing Price  
\$267,492

The metrics above do not contain Stanley Martin Homes Subject Site Data

## SMH Only

Avg Projected Closing Price

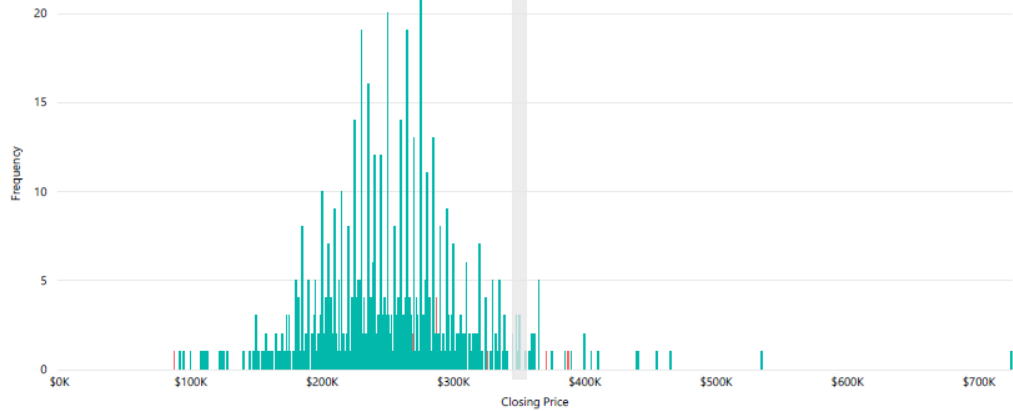
\$351,209

Avg Projected Fin Sq Ft

1,994

## Frequency of Closing Price for All Detached Closings - New and Resale

New vs Resale ● New ● Resale



\*\* SMH projected closing price represented by the shaded areas on the Frequency Graph. \*\*

## Market Summary

\*\* Without SMH \*\*

Avg Market Closing Price  
\$252,274

Avg Market Fin Sq Ft  
1,867

Avg New Closing Price  
\$267,492

Avg New Fin Sq Ft  
2,123

Avg Resale Closing Price  
\$251,496

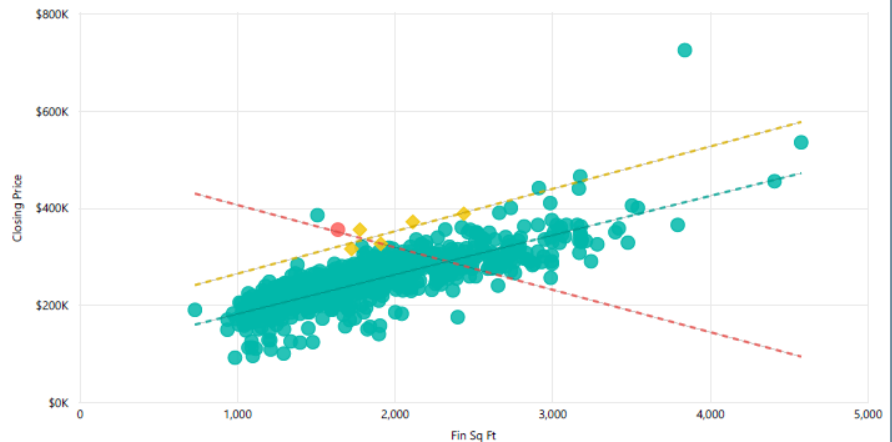
Avg Resale Fin Sq Ft  
1,866

Avg Resale Year Built

1998

## Closing Price by Fin Sq Feet for All Detached Closings - New and Resale

SMH vs Market ● New ● Resale ● SMH



\*\* SMH projected closing price represented by the yellow diamonds on the graph above. \*\*

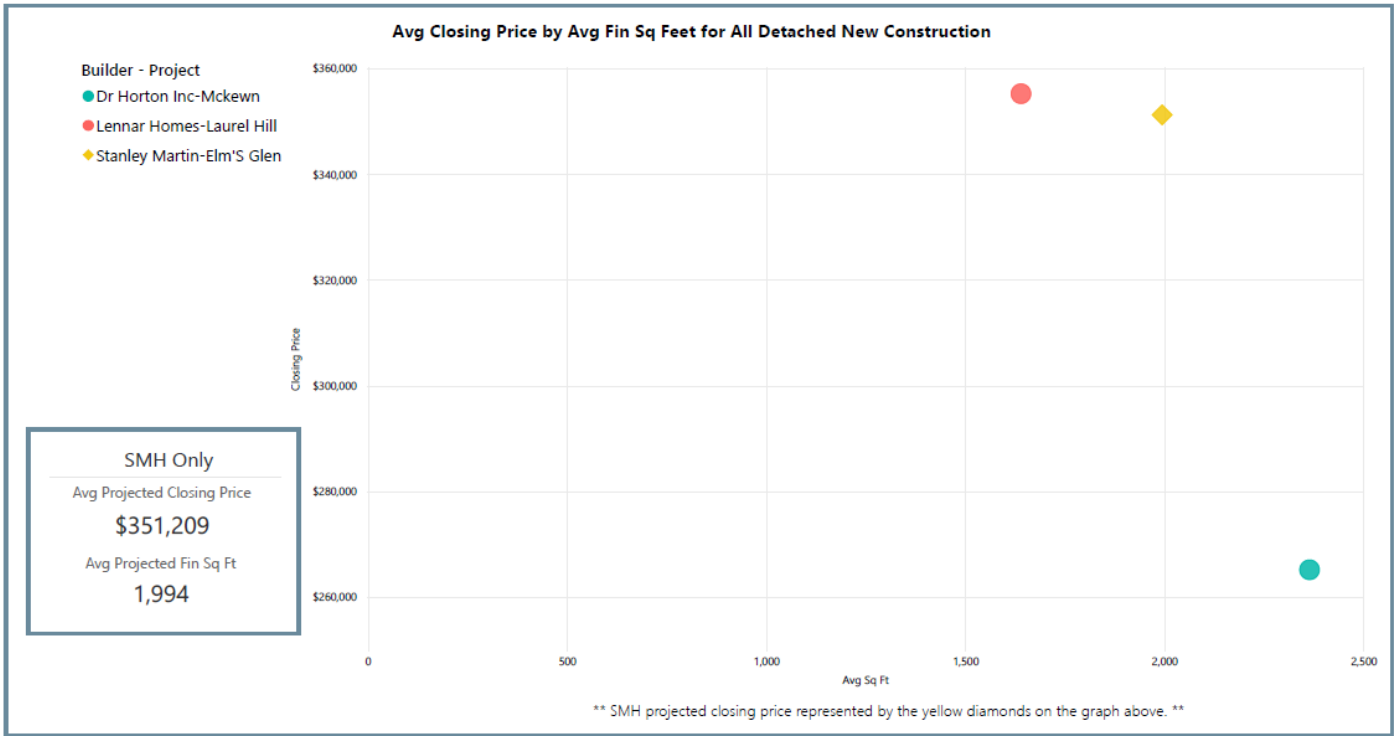


# Relative Affordability Report - Elm's Glen

ALL Detached New Construction Products, TTM ending May 2021

Detached New Construction Subdivision Closings					
Top-Selling Subdivisions by # Closings					
Total Subdivisions					
2					
Builder	Subdivision	Property Type	Total Closings	Closing Price	Avg Sq Ft
Dr Horton Inc	Mckewn	Single Family	38	\$265,184	2,365
Lennar Homes	Laurel Hill	Single Family	1	\$355,210	1,640

Closing Price  
\$267,492



# Relative Affordability Report - Elm's Glen

ALL Attached Products, TTM ending May 2021

Total Sales  
162

Avg Closing Price  
\$178,398

Total Resale Sales  
160

Avg Resale Closing Price  
\$177,708

Total New Con Sales  
2

Avg New Con Closing Price  
\$233,658

The metrics above do not contain Stanley Martin Homes Subject Site Data

## SMH Only

Avg Projected Closing Price

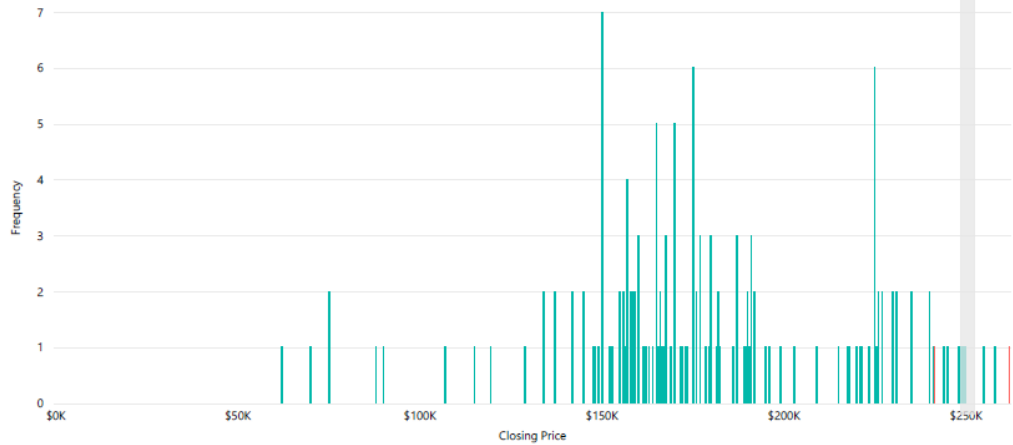
\$251,790

Avg Projected Fin Sq Ft

1,463

## Frequency of Closing Price for All Attached Closings - New and Resale

New vs Resale ● New ● Resale



## Market Summary

\*\* Without SMH \*\*

Avg Market Closing Price  
\$178,398

Avg Market Fin Sq Ft  
1,535

Avg New Closing Price  
\$233,658

Avg New Fin Sq Ft  
1,696

Avg Resale Closing Price  
\$177,708

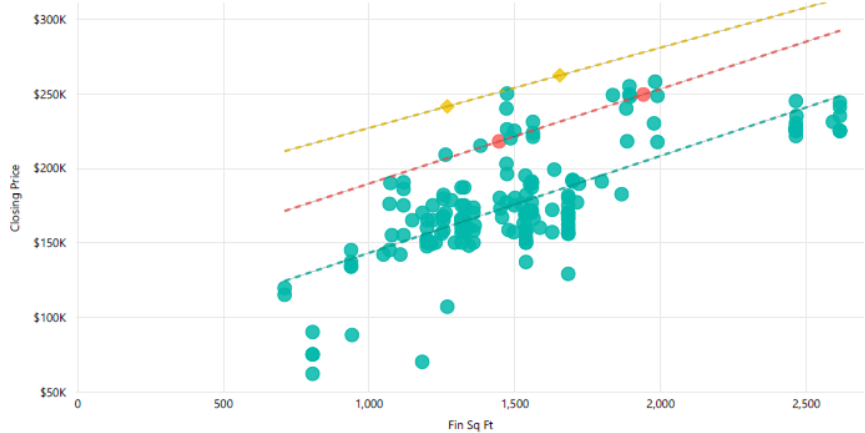
Avg Resale Fin Sq Ft  
1,533

Avg Resale Year Built

2006

## Closing Price by Fin Sq Feet for All Attached Closings - New and Resale

SMH vs Market ● New ● Resale ● SMH





# Relative Affordability Report - Elm's Glen

ALL Attached New Construction Products, TTM ending May 2021

Attached New Construction Subdivision Closings					
Top-Selling Subdivisions by # Closings					
Total Subdivisions					
1					
Builder	Subdivision	Property Type	Total Closings	Closing Price	Avg Sq Ft
Dr Horton Inc	Mckewn	Townhouse	2	\$233,658	1,696

Closing Price

\$233,658

