HouseCanary

The HouseCanary

Valuation Methodology

What is a Valuation?

A home valuation is more than just a number.
For a valuation to be useful, it must include two things:

Home Value

The home's value is the amount of money that we think the home is worth in the current market and is expressed as a range, accommodating for variation in each home that could affect how attractive it is to buyers.

+

Context

Property details, the home's condition, comparable recently sold listings (comps), local market data, and any additional information that can help users understand the why and how of the home's value.

Modeling Principles

Most valuation models today rely on a mixture of a given home's historical sales and a selected set of comps — recent sales of homes that are similar in characteristics and geographically close to the home being valued. HouseCanary's AVM is able to generate more accurate valuations than most traditional models because it considers all previous sales history for both the subject property and its surrounding neighborhood, as well as macroeconomic data, capital markets data, mortgage records, search and social data, and many other data sources.

Primary Assumptions

Our models are based on these primary assumptions about how to calculate an accurate home valuation:

1. **Market data**
   Home prices in a neighborhood tend to move together through time as the market heats up, cools down, or remains stable.

2. **Historical data**
   Using all known historical price events is statistically more efficient than relying only on recent sales.

3. **Sales and home history**
   The best predictor of a home's value is past price occurences of that home, in addition to considering price trends and any changes to the condition and structure of the home.

4. **Machine learning**
   Machine learning-based algorithms are more efficient than classical models when it comes to recognizing complex relationships between a home and its surrounding neighborhood.

5. **Clean data**
   We can further improve valuation accuracy by focusing human effort on enhancing existing datasets to generate cleaner data.

The Valuation Landscape

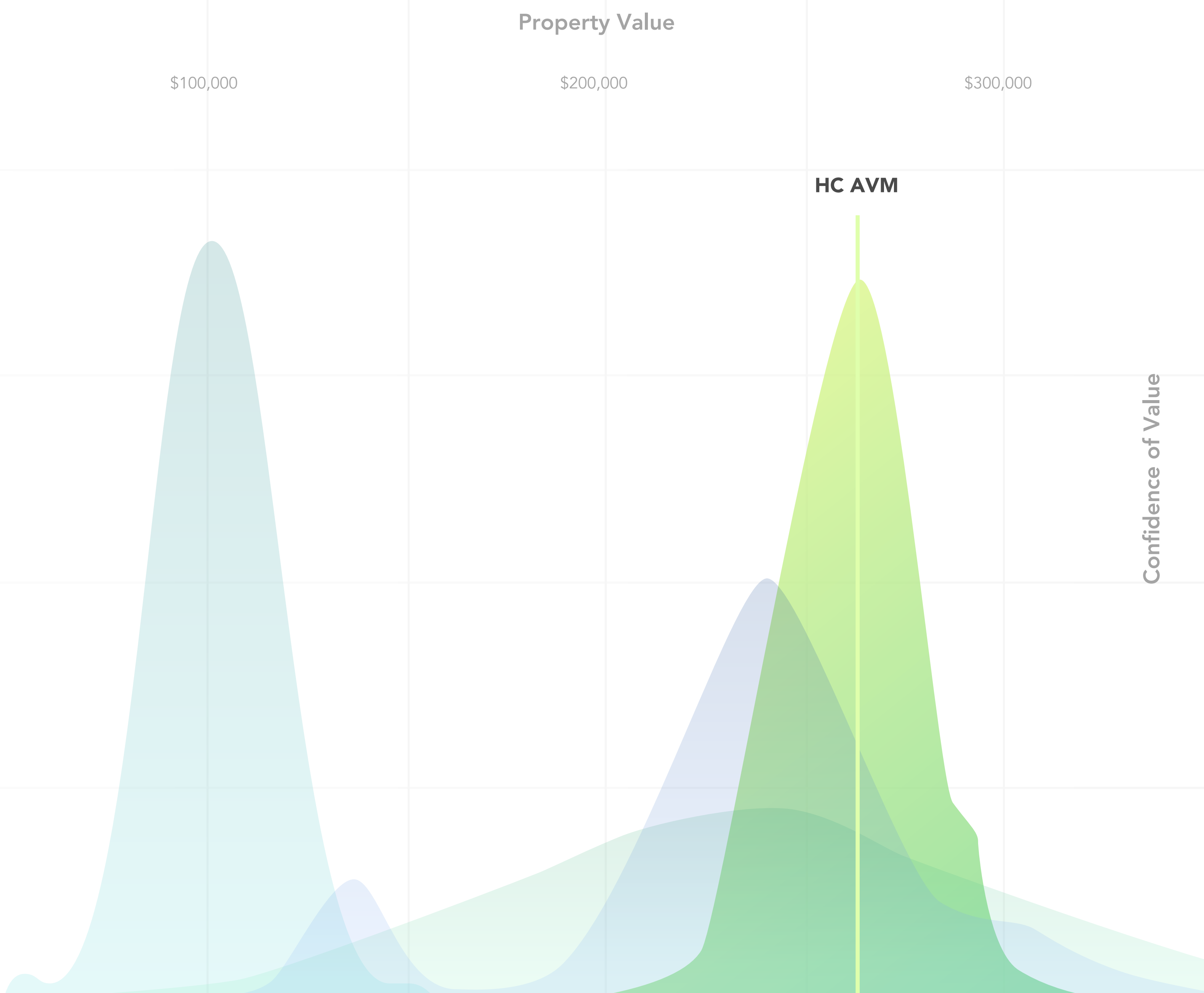
We look at hundreds of data points to pinpoint the most accurate valuation and to build the context around our valuations. This gives users full transparency into why a home is valued within that price range and how confident we feel about those numbers.

To understand the complex relationships between all of these data points, we apply artificial intelligence.

It Starts with Subject Data

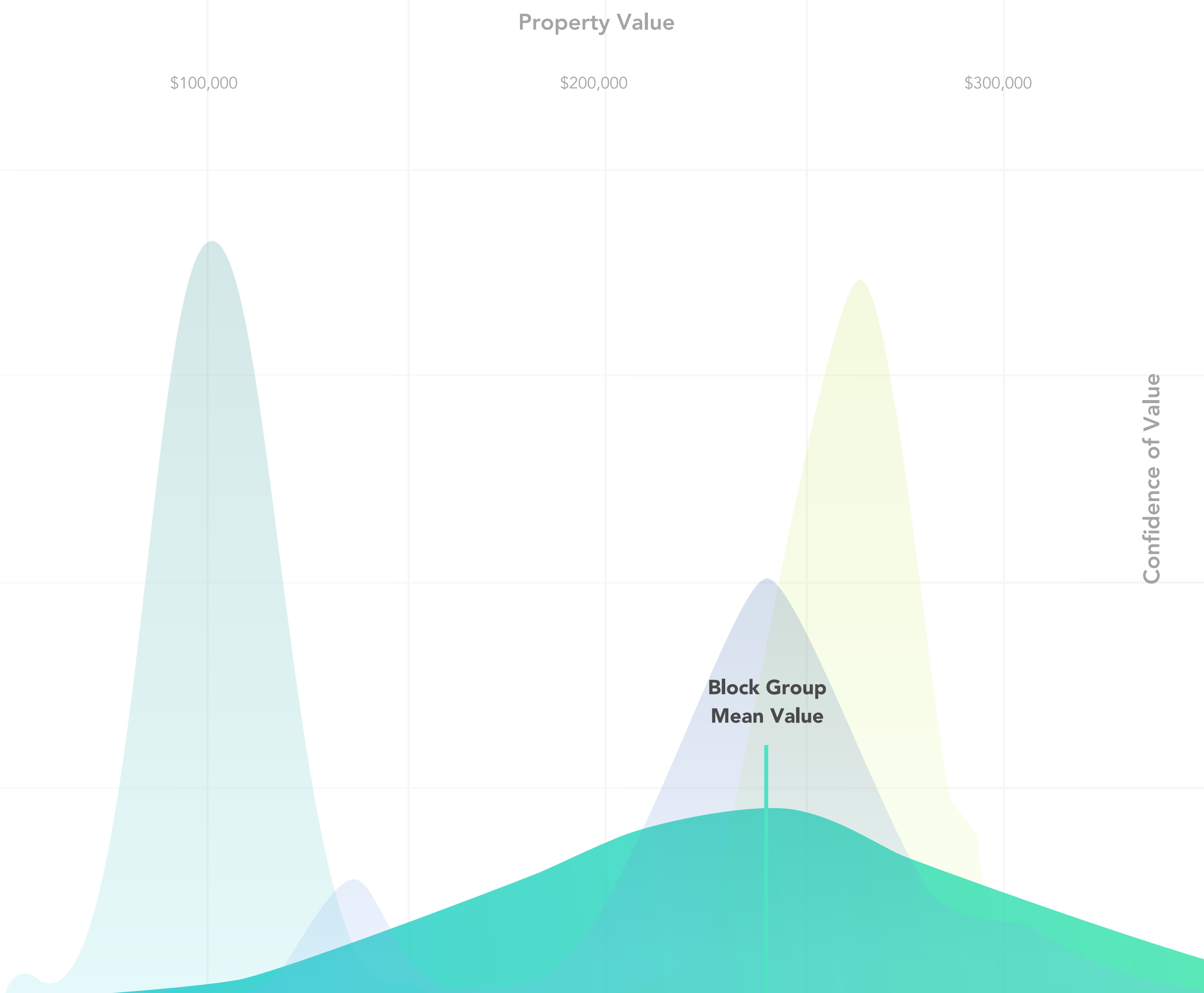
The more we know about the subject property, the better we are able to pinpoint its current fair-market value. Our sources include MLS and county assessor records, but we also include multiple other sources, such as mortgage records, capital markets data, and many others. The very best predictor of value is past price occurrences of that home and how those prices line up with the rest of the market, so we look at more than four decades' worth of sales data.

Depending on the quality of the data and how recently the home was listed or sold, we also provide a confidence rating, which helps the valuation consumer understand how confident we are in each valuation based on the underlying metrics we were able to collect for the subject property.



Blending in the Background

We analyze the subject property details in the larger context of the market, neighborhood, and block where that specific home is located. So we also look at real estate price trends, how the subject property compares to other homes in the area, local housing inventory, local economic and employment trends, consumer demand and buying behavior, and many other factors to help us understand the overarching market and its attributes.



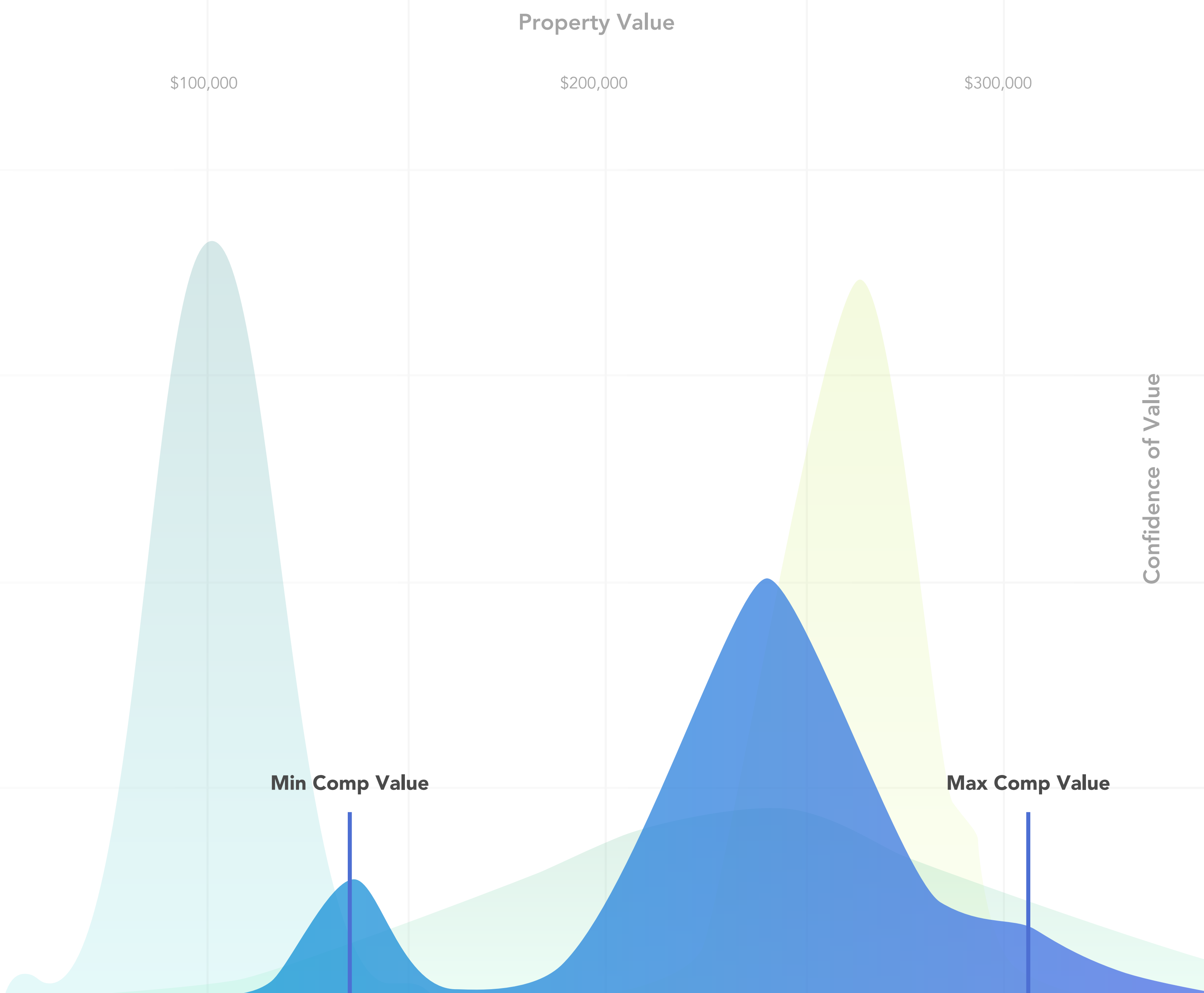
Adding in the Land Data

The home valuation nearly always includes the land on which the home sits. So another factor we consider is land data, which can give us additional insight into how much the property as a whole is worth. For example, in a market with a lot of slopes, hills, and cliffs, a plot of land that's relatively flat might be worth more because the land itself is more usable. HouseCanary's privacy score is another measurement we use to determine property value; it quantifies how private a home is based on the land's slope, nearby neighbors overlooking the property, and other factors that affect privacy, and then calculates how important privacy is (and therefore how much it's worth) in the home's neighborhood and market.

Comp Data Pulls it all Together

Using comp data to pinpoint a home's value is a tried-and-true technique, so HouseCanary applies comps to our valuations. We use advanced technology to find the most relevant recently sold listings that are the closest possible comparisons to the subject property, and then identify ways in which the subject property is different from each so that our comp analysis compares apples-to-apples as much as possible.

We identify between three and five comps for each subject property, all sold within the past six months. We look at how homes on the same block as the subject property compare to the subject property in terms of value, size, bedroom/bathroom count, and other significant details. Depending on the quality of the comps we're able to pull and our confidence in the data, we can pinpoint a confidence level for the valuation.



The Final Result

HouseCanary AVM

All of these data factors combined give us the HouseCanary AVM, which generates both a property value and a range of values that could apply to the subject home — low and high. By combining the subject property data, background data, land data, and comp data, we are able to generate a confidence level in each AVM. The confidence level will be higher if the data in each group is high-quality, and it will be lower if we're missing any data or we're unsure of the quality of the data we've collected.

Home values are best understood as a range of appropriate values instead of a single point estimate, which is why we provide both an upper and lower estimate for our AVM.

Assessing Data Quality

Before we can produce a confident property valuation, we must gather, combine, and filter property data. Our property-level data comes from several sources:

- Public record data includes property characteristics and historical recorded sales prices over the previous 40 years, where available.
- MLS data includes property characteristics, listed prices, and contract prices.
- Other property information includes data on mortgage balances and measures of financial distress, along with other details that impact value, such as proximity to a busy street or a golf course.

Our process starts with a profile of each data source. We create field-by-field, value-by-value rules to map incoming data to existing data. Those rules provide the first layer of control, flagging suspicious data changes, anomalous and unexpected values, and inconsistent data use. We also use a United States Postal Service Coding Accuracy Support System (USPS CASS) certified service to validate, standardize, and match all addresses that feed into our system.

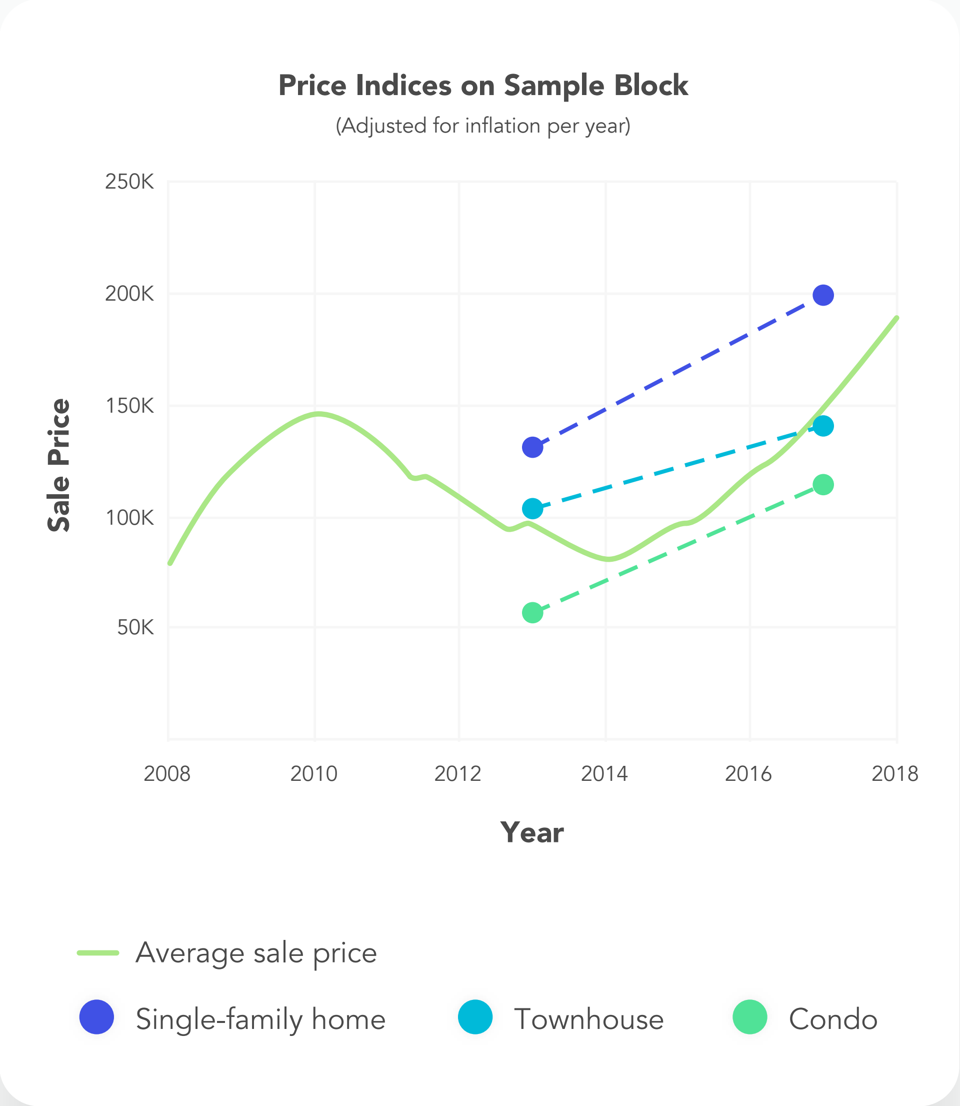
Methodology

Step 1
Build Localized Price Indices

There are approximately 200,000 US Census blocks, each one further subdivided into a series of blocks. We create a price index for every block, which helps us determine how the houses on that block measure up against the neighbors. Our machine learning methods can borrow information from surrounding blocks to estimate a price index for blocks with smaller transaction sample sizes, something that provides lift in rural markets in particular.

First, we update the price of all valid historical sale and list prices to current values. By controlling for both time and location, we get a much larger model dataset than if we limited ourselves to only using closed sales prices over the previous one or two years.

Once we have calculated the time-adjusted prices, we center them around the median block price for each property type: single-family home, condo, or townhouse. We compute the centering by applying housing market growth to both the home we're valuing and the median value of other homes like it on the same block, then determining the home's value in the context of both its growth in value and the growth in value on the block.



Step 2
Train Valuation Models

In our models, we account for many variables beyond property type, block median price, and time. Depending on the data available, our models can include property characteristics, neighborhood characteristics, macro- and microeconomic data, spatial relationships, repeat price observations, and much more.

The three models we use evaluate:

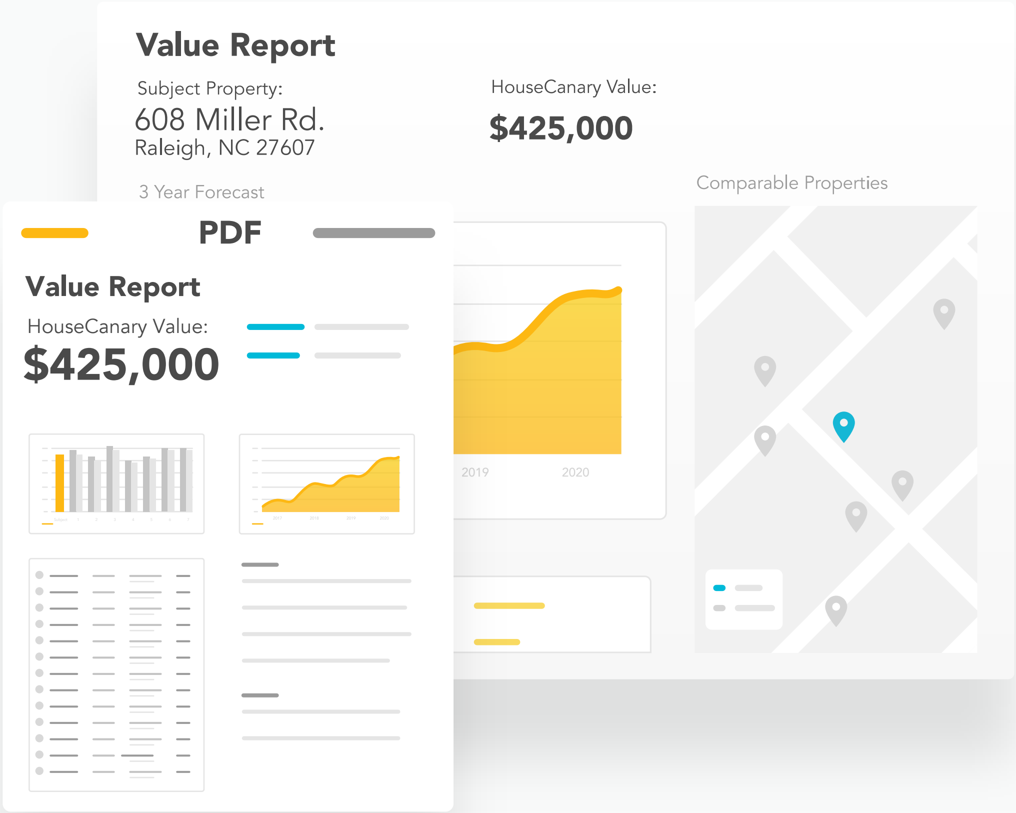
- Price based on market activity
- Price based on the home's individual sale history
- Price per square foot

From these models, we generate an estimated price range for every property. We then take those model estimates and apply them to previous years to yield a historical price estimate. Because we have decades of price history for each home, we can use this historical estimate information to determine the accuracy of our model's valuation. This process is done for all three models within the algorithm.

Step 3
Producing a Final Valuation

We re-assess how well each model is able to determine a certain property's value every time a user runs an AVM on that home. Coming up with a single valuation involves assessing how well each of the three models was able to estimate the observed historical prices of the home and verifying each model against the other two models. Some factors that could influence a model's trustworthiness include data availability, data support, and the level of agreement between the models.

After we have determined how credible each model is for this home in the current market, we use our proprietary algorithms to balance the models against each other and calculate how influential each model should be in terms of the final valuation. This last step allows us to provide the valuation — the price range that we think is closest to the home's current fair-market value.



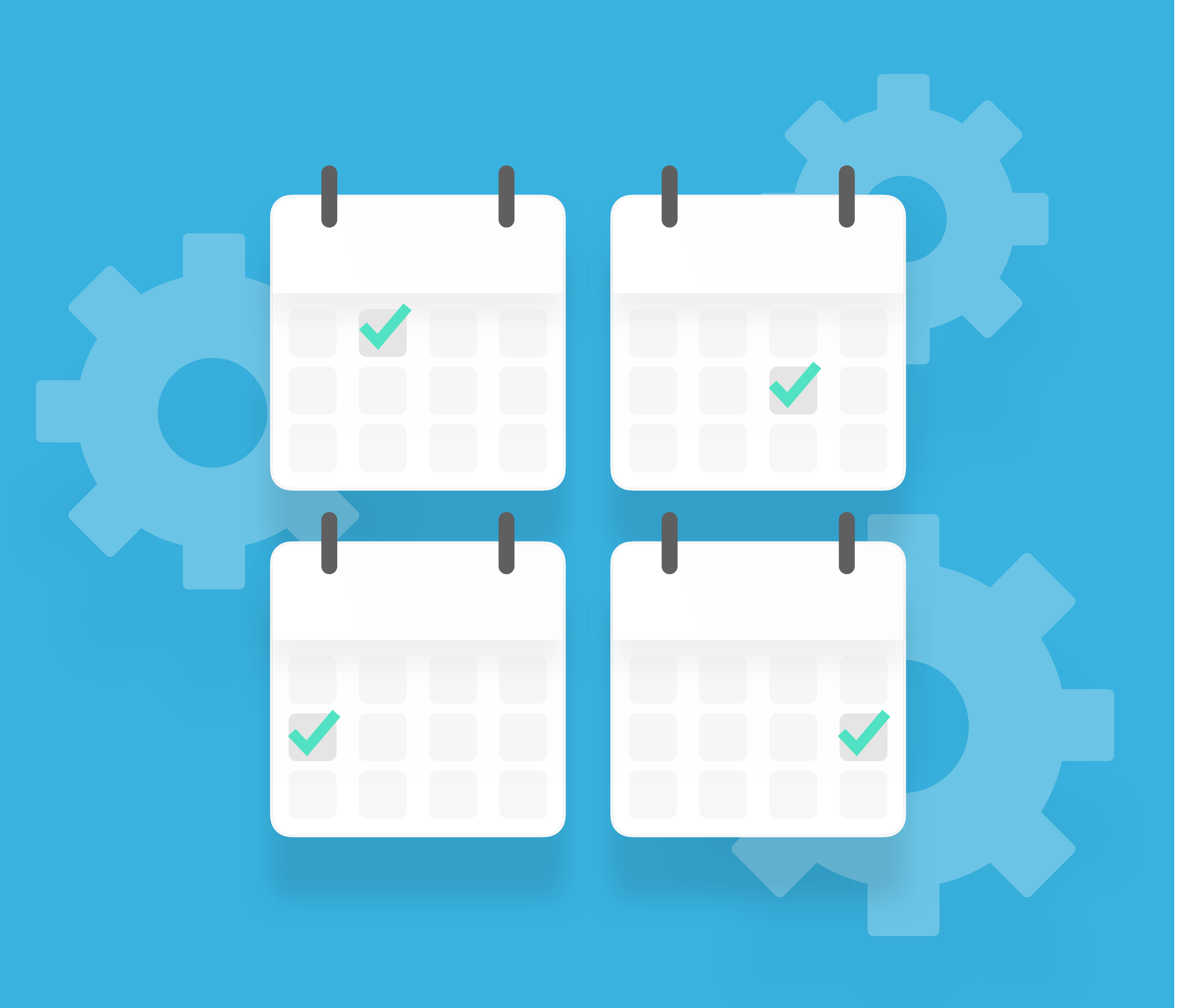
Testing and Validation

Continuous Internal Testing

HouseCanary runs continuous testing on our accuracy and coverage. To make sure we're accommodating for data delay in some states, all our internal tests cover a six-month testing window.

Our internal testing measures how close a value estimate came to the actual sales price of different houses across the country. This helps us understand whether our valuations in some areas are more accurate than valuations in other areas, and to tweak our models where they don't get as close. Our internal tests track all of the performance measures below:

- **hit_rate:** The proportion of sold properties for which we had an estimate of value prior to the sale.
- **Median_Abs_Pct_Err:** The 50th percentile of absolute error in percentage terms. If this value equals 3.0%, then half our estimates were within +/-3.0% of actual sales price.
- **Median_Pct_Err:** The 50th percentile of actual error in percentage terms (not absolute error). Values close to zero imply that the estimator is unbiased.
- **Within X%:** The percent of estimates that fell within +/-X% of actual sales price. HouseCanary produces this value for the 5%, 10%, and 20% bounds.
- **Within_HC_Prediction_Interval:** We provide both an upper and lower estimate for our AVMs. This range is the prediction interval. We aim to generate a prediction interval that includes a certain coverage probability — the percentage of homes with sales prices that fell within the boundaries of our upper and lower estimates of value.



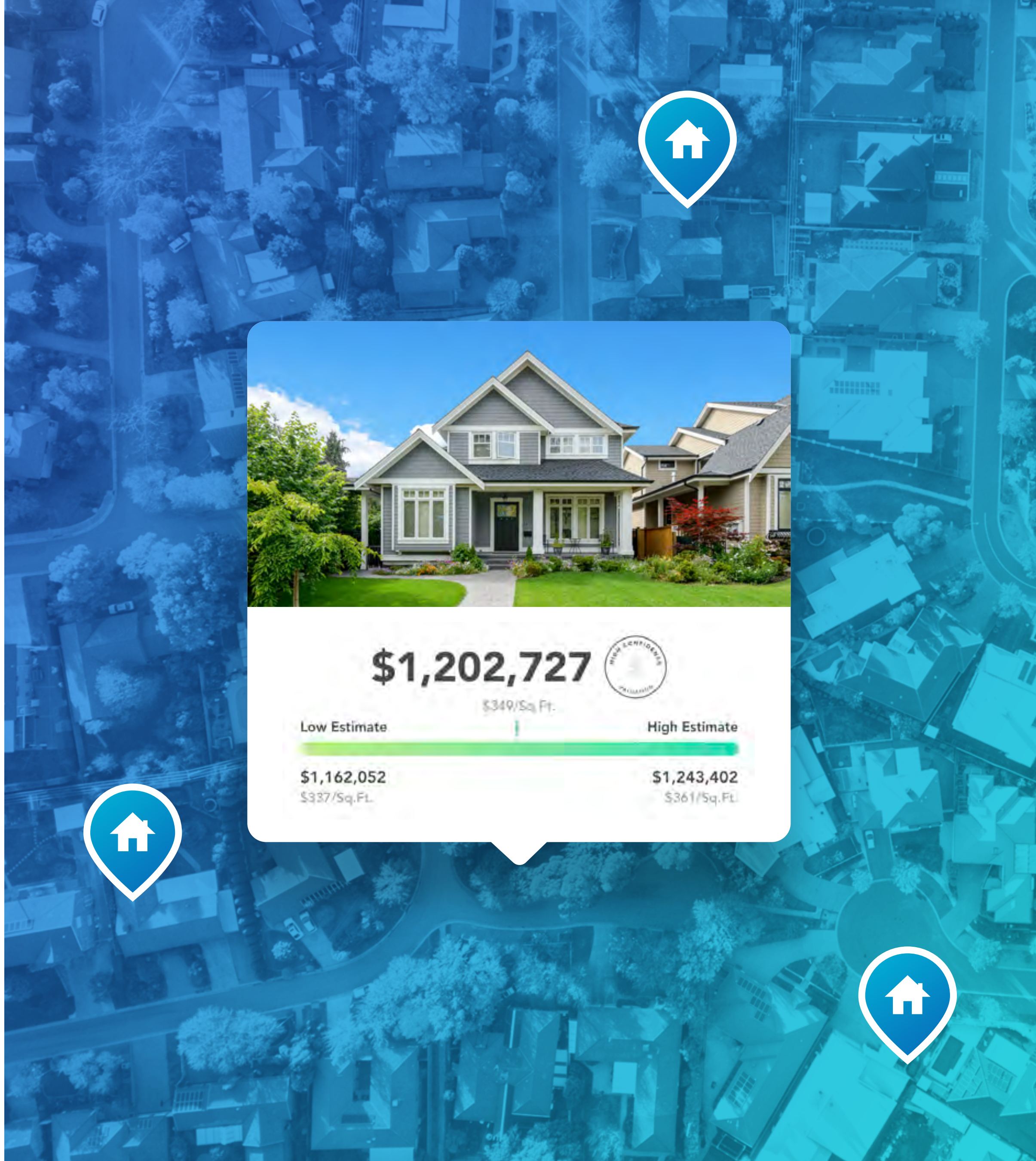
Quarterly Third Party Testing

We don't only test our accuracy internally; we also use independent third parties to evaluate our valuations and give us feedback about how accurate they are in all of the metrics above. If you'd like to see detailed test results for national, state, and MSA levels, we can provide those for you. We're also able to show you our internal test results at a national level if you're interested in seeing how accurate our valuations were for a specific property type and month.

Conclusion

Considering all previous sales for the subject property and neighborhood in addition to macroeconomic data, capital markets data, mortgage records, search and social data, and many other sources of information give HouseCanary a powerful foundation for home valuations. By combining this foundation with proprietary algorithms that use machine learning to recognize and exploit complex relationships between the variables, we're able to generate a value range that can inform real estate decisions both for individual homes and for thousands of homes at once.

If you'd like to learn more about HouseCanary's valuation products, please contact us. We will be happy to share more details about how HouseCanary can give you the data — and confidence — to make better decisions about your real estate assets.



About HouseCanary

Founded in 2013, HouseCanary is a real estate technology company providing the most accurate home valuations to drive smarter decisions across the real estate ecosystem. Clients include some of the largest financial institutions including the top five buyers of residential whole loans on Wall Street, three of the largest Wall Street investment firms and four of the top five single family rental companies.

HouseCanary can be found at **www.housecanary.com**