HouseCanary

The HouseCanary

Dataset

Introduction

The data that informs the real estate industry is notoriously scattered, with hundreds of sources using different standards. The raw data feeds that agents, investors, and analysts have historically depended on are plagued by gaps in coverage, lag from poorly updated sources, and inconsistencies across sources.

HouseCanary quality-checks and normalizes data from around the country, capturing deep histories for added insight, and constantly adding new information to inform your most important business decisions. Our Analytics Platform makes it easy to pull hundreds of significant Data Points covering all U.S. states down to specific micro market dynamics for the over 4 million census blocks and over 100 million properties, with some of the broadest nationwide coverage.

How We Craft, Clean and Verify our Dataset

HouseCanary's data curation and quality checking process involves three basic steps:

1 Combining streams of data from many sources

2 Checking data quality with automated and human systems

3 Verifying address information with USPS CASS service

1. Matching up streams of data from many sources

We fully profile all data sources that feed into our platform. We source information from:

- Public record data including property characteristics sourced from 3,100+ county assessor offices and historical recorded sales prices sourced from 2,700+ county recorder offices over the previous 40 years, where available.

- MLS data including property characteristics, listed prices, and contract prices.

- Other property information including data on mortgage balances, mortgage type, and measures of financial distress, along with other details that impact value, such as neighborhood externalities and distinguishing property features.

Field-by-field, value-by-value rules determine how we organize incoming data in our content management system, providing a first layer of control that flags suspicious data changes, anomalous and unexpected values, and data inconsistencies.

2. Checking data quality with automated and human systems

HouseCanary ensures data quality with systems and processes developed and supported by a team of data engineers, data analysts, and domain experts. Flagged content is quarantined until a human review approves it as valid, the issue is resolved by the source, or we implement new handling logic that resolves the problem. Validated items that enter our content management system are linked and normalized at the address, building, and census block levels. A second quality control pass compares different sources to determine correct data for a given object.

3. Verifying address information with USPS CASS service

We validate, standardize, and match all addresses, whether they come from data feeds or from user inputs. We do this using a United States Postal Service Coding Accuracy Support System (USPS CASS) certified service. Any change of any component of a full address triggers a new validation, standardization, and matching event.

Address Verified

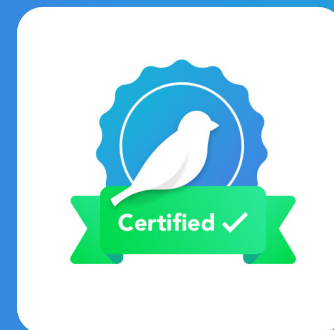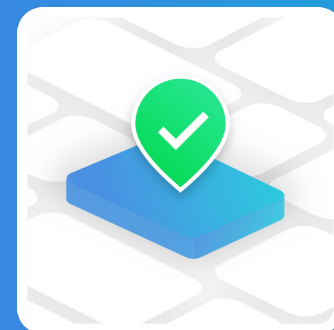3016 Poplar Street
Austin, TX 78702

What we mean by breadth and depth

Some data providers see coverage as a zero-sum game: they'll say they cover a property, or they don't. At HouseCanary, we're not satisfied to only provide an address and basic details. We share how much information is available about a home and our level of confidence in that information:

- **Hit rate:** We can find the home and confirm its address
- **AVM:** We can provide an automated valuation for this home
- **High confidence:** Our valuation for this home has a narrow standard deviation
- **Certified:** We're confident enough in the data and this valuation to insure it

We refresh public records and MLS data daily, and typically compile and add market-level data monthly or quarterly. The only potential delays in our dataset are due to delays from the original source. In non-disclosure states and counties, where public disclosure of transaction details is not required, HouseCanary checks with other sources to complement public record data. Our quality control processes identify which sources are most reliable, and we surface our underlying data so you can compare sources yourself.

How you get more with HouseCanary data

Intuitive tools and visually compelling solutions, and accurate models help users get more out of HouseCanary data. No matter how sophisticated an algorithm is, its results are only as good as its inputs. HouseCanary's machine learning models are robust because we train them with extremely rich, normalized, quality-tested data. We build our AVMs on the census tract level to provide very localized valuation models for each of five property types. Our unique comparable selection system uses model-based adjustment to provide a data-driven Similarity Score relative to a subject property.

We offer access to our dataset through the Analytics Platform, a singular system of insight which you can integrate into your own interface. You can seamlessly shift from studying a market to zooming in on one property, and save time by trusting a single source of truth to make informed decisions.

$247,500

$245,000
$163/Sq. Ft.

Low Estimate   High Estimate
$230,000 $248,000
$153 / Sq. Ft. $165 / Sq. Ft.

↑ 6.3%

↑ 6.2%

What's next: adding sources and continuing to improve

HouseCanary data is more accurate, more consistent, and has better coverage than human or automated alternatives. The artificial intelligence driving our quality checking will continue to improve as it processes more and more data. We are also constantly growing our reach, for example growing our footprint as an MLS-participating brokerage to better serve our clients and their real estate decision-making.

By continuing to gather data and add new data sources, we can increase our coverage and further fine-tune our valuation and forecasting models, achieving better and better results as time passes.

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 leading institutional investors and enterprise lenders who trust HouseCanary's products to fuel acquisition, portfolio management, underwriting, and other processes.

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

Learn more about HouseCanary's industry-leading dataset. Contact us today at **www.housecanary.com/contact-sales**