BABELSTREET

Refined Information for Real Estate

BABEL STREET DATA

The Babel Street Data Refined Information Real Estate product provides critical insights into the housing market. By leveraging our library of raw data collections comprised of multiple real estate and apartment rental datasets with historical data dating back to 2016, users gain a better understanding into market demand, pricing, and areas poised for population growth throughout the United States. Delivered data products are preprocessed and readily available for easy analysis.

Product Details

The Refined Information Real Estate product eliminates the time-consuming process of preparing and combining multiple data sources. Our skilled analysts begin by identifying data collections from online real estate listing sites and apartment rental listing sites that contain selling and rental pricing, home price estimates, property types (single family, townhome, condo), statistical location areas (metropolitan and micropolitan), square footage, bedroom and bathroom counts, and historical age of the property. These identified raw real estate data collections, which have been cleansed, curated, and enriched for Babel Street Data Raw Collection products, are then normalized and aggregated monthly and quarterly to provide a single data source for analysis.

The Refined Information Real Estate data includes many collected fields that provide valuable insights into the housing market that is being affected by economic factors. The ability to filter data according to price, location, property type, square footage, property age, and number of bedrooms and bathrooms allows users to analyze the data to answer specific questions.

Benefits

The Refined Information Real Estate product provides a cost-effective alternative to purchasing multiple raw real estate listing data sets. While purchasing raw data may allow more flexibility to utilize and analyze all underlying data to answer a multitude of questions, it is often expensive and requires significant effort from data analysis teams. Refined Information products provide ready-to-use combined data, freeing up resources to focus on making quick, informed decisions. Accessing data from all leading real estate listing companies helps users discover additional raw collections that are valuable for future detailed analysis.

Uses for Real Estate Data

Investment Analysis

The ability to assess locational risks and evaluate locational trends allows investors to identify new areas for investment and effectively manage their portfolios based on insights that help predict market movement.

Competitive Analysis

Real estate agents and brokerage firms can use average estimated values for areas to set competitive price points for property listings according to area, property type, size, and bed and bath counts.

Mortgage Lenders

Refined Real Estate insights can help mortgage lenders identify geographic areas with a high foreclosure risk and understand areas where home sales are above market estimates to target borrowers interested in refinancing at a lower rate in the future.

Product Methodology

FIELD	DESCRIPTION	EXAMPLE
start_date	Earliest date for the sale of properties included in the aggregation period. For a monthly or quarterly period, it is the first day of the respective month.	1/1/2018
end_date	Latest date for the sale of properties included in the aggregation period. For a weekly refinement, this is a Sunday. For a monthly or quarterly period, it is the last day of the respective month.	9/1/2023
cbsa_title	Name of the CBSA's Metropolitan Statistical Area (MSA) or Micropolitan Statistical Area (μSA) where the properties are located. Also includes groupings of properties that do not exist within any CBSA ("Non-MSA").	Aberdeen, SD
property_type	The kind of properties aggregated; one of: Single Family, Condo, or Townhouse.	single_family
sum_raw_sold_listing	Count of all property sales observed.	33
sum_adj_sold_listing	Count of property sales observed, excluding sales with data that strongly deviates from the majority of the area's sales. All remaining averages use these adjusted sales records.	33
avg_sold_price	Average sold price of properties, in USD	147407.2121
avg_price_estimate	Average estimated price of sold properties, in USD	142025.5455
price_estimate_difference	Average percentage difference from the estimated prices to the actual sold prices	0.03789224
avg_sqft	Average square footage size of the properties	96.85134593
avg_price_per_sqft	Average USD cost per square foot for the sold properties	1838.575758
avg_bedrooms	Average number of bedrooms in the sold properties	3.454545455
avg_bathrooms	Average number of bathrooms in the sold properties	1.845454545
prop_age_X	Count of properties within age ranges measured from the sold date. Age ranges are: Less than 1 year old, 1–2 years old, 3–8 years, 9–14 years, 15–20 years, 21–36 years, 37–62 years, 63–105 years, and greater than 106 years old.	2
no_prop_age	Count of properties with no observable age	2

Aggregation Procedure

The Refined Real Estate product aggregates raw property sale data into geographic areas and standard calendar periods. For example, an individual output may be grouped by CBSA and calendar month. Babel Street first identifies properties that were observed to have sold in that area in that calendar month. Only residential properties with structures are included; land, vacant lots, or commercial or industrial properties are not included.

The physical address of each property is mapped to its corresponding MSA or µSA. If the address is not located within a CBSA, then the property is aggregated into a "Non-MSA" group for that state.

Babel Street cleans the raw data to exclude properties that exhibit an anomaly in one or more of their key data points. This may include inaccurate bedroom counts, bathroom counts, or square footage measurements — for example, the sale of a property that is one square foot in total size.

We also build a normal distribution of sales data for trailing three-month periods in each area for each property type. Where sufficient volume exists, any properties with sale prices that fall outside two standard deviations from that area's average are excluded from the remaining aggregation. This prevents exceptionally high or low sales prices from misrepresenting the market in that region while still covering more than 95% of property sales.

When calculating the average price per square foot, Babel Street Data may include some data points from a property sale while excluding others. For example, if a property's sold price is determined to be valid based on the above analysis but its square footage is not, then it will still be counted as a sale in that MSA, and its price will be combined into the avg_sold_price calculation. However, its square footage will not be combined into the avg_sqft calculation. The reverse is also possible, with valid square footage and invalid price, or both may be invalid. These partial inclusions account for cases where the direct quotient of avg_sold_price and avg_sqft does not equal avg_price_per_sqft.

For the remaining properties, their sale prices, price estimates, and size data are averaged, and other fields are calculated.

Babel Street. Unlock the Most Insights that Matter.

Babel Street is the trusted technology partner for the world's most advanced identity intelligence and risk operations. The Babel Street Insights platform delivers advanced AI and data analytics solutions to close the Risk-Confidence Gap.

Babel Street provides unmatched, analysis-ready data regardless of language, proactive risk identification, 360-degree insights, high-speed automation, and seamless integration into existing systems. We empower government and commercial organizations to transform high-stakes identity and risk operations into a strategic advantage.

Learn more at **babelstreet.com**

BABELSTREET