San Francisco Market Area Analysis

Single-family detached houses

Guide to the site files and the city page

William Bert Craytor
Valuation Engineer Research Labs

Site folder `20261101_city`, built 1 November 2026

1 What the folder is

The folder `ValEngr-Data/sites/san-francisco/20261101_city` is a complete website made by the program `market_report` from the San Francisco run's model files. The city is fitted whole, as one market area, and the page shows what drives the value of single-family detached houses across the city, as fitted by three regression programs, with the adjustments each model implies.

No sale, address, parcel number, MLS number, coordinates or remarks is in these files. They hold fitted models, curves, surfaces and summary measures only.

One market area, the district as a factor. The San Francisco Association of Realtors divides the city into ten districts (District 1 Northwest to District 10 Southeast), each broken into subdistricts (8A Downtown, 8B Financial District, 8C Nob Hill, and so on); the MLS area of a listing is its subdistrict. The subdistricts are too many (over ninety with sales since 2015) and, several of them, too small (Downtown 4 sales, Mission Bay 1) to fit one by one, so the whole city is fitted at once with the ten districts as a factor: one shift in value for each district, fitted as a random effect in the GAM beside the smooth curves in latitude and longitude. The *Neighborhoods* bar chart on the page gives the value change when the typical sale is placed in each district. Listings coded with no neighbourhood were given the neighbourhood of their neighbours before grouping.

The run. Closed single-family detached sales from 1 January 2015 to 30 September 2026, effective date 1 October 2026, 13,321 fitting sales after the sale type factor, the duress rule, the extreme price filter (the lowest and highest 0.5% of prices left out) and the reviewed listing corrections. Elevation and the bay and ocean view, measured from the lidar surface, are offered to the models (views on). Ten outer folds with the whole pipeline run again in each.

The models. The log model's best model is the GAM with the district factor: R^2 0.877, CV R^2 0.831, OOF R^2 0.832, RMSE* \$535,658, MdAPE* 10.7%. The \$ model's best model is the same GAM: R^2 0.843, CV R^2 0.827, OOF R^2 0.827, RMSE* \$543,686, MdAPE* 12.9%. Two of the log model's check models (MARS degree 1 and the elastic net on its basis) have a negative held-out R^2: 15 held-out sales outside the fitting range carry over 90% of their squared residual. The GAM is not affected; the comparison table on the page shows all three programs.

Not an appraisal. These pages show market area models: how features affect the value of single-family houses in the city, for adjustments and market analysis. They are not an appraisal, give no opinion of value for any property and are provided without warranty of any kind.

2 Opening the site

- **Online:** `https://research.valuation-engineer.com/CountyMktAnalysis/SanFrancisco/SFR_Detached/AsOf_20261030/`, or from the menu *California County Models* on research.valuation-engineer.com. The path names the property class (`SFR_Detached`): other classes, such as attached houses, condominiums, multi-family and mobile homes, get their own folders.
- **On your own computer:** double-click `index.html`. Every page works from the folder itself, with no server and no internet connection. Keep the folder together: the pages read `assets/` and `data/` beside them.
- **On a web server:** copy the whole folder as it is. Plain file hosting is enough.
- A current browser (Chrome, Edge, Firefox or Safari) is needed for the 3D graphs.

3 The files

File or folder	What it holds
`index.html`	The city page: a map of the city with the ten districts, each shaded by the value of the typical sale placed in that district (the district factor of the log model), the subdistricts (the MLS areas) drawn inside them with their names on hover, and a one-row table for the city-wide market area. Clicking the map opens the city page. Both sets of boundaries are derived from the MLS area code of every listing with coordinates (a 90 m grid, each cell labelled by its nearest listings), so they follow the Association's current coding; Golden Gate Park, the Presidio and Lake Merced are left blank.
`<town>.html`	One page per town, here `san-francisco-districts.html`. Section 4 describes how to use it.
`data/county.js`	The city page's data: the city line, the district and subdistrict boundaries, the district factor and the summary figures.
`data/<town>.js`	One town's data: the fitted models of the log model and the $ model, the curves, surfaces and comparison measures.
`assets/site.js`, `assets/site.css`	The page program and its styles, shared by every page.
`assets/plotly-4.1.2.min.js`	Plotly.js, the graph library (MIT license, `assets/plotly-LICENSE.txt`).

The county page's table

One row, the whole city: the number of fitting sales, the value of the typical sale (log model), and R^2, OOF R^2 and MdAPE* for the log model and for the $ model. A town that was not analysed shows the reason in its row.

4 Using a town page

4.1 The bar at the top

- ← **San Francisco** returns to the city page.
- **Log model / $ model** switches every table and graph on the page between the two models. The log model gives effects as percentages of value; the $ model gives effects in dollars.
- **Dark mode / Light mode** changes the colors. The browser remembers the choice.

The row of links under the title jumps to each section of the page.

4.2 The sections

1. **At a glance.** Tiles for the log model and the $ model side by side: the value of the typical sale at the effective date, the best model, the number of fitting sales, R^2, OOF R^2 and the typical residual on held-out sales (MdAPE*): the part of price the features do not carry, mostly condition, quality and appeal. A note appears when a model's OOF R^2 is below 0.65. A table under the tiles lists the typical sale's features.
2. **The models.** The best model from each of the three programs: the GAM (mgcv-rs), MARS (earth-rs) and the elastic net (glmnet-rs). The model with the highest OOF R^2 is marked as the best model. The tiles, the solid line on each graph and the county table use the best model.
3. **What drives value.** One graph per feature: living area, lot size, age, bedrooms, bathrooms, garage, elevation, bay view, lake or river view and days (sale_age) before Eff. Date, where the town's models use them. Each graph changes that one feature of the typical sale and shows the change in value. The solid line is the best model; the shaded band is the GAM's 95% range.
    - Move the pointer over a line to see the value at that point and the adjustment per step (for example, per 100 sq ft of living area or per 30 days of sale age).
    - Tick **Show the other models on the graphs** to draw the other two programs' lines as well, with a legend.
4. **Neighborhoods and remarks features.** Bar charts for the features that take levels instead of numbers:
    - *Realtor district:* each bar is the value change when the typical sale is placed in that district instead of its own. The eleventh level, "No district", holds the sales still coded 9909 San Francisco County after the corrections.
    - *Fixer:* the fixer discount, the value change when the listing remarks describe the house as a fixer.
    - *Sale type:* the value change from a standard sale for a trust/probate sale, a short sale and a foreclosure sale. A sale type with fewer than 3 sales in the town counts as standard, and the notes under the chart say so.
5. **Surfaces (3D).** Choose a surface from the list:
    - *Location (latitude × longitude):* how the typical sale's value changes across the town. The surface is left blank outside the town limit.

- *Pairs of features:* where the best model lets two features act together (an interaction pair).
- *Living area × CQA score:* the value of the typical sale at each CQA score from 0.00 to 10.00.

Drag to turn a 3D view and scroll to zoom. The contour map beside it shows the same surface flat, with lines of equal value.

6. **Equations.** Each fitted model written out. MARS and the elastic net are sums of straight-line pieces (hinges); the GAM is a sum of smooth curves, drawn in the graphs above.
7. **Comparison of the models.** Two tables: accuracy on the fitting sales and on sales held out of each fit (R^2, CV R^2, OOF R^2, RMSE*, MdAPE*), and whether the three models agree on the adjustments at the typical sale.
8. **Definitions.** Every term used on the page.

4.3 On every graph

- Hovering shows the exact values.
- The buttons in a graph's top corner zoom, pan, reset the view and save the graph as a PNG image.
- Where a legend is shown, clicking a name hides or shows that line; double-clicking shows that line alone.

5 Reading the figures

- **R^2 first, then the held-out measures.** R^2 is measured on the fitting sales. CV R^2, OOF R^2, RMSE* and MdAPE* are measured on sales left out of each fit, so they show how well a model values sales it has not seen. A model fits its own sales better than it values new ones, so R^2 runs above the held-out measures.
- **Typical sale.** A typical sale, not a typical house of the town: houses that sell are often more updated than the houses that do not.
- **Residual.** The sale price less the model's value. The residual holds mostly what the features do not measure, such as condition, quality and appeal.
- An asterisk (*) marks a standard statistical name.

Definitions

Best model
Of the three programs' models (GAM, MARS, elastic net), the one with the highest OOF R^2.

CQA score
Condition, quality and appeal: the percentile of a house's market appeal divided by 10, from 0.00 to 10.00, read from the residual and refined by the appraiser's judgment.

CV R^2
The average, over the ten groups of sales, of the share of price variation explained for each group when the model is fitted without that group (cross-validated R^2).

Effective date
The date the values refer to; every graph shows the market at this date.

Fitting sales (training sample)
The town's sales the models are fitted on, after the data rules have removed sales that cannot be used.

Fixer discount
The value change when the listing remarks describe the house as a fixer.

Interaction pair
Two features the model is allowed to combine, so the effect of one depends on the level of the other.

Log model (log-transformed model), $ model (untransformed model)
The two models of each program. The log model fits the logarithm of the price, so effects are percentages of value; the $ model fits the price itself, so effects are dollars.

MdAPE* (median absolute percentage error)
The middle value of the held-out sales' percentage differences between price and the model's value, ignoring sign. Lower is better.

OOF R^2
The share of price variation explained when every sale is valued by a model fitted without that sale's group (pooled out-of-fold R^2).

R^2
The share of the variation in sale prices that the model explains, on the sales it was fitted on (coefficient of determination). Higher is better.

RMSE* (root mean square error)
The square root of the average squared difference between price and the model's value on held-out sales, in dollars. Lower is better.

Sale type
Who sold the house: a standard seller, a trustee or executor (trust/probate), an owner in a short sale, or a lender after foreclosure, read from the listing remarks.

Unincorporated market
A town page for a place that is not a city: Alamo, Bay Point, El Sobrante, Kensington, Rodeo, Crockett, Bethel Island and Pacheco. On the map each is drawn as its Census place plus the 400 m grid cells that hold at least 3 of its sales. The cities are drawn the same way from their city limits, since many of their sales lie in unincorporated areas listed under the city's name.

Town limit
The incorporated boundary of the town, from the US Census Bureau's 2024 TIGER/Line places.

Typical sale
The middle (median) of each feature over the town's fitting sales, valued by the town's best model at the effective date.

© 2026 William Bert Craytor · Valuation Engineer Research Labs. Town limits and the county line: US Census Bureau TIGER/Line and cartographic boundary files 2024 (public domain). Elevation and ocean views from USGS 3DEP data (public domain). Building outlines from Microsoft Building Footprints (ODbL). Graphs drawn with Plotly.js 4.1.2 (MIT license).