NorCal Home Insurance Company

Rohan Agarwal - 3035578142
Chloe Chung - 3037565517
Emir Dinc - 3036718832
Qichao Aoc - 3035094492
Vishal Koganti - 3037381475

Background

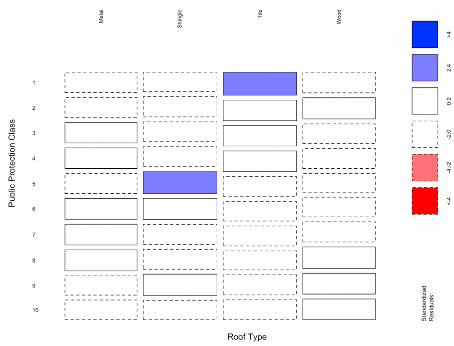
As a result of the large wildfires in 2017 and their severe impact on NCHIC's book of business, the company has taken action to adjust the premium of its homeowners product. The data analytics team has conducted modeling for this project and provided statistically sound rating variables and relativities to the actuarial team to provide new risk management solutions against wildfires and other catastrophes.

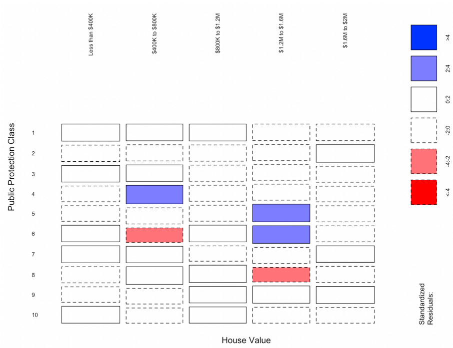
In insurance, exposure refers to the amount of risk that an insurer takes on in insuring a particular individual or group. Exposure is the probability of loss, which is determined by the severity and frequency of losses. It is important in terms of pricing for insurance because insurers need to know how likely it is that a policyholder will make a claim. It is necessary for insurance because it directly affects the premium that an insurer charges. The higher the exposure, the higher the premium. Insurance companies use exposure to determine the amount of risk they are willing to take on in insuring a particular individual or group. The amount of risk an insurer is willing to take on is directly related to the premium that they charge for coverage. One way to think of exposure is the amount of money an insurer would have to pay out in the event of a covered loss. For example, if an insurer has a policy with a \$100,000 limit and a covered loss occurs that costs \$150,000, the insurer would have to pay \$50,000 out of their own pocket. This \$50,000 is the insurer's exposure. The most commonly used exposure base for homeowner insurance is the replacement cost of the home. This is because it provides the most coverage for the home and its contents.

Preliminary Analysis

Correlated Factors

In order to determine the relatedness between two categorical variables, a chi-squared test is used against each pair of property characteristics. We found that the correlation between roof type and public protection class is high with a chi-squared of 39.428 and a p-value of 0.0579. Similarly, there is a high correlation between house value and public protection class with a chi-squared of 63.88 and a p-value of 0.002853. The mosaic plots allow us to better visualize the correlation between variables. We see that tile roofs are correlated with a public protection class of 1 meaning that tile roofs can better protect property against fire, while shingle roofs perform mediocrely in property fire protection. Moreover, properties with lower house value are associated with a higher public protection class. Before considering whether to include these correlated factors in the premium calculation, we must also observe their interactions with claim frequency and severity which determines the claim losses the company has to pay in the end.



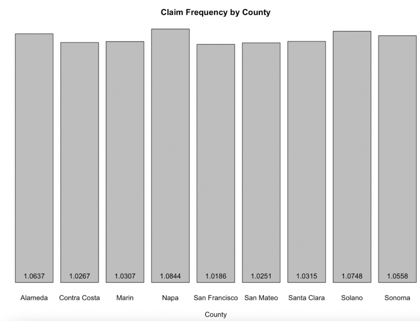


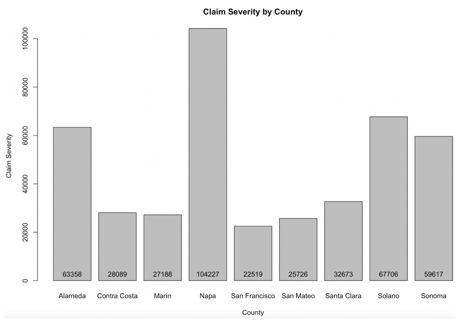
Pricing Factors

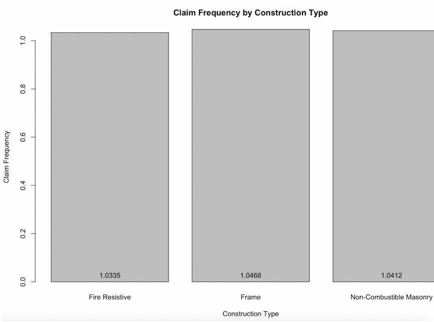
To model the claim frequency and severity using generalized linear regressions, we can consider Poisson and Gamma distribution respectively. Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time, so Poisson's distribution is a method in the frequency distribution. On the other hand, Gamma distribution can use the shape parameter to manipulate the graph to normal as the sample size increases. Moreover, the Gamma distribution is skewed to the right, which aligns with the nature of claim losses where an insurance company would usually have more small claims and fewer large claims. Moreover, the Tweedie distribution is often used to describe pure premium in insurance. As a family of probability distributions that include the Normal, Gamma, and Poisson distributions as special cases, the Tweedie distribution is able to model both the variability of claim frequency and the variability of claim severity.

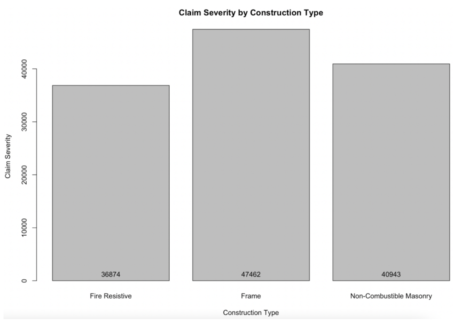
To find out which property characteristics affect claim frequency and severity the most, we test each variable against claim frequency and severity using Poisson and Gamma generalized linear regressions respectively. In the end, we find that county, construction type, and house value are the most relevant factors with statistically significant results, and we can use bar graphs to visualize the effects of each relevant variable on claim frequency and severity.

In the county graphs, we can observe significant differences in claim frequency and severity among the different counties, where Napa has the highest claim frequency and severity while San Francisco has the lowest.



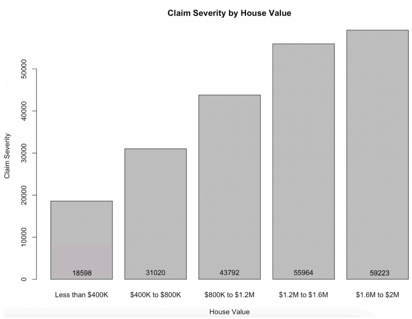


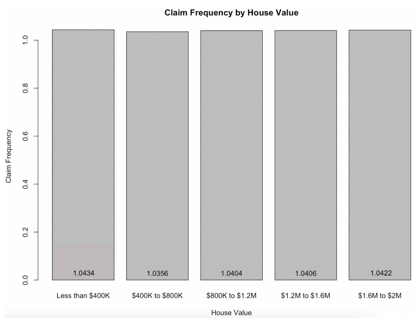




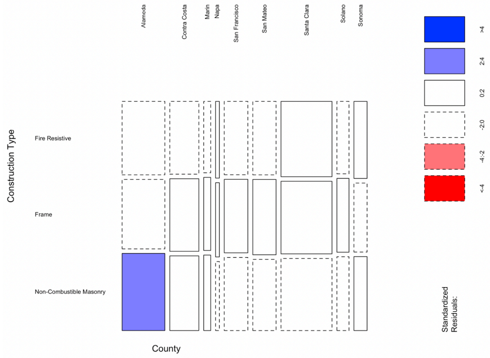
Similarly, we see using frame as the construction type has the highest claim frequency and severity, but houses with fire resistive construction have the lowest frequency and severity.

In the house value graphs, we can observe the trend that higher house value is associated with higher claim frequency and severity, which is expected as more expensive housing is prone to more man-made destructions including burglary.





To further investigate the relationship among these three variables chi-square test and mosaic plots are conducted again and we see a correlation between county and construction type where properties located in Alameda mainly use non-combustible masonry as construction type. Therefore, when calculating the rate relativities in pricing insurance, we can consider including this correlation between county and construction type in the calculation.



Rate Relativities

To calculate the rate relativities for counties, construction type, and house value, we first divide the claim frequency and claim severity of each category by the base level given to get the frequency rate and severity rate respectively. By multiplying these two rates, we then have the pure premium rate. The table shows the indicated rates for house value using this method.

House Value	Claim Amount	Claim Count	Total Exposure	Claim Frequency	Claim Severity	Frequency Rate	Severity Rate	Pure Premium Rate
Less than $400K	115867574	6230	5971	1.0434	18598	1.0075	0.5996	0.6041
$400K to $800K	195087339	6289	6073	1.0356	31020	1.0000	1.0000	1.0000
$800K to $1.2M	269848384	6162	5923	1.0404	43792	1.0046	1.4117	1.4182
$1.2M to $1.6M	347203257	6204	5962	1.0406	55964	1.0049	1.8041	1.8129
$1.6M to $2M	374701518	6327	6071	1.0422	59223	1.0064	1.9092	1.9213

County	Construction Type	Frequency Rate	Severity Rate	Pure Premium Rate	Simple Pure Premium Rate	Difference
Alameda	Fire Resistive	0.9505	0.6340	0.6026	0.7105	-0.1079
Alameda	Frame	0.9737	0.8479	0.8257	0.9261	-0.1005
Alameda	Non-Combustible Masonry	0.9648	0.7236	0.6982	0.7948	-0.0966
Contra Costa	Fire Resistive	0.9285	0.3139	0.2914	0.3041	-0.0126
Contra Costa	Frame	0.9363	0.3532	0.3307	0.3963	-0.0656
Contra Costa	Non-Combustible Masonry	0.9235	0.3090	0.2853	0.3401	-0.0548
Marin	Fire Resistive	0.9245	0.2946	0.2724	0.2954	-0.0231
Marin	Frame	0.9268	0.3273	0.3033	0.3851	-0.0818
Marin	Non-Combustible Masonry	0.9469	0.3218	0.3047	0.3305	-0.0258
Napa	Fire Resistive	0.9719	1.0483	1.0189	1.1916	-0.1727
Napa	Frame	0.9888	1.3519	1.3368	1.5532	-0.2163
Napa	Non-Combustible Masonry	0.9846	1.2245	1.2057	1.3329	-0.1272
San Francisco	Fire Resistive	0.9210	0.2432	0.2240	0.2418	-0.0179
San Francisco	Frame	0.9186	0.2847	0.2615	0.3152	-0.0537
San Francisco	Non-Combustible Masonry	0.9266	0.2545	0.2358	0.2705	-0.0347
San Mateo	Fire Resistive	0.9252	0.2721	0.2517	0.2780	-0.0263
San Mateo	Frame	0.9301	0.3509	0.3264	0.3624	-0.0360
San Mateo	Non-Combustible Masonry	0.9283	0.2683	0.2491	0.3110	-0.0619
Santa Clara	Fire Resistive	0.9284	0.3333	0.3094	0.3553	-0.0459
Santa Clara	Frame	0.9392	0.4347	0.4083	0.4631	-0.0549
Santa Clara	Non-Combustible Masonry	0.9338	0.3682	0.3438	0.3975	-0.0536
Solano	Fire Resistive	0.9551	0.5829	0.5567	0.7672	-0.2105
Solano	Frame	1.0000	1.0000	1.0000	1.0000	0.0000
Solano	Non-Combustible Masonry	0.9633	0.7557	0.7280	0.8582	-0.1302
Sonoma	Fire Resistive	0.9511	0.6638	0.6313	0.6636	-0.0323
Sonoma	Frame	0.9549	0.7175	0.6851	0.8650	-0.1799
Sonoma	Non-Combustible Masonry	0.9614	0.6928	0.6661	0.7424	-0.0763

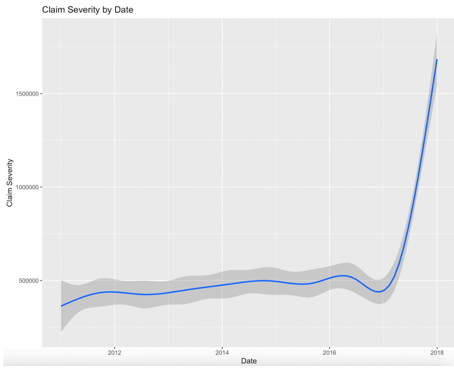
Number of Policies	Fire Resistive	Frame	Non-Combustible Masonry
Alameda	2097	1982	2202
Contra Costa	1398	1395	1435
Marin	328	336	347
Napa	190	184	171
San Francisco	1147	1149	1144
San Mateo	1132	1161	1097
Santa Clara	2541	2449	2434
Solano	581	592	592
Sonoma	671	601	644

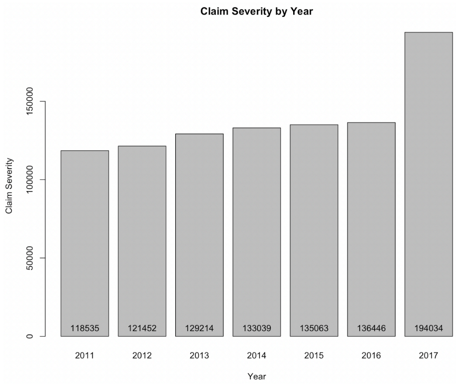
However, special consideration is given to county and construction type as we observe a correlation between these two property characteristics from the mosaic plot. Therefore, we combine county and construction to create 27 categories and recalculate the frequency, severity, and pure premium rates. Compared to the simple pure premium rate, where we simply multiply the pure premium rates of county and construction type calculated separately, we see the pure premium rate is lower after considering correlation.

When calculating the simple pure premium rate, we assume there is a uniform distribution of exposures among all 27 categories. Yet, looking at the total number of policies in each category, we can tell the distribution of policies is not even among all categories. Because of the disproportionate number of exposures, ignoring the correlation between county and construction type would lead to distortion in the indicated rate relativities, resulting in the "double counting" effect. Therefore, by creating 27 categories instead, we can account for the distributional bias, prevent overcharging policyholders, and avoid potential adverse selection.

Catastrophe Risk Analysis

Catastrophe risk management is different from traditional risk management because catastrophes such as hurricanes, tornadoes, floods, and earthquakes are extremely difficult to predict. We cannot model catastrophic losses using the same method as car accidents or burglaries due to the small sample size and rare occurrence. Also, catastrophes would affect a large group of policyholders at once and insurance companies need to pay out claim losses at the same time, which requires a strong and stable reserve. Due to these two reasons, there can be certain years in which a catastrophe occurs in which the claim severity is extremely high compared to other years. This means that during the years in which a catastrophe occurs, the insurer will either need to be able to pay the increase in the number of claims using existing reserves or develop a catastrophe management plan to account for these special occasions.





To see whether there are catastrophic events in the past seven years, we put claim severity against time and found out the claim severity in 2017 is extremely high compared to other years due to the wildfires. In order to investigate the significant increase in catastrophic claims, the 2017 data should be singled out and further analyzed against each property characteristic to see which characteristics are the most relevant to catastrophic claims. Therefore, the generalized linear regressions are done again using only 2017 claims data. Yet, we found no significant deviations from the aggregate data, and the three relevant factors including county, construction type, and house value remain true. Therefore, to address future potential catastrophes, an overall catastrophe load can be added to the premium calculation to compensate for future catastrophe losses. The catastrophe load can be estimated by dividing the claim severity in 2017 by the claim severity in 2016 to see how much more claims are needed to be paid because of the catastrophe, which is 194033.8/136445.8 = 1.4221.

Recommendation

To calculate the premium charged for each policy, we first calculate the adjusted pure premium according to the different property characteristics of each policy using the following formula.

Adjusted Pure Premium = Base Class Pure Premium * Rate Relativity (House Value) * Rate Relativity (County & Construction Type) * Catastrophe Load

For Base Class Pure Premium = 5000 and Catastrophe Load = 1.4221, where Base Class Pure Premium is for a property located in Solano, having frame as the construction type, and valued at $400K to $800K.

Then, the premium charged for each policy is calculated using the pure premium approach.

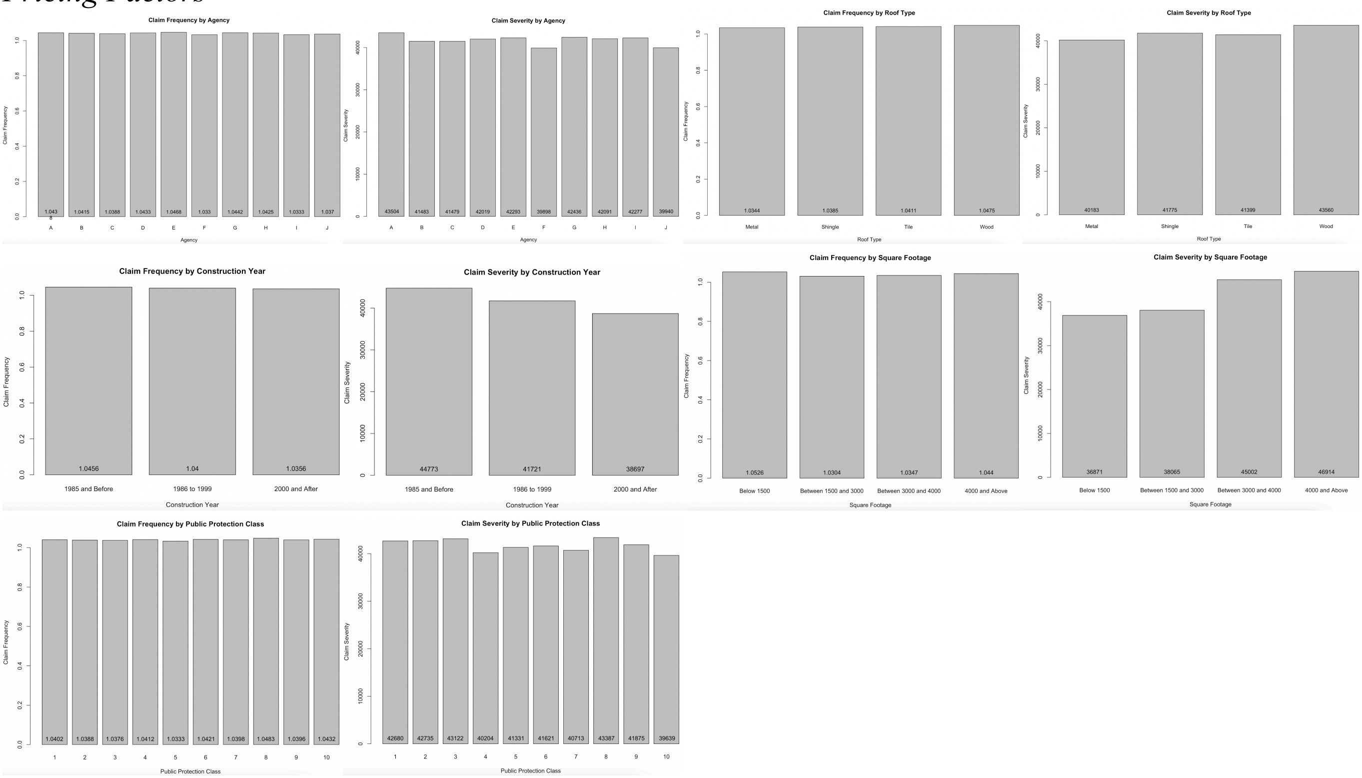
Premium Charged = Adjusted Pure Premium / (1- (Variable Expense Ratio + Profit Load))

For Variable Expense Ratio = 45% and Profit Load = 5%.

As catastrophes can be extremely unpredictable and damaging, they can cost an insurer a lot of money with proper management. Apart from including a catastrophes load and using internal reserves to absorb such losses, there are other ways for an insurer to be protected against large losses from catastrophes. One possible method is through reinsurance, where an insurer transfers part of their risk portfolio to another party, mostly a reinsurance company. This way, direct insurers are not responsible to pay all of the claims and lower their financial and liquidity strain. Although the insurer will need to pay a share of the insurance premium to the reinsurer, it allows the insurer to survive the financial burden during catastrophes and other events. Another possible solution for insurers to manage catastrophe risk is to offer separate insurance specifically for catastrophe damages, or an optional rider attached to the homeowner insurance. Insurers will only be responsible to cover catastrophe damages for the policyholders with the specific catastrophe policy or rider, lowering several claim losses due to catastrophes.

<u>Appendix</u>

Pricing Factors



Rate Relativities

County	Claim Amount	Claim Count	Total Exposure	Claim Frequency	Claim Severity	Frequency Rate	Severity Rate	Pure Premium Rate
Alameda	423292887	6681	6281	1.0637	63357.71	0.9897	0.9358	0.9261
Contra Costa	121934953	4341	4228	1.0267	28089.14	0.9553	0.4149	0.3963
Marin	28330270	1042	1011	1.0307	27188.36	0.9589	0.4016	0.3851
Napa	61597945	591	545	1.0844	104226.64	1.0089	1.5394	1.5532
San Francisco	78905295	3504	3440	1.0186	22518.63	0.9477	0.3326	0.3152
San Mateo	89396147	3475	3390	1.0251	25725.51	0.9537	0.3800	0.3624
Santa Clara	250206788	7658	7424	1.0315	32672.60	0.9597	0.4826	0.4631
Solano	128438607	1897	1765	1.0748	67706.17	1.0000	1.0000	1.0000
Sonoma	120605180	2023	1916	1.0558	59616.99	0.9824	0.8805	0.8650

Construction Type	Claim Amount	Claim Count	Total Exposure	Claim Frequency	Claim Severity	Frequency Rate	Severity Rate	Pure Premium Rate
Fire Resistive	384342667	10423	10085	1.0335	36874.48	0.9875	0.7769	0.7672
Frame	489242600	10308	9849	1.0466	47462.42	1.0000	1.0000	1.0000
Non-Combustible Masonry	429122805	10481	10066	1.0412	40942.93	0.9949	0.8626	0.8582

County	Construction Type	Claim Amount	Claim Count	Total Exposure	Claim Frequency	Claim Severity	Frequency Rate	Severity Rate	Pure Premium Rate	Simple Pure Premium Rate	Difference
Alameda	Fire Resistive	2202	2097	120538828	1.0501	54741	0.9505	0.6340	0.6026	0.7105	-0.1079
Alameda	Frame	2132	1982	156102344	1.0757	73219	0.9737	0.8479	0.8257	0.9261	-0.1005
Alameda	Non-Combustible Masonry	2347	2202	146651716	1.0658	62485	0.9648	0.7236	0.6982	0.7948	-0.0966
Contra Costa	Fire Resistive	1434	1398	38862806	1.0258	27101	0.9285	0.3139	0.2914	0.3041	-0.0126
Contra Costa	Frame	1443	1395	44012558	1.0344	30501	0.9363	0.3532	0.3307	0.3963	-0.0656
Contra Costa	Non-Combustible Masonry	1464	1435	39059589	1.0202	26680	0.9235	0.3090	0.2853	0.3401	-0.0548
Marin	Fire Resistive	335	328	8522128	1.0213	25439	0.9245	0.2946	0.2724	0.2954	-0.0231
Marin	Frame	344	336	9721924	1.0238	28261	0.9268	0.3273	0.3033	0.3851	-0.0818
Marin	Non-Combustible Masonry	363	347	10086218	1.0461	27786	0.9469	0.3218	0.3047	0.3305	-0.0258
Napa	Fire Resistive	204	190	18466601	1.0737	90523	0.9719	1.0483	1.0189	1.1916	-0.1727
Napa	Frame	201	184	23464175	1.0924	116737	0.9888	1.3519	1.3368	1.5532	-0.2163
Napa	Non-Combustible Masonry	186	171	19667169	1.0877	105737	0.9846	1.2245	1.2057	1.3329	-0.1272
San Francisco	Fire Resistive	1167	1147	24504288	1.0174	20998	0.9210	0.2432	0.2240	0.2418	-0.0179
San Francisco	Frame	1166	1149	28666992	1.0148	24586	0.9186	0.2847	0.2615	0.3152	-0.0537
San Francisco	Non-Combustible Masonry	1171	1144	25734015	1.0236	21976	0.9266	0.2545	0.2358	0.2705	-0.0347
San Mateo	Fire Resistive	1157	1132	27183241	1.0221	23495	0.9252	0.2721	0.2517	0.2780	-0.0263
San Mateo	Frame	1193	1161	36147594	1.0276	30300	0.9301	0.3509	0.3264	0.3624	-0.0360
San Mateo	Non-Combustible Masonry	1125	1097	26065312	1.0255	23169	0.9283	0.2683	0.2491	0.3110	-0.0619
Santa Clara	Fire Resistive	2606	2541	74999975	1.0256	28780	0.9284	0.3333	0.3094	0.3553	-0.0459
Santa Clara	Frame	2541	2449	95376246	1.0376	37535	0.9392	0.4347	0.4083	0.4631	-0.0549
Santa Clara	Non-Combustible Masonry	2511	2434	79830568	1.0316	31792	0.9338	0.3682	0.3438	0.3975	-0.0536
Solano	Fire Resistive	613	581	30855577	1.0551	50335	0.9551	0.5829	0.5567	0.7672	-0.2105
Solano	Frame	654	592	56471675	1.1047	86348	1.0000	1.0000	1.0000	1.0000	0.0000
Solano	Non-Combustible Masonry	630	592	41111356	1.0642	65256	0.9633	0.7557	0.7280	0.8582	-0.1302
Sonoma	Fire Resistive	705	671	40409223	1.0507	57318	0.9511	0.6638	0.6313	0.6636	-0.0323
Sonoma	Frame	634	601	39279093	1.0549	61954	0.9549	0.7175	0.6851	0.8650	-0.1799
Sonoma	Non-Combustible Masonry	684	644	40916863	1.0621	59820	0.9614	0.6928	0.6661	0.7424	-0.0763