



CA ZIP Premium Predictor

By: Sandra Srinivasan, Gabriel Zamani,
Maria Aragon

About Us



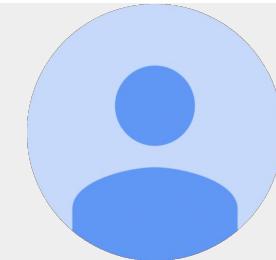
Sandra Srinivasan

Linkedin:
https://www.linkedin.com/in/saipr_aharshitha-srinivasan/
Github:
<https://github.com/sandrasri>



Gabriel Zamani

Linkedin:
<https://www.linkedin.com/in/gabriel-zamani-834408299>



Maria Aragon

Linkedin:
<http://linkedin.com/in/maria-juliana-aragon-1388ba259>

California's Home Insurance Crisis

The 2025 LA wildfires caused \$250 billion in damages and \$40 billion in insurance claims, prompting insurers to retreat and largely stop issuing home policies in California

WHO, WHAT, WHY



WHO

Insurance Companies

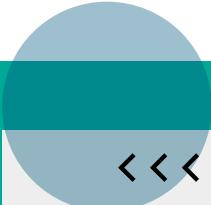
Determining ZIP Code premiums and whether its underpriced, overpriced or adequately priced based on historical risk

WHAT



WHY

To aid in future adequate pricing and reduce losses



<<<

01.

Hypothesis

Can we predict premiums accurately?
Do these premiums help us classify
underpriced or overpriced zip codes?



Understanding premiums



<<<



Premium

The amount charged to insure a home for one year
Example: covers wildfire damage



Exposure

The amount of risk the insurer is covering.
Usually: 1 home = 1 exposure



Why these terms matter?

We are determining if the insurer earns the equivalent amount of premium for each unit of risk



What factors do insurers typically use?



Property

Home value - age - prices



Location

Wildfire risk - fire protection



Experience

Loss history - severity trends



Census

Housing age, pop density,
median income



Public Hazard

CALFire, fault line locations



Premiums

Revenue per exposure,
ZHVI scores

Data Collection

WildFire Data

- GIS Fire.CA.gov mapping data
- Convert from geo latitude/ longitude to numerical factors

Census Data

- Yearly data from census bureau, needed to be filtered to state and compatible zip-codes

Earthquake Data

- USGS fault activity map
- Convert latitude/longitude to zipcode data



ZHVI Score

- Zillow databases filtered by country, further filtration required to convert location to zip-codes

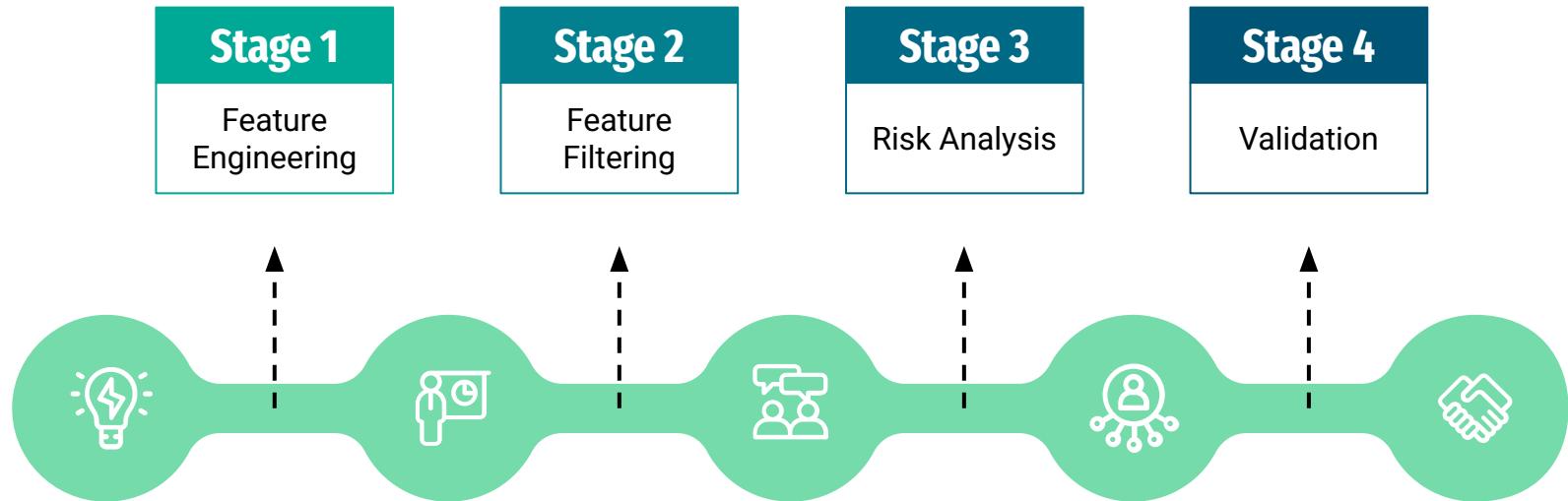
Fire Station Mapping

- Convert wildfire first responder data to their zip-code locations for more accurate localization

Premiums / Exposures

- Found clean, well-formatted California insurance data on Kaggle

Processing Data



Converting

Converting data into numerical features, relevant to premiums

VIF

Removed any factors that had multicollinearity issues

SHAP Analysis

Provides a single, interpretable risk measure

Modeling

Using regression and to determine premiums

Backtesting

Ran through the year 2021 to check if data predicted accurately

Models Comparison

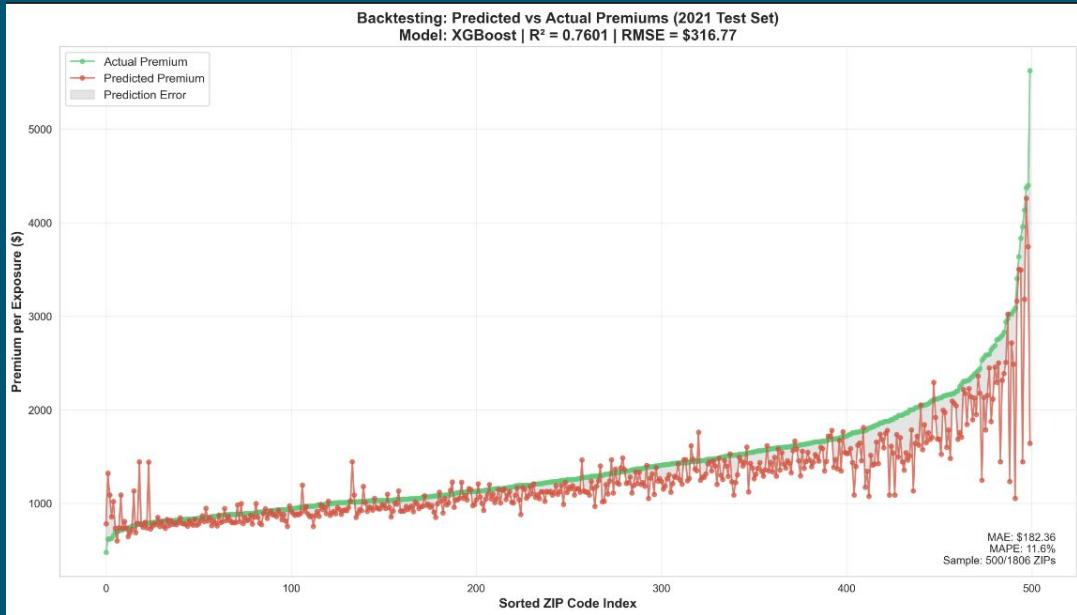


Random Forest	Models	MAE	RMSE	R2	
	RF		\$194.54	\$326.27	0.7454
	XGB		\$182.36	\$316.77	0.7602
	SVR		\$341.93	\$519.24	0.1643

Which is the best model?

XGBoost, it combines gradient boosting with regularization so its the best at handling nonlinearity and feature interactions

How Do We Verify?



Predicted vs
actual
premium trend
closely
matched

02.

PRODUCT DEMO

California zip premium predictor

Limitations & Future Improvements



Data Limitations

Public data only (no homeowners level features)

Why? California privacy laws limit how much data is publicly available especially to the individual scale



Model Assumptions

Predicts expected premium per exposure rather than losses

Similarly, finding loss data was limited due to privacy restrictions



Next Steps

If given, add internal insurer loss and policy data

Along with extending and including recent year data

03.

BUSINESS APPLICATIONS

What insurers can do?

< < <

SUPPORT RATE FILINGS

Use zip-level mispricing data for regulatory filings or rate changes

OPTIMIZE PORTFOLIO

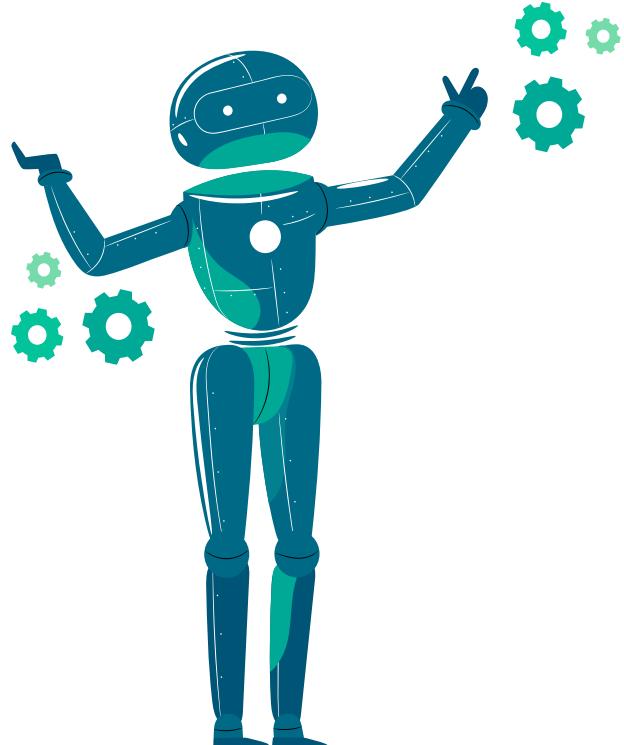
Grow in fairly pricing ZIPs, reprice or exit underpriced high-risk ZIPs

MONITOR RISK PROACTIVELY

Update as hazard and economic data change to stay ahead of losses

GAIN COMPETITIVE EDGE

Offer better risk-aligning pricing, before competitors



Thank you!!

Any questions!

Project Github: <https://github.com/sandrasri/CaliforniaHousingPremiums.git>

SOURCES

<<<<

Merlin Law Group. (2024). "Insurance Companies Cancel Fire Insurance in California."

<https://www.merlinlawgroup.com/insurance-companies-cancel-fire-insurance/>

Yale Law Journal. "The Uninsurable Future: The Climate Threat to Property Insurance and How to Stop It."

<https://yalelawjournal.org/essay/the-uninsurable-future-the-climate-threat-to-property-insurance-and-how-to-stop-it>

CAL FIRE. Fire Resource Assessment Program (FRAP) - GIS Mapping and Data Analytics.

<https://www.fire.ca.gov/Home/What-We-Do/Fire-Resource-Assessment-Program/GIS-Mapping-and-Data-Analytics>

CAL FIRE. Historic Fire Perimeters Data Dictionary.

<https://34c031f8-c9fd-4018-8c5a-4159cdff6b0d-cdn-endpoint.azureedge.net/-/media/calfire-website/what-we-do/fire-resource-assessment-program---frap/historic-fire-perimeters-data-dictionary.pdf>

National Interagency Fire Center (NIFC). California Wildland Fire Direct Protection Areas (2023).

<https://data-nifc.opendata.arcgis.com/datasets/nifc::california-wildland-fire-direct-protection-areas-2023/about>

CAL FIRE Forestry. Fire Hazard Severity Zones (FHSZ) GIS Data.

<https://calfire-forestry.maps.arcgis.com/home/item.html?id=93a1f8cc1456497f86ecd25933e6c9b9>

California Geological Survey (CGS). Seismic Hazards Program - Alquist-Priolo Fault Hazard Zones.

<https://lab.data.ca.gov/dataset/cgs-seismic-hazards-program-alquist-priolo-fault-hazard-zones1>

U.S. Census Bureau. American Community Survey (ACS) 5-Year Estimates (2018-2021). <https://data.census.gov/> - California ZIP Code Tabulation Areas (ZCTAs)

Zillow Research. Zillow Home Value Index (ZHVI) - ZIP Code Level Data. <https://www.zillow.com/research/data/>

Kaggle Dataset. California Home Insurance Premium Data (2018-2021).

<https://www.kaggle.com/datasets/nathancho1234/masterdata2018-2021>