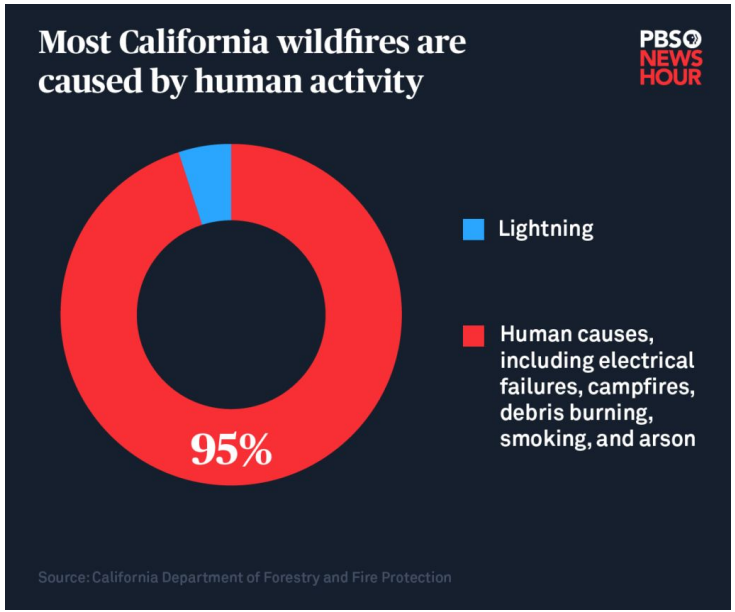
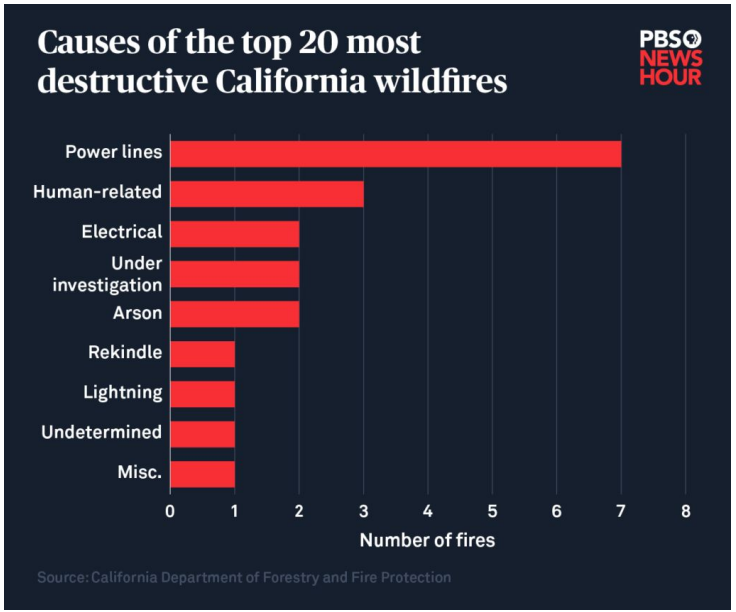
Can we predict the population increase inside California's high-risk areas of wildfire using demographic and home price data?

Shenyue Jia

February 2023

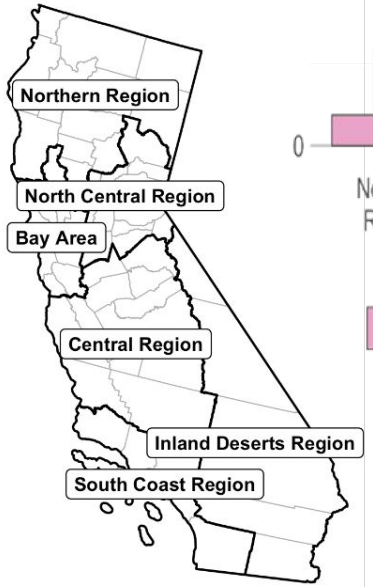
Why does population increase inside wildland-urban interface (WUI, a high risk area of wildfires) matter?

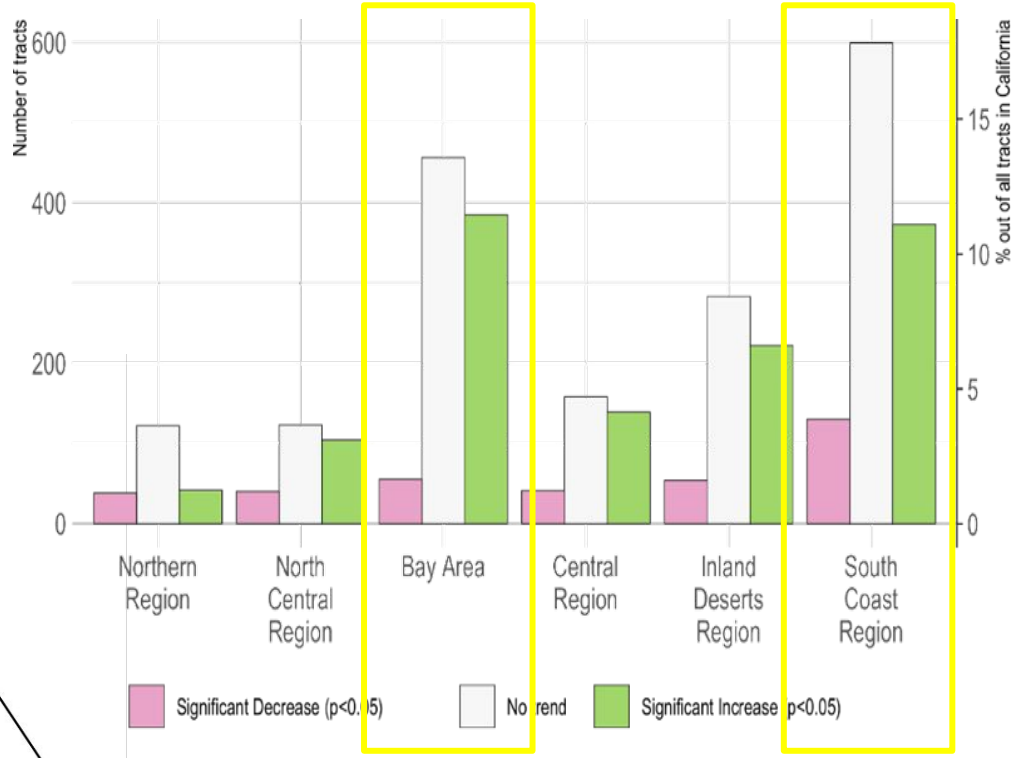




California's catastrophic wildfires in 3 charts, PBS News, Sep 14, 2020

Where did the population increase the most from 2010 to 2019?





Project Description

Predict the **rate of population increase** from 2010 to 2019 in California at the census tract level based on

- Home values and home affordability (income, income to mortgage ratio)
- Demographic indicators (ethnic diversity, renting/owning ratio, etc.)

Feature Dataset

Feature Name	Description	Feature Name	Description
House affordability index	Derived based on income to mortgage ratio	**Median home value**	Median of home value inside a census tract
Rent-own ratio	# of residents renting vs. # residents owning	**Diversity index (DI)**	How ethnically diverse is a census tract (high DI, highly diverse)
Wildfire hazard potential (WHP)	Multi-year average of wildfire risk	**Per capita income**	Mean income for every person
Wildfire risk to homes	Multi-year average wildfire risk to homes and properties	**Median household income**	Median of household income inside a census tract

Stakeholders I

This analysis can be utilized by

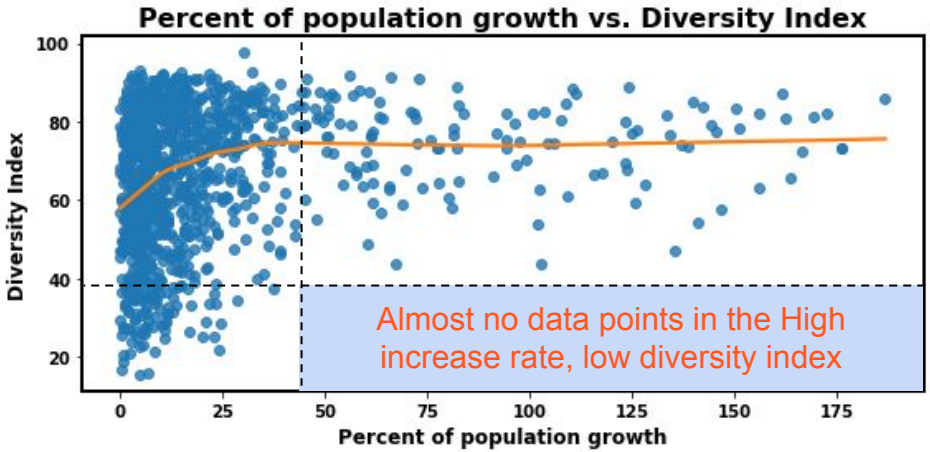
- **Insurance companies** to decide the potential area for marketing as more people are now living in high-risk wildfire area
- **Emergency response agencies** to adjust responding plans to meet the needs of communities with a fast pace of population increase, especially those with high wildfire risk

Stakeholders II

This analysis can also be utilized by

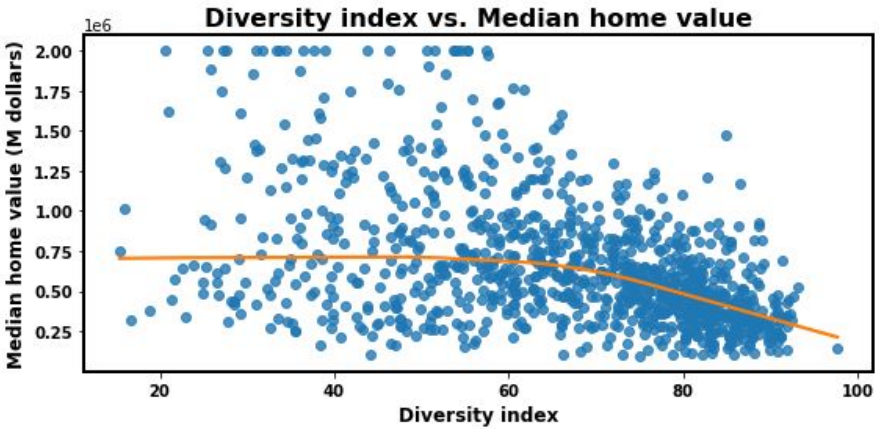
- **Real-estate developers** to determine new developing projects
  - Option 1: develop in area where median home values are high to make more profit in a relatively competitive market
  - Option 2: develop multi-level apartment complex or condos in area where rent-own ratio is high
  - Option 3: develop in area (or purchase land) where homes are more affordable (high House Affordability Index) to meet needs in a currently less competitive, but with high potential of growth in the future

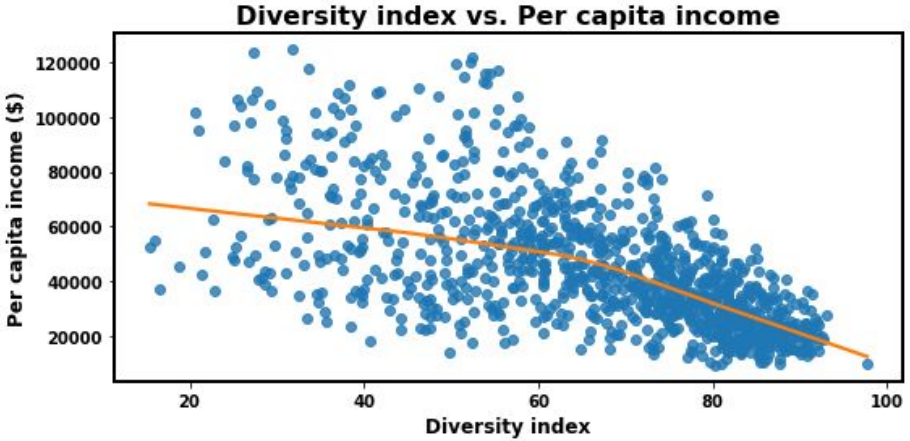
Key findings from data



- Census tracts with the greatest population increase often have a highly ethnically diverse population (high Diversity Index)
- In contrast, ethnically homogeneous tracts often have a much lower population growth rate

Key findings from data





- In highly ethnically diverse tracts with significant population increase, median home values are generally lower
  - Infer: home price is a major driver of population increase in these tracts
  - This pattern can also be confirmed by Diversity Index vs. Per Capita Income

Modeling Results

- Preprocess: 1) outlier removal, 2) min_max scaler, 3) PCA
- Model format: random forest with hyperparameter tuning

	Train	Test
R^2		
MAE		
RMSE		

Summary

- Data used in this project (home affordability and demographics) cannot successfully predict population increase rate in California
- Why
  - This question is highly related with geographic location, which was missing in the features
  - Only census tracts with the biggest increase responses well with the change of features
- Possible improvements
  - Geographic location needs to be included in the future model
  - Further filter or manipulate the target variable to focus on those with the biggest increase only

	Train	Test
R^2	0.49	0.12
MAE	14.29	17.91
RMSE	21.50	29.46