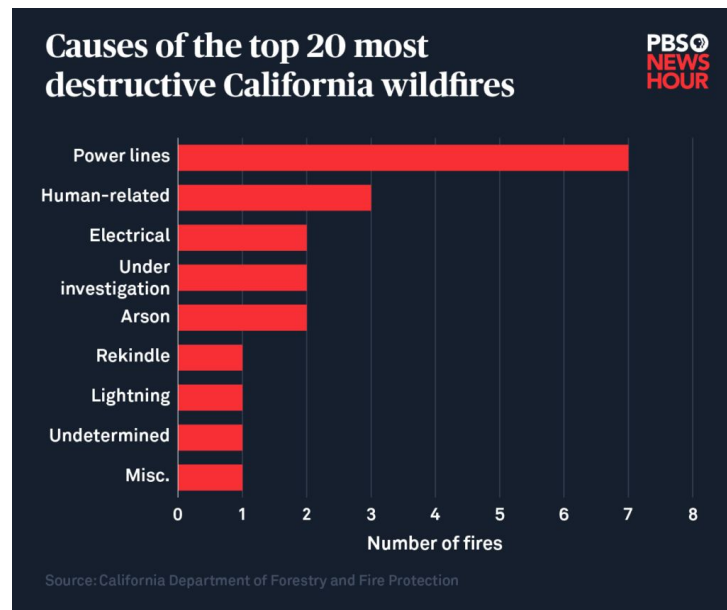
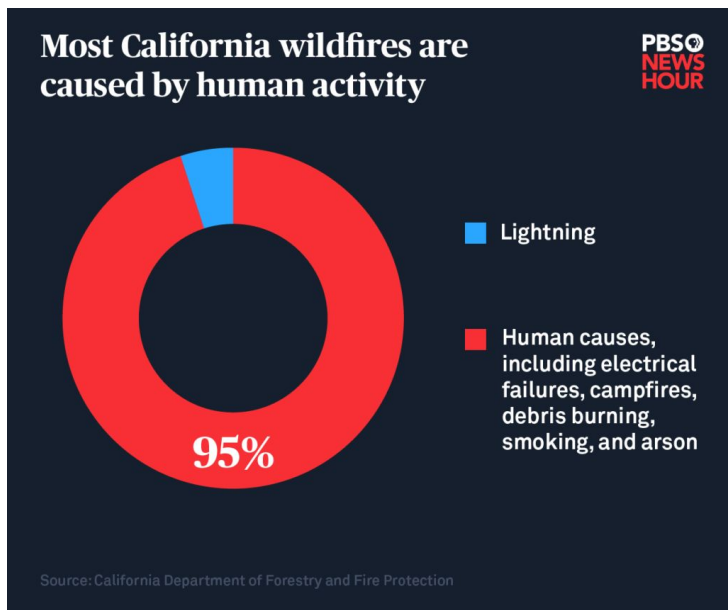


**Can we predict  
the rate of population increase  
inside California's high-risk areas of wildfire  
with demographic and home affordability data?**

Shenyue Jia

February 2023

# Why does population increase inside high-risk areas of wildfire in California matter?



# 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.)
- Target variable ONLY contains census tracts with an **increase** in population

# Stakeholders I

This analysis can be utilized by

- **Insurance companies** to identify 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

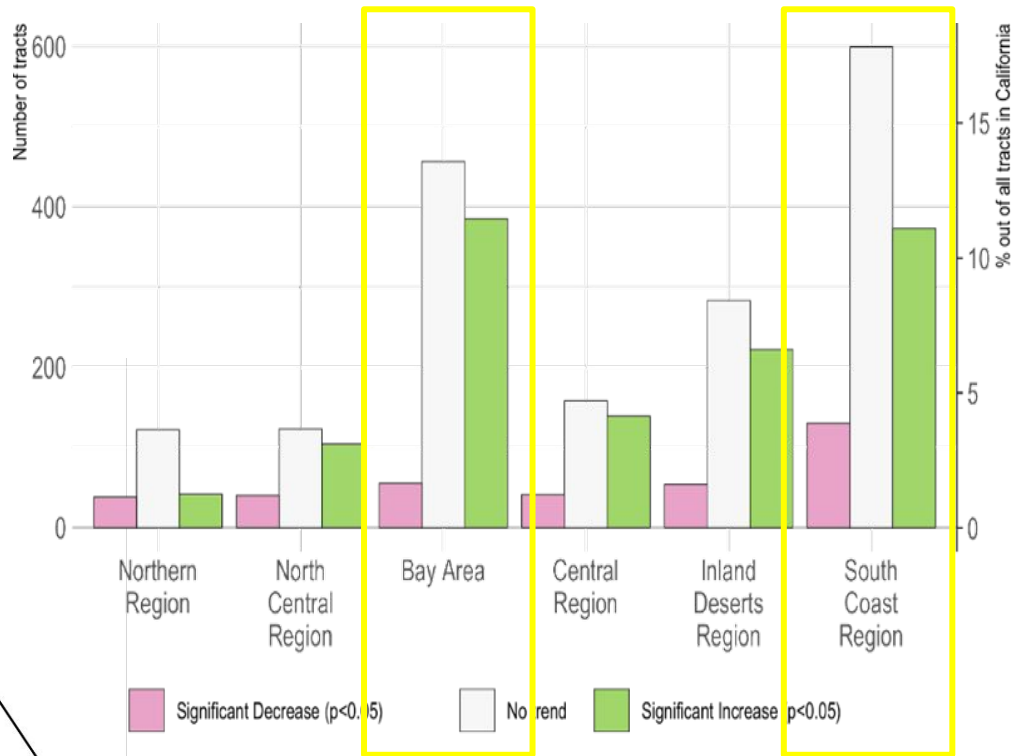
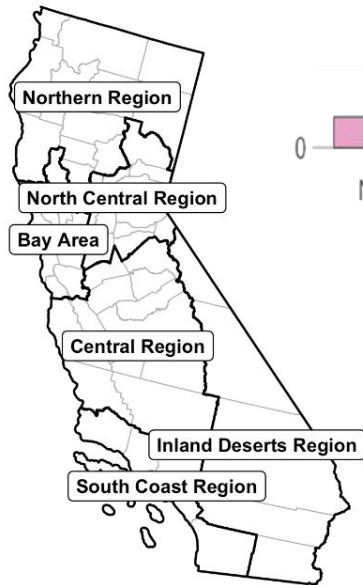
This analysis can also be utilized by

- **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

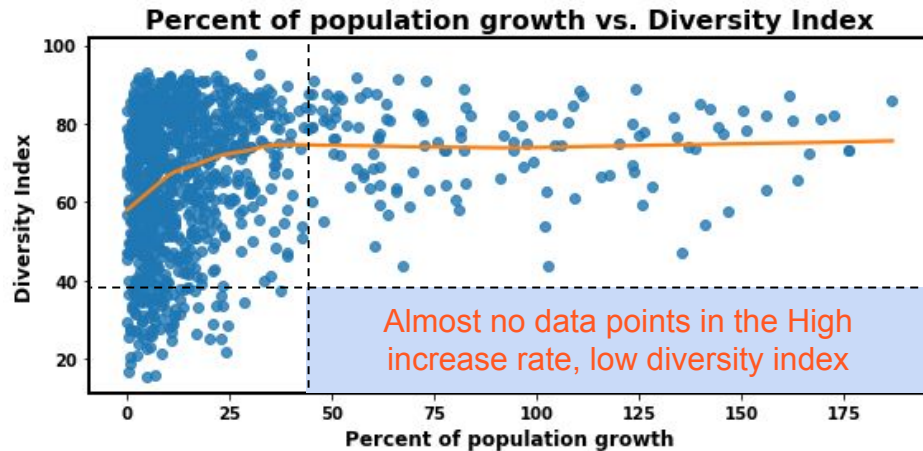
# Features (approx. 1,000 census tracts, 27 variables)

Feature Name	Description	Feature Name	Description
House affordability index	Derived based on the ratio between median income and mortgage	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

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



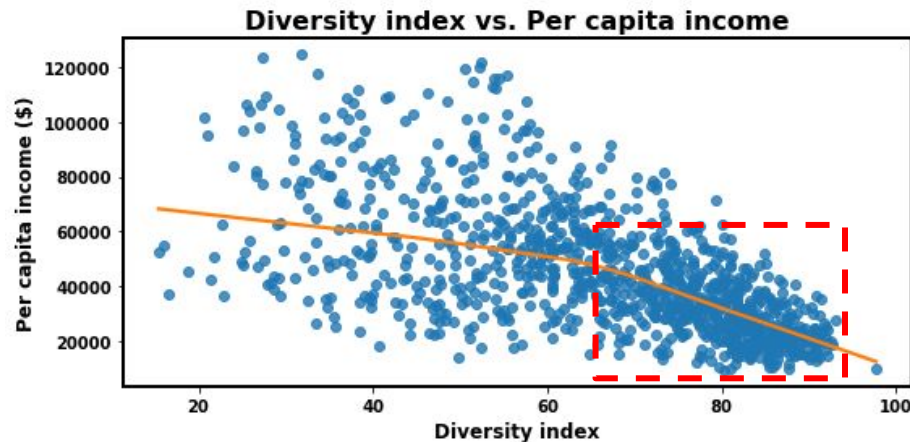
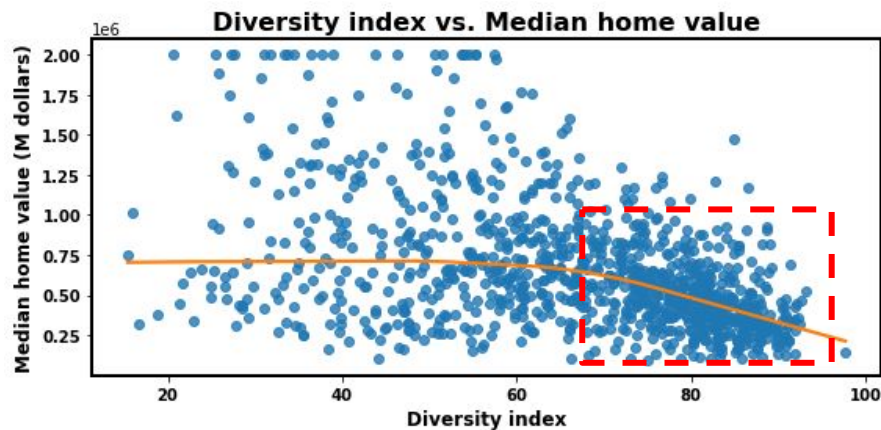
# Does ethnic diversity affect population growth



- Census tracts with the greatest population increase often have a highly ethnically diverse population (high Diversity Index)



# Ethnic Diversity and Home Affordability



- In highly ethnically diverse tracts with significant population **increase**, median home values are generally **lower**

# Modeling Results

- Preprocess: 1) outlier removal, 2) standard scaler, 3) PCA
- Model format: **Random Forest** regression model with hyperparameter tuning

	Train	Test
<b>R<sup>2</sup></b>	0.53	0.22
<b>MAE</b>	13.86	17.08
<b>RMSE</b>	20.58	27.62

# Recommendations

- Data used in this project (home affordability and demographics) cannot successfully predict population increase rate in California (yet)
- Why
  - This question is highly related with geographic location, which was missing in the features
- Possible improvements
  - Geographic location needs to be included in the future model (use different regions)
  - Separate model for each region (no one-size-fits-all solution)
  - A better target variable (e.g. **% of people county pop living in census tract with significant population increase**)

