

CMDA-4654

Xavier Akers

2025-02-18

Teammate Introduction

Meet Yusi Yao! Born in Nanjing, China, Yusi is always on the go-whether it is heading back to D.C. on weekends or enjoying the outdoors. When in Blacksburg, Rainbowl is the go-to spot for a great meal. In free time, Yusi enjoys fishing and playing soccer, making the most of both nature and sports.

Data Introduction

Dataset Overview

Our data is the **Zillow Observed Rent Index (ZORI)**. The ZORI score tracks typical rental prices in a given area. This index attempts to represent an accurate rental housing stock by focusing on the middle range of rents, excluding very high and very low prices. It is designed to depict rental housing prices for all homes, not only homes currently listed for-rent. Additionally, it is smoothed out to remove short-term spikes and provide a better understanding of long-term trends. The data covers monthly rent values from **January 2015 to January 2025**.

The ZORI score is taken for the categories: All homes, Single Family Residences, and Multi-Family Residences. We have chosen to analyze the [All Homes Plus Multifamily Time Series \(\\$\)](#) dataset.

Data Category

This dataset belongs to category 8 **Housing**. In this project, we tend to discover the LA wildfire's impact on the housing market in the ZIP codes of Los Angeles.

Data Dictionary

Column Name	Description
RegionID	Unique ID for each ZIP code.
SizeRank	Ranking of ZIP code by housing market size.
RegionName	ZIP code number.
RegionType	Type of region (e.g., "zip").
StateName	Full name of the state.
State	Two-letter state abbreviation.
City	City name.
Metro	Metro area including the ZIP code.
CountyName	County name.
2015-01-31, ..., 2025-01-31	Monthly rent estimates in dollars.

Data Source

This dataset comes from **Zillow's public data**. More details can be found at:
[Zillow Research Data](#)

Analysis & Discussion

We aim to analyze the **Rental Price Index** in areas of wildfires. Specifically, we will be looking at the location of the Easy Fire (October 2019), Getty Fire (October 2019), and the Saddle Ridge Fire (October 2019). The ZIP codes of each fire are stored in `easy_fire_zips`, `getty_fire_zips`, and `saddle_ridge_zips`.

```
easy_fire_zips = c(93065, 91360, 93021, 93063, 91320, 93012, 91362, 91361, 91307, 93015, 93066)
getty_fire_zips = c(90049, 90025, 90024, 90272, 90403, 91403, 91436, 90402, 90077, 90095, 90073)
saddle_ridge_fire_zips = c(91342, 91344, 91326, 91311, 91321, 91381, 93063, 91331, 91350, 91355, 91387, 91304,
fire_zips = c(easy_fire_zips, getty_fire_zips, saddle_ridge_fire_zips)
```

A sample of the Los Angeles Fire Housing data for only these ZIP codes is shown below.

	RegionID	SizeRank	RegionName	RegionType	StateName	State	City
1	96368	37	91342	zip	CA	CA	Los Angeles
2	96368	37	91342	zip	CA	CA	Los Angeles
3	96368	37	91342	zip	CA	CA	Los Angeles

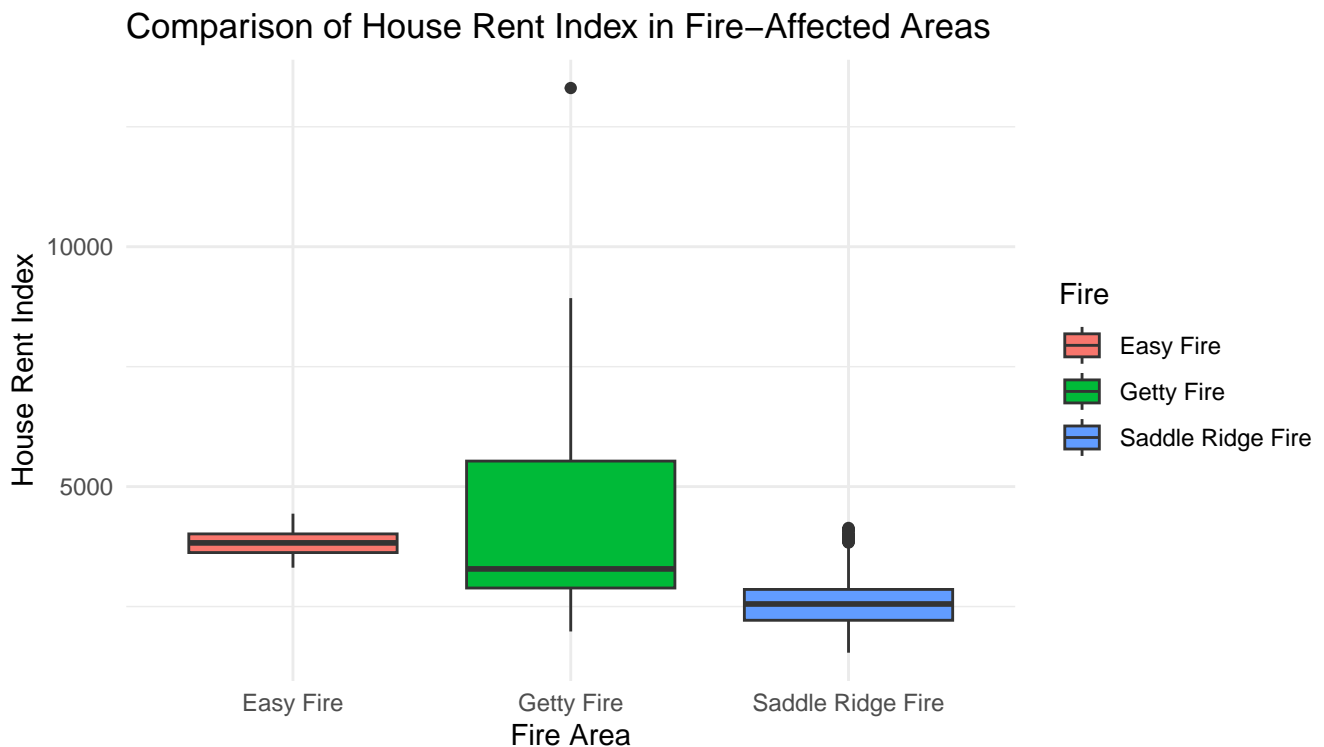
	Metro	CountyName	Date	Rent_Price
1	Los Angeles-Long Beach-Anaheim, CA	Los Angeles County	2022-02-28	2588.083
2	Los Angeles-Long Beach-Anaheim, CA	Los Angeles County	2022-03-31	2592.541
3	Los Angeles-Long Beach-Anaheim, CA	Los Angeles County	2022-04-30	2597.000

	Fire
1	Saddle Ridge Fire
2	Saddle Ridge Fire
3	Saddle Ridge Fire

Location Comparisons

The following boxplot displays the distribution of rent prices in ZIP codes affected by the different wildfires. From the distribution of rent prices, we can make educated assumptions about each location's rental market.

```
ggplot(fire_housing_data, aes(x = Fire, y = Rent_Price, fill = Fire)) +
  geom_boxplot() +
  labs(
    title = "Comparison of House Rent Index in Fire-Affected Areas",
    x = "Fire Area",
    y = "House Rent Index"
  ) +
  theme_minimal()
```

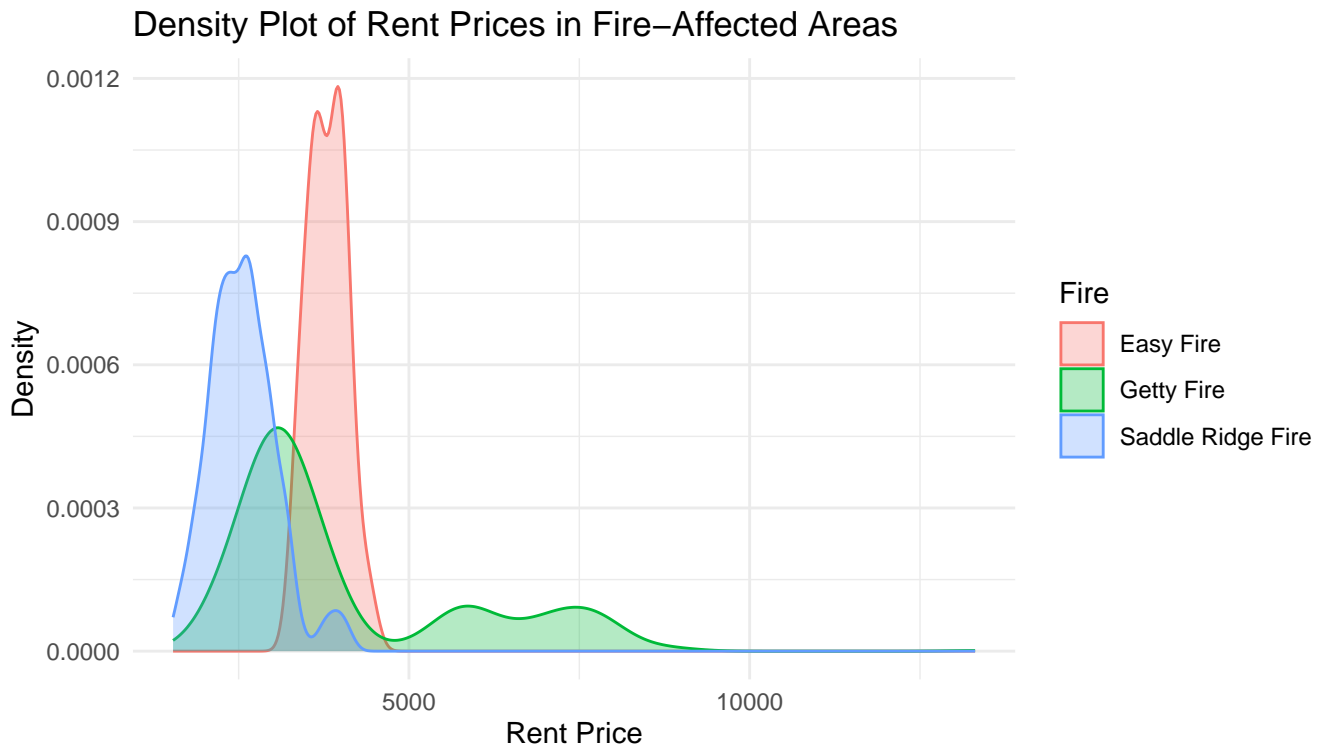


Rental indexes in the Easy Fire region were generally higher than the other regions, with very few extreme values. The Getty Fire region had a wider range of rental prices, including some outliers. This suggests the Getty Fire also included some wealthier or more in demand areas. The Saddle Ridge Fire region had a more concentrated range of rental prices, with a few high-end outliers.

The following density plot visualizes each of the wildfire locations. This helps us better understand the underlying distribution of the data.

```
ggplot(fire_housing_data, aes(x = Rent_Price, color = Fire, fill = Fire)) +
  geom_density(alpha = 0.3) +
  labs(
    title = "Density Plot of Rent Prices in Fire-Affected Areas",
    x = "Rent Price",
    y = "Density"
  )
```

```
) +  
theme_minimal()
```



This visual reinforces our assumptions from the prior boxplot. Residences in the Easy Fire region experience lower variance in rental price index. Houses in the Saddle Ridge fire are generally lower than the other two regions, with a few outliers, potentially indicating a more expensive or in demand area nearby. The Getty fire region has the largest variance, with one large peak and two smaller peaks on the high end. This indicates that the Saddle Ridge Fire spread to a broader range of communities.

Below is a table displaying the summary statistics of the rental prices in the fire-affected areas.

```
summary_table <- fire_housing_data %>%  
  group_by(Fire) %>%  
  summarize(  
    Mean_Rent = mean(Rent_Price, na.rm = TRUE),  
    Median_Rent = median(Rent_Price, na.rm = TRUE),  
    Min_Rent = min(Rent_Price, na.rm = TRUE),  
    Max_Rent = max(Rent_Price, na.rm = TRUE),  
    Std_Dev = sd(Rent_Price, na.rm = TRUE)  
  )  
  
print(summary_table)
```

A tibble: 3 x 6

Fire	Mean_Rent	Median_Rent	Min_Rent	Max_Rent	Std_Dev
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1 Easy Fire	3800.	3827.	3310.	4436.	278.
2 Getty Fire	4067.	3284.	1980.	13308.	1766.
3 Saddle Ridge Fire	2563.	2553.	1538.	4138.	487.

Predictive Analysis