# Predicting the Price of a House for Affordable Housing Policy Implementation by the Government.

—

By Smith Kwabena Agyei (ITC, Data Scientist)

# The Housing Affordability Crisis

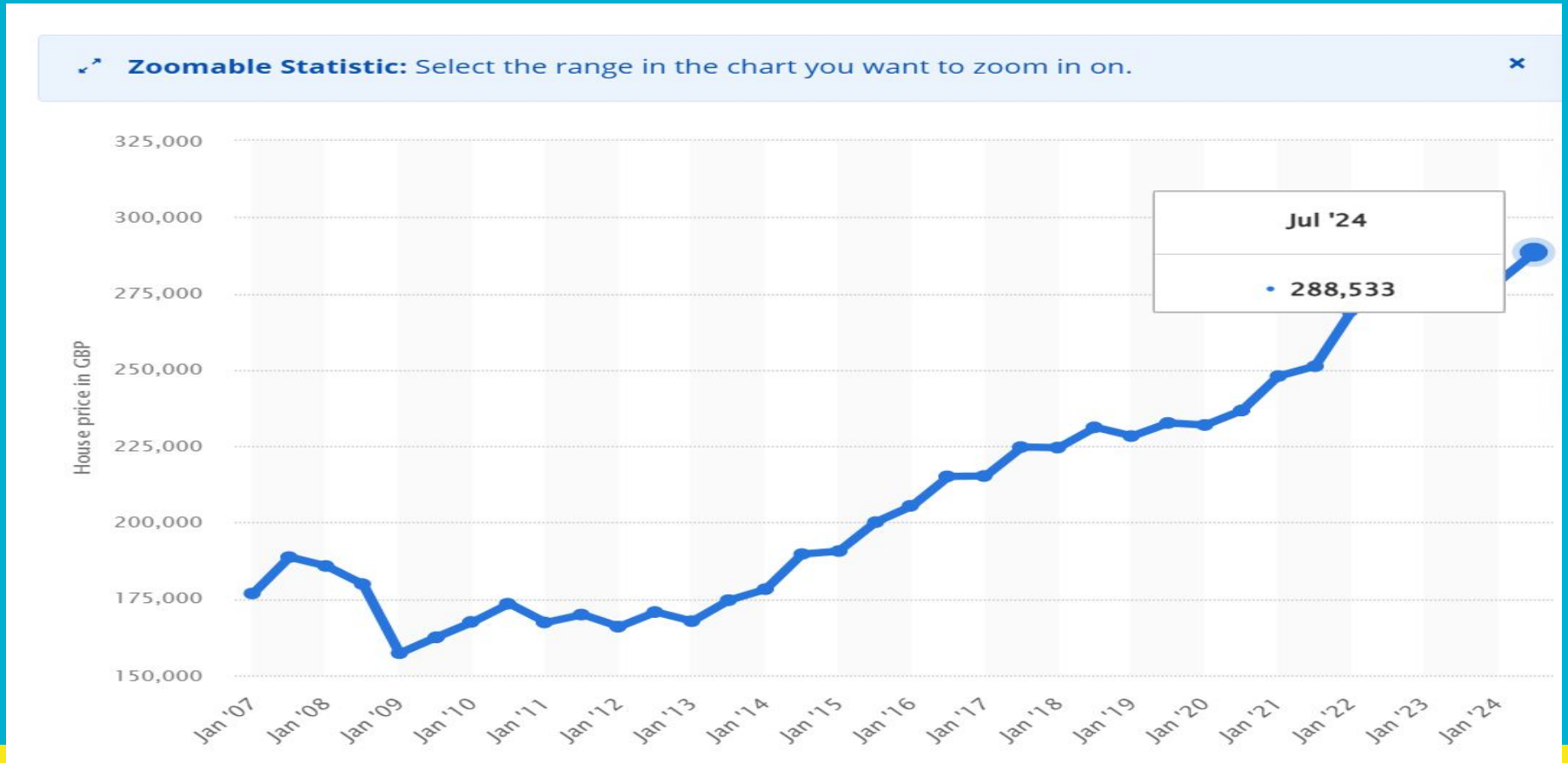
Problem Statement: Rising housing costs are outpacing incomes, leading to affordability challenges.

The housing affordability crisis is driven by rising home prices and limited affordable space, especially in urban and high-growth areas. Housing costs have increased faster than incomes, making homeownership difficult for middle- and low-income families. High prices force people to rent longer, commute farther, and miss out on wealth-building opportunities through homeownership. This crisis contributes to growing inequality and requires policy changes, more affordable housing options, and creative financing solutions to address.

# Current Housing Prices: 2024



Statistics: National average home price increase over the past decade.



# The Housing Affordability Crisis



Goal: Explore data-driven insights to understand number of rooms, the square feet and lot acres covered and predict a Sold Price for the house

Predicting the sold price helps policymakers understand key factors driving housing costs, such as location, property size, and amenities. By identifying these drivers, they can create targeted policies, like promoting compact housing or offering incentives in high-cost areas, to enhance housing affordability. This data-driven approach provides actionable insights to make housing more accessible.

**Tools - Python Libraries**

- Pandas
- Numpy
- MatplotLib
- Seaborn

# Overview / Structure of Dataset in Use



| MLS | Sold_Price | Zipcode | Longitude | Latitude | Lot_Acres | Taxes | Year_Built | Bedrooms | Bathrooms | Sqrt_ft | Garage | Kitchen_Features | Fireplaces | Floor_Covering | HOA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 21530491 | 5300000.0 | 85637 | -110.378200 | 31.356362 | 2154.0 | 5272.00 | 1941 | 13 | 10.0 | 10500.0 | 0.0 | Dishwasher, Freezer, Refrigerator, Oven | 6 | Mexican Tile, Wood | 0.0 |
| 21529082 | 4200000.0 | 85646 | -111.045371 | 31.594213 | 1707.0 | 10422.36 | 1997 | 2 | 2.0 | 7300.0 | 0.0 | Dishwasher, Garbage Disposal | 5 | Natural Stone, Other | 0.0 |

**MLS (Multiple Listing Service)**
**Sold_Price**: The final sale price of the property, indicating the amount for which the property was actually sold.
**Zipcode**: A postal code of the property's location, useful for determining local market trends, amenities, and schools.
**Longitude & Latitude**: Geographic coordinates providing the exact location of the property.
**Lot_Acres**: The size of the property lot in acres, which can affect the property's value and appeal, especially in rural areas.
**Taxes:** Annual property taxes, usually based on the assessed value of the home and local tax rates.

**Year_Built**: The year the property was originally constructed, helpful in assessing the property's age and potential need for updates or renovations.
**Bedrooms**: Number of bedrooms, which is a critical factor in determining the home's suitability for buyers based on family size or space needs.
**Bathrooms**: The number of bathrooms in the property is another key metric for homebuyers when assessing convenience and comfort.
**Sqrt_ft**: The total square footage of the property, representing the usable floor area and influencing the home's market value.
**Garage**: Information about garage availability, size, or type, which adds value and convenience.
**Kitchen_Features**: Special features of the kitchen, such as modern appliances, countertops, or layout styles (e.g., open or galley kitchen), are important in appealing to buyers.
**Fireplaces**: number of fireplaces, adding value, especially in regions where heating is essential.
**Floor_Covering**: The type of flooring (e.g., hardwood, carpet, tile) throughout the home, which can impact both aesthetics and maintenance needs.
**HOA (Homeowners Association)**: Indicates if the property is part of an HOA, including dues and fees, which affect the cost of living in the home. It may also reflect amenities or neighborhood regulations.

# Dataset after Categorization based on price/square feet

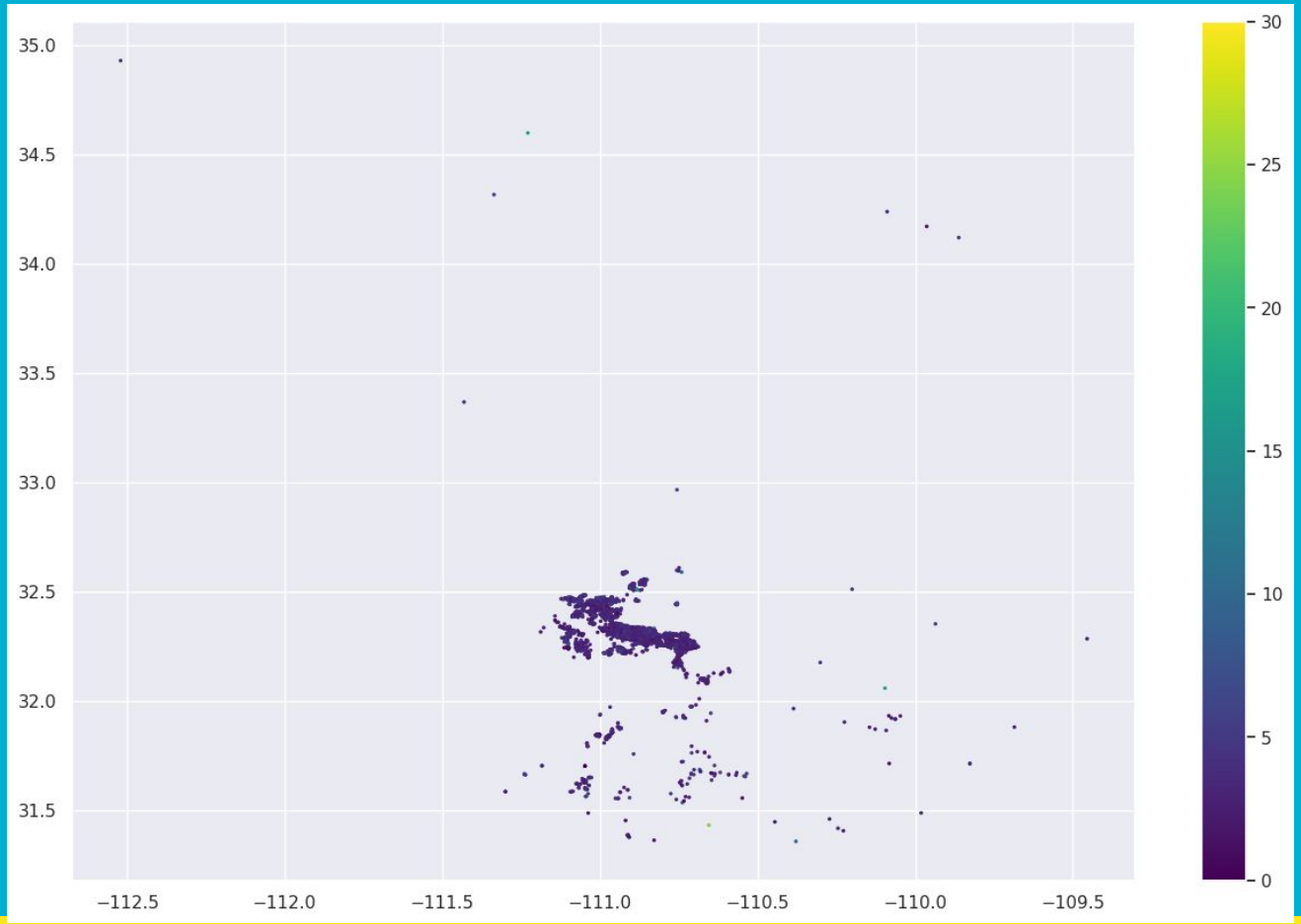


| | MLS | Sold_Price | Zipcode | Longitude | Latitude | Lot_Acres | Taxes | Year_Built | Bedrooms | ... | Sqrt_ft | Garage | Kitchen_Features | Fireplaces | Floor_Covering | HOA | Price_Per_Sqrft | Price_Cat |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 21530491 | 5300000.0 | 85637 | -110.378200 | 31.356362 | 2154.0 | 5272.00 | 1941 | 13 | ... | 10500.0 | 0.0 | DISHWASHER, FREEZER, REFRIGERATOR, OVEN | 6 | MEXICAN TILE, WOOD | 0.0 | 504.761905 | 10 |
| 21529082 | 4200000.0 | 85646 | -111.045371 | 31.594213 | 1707.0 | 10422.36 | 1997 | 2 | ... | 7300.0 | 0.0 | DISHWASHER, GARBAGE DISPOSAL | 5 | NATURAL STONE, OTHER | 0.0 | 575.342466 | 11 |
| 3054672 | 4200000.0 | 85646 | -111.040707 | 31.594844 | 1707.0 | 10482.00 | 1997 | 2 | ... | 2767.0 | 0.0 | DISHWASHER, GARBAGE DISPOSAL, REFRIGERATOR | 5 | NATURAL STONE, OTHER: ROCK | 0.0 | 1517.889411 | 30 |

**Price per square feet is categorized into bins: ⇒ Price_Per_Sqrft // 50**

# K-Nearest Neighbors (KNN) for Classification






Scatter Plot showing the KNN Classification of Location and Price Categories

KNN Classification Accuracy ⇒ 97%

# Dataset after Categorization based on price/square feet Square Feet



| | Sold_Price | Sqrt_ft | Lot_Acres | Rooms | Knn_Pr |
|---|---|---|---|---|---|
| 0 | 5300000.0 | 10500.0 | 2154.00 | 23.0 | 10 |
| 1 | 4200000.0 | 7300.0 | 1707.00 | 4.0 | 11 |
| 2 | 4200000.0 | 2767.0 | 1707.00 | 5.0 | 30 |
| 3 | 4500000.0 | 9019.0 | 636.67 | 12.0 | 9 |
| 4 | 3411450.0 | 6396.0 | 3.21 | 10.0 | 10 |

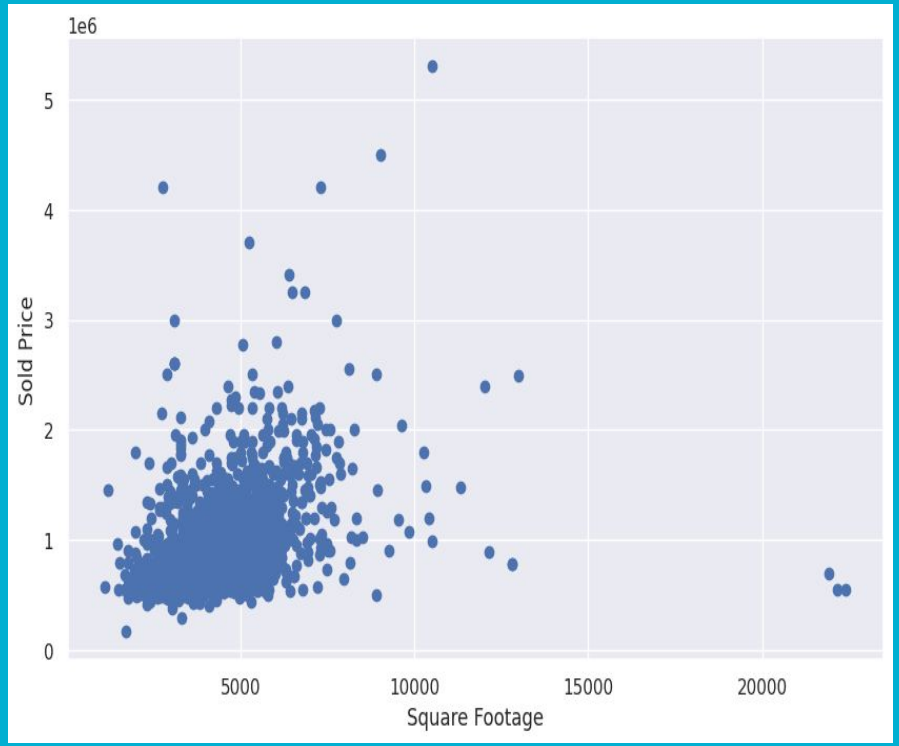
1. Feature engineering performed to add bathrooms and bedrooms together to create a feature called Rooms.
   Dataframe is having 4990 rows x 5 columns

# Plots for insights on the Data

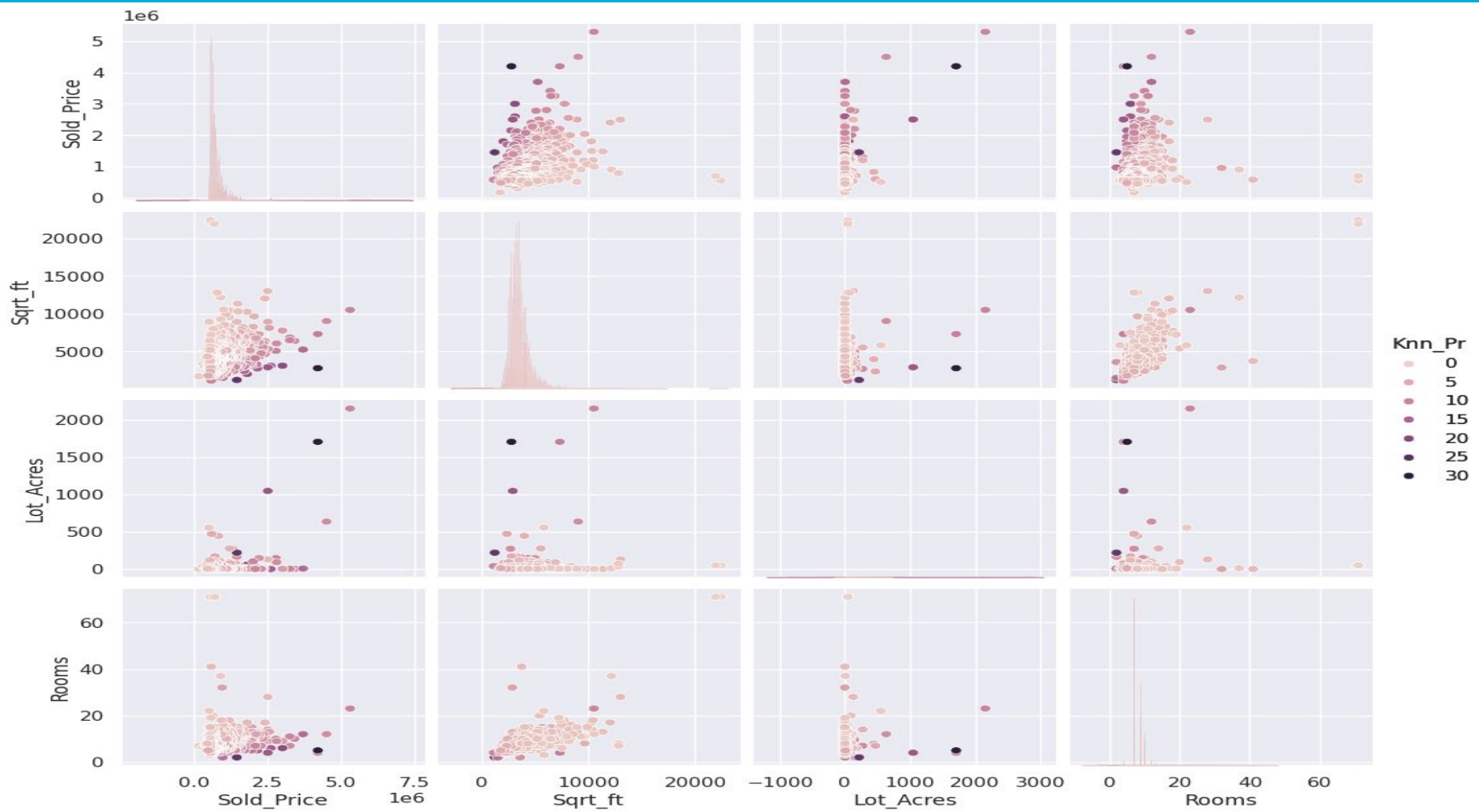


## Heatmap.


## Square Footage vs. Sold Price to show the impact of home size on



# Plots for insights Continued



# 3 things I learned

## 1 High-density zoning changes

Larger homes with more rooms are more expensive; reducing the number of these features in new affordable developments may help lower prices.
(Sq Ft vs. Sold Price Plot)

## 2 Tax incentives

Offer incentives to developers who build affordable housing in high-demand areas.

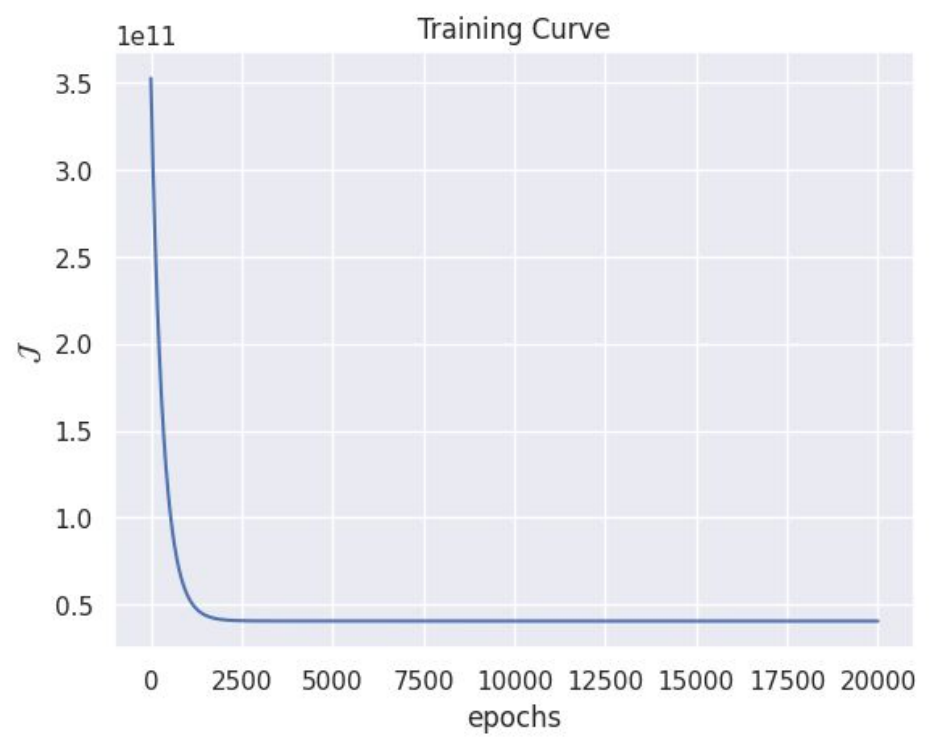
## 3 Home Features

Encourage compact home designs with fewer high-cost features to improve affordability.

# Model implementation (OLS Multivariate (mv) Linear Regression Class)



## Training Curve




## Sample Prediction Results

```
# Assign Test values (Square Feet,Lot acres, Rooms, Price Cat)
Xtest = np.array([[4900,2.4,4,3],[3160,5.6,4,20],[4317,5.6,4,20]])

# Predict Square foot
my_reg.predict(Xtest)
```

### Results of prediction

```
array([992858.56001153, 640386.22156363, 874814.60209192])
```

### Mean Absolute Error (MAE)

```
1159.6646445628526
```

# Impact of the proposed Solution



## 1 Job Creation and Economic Growth

Larger homes with more rooms are more expensive; reducing the number of these features in new affordable developments may help lower prices.
(Sq Ft vs. Sold Price Plot)

## 2 Financial Stability

Market Volatility: As demonstrated during the 2008 financial crisis, housing bubbles, in which values quickly rise before plummeting, have the potential to destabilize an economy. Recessions and banking crises can result from an over-reliance on housing debt.
Household Debt: A significant portion of household debt is frequently related to housing. High debt levels might make people less able to spend, which puts the economy at risk.

## 3 Summary

In summary, housing impacts the economy by driving employment, consumer spending, public revenue, and financial stability. It also influences inflation, interest rates, and social stability, making it a critical area for policymakers to consider.

# Thank you

Together, let us eradicate homelessness