**Interpretation and Recommendations**

**5.1: Compare Ridge and Elastic Net Performance**

| **Model** | **RMSE** | **R² Score** |
| --- | --- | --- |
| **Ridge** | 32.69 | 0.7794 |
| **Elastic Net** | **32.49** | **0.7820** |

📌 **Key Takeaways:**

* **Elastic Net slightly outperforms Ridge** with **lower RMSE (32.49)** and **higher R² (0.7820)**.
* **Both models perform well**, capturing **~78% of variance** in the test set.
* **Feature selection with Elastic Net helped**, making it **more interpretable**.

**5.2: Trade-offs Between Ridge and Elastic Net**

✅ **Ridge Regression**

* Keeps **all features** but **shrinks** their coefficients.
* Useful when **all features are important**, preventing overfitting.

✅ **Elastic Net Regression**

* **Selects key features** while shrinking others.
* Works best when **some features are irrelevant or highly correlated**.
* Helps with **automatic feature selection**, improving interpretability.

📌 **Elastic Net is the better choice** because it **balances feature selection and regularization**.

**5.3: Key Findings from Feature Importance**

The **Top 10 Features** influencing **PricePerSq** were:

* 🔴 **Most important features had large absolute coefficients** (from Elastic Net).
* **Larger homes (Size in sqft)** and **higher property tax** were key factors.
* **Demographic & economic factors (e.g., median income, crime rate) also played a role**.

📌 **Real-World Insight:**

* **Buyers should focus on neighborhoods with strong economic indicators (high median income, low crime).**
* **Homes with high property taxes often have a higher price per square foot.**
* **Lot size, number of bedrooms, and year built were less impactful than expected.**

**5.4: Final Recommendations**

🔹 **For Investors & Buyers:**

* **Look for areas with high median income and low crime rates** → Strong indicators of property value growth.
* **Prioritize homes with optimal price per square foot** → Avoid overpaying for larger lot sizes that may not justify a high price.

🔹 **For Further Model Improvements:**

* **Try additional nonlinear models (e.g., Gradient Boosting, Random Forest).**
* **Test additional features (e.g., school district ratings, proximity to amenities).**
* **Perform deeper feature engineering** (e.g., interaction terms between variables).

**✅ Final Step: Wrap-Up**

You've successfully completed:

* 📊 **Data Preparation & Scaling**
* 🔎 **Nested Cross-Validation for Ridge & Elastic Net**
* 📈 **Hyperparameter Tuning & Model Evaluation**
* 🎨 **Improved Visualizations & Feature Analysis**
* 🏡 **Real-World Interpretation for Housing Prices**

**🎯 Final Decision:**

**Elastic Net is the best model for this dataset, balancing predictive power and interpretability.** 🚀

**Predicting House Prices per Square Foot Using Ridge and Elastic Net Regression**

**1. Introduction**

The objective of this analysis was to develop a predictive model for **housing prices per square foot (PricePerSq)** using **Ridge Regression** and **Elastic Net Regression**. The dataset consists of various factors, including property characteristics, economic indicators, and demographic attributes. A **nested cross-validation approach** was implemented to ensure robust model selection.

**2. Data Preparation**

**2.1 Dataset Overview**

The dataset includes **600 records** with **30 features**, covering property details (size, tax, bedrooms, etc.) and economic/demographic indicators (median income, crime rate, commute time, etc.). The target variable for prediction is **PricePerSq**.

**2.2 Preprocessing Steps**

* **Feature Scaling:** Standardization (Z-score normalization) was applied to numeric features to ensure uniformity.
* **Multicollinearity Check:** Highly correlated features were identified using a correlation matrix.
* **Train-Test Split:** The dataset was divided into **80% training (480 samples)** and **20% testing (120 samples)**.

**3. Model Development**

**3.1 Nested Cross-Validation**

A **nested cross-validation strategy** was used:

* **Outer loop (10-fold CV):** Estimates model performance on unseen data.
* **Inner loop (5-fold GridSearchCV):** Tunes hyperparameters for Ridge and Elastic Net models.

**3.2 Hyperparameter Tuning**

* **Ridge Regression:** Tuned **alpha (regularization strength)**.
* **Elastic Net Regression:** Tuned **alpha** and **l1\_ratio** (mix of L1 & L2 regularization).
* **Final Best Hyperparameters:**
  + **Ridge:** Best alpha = **10**
  + **Elastic Net:** Best alpha = **0.1**, l1\_ratio = **0.5**

**4. Model Evaluation**

Both models were trained on the full **80% training data** and tested on **20% test data**.

| **Model** | **MSE** | **RMSE** | **R² Score** |
| --- | --- | --- | --- |
| **Ridge** | 1068.54 | 32.69 | 0.7794 |
| **Elastic Net** | 1055.70 | **32.49** | **0.7820** |

**5. Model Interpretation & Insights**

**5.1 Feature Importance (Elastic Net)**

The **top 10 most influential features** on housing price per square foot were:

1. **Size(sqft)**
2. **Property Tax**
3. **Median Income**
4. **Crime Rate**
5. **Last Sold Price**
6. **Lot Size**
7. **Commute Time**
8. **% College Educated**
9. **Year to Election**
10. **High School Quality**

**5.2 Visual Analysis**

* **Actual vs. Predicted Scatter Plot:** Showed good alignment between predicted and actual prices.
* **Residual Plot:** Errors were symmetrically distributed, indicating no major biases.
* **Feature Importance:** Elastic Net effectively selected relevant features while shrinking others.

**6. Final Recommendations**

**6.1 Model Selection**

* **Elastic Net performed slightly better** (lower RMSE, higher R²) than Ridge, making it the preferred model.
* **Elastic Net also helps in feature selection**, making it more interpretable for real estate predictions.

**6.2 Business Insights**

* **Homes in areas with high median income and low crime rates** tend to have higher PricePerSq.
* **Larger homes with high property taxes** often have higher values.
* **Commute time and local amenities** play a role in house pricing trends.

**7. Future Work**

To improve predictions further, we recommend:

* **Exploring Non-Linear Models** (e.g., Random Forest, Gradient Boosting).
* **Incorporating External Factors** (e.g., school district ratings, proximity to public transport).
* **Feature Engineering** (e.g., interaction terms between economic and property variables).

**8. Conclusion**

This study successfully applied **Ridge and Elastic Net regression** to predict housing prices per square foot. **Elastic Net emerged as the best model** due to its balance between predictive power and interpretability. These insights can guide investors and home buyers in making data-driven real estate decisions.

## Housing Price Prediction Using Ridge and Elastic Net Regression

**1. Introduction**

The goal of this project was to analyze housing prices in Jacksonville, FL, by using **Ridge Regression** and **Elastic Net Regression**. The target variable was **PricePerSq**, which represents the price per square foot of a house. The dataset included various features such as **size, property tax, median income, crime rate, and other economic and demographic indicators**.

To ensure robust predictions, we applied **nested cross-validation**, hyperparameter tuning, and feature scaling. The final models were evaluated based on **Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R² Score**.

**2. Data Preparation**

**2.1 Data Cleaning & Preprocessing**

* The dataset was loaded from C:\Users\asraf\Downloads\Project MAA\Data set 2022-2023.xlsx.
* Non-numeric values (such as $ in the "Last Sold Price" column) were cleaned.
* Features were scaled using **StandardScaler** to improve model performance.

**2.2 Feature Engineering**

* Highly correlated features were identified to avoid multicollinearity.
* **Top correlated pairs included**:
  + **Year Built vs. Age Sold**
  + **Last Sold Price vs. Size (sqft)**
  + **Median Income vs. Crime Rate**

**2.3 Train-Test Split**

* The dataset was split into **80% training (480 samples) and 20% testing (120 samples).**

**3. Model Selection and Training**

**3.1 Nested Cross-Validation**

* **Outer Loop:** **10-Fold Cross-Validation** was used for model evaluation.
* **Inner Loop:** **5-Fold GridSearchCV** optimized hyperparameters for **Ridge (alpha) and Elastic Net (alpha & l1\_ratio).**

**3.2 Hyperparameter Tuning**

* **Best Ridge Alpha:** **10**
* **Best Elastic Net Alpha:** **0.1**, **L1 Ratio:** **0.5**

**4. Model Performance Evaluation**

**4.1 Final Test Performance**

| **Model** | **RMSE** | **R² Score** |
| --- | --- | --- |
| **Ridge Regression** | 32.69 | 0.7794 |
| **Elastic Net Regression** | **32.49** | **0.7820** |

🔹 **Elastic Net slightly outperformed Ridge**, showing lower RMSE and a higher R² score.  
🔹 **Both models explain ~78% of the variance** in housing price per square foot.

**4.2 Visualizations**

* **Actual vs. Predicted Prices:** Elastic Net had slightly better alignment with actual values.
* **Residual Plot:** Errors were evenly distributed, confirming no major bias.
* **Feature Importance:** The most influential features were:
  + **Size (sqft)**
  + **Property Tax**
  + **Median Income**
  + **Crime Rate**
  + **Lot Size**

**5. Key Findings and Recommendations**

**5.1 Model Interpretation**

✅ **Ridge Regression** retains all features but shrinks coefficients to reduce overfitting.  
✅ **Elastic Net** selects **important features**, balancing Ridge (L2) and Lasso (L1) penalties.  
✅ **Elastic Net is the preferred model** as it provides interpretability while keeping strong performance.

**5.2 Business Insights**

🏡 **For Real Estate Investors & Buyers**:

* Focus on **areas with high median income and low crime rates** for better price appreciation.
* Property tax is a strong indicator of **price per square foot**—higher taxes often mean **higher-valued properties**.
* Lot size and number of bedrooms were **less influential** than expected.

🔍 **For Further Improvements**:

* Test **non-linear models** like Gradient Boosting or Random Forest.
* Add external data (e.g., **school ratings, proximity to shopping centers**).
* Perform **deeper feature engineering** (e.g., **interaction effects** between variables).

**6. Conclusion**

This project successfully **developed a predictive model for housing prices per square foot**, using **regularized regression techniques**. Elastic Net proved to be the best model, offering both high accuracy and interpretability. Future work could involve incorporating additional data sources and testing more complex models.

🚀 **Final Decision:**  
**Elastic Net is the best model for predicting housing prices per square foot in Jacksonville, FL.**

**Appendices**

📂 **Dataset Path:** C:\Users\asraf\Downloads\Project MAA\Data set 2022-2023.xlsx  
🛠 **Libraries Used:** pandas, sklearn, matplotlib, seaborn, numpy  
👤 **Author:**