# California Housing Price Prediction 🏠

## Project Overview

This project applies the **OSEMN framework** (Obtain, Scrub, Explore, Model, Interpret) to predict housing prices in California using 1990 census data.

I compared a baseline **Linear Regression** model against a **Random Forest Regressor** to demonstrate the power of ensemble learning on complex, non-linear datasets.

## Key Results

| **Model** | **R2 Score** | **RMSE (Average Error)** |
| --- | --- | --- |
| **Linear Regression** | 0.625 | ~$70,060 |
| **Random Forest** | **0.816** | **~$49,038** |

**Conclusion:** The Random Forest model outperformed the linear baseline by **19%**, reducing the prediction error by over **$21,000** per house.

## Technologies Used

* **Pandas:** Data cleaning (Imputing missing values with median, One-Hot Encoding categorical features).
* **Seaborn:** Correlation Heatmaps to identify key price drivers (e.g., Median Income).
* **Scikit-Learn:** \* LinearRegression (Baseline)
  + RandomForestRegressor (Final Model)
  + train\_test\_split, r2\_score, mean\_squared\_error

## How to Run

1. Clone the repository.
2. Install dependencies: pip install pandas scikit-learn seaborn matplotlib
3. Run the notebook California\_Housing\_Prediction.ipynb.