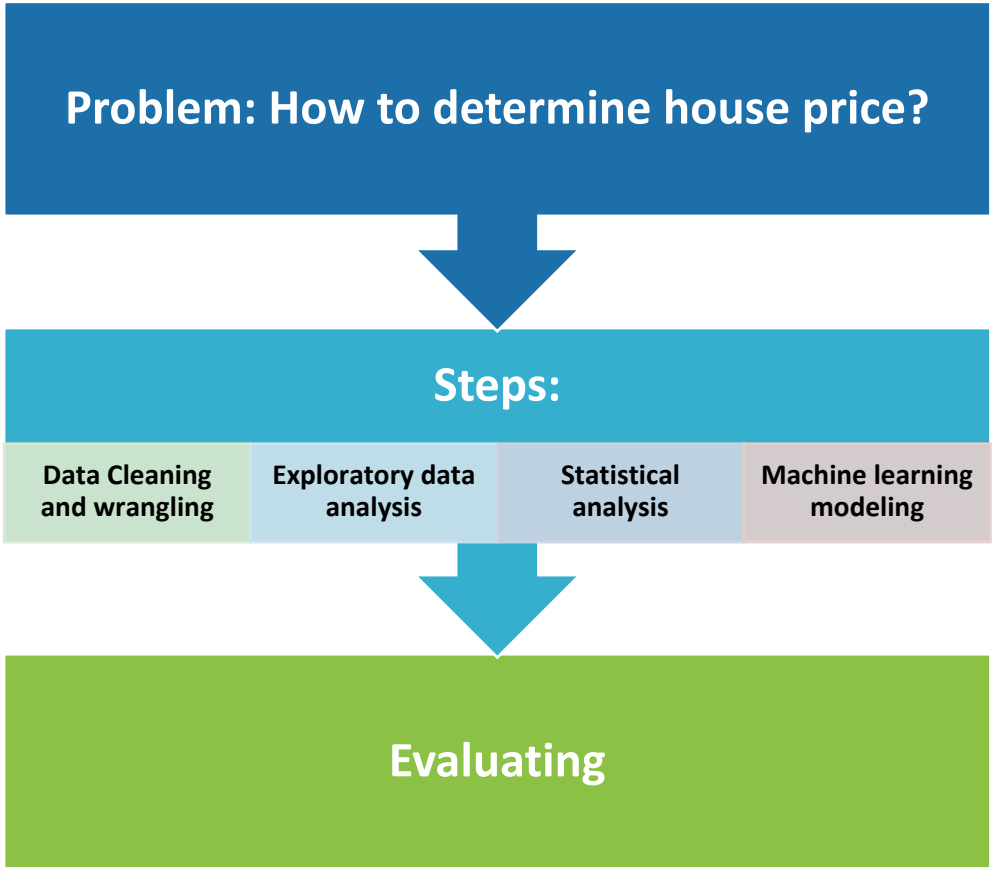
House Price Prediction

NEIL SONALKAR

Overview


Problem: How to determine house price?
Steps:
Data Cleaning and wrangling
Exploratory data analysis
Statistical analysis
Machine learning modeling
Evaluating

House Price Prediction is essential for Realtors, Real Estate Investors and Homeowners

Solutions

Data Wrangling	EDA	Statistical Analysis
• Process Missing Data • Preprocess categorical data to dummy table • Normalize data • Perform Outlier Treatment	• Create Visualizations • Check for correlated features • Get data structured for analysis and modeling	• Hypothesis Testing • Analyze Trends

Solutions (Cont.)

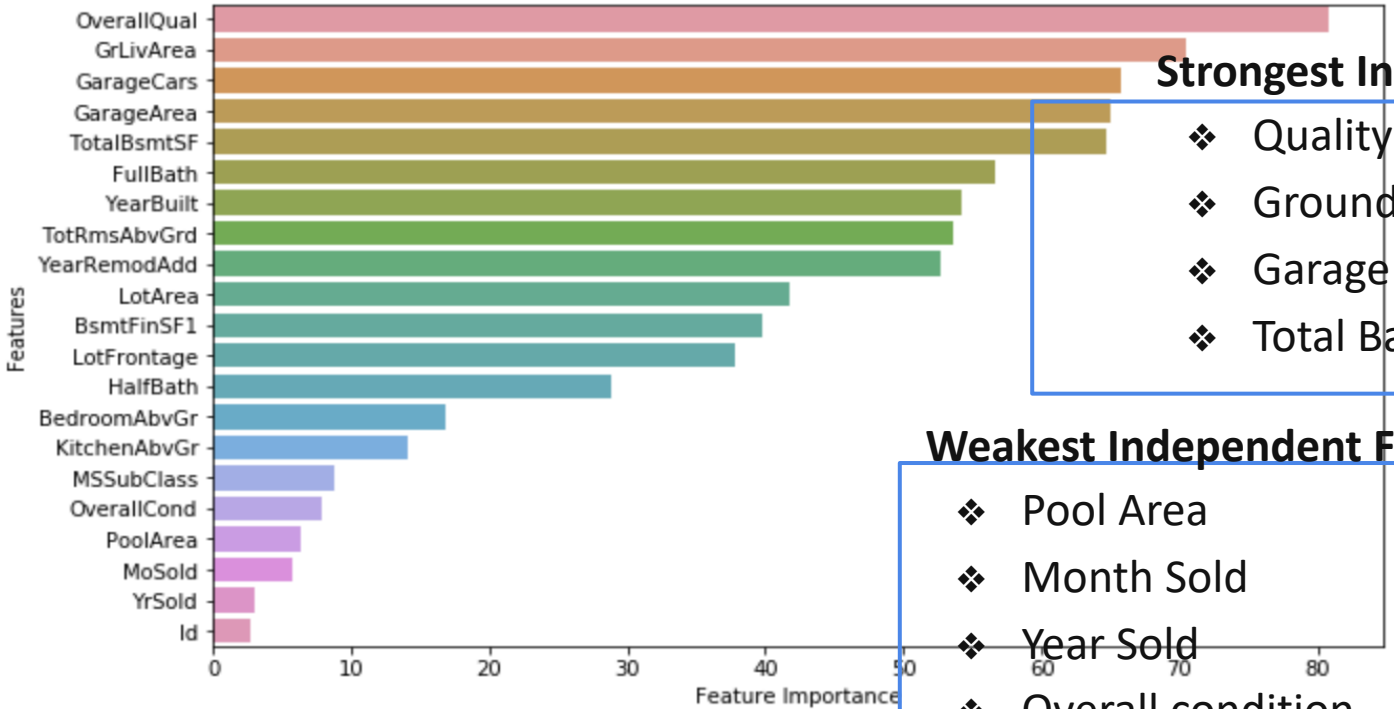
Machine Learning	Evaluation	Data Analysis
• Run Regression Algorithms on Data • Cross Validate • Hyperparameter Tuning • Predict Prices	• R-Squared Adjusted • RMSE • Compare against all algorithms	• Recommend Algorithm that performed the best (Explain Process) • Recommend what to look for when buying a house (Important Features)

Results

COMPARISON OF DIFFERENT REGRESSION MODELS:

Regression Algorithms	R-Squared Adjusted
Linear Regression	.789
SVM	.849
Random Forest	.874
Gradient Boosting	.884

- **Feature importance hierarchy:**



Strongest Independent Features

- ❖ Quality
- ❖ Ground Living Area
- ❖ Garage Area
- ❖ Total Basement Sqft

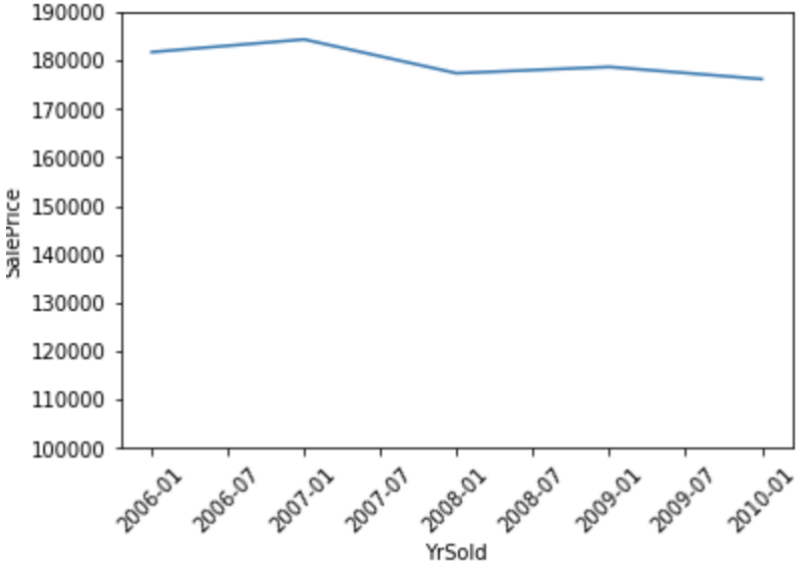
Weakest Independent Features

- ❖ Pool Area
- ❖ Month Sold
- ❖ Year Sold
- ❖ Overall condition

Price Check

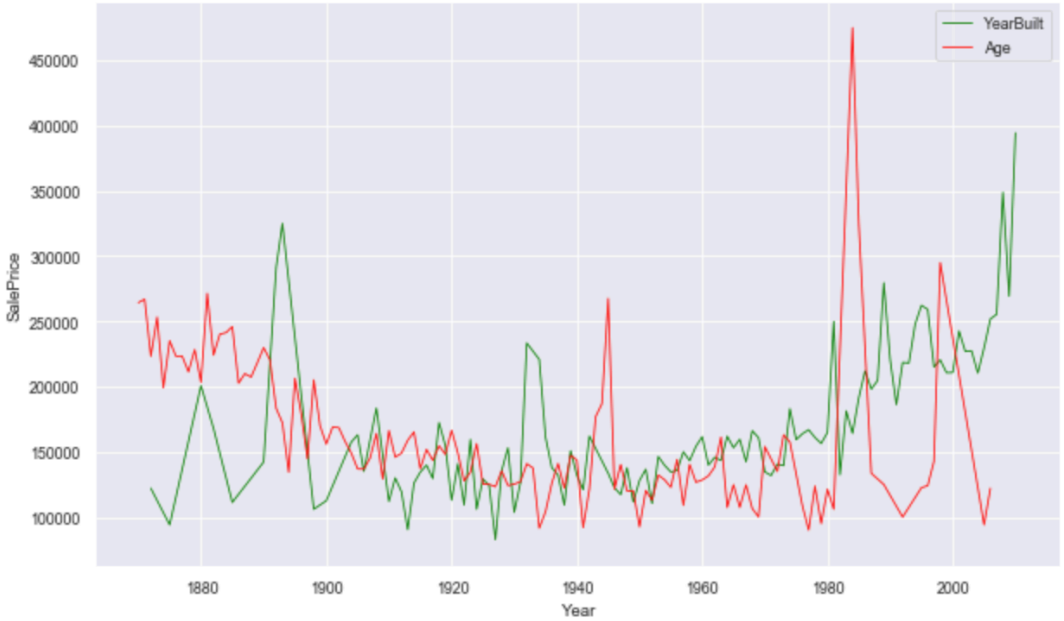
We see a slight decrease in the mean house price of houses sold in 2010

	YrSold	SalePrice
0	2006-01-01	181726.668790
1	2007-01-01	184335.872340
2	2008-01-01	177360.838816
3	2009-01-01	178661.482249
4	2010-01-01	176150.777143



Time series view:

- Age of house and Sale Price correlation over time.
- We see in green the sales price of each house that was built in that year. We see in red, the age of the house before being sold.
- We see a negative correlation along time between the two.



Solutions:

- **Real Estate Investor:**
  - **If a real estate investor was looking to sell one of their properties, avoid selling in 2010. Buyers Market. Low risk properties**
  - **Important features:**
    - **Area of: Basement, Living, Garage**
    - **Overall Quality**
- **Real Estate Agents:**
  - **Know how to determine a price point based on the important features of the house.**
- **Home Buyers:**
  - **Level of transparency on house costs when buying and selling homes.**

Recommendations and Future Work :

- **Primary Recommendations:**
  - **Find external features that might correlate with House Price. Ex) Crime rate, school district, population rate, etc.**
  - **Find more instances of Townhomes and Condos. Majority of data was single family homes.**
- **Future Work:**
  - **Create a program that allows clients to enter the important features of a house to determine the house price.**
  - **Find more recent data.**