



REAL ESTATE PRICE PREDICTION AND ANALYSIS

Bolatbekov Adilet

Muratbek Darmen





IMPORTANT FOR

For people working in real estate industry

For investors

For regular people planning on buying or selling house

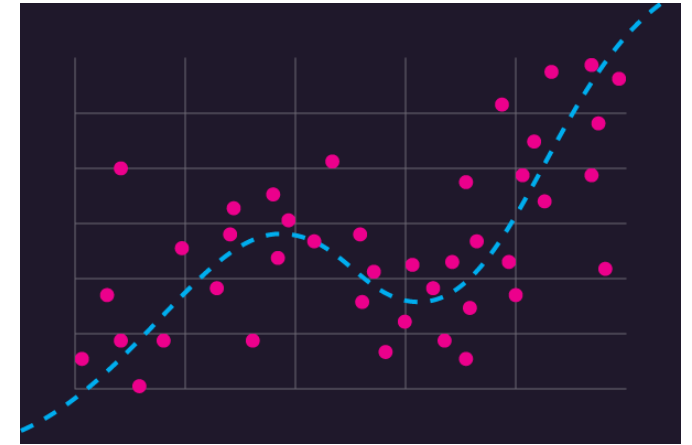
Real Estate price prediction and analysis is a crucial field of research

DATA ANALYTICS AND REAL ESTATE

- Housing costs are a critical area of research for economists and financiers
- Data-driven methodologies for estimating property prices have grown in popularity
- I used the Boston Houses dataset to explore variables that affect a house's price
- Goal was to create regression models to accurately calculate a house's worth

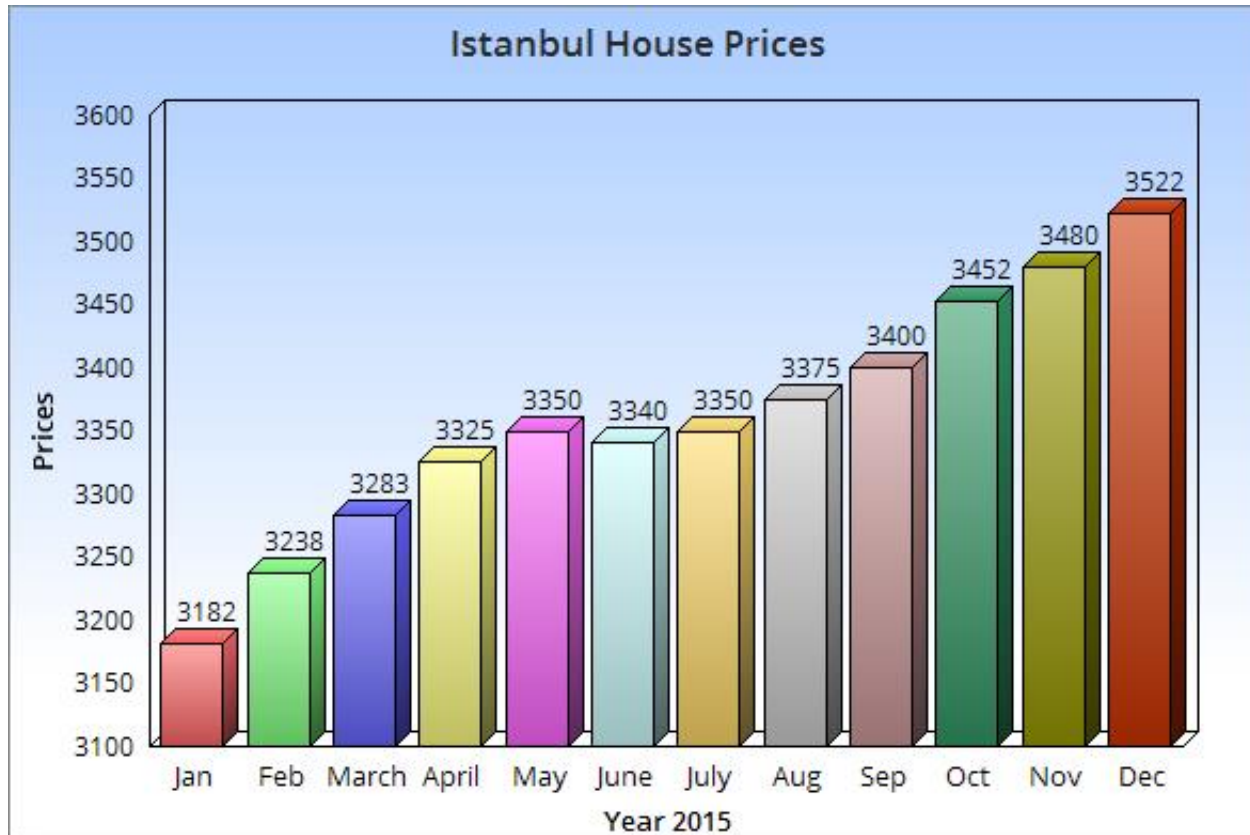


ANALYZING HOUSING PRICES WITH REGRESSION MODELS



- By analyzing the Boston Houses dataset, we can identify factors that impact housing prices
- Using regression models, we can estimate a house's worth based on those variables
- This information can be useful for buyers, sellers, and investors looking to make informed decisions about housing.

RELATED WORK: REAL ESTATE PRICE PREDICTION STUDIES



- Real estate price prediction is a topic of interest for scholars
- Previous studies have used regression analysis, machine learning, neural networks, and other approaches
- "Real Estate Market Price Prediction Model of Istanbul" by Tekin and Sari used regression analysis and machine learning to anticipate Istanbul's real estate market prices
- "House Price Prediction: Hedonic Price Model vs. Artificial Neural Network" by Limsombunchai compared the efficacy of the hedonic pricing model with the artificial neural network
- "Real Estate Prediction" by Dabreo et al. investigated multiple methods for predicting real estate prices, including regression analysis, machine learning, and neural networks
- Another study by Dabreo et al., titled "Real Estate Price Prediction," anticipated real estate values using the Random Forest method

DATA

- The Boston Housing Dataset is a well-known and often used dataset in the fields of machine learning and predictive analytics.
- It offers a complete collection of variables and observations to study and forecast housing prices in the Boston region.
- The dataset has 511 observations of 13 distinct variables, including crime rate, land zoning, nitric oxide concentration, proximity to roads, and median value of owner-occupied residences.



DATA CONTENT

The dataset has 13 features:
CRIM, ZN, INDUS, CHAS, NOX, RM, AGE, DIS, RAD, TAX, PTRATIO, LSTAT, and MEDV.

CRIM: per capita crime rate by town

ZN: proportion of residential land zoned for lots over 25,000 sq.ft.

INDUS: proportion of non-retail business acres per town

CHAS: Charles River dummy variable (1 if tract bounds river; 0 otherwise)

NOX: nitric oxides concentration (parts per 10 million)

RM: average number of rooms per dwelling

AGE: age and proportion of owner-occupied units built

DIS: weighted distances to five Boston employment centres

RAD: index of accessibility to radial highways

TAX: full-value property-tax rate per \$10,000

PTRATIO: pupil-teacher ratio by town

LSTAT: % lower status of the population

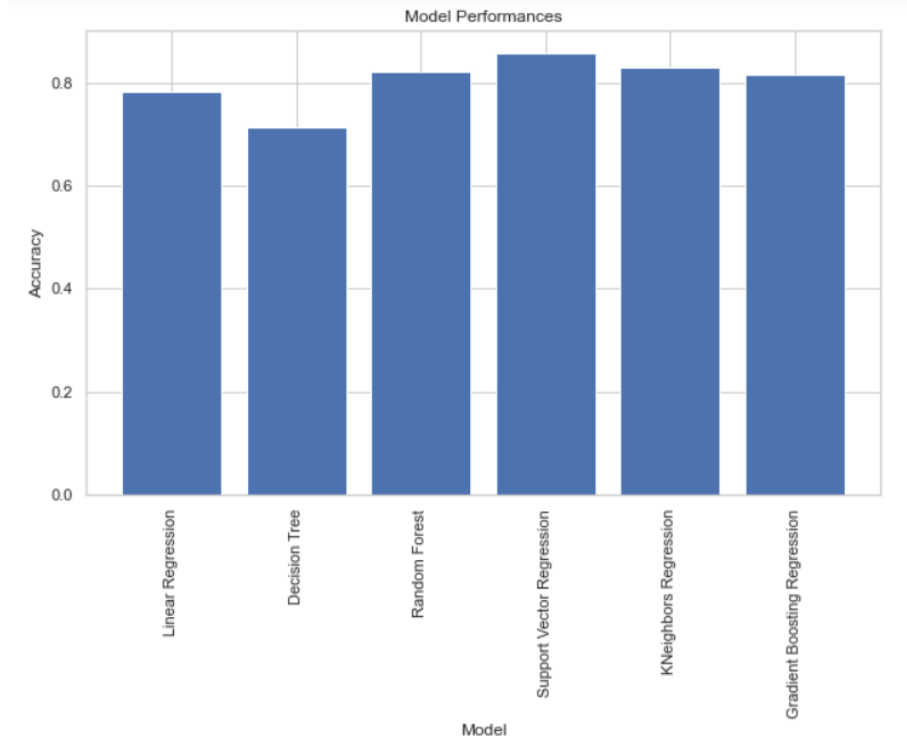
MEDV: Median value of owner-occupied homes in \$1000's

ABOUT THE DATASET

- The dataset was initially made available by Harrison, D., and Rubinfeld, D. L. as a benchmark dataset for regression issues in machine learning.
- It was gathered by the U.S. Census Service.
- The dataset has both quantitative and categorical variables, therefore pretreatment or filtering was not required for this project. However, data preprocessing was used to normalize the data and eliminate any outliers in order to increase the model's accuracy.
- Only four important features—RM, LSTAT, PRATIO, and MEDV—were chosen for investigation in this research.



Gradient Boosting Regression testing Accuracy: 0.8165560551586224
Gradient Boosting Regression MAE: 0.11, MSE: 0.02, R2: 0.82



METHODS

The background of the slide features a light teal gradient. At the top, two hands in dark grey suits with white cuffs are shaking. Below the handshake is a stylized illustration of a two-story yellow house with a brown roof, a chimney, and several windows. In the foreground, to the left of the house, is a wooden signpost with a sign that says 'SOLD' in large, yellow, block letters. There are also some small, stylized trees in brown and yellow colors at the bottom of the scene.

CONCLUSION

- Objective: To pinpoint the most significant elements that have an effect on the median price of owner-occupied homes in Boston
- Pre-processing: Selected important attributes and removed certain data points to prevent inaccurate findings
- Frameworks: numpy, pandas, matplotlib, seaborn, and scikit-learn
- Methodologies: Exploratory data analysis (EDA), statistical analysis, and predictive modeling
- Checks: Missing values, dataset distribution, and visualization to find patterns, trends, and outliers
- Findings: Factors that significantly influence the median price of owner-occupied residences in Boston include the number of rooms, percentage of lower socioeconomic population, student-teacher ratio, and weighted distances to job areas
- Machine learning model: Created a trustworthy model using scikit-learn to precisely forecast median prices
- Implications: Our work offers insightful information and a solid framework for future research, examinations, and prediction models aimed at enhancing Boston's housing accessibility and affordability.