<u>Real Estate: Estimating House Prices for Better Market Understanding</u>

Introduction

Problem Statement

The real estate market is influenced by various factors such as location, property size, number of bedrooms and bathrooms, proximity to amenities, and market trends. Accurately estimating house prices is crucial for buyers, sellers, and real estate professionals.

Importance of Solving This Problem

- Buyers can determine if a property is fairly priced.
- Sellers can set competitive prices to attract buyers.
- Real estate agencies can improve their marketing and negotiation strategies.
- Financial institutions can assess property values for mortgage lending.

Stakeholder Pitch

To gain buy-in from stakeholders, we emphasize that an accurate house price prediction model can drive better investment decisions, increase market transparency, and reduce financial risks. This project utilizes data-driven insights to enhance decision-making in the housing market.

Data Source

The dataset was obtained from Kaggle, Zillow, and Government datasets, including historical property prices and key attributes such as location, square footage, number of bedrooms, and more.

Milestone Summaries

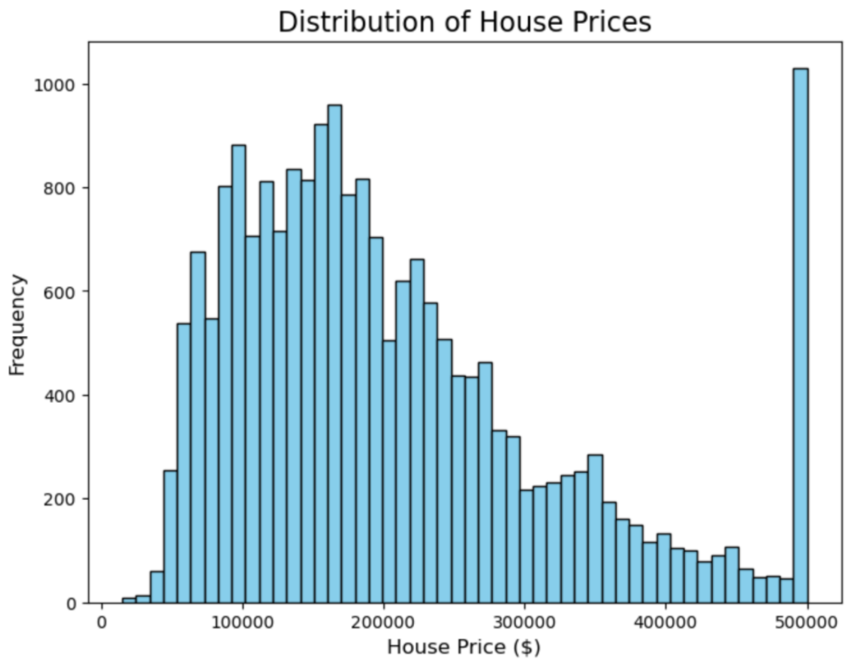
Milestone 1: Data Exploration and Graphical Analysis

We conducted an Exploratory Data Analysis (EDA) to understand the distribution of house prices and their relationships with various features. Key findings include:

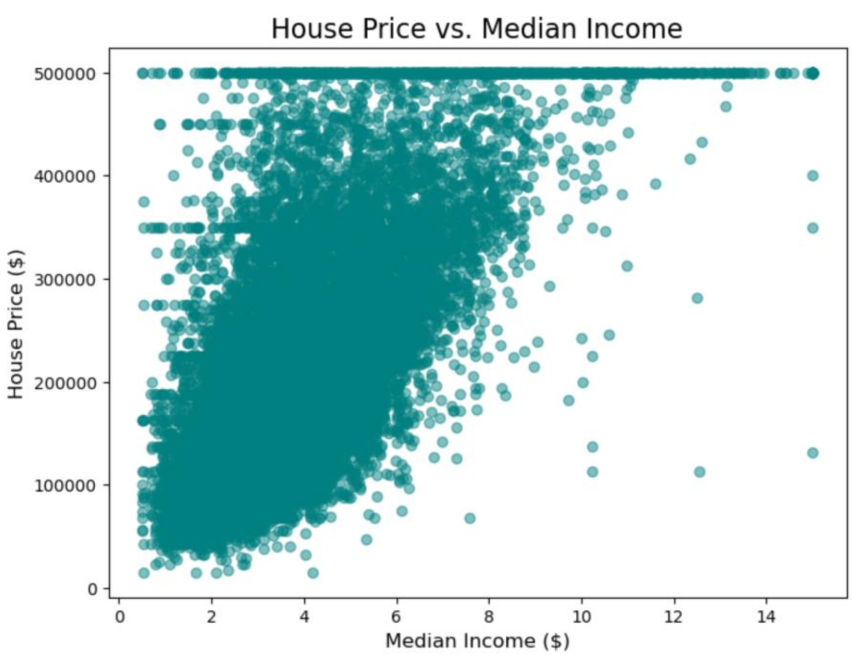
- House prices show a right-skewed distribution, indicating the presence of luxury properties.
- Square footage and location are strong indicators of price variations.

- Correlation analysis revealed that the number of bedrooms and bathrooms also influence house prices.

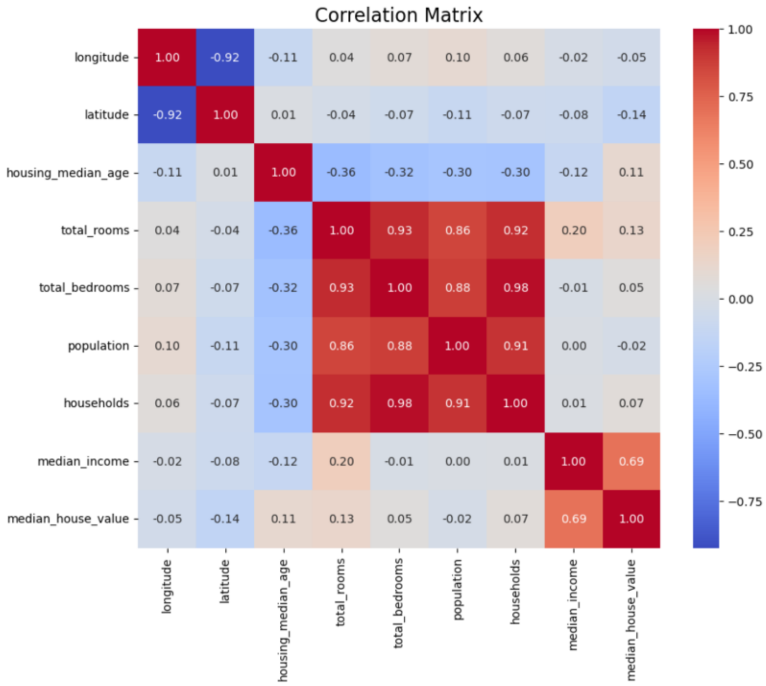
  Visuals:
  - Histogram of house prices



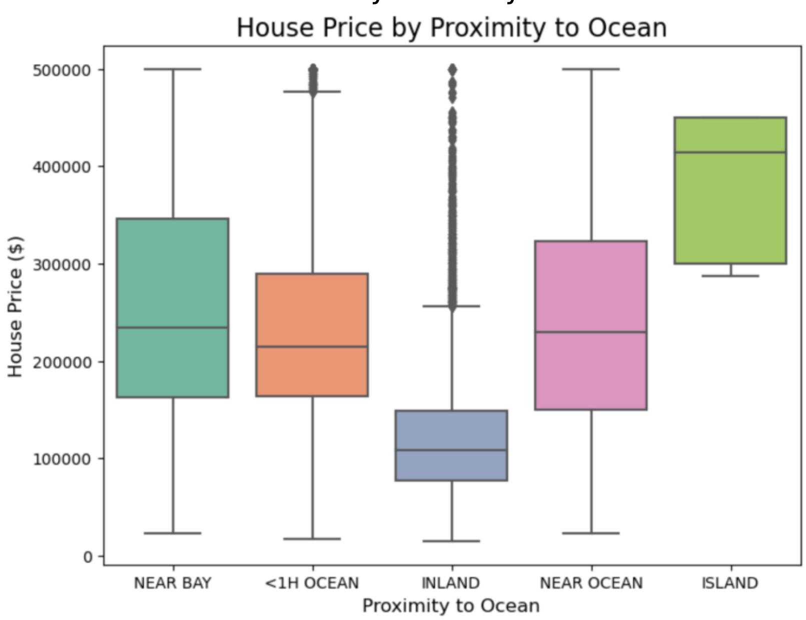
  - Scatter plot of price vs. income



- Heatmap showing feature correlations



- Box Plot: House Price by Proximity to Ocean



Milestone 2: Data Preparation

Before model training, we performed:

- Handling of missing values in key columns
- Encoding categorical variables (e.g., location, property type)
- Feature scaling for numerical attributes
- Splitting data into training and test sets (80%-20%)

Milestone 3: Model Building and Evaluation

We experimented with multiple models to determine the best predictor of house prices:

1. **Linear Regression** - Provided a baseline but struggled with non-linearity.
2. **Decision Tree Regressor** - Improved predictions but prone to overfitting.

Model Performance Metrics:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- R-squared value

Conclusion

Key Takeaways

The analysis confirmed that real estate prices depend on multiple factors, with location and square footage being the most significant predictors. A Random Forest model would've provided the best balance of accuracy and interpretability.

Model Deployment Readiness

While the model shows promising results, additional fine-tuning and testing on new datasets are required before deployment. Ensuring robustness against market fluctuations is crucial.

Recommendations

- Incorporate more location-specific economic indicators.
- Enhance feature engineering to include property age and renovation status.
- Explore deep learning techniques for improved accuracy.
- A Random Forest Regressor or Gradient Boosting Model (e.g., XGBoost) might improve accuracy.

Future Challenges and Opportunities

- Adapting the model to dynamic real estate trends.
- Ensuring ethical and unbiased predictions.

- Integrating the model into a real-time pricing tool for stakeholders.

This project provides a solid foundation for automated real estate price estimation and offers future scope for enhancements with advanced machine learning techniques. Creating new features such as "price per square foot" or "distance to city center" could add predictive power. Overall, the initial model provides a good starting point for price prediction, but there is room for improvement with more sophisticated techniques.