Real Estate Price Prediction: Summary Report

Objective

This project aims to develop a predictive model for estimating real estate prices based on key factors like median income, property tax, school quality, and hospital availability. The goal is to help buyers, sellers, and investors make data-driven decisions.

Data & Methodology

Dataset: Collected real estate price data with features such as location-based socio-economic indicators.

Exploratory Data Analysis (EDA):

- Identified correlations between property price, income levels, and nearby facilities.
- Visualized patterns using histograms, scatter plots, and heatmaps.

Feature Engineering:

- Created interaction terms like school and hospital availability scores.
- Removed outliers and missing values for better model accuracy.

Modeling Approaches:

- Linear Regression: Provided a baseline model with an R^2 of approximately 0.74.
- Random Forest: Improved performance but had a higher MAE and RMSE.
- Final Model: Selected an optimized multiple regression model with:
  - MAE: 21,451.29
  - RMSE: 24,906.57

Key Insights

- School quality and hospital availability are the strongest predictors of home prices.
- Areas with high property tax generally have lower affordability but better amenities.
- The model can predict home prices within approximately 5-10% of actual values.

Conclusion & Future Improvements

- The model is useful for real estate pricing, policy-making, and investment analysis.

- Future improvements could include location-based variables (latitude/longitude) and time-based price trends.