August 11th, 2023

House Price Prediction and Analysis in King County

MEET OUR TEAM

NITIN CHANDRA

ARGHYA KHANRA

PARTHIBAN R

SANAA

SANSKRUTEE

INTRODUCTION

Purpose:-

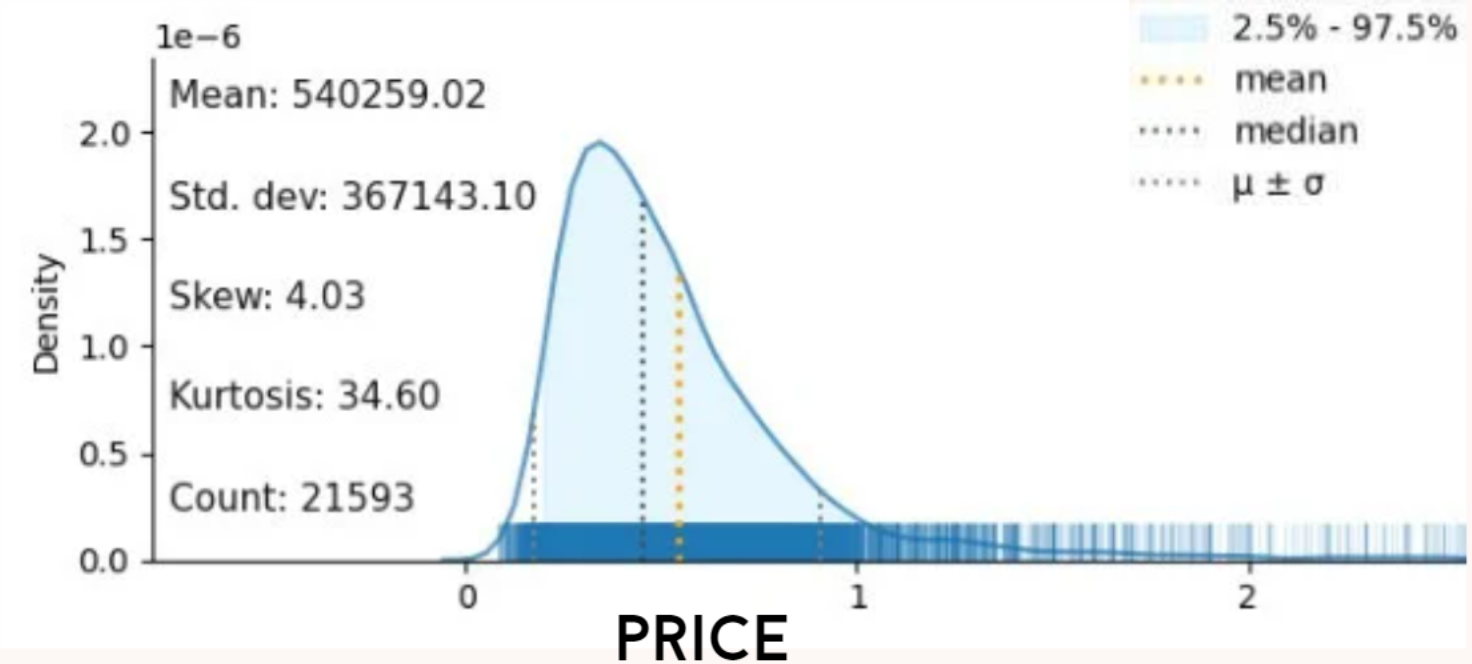
- The purpose of the analysis was to **PREDICTING HOUSE PRICES** via different models and the **Variables** included in the dataset.

Scope:-

- The scope of this project is to build a regression model to predict the price of houses in King County, Washington. The dataset used for this project contains house sale prices for King County between May 2014 and May 2015.
- To build a regression model that can accurately predict the price of houses in King County.
- To identify the most important features that affect the price of houses in King County.
- To compare the performance of different regression models.

DATA PREPROCESSING

- The DATA given didn't require a lot of cleaning only some columns like **"ID" & "DATE"** were cleaned to extract the relevant information required
- **Zip code** was taken for one hot as it was a categorical column
- **Univariant Analysis** was done to visualize relationships between numerical variables.



MODEL WE PERFORMED & USAGE

Linear Regression: Suitable for basic house price estimates based on key features like square footage and bedrooms.

Decision Tree Regression: Captures complex feature interactions, useful for understanding neighborhood-specific influences on house prices.

Random Forest Regression: Balances accuracy and stability by combining decision trees, delivering reliable house price predictions.

Support Vector Regression (SVR): Excels at capturing intricate relationships and non-linear patterns for precise house price predictions

ACCURACY SCORES OF EACH MODEL

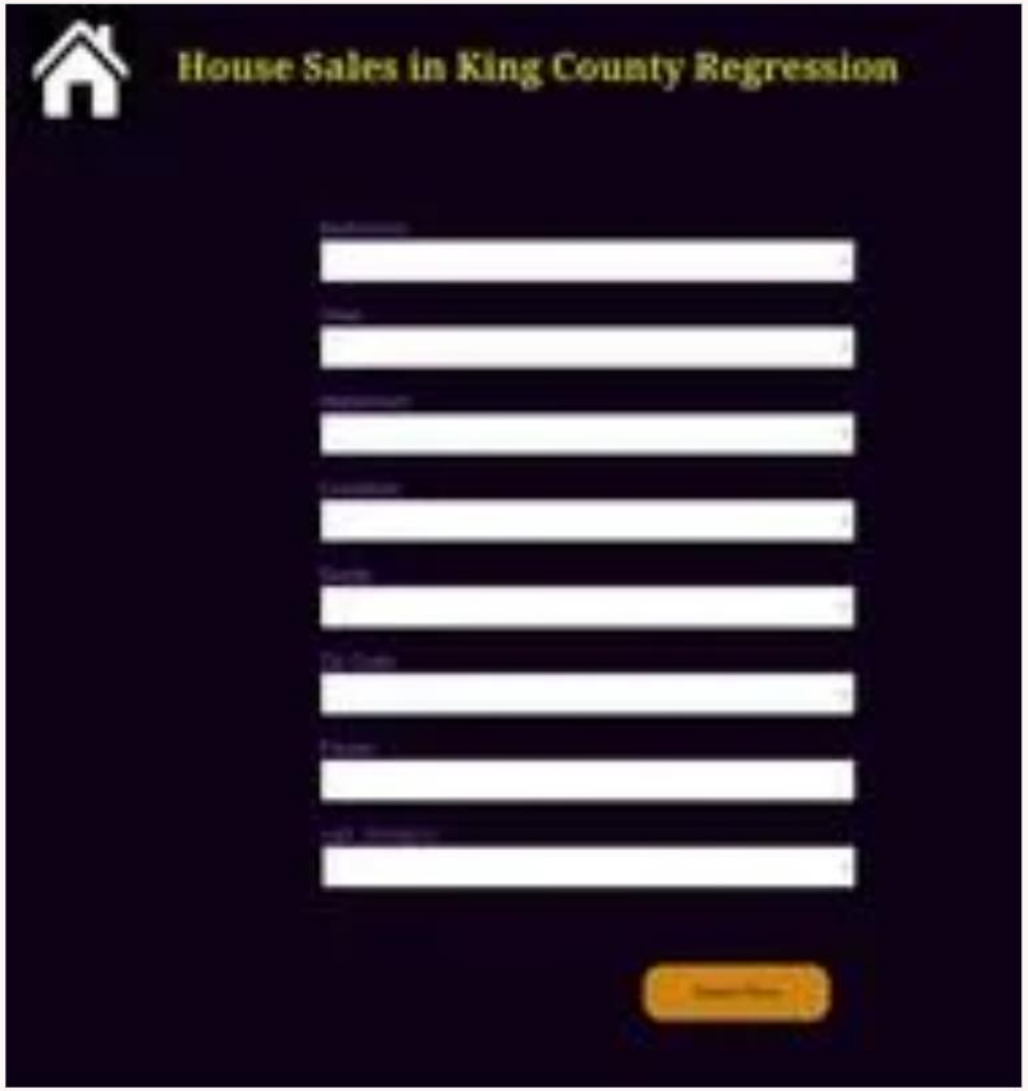
Random Forest: "Outperforms with **lowest MSE (19.04B)** and **highest R2 (0.84)**, excelling in house price prediction."

Linear Regression: "Respectable **MSE (23.70B)** and **R2 (0.81)** offer reasonable estimates based on simpler features."

Decision Tree: "Moderate performance **MSE (39.72B)** and **R2(0.67)** capturing some complexity, but lags behind Random Forest's accuracy."

SVR: "Struggles with complex patterns, displaying **high MSE (128.30B)** and negative **R2 (-0.05),** needing improvement."

WEBSITE OVERVIEW & FEATURES



https://house-price-predictions.netlify.app

CONCLUSION

Our study of house price prediction models highlights **'Random Forest'** as the **standout performer with the lowest Mean Squared Error (19.04 billion)** and **highest R-squared (0.84)**, indicating its superior accuracy.
'Linear Regression' offers solid estimates **(MSE: 23.70 billion, R-squared: 0.81)** based on basic features.
'**Decision Tree'** has moderate performance **(MSE: 39.72 billion, R-squared: 0.67)**, while 'SVR' struggles **(MSE: 128.30 billion, R-squared: -0.05)**, suggesting refinement. Overall,
'Random Forest' excels for robust house price predictions, while other models provide insights and potential for exploration.

THANKYOU