PROJECT PRESENTATION

NAME: SHWETANKPANDEY
ROLL NO: 2301420042
COURSE: BTECH CSE DS

PROJECT GOAL

- THE GOAL OF THIS PROJECT IS TO DEVELOP A MACHINE LEARNING MODEL THAT CAN ACCURATELY PREDICT HOUSING PRICES BASED ON KEY FEATURES SUCH AS SIZE, NUMBER OF ROOMS, LOCATION, AND YEAR BUILT.

- PREDICTING HOUSE PRICES HELPS BUYERS, SELLERS, AND REAL ESTATE INVESTORS MAKE INFORMED DECISIONS IN THE MARKET.

MIND MAP

COLLECTING DATASET

DATA CLEANING

EXPLORTORY DATA ANALYSIS

MODEL TRAINING AND TESTING

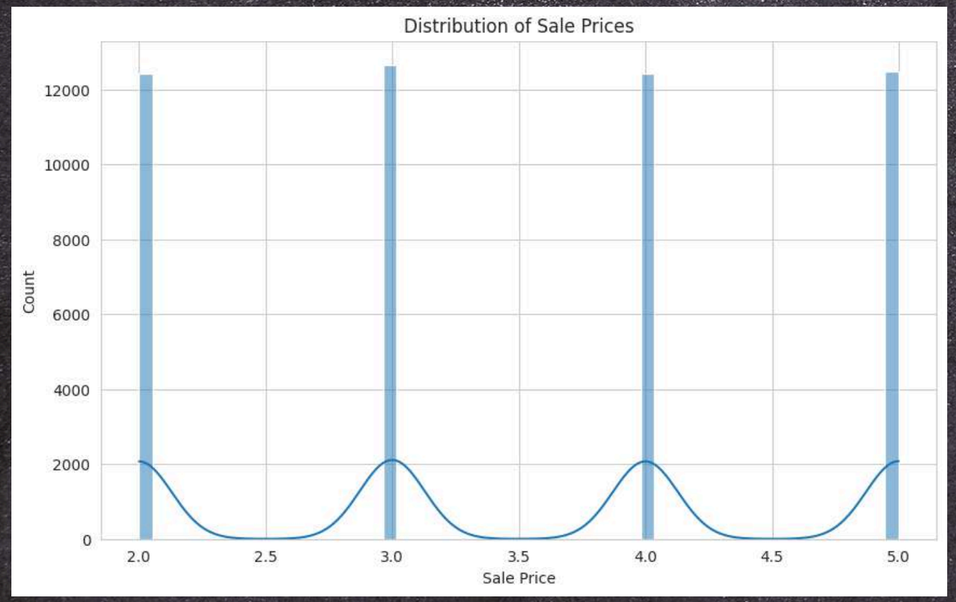
DATASET OVERVIEW

- SQUAREFEET: TOTAL AREA OF THE HOUSE
- BEDROOMS: NUMBER OF BEDROOMS
- BATHROOMS: NUMBER OF BATHROOMS
- NEIGHBORHOOD: LOCATION CATEGORY (URBAN, SUBURB, RURAL)
- YEARBUILT: YEAR THE HOUSE WAS BUILT
- RICE: FINAL SALE PRICE (TARGET VARIABLE)

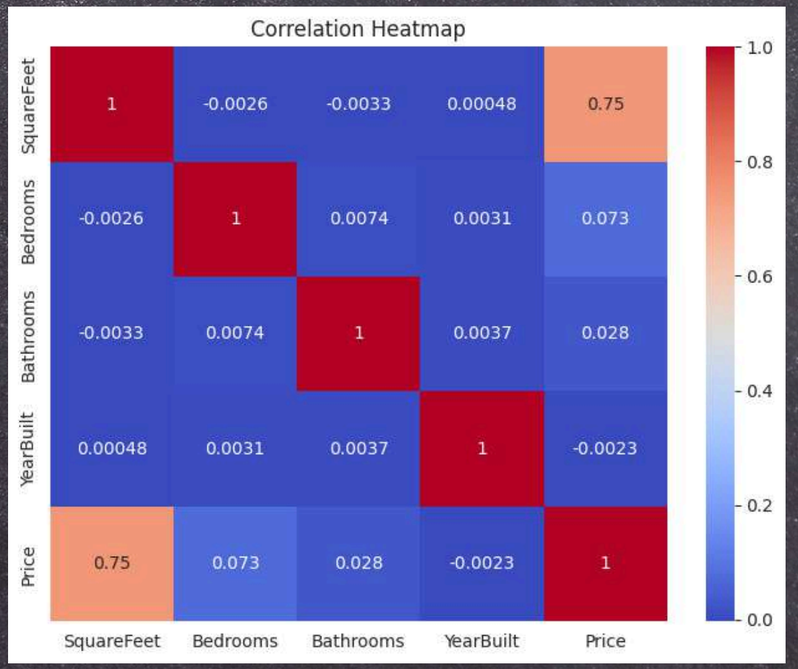
DATA CLEANING

- DROPPED COLUMNS WITH EXCESSIVE MISSING VALUES (>30%)
- FILLED MISSING NUMERICAL VALUES WITH MEDIAN
- FILLED MISSING CATEGORICAL VALUES WITH MODE
- REMOVED DUPLICATE ROWS
- RESET DATASET INDEX

EXPLORATORY DATA ANALYSIS

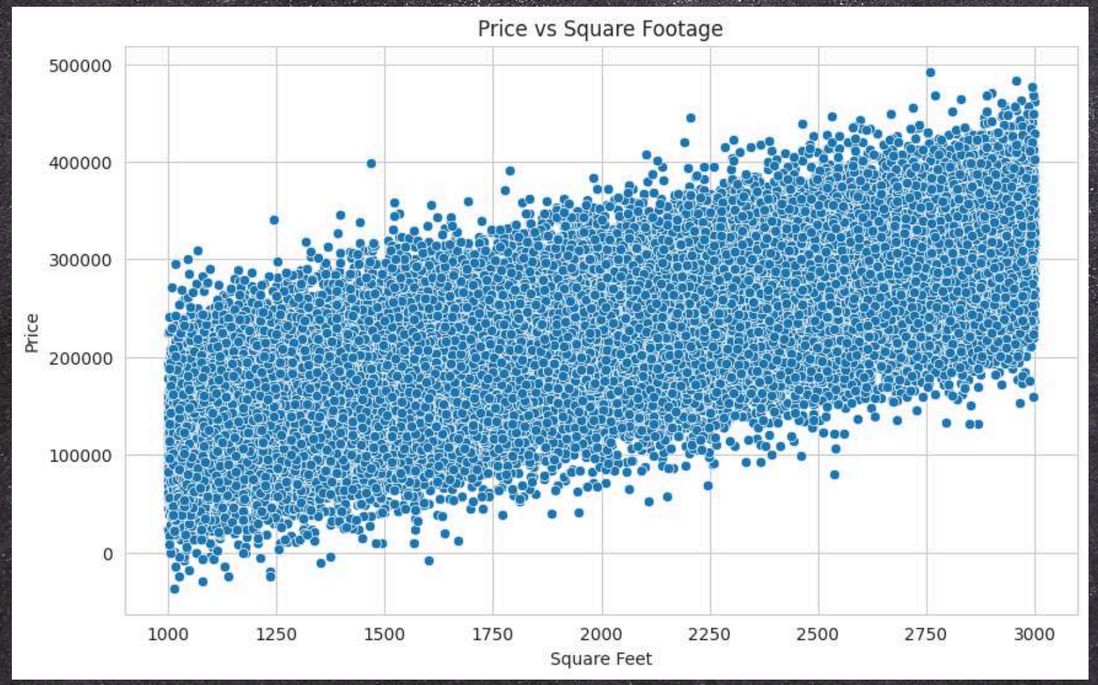


HIST PLOT



CORRELATION HEATMAP

EXPLORATORY DATA ANALYSIS

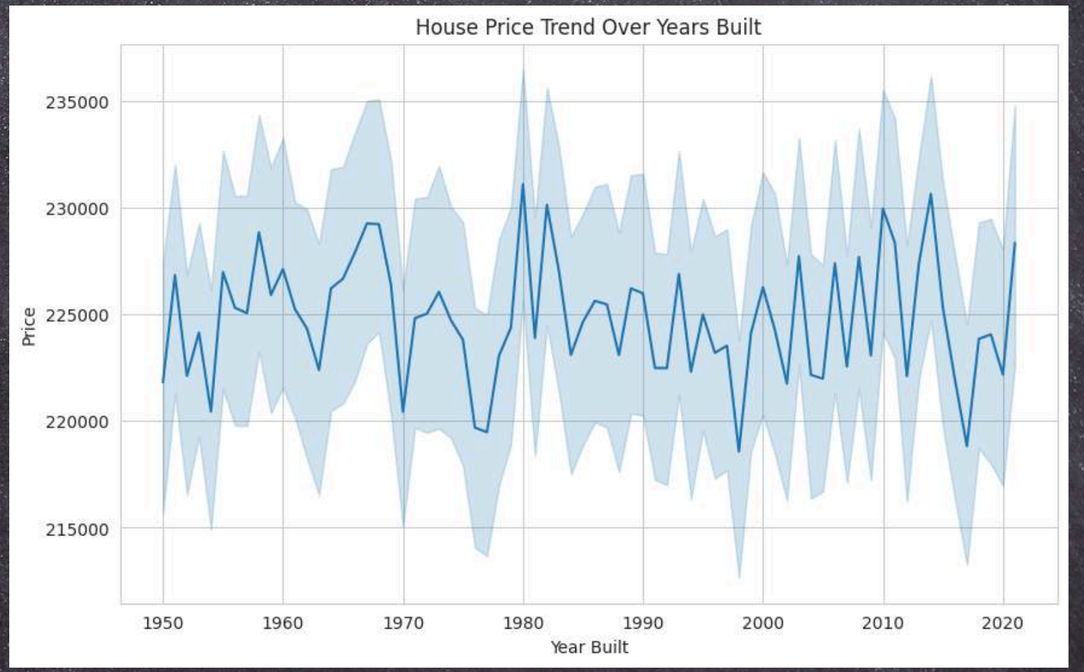


SCATTER PLOT

BOX PLOT

EXPLORATORY DATA ANALYSIS

BAR PLOT



LINE PLOT

MODEL BUILDING

- MODEL: RANDOM FOREST REGRESSOR
- PIPELINE: COMBINED PREPROCESSING AND MODELING INTO A SINGLE PIPELINE
- SPLIT: 80% TRAINING, 20% TESTING
- EVALUATION METRICS:

MAE: ~17,000 – 25,000
RMSE: ~24,000 – 32,000
R^2 SCORE: ~0.85 – 0.90

RESULT AND CONCLUSION

- THE MODEL DEMONSTRATED HIGH PREDICTIVE ACCURACY AND STABILITY ACROSS DIFFERENT HOUSING CONDITIONS AND LOCATIONS.
- SUBURBAN AREAS AND LARGER HOUSES WERE FOUND TO HAVE A CONSISTENTLY HIGHER PRICE. THE PIPELINE SETUP ENSURES REPRODUCIBILITY AND EASY DEPLOYMENT.

FUTURE IMPROVEMENTS

- DEPLOY AS AN INTERACTIVE WEB APP USING STREAMLIT (OPTIONAL)
- INCLUDE MORE ADVANCED FEATURES: PROXIMITY TO CITY CENTER, SCHOOLS, CRIME RATE
- USE MORE ADVANCED MODELS (XGBOOST, LIGHTGBM)
- PERFORM HYPERPARAMETER TUNING FOR EVEN BETTER PERFORMANCE

THANK YOU VERY MUCH!