**House Price Prediction - Report & Documentation**

**1. Data Preprocessing & Feature Engineering**

**Steps Taken:**

✔ **Handled Missing Values:** Imputed missing values using median/mode based on feature type.  
✔ **Feature Scaling:** Standardized numerical features using **MinMaxScaler** for consistency.  
✔ **Feature Encoding:** Converted categorical features (e.g., "Waterfront Present") into numerical format.  
✔ **Feature Engineering**

**2. Model Selection & Optimization**

**Chosen Models:**

🔹 **LinearRegressor** –Classical model.🔹 **DecisionTreeRegressor** – Simple and interpretable but prone to overfitting.  
🔹 **RandomForestRegressor** – More robust with better generalization.  
🔹 **XGBoost** – Used for fine-tuned optimization due to its performance on structured data.

**Hyperparameter Tuning:**

🔹 Used **GridSearchCV** to tune:

* n\_estimators (Number of trees)
* max\_depth (Tree depth)
* learning\_rate (for XGBoost)
* min\_samples\_split (Min data points per split)  
  🔹 Optimized for **R² score**.  
  🔹 Final Model: **RandomForestRegressor with optimized hyperparameters** (best trade-off between performance & interpretability).

**3. Deployment Strategy & API Usage**

**Backend - FastAPI Deployment:**

🔹 **FastAPI** used to expose the model via an HTTP API.  
🔹 **Model serialized with joblib** for efficient loading.  
🔹 Hosted on **Render.com** for public access.

**4. API Usage Guide**

****

🔹 **Response:** JSON object with predicted\_price.  
🔹 **Errors:** Returns "Expected 19 features, but got X" if input is incorrect.