House Price Prediction Project Report

1. Introduction
This project aims to predict house prices using Linear Regression.
It analyzes how different features such as area, bedrooms, bathrooms, location, and condition affect the price of a property.

2. Dataset Description
- Total records:2000 houses
- Total features10
- Main features: `Area`, `Bedrooms`, `Bathrooms`, `Floors`, `YearBuilt`, `Location`, `Condition`, `Garage`, and `Price`.
- No missing values were found in the dataset.

3. Data Preprocessing
- Converted text data (`Location`, `Condition`, `Garage`) into numeric form using one-hot encoding.
- Verified data types and ensured all numerical features were ready for training.
- Split the data into training (80%) and testing (20%) sets.

4. Exploratory Data Analysis (EDA)
- Area showed the highest positive correlation with Price.
- Houses with more bedroom an bathrooms are costlier.
- Urban and Downtown areas have higher average prices.
- Visualized data using scatter plots, box plots, and a correlation heatmap.

5. Model Building
- Used Linear Regression from scikit-learn.
- Trained the model on the training data and predicted house prices for the test set.
- Achieved a good R^2 Score, showing strong prediction accuracy.

6. Model Optimization
- Implemented:
- Polynomial Regression(for non-linear relationships)
- Polynomial Regression achieved slightly better performance.

7. Findings & Conclusions
- Area,Bedrooms and Bathrooms are key predictors of house price.
- Houses in better condition and prime locations are more expensive.
- Linear Regression is a simple but powerful model for this type of prediction.

8. Learning Outcomes

- Learned how to clean, analyze, and visualize data effectively.
- Understood how regression algorithms work in real-world prediction problems.
- Improved knowledge of model evaluation using metrics like MAE, MSE, and R^2.
- Gained experience in optimizing models using advanced regression techniques.

9. Future Work
- Include more features such as proximity to schools, public transport, or amenities.
- Try advanced models like Random Forest or Gradient Boosting for better accuracy.
- Build a web-based prediction app using the trained model.

10. Conclusion
- The project demonstrates how machine learning and data analysis   can help estimate house prices accurately.
- It highlights the importance of data quality, preprocessing, and model selection in achieving reliable predictions.