**🏠 House Price Prediction Project Report**

-by **KEERTHANA.R**

**🎯 Learning Objective**

To understand and apply regression modelling and feature engineering techniques to predict house prices using multiple linear regression.

**📌 1. Project Overview**

This project aims to predict house prices based on features such as area, location, and number of rooms. By leveraging multiple linear regression, we explore how different variables influence pricing and how preprocessing, feature selection, and model evaluation contribute to building a robust predictive model.

**🧰 2. Tools & Technologies**

* **Programming Language**: Python
* **Libraries**: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
* **Model**: Multiple Linear Regression

**🧪 3. Data Preprocessing**

Key preprocessing steps include:

* Handling missing values
* Encoding categorical variables (e.g., location)
* Feature scaling
* Outlier detection
* Multicollinearity check using VIF

**📊 4. Exploratory Data Analysis**

* Correlation matrix
* Distribution plots
* Scatter plots (e.g., Area vs Price)
* Boxplots for location-based price variation

**🔍 5. Feature Importance**

* Most influential features: Area, Location, Number of Rooms
* Visualized using regression coefficients and heatmaps

**📐 6. Model Evaluation**

* **RMSE**: Measures average prediction error
* **R² Score**: Indicates model fit
* Lower RMSE and higher R² suggest better performance

**⚠️ 7. Multicollinearity**

* Can distort coefficient estimates
* Addressed by removing or combining correlated features

**🔁 8. Polynomial Regression**

* Used to capture non-linear relationships
* May improve accuracy but risks overfitting

**📍 9. Location-Based Analysis**

* Location encoded and analyzed
* Significant impact on price due to demand and amenities

**🚨 10. Outlier Impact**

* Detected using boxplots and Z-scores
* Outliers can skew predictions and increase error

**📈 11. Assumptions of Linear Regression**

* Linearity
* Independence of errors
* Homoscedasticity
* Normal distribution of residuals
* No multicollinearity

**🚀 12. Model Deployment**

* Save model using joblib or pickle
* Deploy via Flask or Streamlit
* Optionally create a REST API with FastAPI

**✅ Conclusion**

This project demonstrates how multiple linear regression can be used to predict house prices effectively. With proper preprocessing, feature selection, and evaluation, the model provides valuable insights into the housing market. Future improvements could include ensemble models, deeper location analysis, and real-time deployment.

***THANK YOU!***