Project Title	**Real Estate Investment Advisor: Predicting Property Profitability & Future Value**
Skills take away From This Project	Python, Machine Learning, EDA, Data Analysis, Feature Engineering, Regression, Classification, Streamlit, MLflow, Model Evaluation, Feature Scaling, Domain Understanding.
Domain	Real Estate / Investment / Financial Analytics

Problem Statement

Develop a machine learning application to assist potential investors in making real estate decisions. The system should:

1. Classify whether a property is a "**Good Investment**" (Classification).
2. Predict the estimated **property price after 5 years** (Regression).

Use the provided dataset to preprocess and analyze the data, engineer relevant features, and deploy a user-interactive application using Streamlit that provides investment recommendations and price forecasts. MLflow will be used for experiment tracking.

Business Use Cases

✅ Empower real estate investors with intelligent tools to assess long-term returns.

✅ Support buyers in choosing high-return properties in developing areas.

✅ Help real estate companies automate investment analysis for listings.

✅ Improve customer trust in real estate platforms with data-backed predictions.

Approach

- **Step 1: Data Preprocessing**

  - Handle missing values and duplicates.
  - Normalize or scale numerical features.
  - Encode categorical features like `Location` and `Property_Type`.
  - Create new features like Price per Sqft, School Density Score, etc.
  - Create a binary label "Good Investment" based on domain rules (e.g., appreciation rate > threshold).

- **Step 2: Exploratory Data Analysis (EDA)**

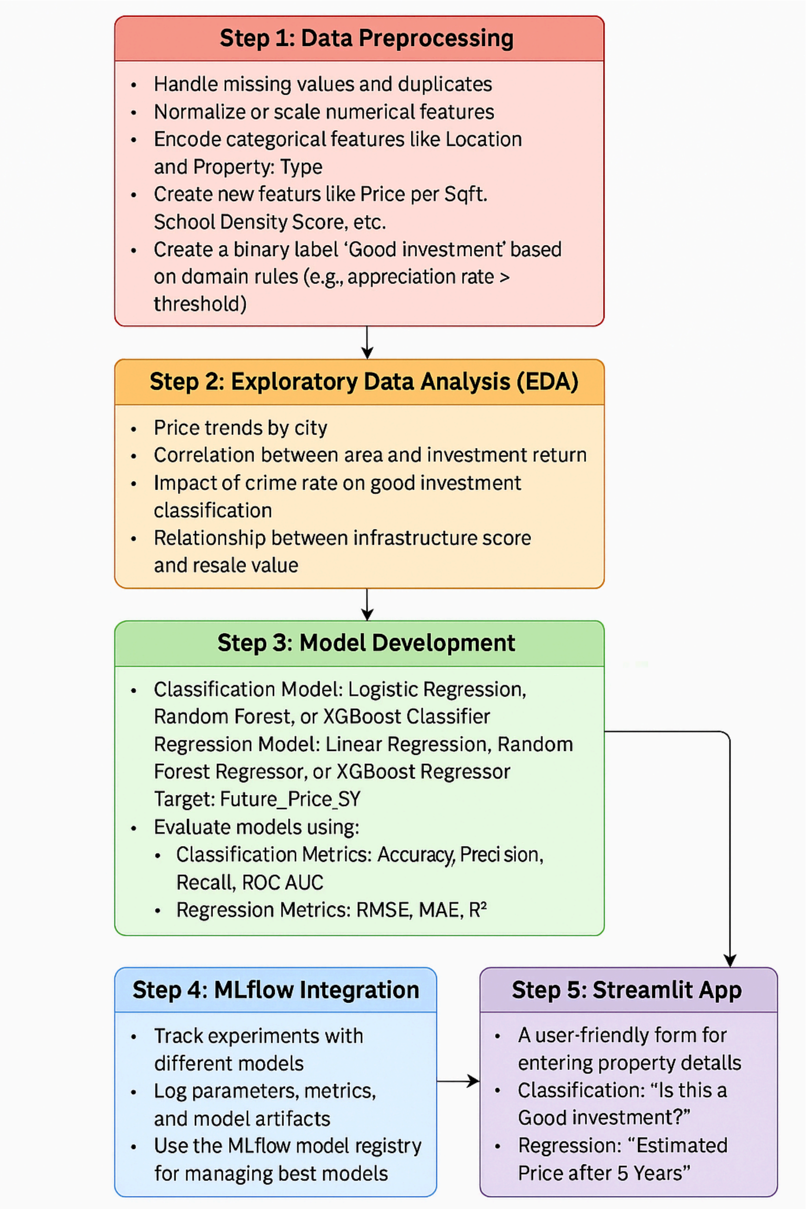
  - Price trends by city
  - Correlation between area and investment return
  - Impact of crime rate on good investment classification
  - Relationship between infrastructure score and resale value

- **Step 5: Streamlit App**

  - A user-friendly form for entering property details, filtering out properties by area, price, BHK etc.
  - Show:
    - Classification: “Is this a Good Investment?”
    - Regression: “Estimated Price after 5 Years”
  - Add visual insights (e.g., location-wise heatmaps, trend charts)

- Show model confidence scores & feature importance

Data Flow and Architecture



Step 1: Data Preprocessing

- Handle missing values and duplicates
- Normalize or scale numerical features
- Encode categorical features like Location and Property: Type
- Create new featurs like Price per Sqft. School Density Score, etc.
- Create a binary label 'Good investment' based on damain rules (e.g., appreciation rate > threshold)

Step 2: Exploratory Data Analysis (EDA)

- Price trends by city
- Correlation between area and investment return
- Impact of crime rate on good investment classification
- Relationship between infrastructure score and resale value

Step 3: Model Development

- Classification Model: Logistic Regression, Random Forest, or XGBoost Classifier Regression Model: Linear Regression, Random Forest Regressor, or XGBoost Regressor Target: Future_Price_SY
- Evaluate models using:
  - Classification Metrics: Accuracy, Preci sion, Recall, ROC AUC
  - Regression Metrics: RMSE, MAE, R^2

Step 4: MLflow Integration

- Track experiments with different models
- Log parameters, metrics, and model artifacts
- Use the MLflow model registry for managing best models

Step 5: Streamlit App

- A user-friendly form for entering property detalls
- Classification: "Is this a Good investment?"
- Regression: "Estimated Price after 5 Years"

Target variables:

🔹 Regression: Future Price (5 yrs)

Goal: Predict property price in 5 years.

- **Fixed Rate:** Future = Current * (1 + r)^t (e.g., r=8%)
- **By Location/Type:** Use different growth rates per city/property.
- **Feature-based:** Predict growth using property features (BHK, sqft, RERA, etc.)

🔹 **Classification: Good Investment**

Goal: Decide if property is worth buying.

- **Price vs Median:** Price ≤ median → good investment
- **Price per Sq Ft:** Cheaper than median → good
- **Multi-factor Score:** Combine features (BHK≥3, RERA, ready-to-move) → threshold = good

Dataset

Data set link : india_housing_prices.csv

Dataset Description

Feature	Description
ID	Unique identifier for each property record
State	State where the property is located
City	City of the property
Locality	Specific neighborhood or locality
Property_Type	Type of property (Apartment, Villa, House, etc.)
BHK	Number of bedrooms, hall, kitchen

Size_in_SqFt	Area of the property in square feet
Price_in_Lakhs	Price of the property in lakhs (local currency)
Price_per_SqFt	Price divided by area; normalized price metric
Year_Built	Year when the property was constructed
Furnished_Status	Furnishing level (Unfurnished, Semi, Fully)
Floor_No	Floor number of the property
Total_Floors	Total number of floors in the building
Age_of_Property	Age of the property (Current Year - Year_Built)
Nearby_Schools	Number or rating of nearby schools
Nearby_Hospitals	Number of nearby hospitals
Public_Transport_Accessibility	Access to buses/metro/train
Parking_Space	Number of parking spots available
Security	Security features (Gated, CCTV, Guard)
Amenities	Amenities available (Gym, Pool, Clubhouse)
Facing	Direction the property faces (North, South, etc.)
Owner_Type	Owner type (Individual, Builder, Agent)
Availability_Status	Current status (Available, Under Construction, Sold)

Exploratory Data Analysis (EDA)

1–5: Price & Size Analysis

1. What is the distribution of property prices?
2. What is the distribution of property sizes?
3. How does price per sq ft vary by property type?
4. Is there a relationship between property size and price?
5. Are there any outliers in price per sq ft or property size?

6–10: Location-based Analysis

6. What is the average price per sq ft by state?
7. What is the average property price by city?
8. What is the median age of properties by locality?
9. How is BHK distributed across cities?
10. What are the price trends for the top 5 most expensive localities?

11–15: Feature Relationship & Correlation

11. How are numeric features correlated with each other?
12. How do nearby schools relate to price per sq ft?
13. How do nearby hospitals relate to price per sq ft?
14. How does price vary by furnished status?
15. How does price per sq ft vary by property facing direction?

16–20: Investment / Amenities / Ownership Analysis

16. How many properties belong to each owner type?
17. How many properties are available under each availability status?

18. Does parking space affect property price?

19. How do amenities affect price per sq ft?

20. How does public transport accessibility relate to price per sq ft or investment potential?

Results

- Cleaned and analyzed real estate dataset.
- Trained classification model with high accuracy for investment prediction.
- Regression model with low RMSE for price forecasting.
- Streamlit dashboard deployed with user input form and visual results.
- MLflow integrated to manage model tracking and performance comparison.

Project Evaluation Metrics

- **Data Handling**: Imputation, encoding, and outlier detection.
- **Model Evaluation**:
  - Classification: Accuracy, F1-score, Confusion Matrix
  - Regression: RMSE, MAE, R^2
- **Deployment**: Functional Streamlit UI for predictions and analytics.
- **Experimentation**: Version-controlled and logged using MLflow.

Technical Tags:

Python, Pandas, Scikit-learn, XGBoost, Regression, Classification, Streamlit, MLflow, Real Estate Analytics, Data Visualization

Deliverables:

- Cleaned & processed dataset (CSV or SQL)

- Python scripts for EDA, model training, and evaluation
- MLflow experiment logs + registered models
- Streamlit application (with prediction + insights)
- Project documentation (Methodology, Findings, Use Cases)

Timeline:

Refer to your Zen portal or email for deadlines.

References:

Project Live Evaluation Metrics	Project Live Evaluation
EDA Guide	Exploratory Data Analysis (EDA) Guide
Capstone Explanation Guideline	Capstone Explanation Guideline
GitHub Reference	How to Use GitHub.pptx
Streamlit recordings (English)	Special session for STREAMLIT(11/...
Streamlit recordings (Tamil)	
Streamlit documentation	Install Streamlit

Streamlit deployment document	Streamlit App on Streamlit Cloud
Mlflow documentation	An Introduction to MLflow Tracking, …
Mlflow recordings	Day 1 : Time to Learn - 2024_12_24 10_28 G… Day 2: Time to Learn - 2024_12_23 10_28 G…
Project Orientation (English)	POS : Real Estate Investment Adviso…
Project Orientation (Tamil)	