**CA 2: Predicting Housing Prices Using Machine Learning**

# 1. Strategic Overview of the Business Problem

The housing market plays an essential role in economic stability and development. Housing prices directly influence household wealth, investment decisions, and urban planning. Fluctuations in housing prices affect not only individual homeowners but also investors and governments, making the prediction of housing prices a critical task for many sectors. Predicting housing prices with accuracy can help stakeholders make informed decisions about investments, development, and economic policies. Traditional statistical methods have long been used to forecast housing prices, but these techniques often fall short when it comes to capturing the complex and dynamic nature of the housing market. Advances in machine learning offer new opportunities to improve the accuracy and reliability of prediction models by analyzing large datasets and identifying patterns in housing prices more efficiently than traditional methods (Case and Quigley, 1991).

This project builds upon the foundations established in CA 1, leveraging the increased availability of housing data to implement a robust machine learning model. The objective remains to develop a reliable solution for forecasting housing prices based on various features such as property size, location, nearby amenities, and infrastructure. The outcomes have practical applications for real estate companies, investors, urban planners, and policymakers who need data-driven insights for decision-making in the ever-changing housing market.

Housing prices are shaped by a wide array of factors, many of which are interdependent. These include macroeconomic indicators such as GDP growth and unemployment rates, microeconomic factors such as household income levels, and environmental elements like urbanization rates. Understanding how these variables interact provides invaluable insights for crafting policies that can promote economic growth and social stability. The integration of machine learning models into this analysis represents a significant leap forward in identifying nuanced relationships that might otherwise remain obscured.

# 2. Project Plan

Timeline and Milestones

The project was structured to align with a systematic project management methodology, ensuring effective prioritization of tasks and monitoring of progress. Key phases and milestones included:

1. Data Understanding and Preprocessing:
   * Week 1-3: Data loading, cleaning, and feature selection.
2. Exploratory Data Analysis (EDA):
   * Week 4-5: Statistical analysis and visualization of key trends.
3. Machine Learning Implementation:
   * Week 6-8: Model training, evaluation, and optimization.
4. Findings and Reporting:
   * Week 9-10: Synthesizing insights and drafting conclusions.

Challenges and Monitoring

Data Quality: Addressed through rigorous preprocessing steps, ensuring consistency and reliability in the dataset. This included resolving data imbalances and confirming uniformity across time-series entries.

Model Optimization: Iterative experimentation with algorithms and hyperparameters to maximize performance. Particular attention was given to balancing model complexity and interpretability.

Time Constraints: Weekly reviews ensured adherence to the timeline while maintaining flexibility for adjustments. Any delays in specific phases, such as EDA, were mitigated by overlapping tasks with subsequent phases.