A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Robert Jonjic |
| *Student Number* | sba24456 |
| *Module Title* | Strategic Thinking |
| *Assessment Title* | CA 1 – Capstone Project Proposal |
| *Assessment Due Date* | 27/10/2024 |
| *Date of Submission* | 26/10/2024 |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Contents

[Introduction 2](#_Toc180089690)

[Objectives 2](#_Toc180089691)

[Problem Definition 3](#_Toc180089692)

[Scope 3](#_Toc180089693)

[Data Sources 4](#_Toc180089694)

[Ethical Considerations 4](#_Toc180089695)

[References 4](#_Toc180089696)

**Assessment Task: Capstone Project Proposal**

**Title:** *Predicting Housing Prices Using Machine Learning*

# Introduction

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. However, 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).

The increased availability of housing data allows machine learning techniques to provide a reliable solution for forecasting housing prices based on various features such as property size, location, nearby amenities, and infrastructure. This project aims to develop a machine learning model capable of accurately predicting housing prices by analyzing data from European Union sources, focusing on the factors that drive price fluctuations. The outcomes will 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.

# Objectives

The primary objective of this project is to develop a machine learning-based prediction model that can accurately predict housing prices based on key factors. This will involve the design and implementation of various machine learning algorithms, such as regression models, decision trees, and random forests. A significant part of the project will focus on evaluating the performance of these algorithms to determine which is most suitable for predicting housing prices. Additionally, the project will explore real-world applications, providing a tool that helps real estate companies, investors, and urban planners make informed decisions on property investments by forecasting future price trends.

Another key objective is to create an interactive visualization tool that offers real-time insights into housing price predictions and market trends. This dashboard will provide stakeholders with user-friendly access to the model’s predictions, making it easier to interpret the results and apply them to their decision-making processes. Finally, the model will be designed with scalability in mind, ensuring that it can handle large datasets and be adapted for use in different regions within Europe. This adaptability will broaden the model’s applicability and allow it to serve a wider audience.

# Problem Definition

The housing market is subject to constant fluctuations due to various factors, including economic conditions, supply and demand, and location-specific variables. Traditional statistical models used for predicting housing prices often struggle to capture the complex relationships between these variables, resulting in inaccurate or incomplete forecasts. As a result, there is a growing need for more sophisticated prediction techniques that can account for these complexities and offer better accuracy. The goal of this project is to address this need by utilizing machine learning techniques that can automatically learn from historical data and improve predictions over time (Mueller and Tibshirani, 2020).

Machine learning algorithms are well-suited for this task because they can process large amounts of data, identify patterns, and adapt to changing trends in the market. By incorporating various factors such as economic conditions, property features, and location characteristics, the machine learning model developed in this project will offer a more accurate and timely prediction of housing prices. This model will be particularly beneficial for stakeholders such as real estate companies, urban planners, and investors, who rely on accurate forecasts to make data-driven decisions in a dynamic market.

# Scope

The project will be executed over two semesters and will involve several phases. The first phase will focus on data collection and preprocessing. The data will be sourced from the European Union's Eurostat website, which provides comprehensive information on housing prices across EU countries. This dataset will include variables such as housing price indexes, property size, location data, and economic indicators. Preprocessing the data will ensure that it is clean and suitable for use in machine learning models, with steps taken to handle missing values and outliers.

In the second phase, several machine learning algorithms will be developed to predict housing prices. These will include linear regression, decision trees, and random forests. The performance of each algorithm will be evaluated using training and testing datasets to ensure that the models are accurate and robust. In the final phase of the project, an interactive dashboard will be implemented to provide real-time monitoring and visualization of housing price predictions. The dashboard will offer stakeholders an intuitive interface for exploring the model's predictions and understanding market trends.

The scope of this project is focused on the prediction of housing prices within the European Union, and the model will primarily consider factors such as location, infrastructure, and market demand. Commercial real estate and rental markets are excluded from the scope, as they are influenced by different factors. Similarly, broader economic indicators such as unemployment rates or inflation, while important, will not be considered beyond their direct impact on the housing market.

# Data Sources

The primary dataset for this project will be sourced from the European Union's Eurostat database, which offers detailed housing price statistics across EU countries (Eurostat, 2024). This dataset provides a wide range of variables relevant to housing price predictions, including property size, location data, and market conditions. Additional datasets may be sourced from Kaggle or other publicly available platforms to enhance the model's prediction capabilities (Kaggle, 2024). These supplementary datasets will help ensure that the model is trained on a diverse set of data, improving its ability to generalize across different regions and housing markets.

# Ethical Considerations

This project will rely on publicly available datasets, ensuring that no personally identifiable information (PII) is used or processed. The data will be anonymized, and all data processing will comply with GDPR regulations, which govern the handling of personal data within the European Union. Data privacy and security will be prioritized throughout the project, ensuring that the data used is ethically sourced and handled.

Additionally, housing data is often subject to social and economic biases, such as disparities in pricing based on neighborhood characteristics or socioeconomic status. Machine learning models are particularly vulnerable to reinforcing these biases if they are not carefully monitored (O’Neil, 2016). Therefore, this project will implement fairness checks to ensure that the model provides unbiased predictions across different demographic and geographic areas. By addressing these ethical considerations, the project will ensure that the developed model not only meets technical standards but also upholds the principles of fairness and transparency.

# References

Case, K. and Quigley, J. (1991). The Dynamics of Real Estate Prices. *The Review of Economics and Statistics*, 73(1), pp.50–58. Available at: <https://www.jstor.org/stable/2109686>

Eurostat (2024). Housing Price Statistics. [online] Eurostat. Available at: <https://ec.europa.eu/eurostat/web/housing-price-statistics/database>

Kaggle (2024). Real Estate Price Prediction Datasets. [online] Kaggle. Available at: <https://www.kaggle.com/datasets>

Mueller, A. and Tibshirani, R. (2020). *Introduction to Machine Learning with Python: A Guide for Data Scientists*. 1st ed. O’Reilly Media.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. 1st ed. Penguin Random House.

GitHub link: https://github.com/CCT-Dublin/ca-1-Robert-Jonjic