**Business Description**

**Hypothesis**

**General Goal**

**Technologies Used**

**Libraries**

**Accomplishments**

**Data Source, Attributes & Dimensions**

**Descriptive Statistics and Data Visualization**

**Data Preparation and Preprocessing**

**Preliminary EDA (Exploratory Data Analysis)**

**Dealing with Missing Values**

**Possible Classification Features**

**Clustering and Text Analysis**

**Machine Learning Models (Random Forest, KNN, SVM, etc.)**

**Model Evaluation**

**Data Dictionary**

# **Using Machine Learning to Analyze and Predict Ireland's Housing Crisis**

## **Introduction**

People dream of having their own home. Everyone deserves a suitable place to live, and some would even say that it's a basic right to have a house to call your own. We are currently witnessing a housing crisis in Ireland that has only worsened over the years, despite the government introducing several programs to assist the population, such as the Help-to-Buy scheme and the First Home Scheme.

The pressing issues of housing availability and affordability have created significant barriers for potential homeowners. Property prices continue to rise at rates that outpace wage growth, and the rental market offers little relief, with monthly payments often exceeding potential mortgage payments. There aren't enough houses for everyone, and it is a big challenge to save enough money when faced with exorbitant prices, leaving many, especially younger people, without hope of purchasing a home.

The impacts of the housing crisis are numerous: Irish people leaving their own country, people living in poor conditions, and homelessness, along with the social and psychological effects of being unable to own a home. The crisis has led to increased mental health concerns, delayed family formation, and reduced economic mobility. Additionally, younger generations face an uncertain future, where owning a home feels more like a distant dream than a realistic goal.

The ripple effects extend beyond housing itself, affecting Ireland's economic competitiveness, social fabric, and demographic patterns. Many skilled professionals are choosing to relocate abroad, creating a brain drain that could have long-term implications for the country's development. The crisis disproportionately affects certain groups, including young professionals, single-parent households, and essential workers who often find themselves priced out of the communities they serve.

Addressing this issue requires innovative solutions and different approaches to ensure that the basic human need for housing is met. This includes exploring new construction technologies, alternative financing models, and data-driven policy making. This project aims to predict future house prices and utilize machine learning to help people prepare for the future of something so essential: having a roof over their heads and a place to call home.

## **Strategic Overview of the Business Problem**

Housing is a fundamental need, yet the current housing crisis in Ireland has left many without access to affordable homes. Despite efforts such as the Help-to-Buy scheme, challenges persist. Insufficient housing supply, prices that keep increasing, and economic pressures have created significant issues for first-time buyers and renters. This crisis has long term consequences, including homelessness, poor living conditions, mental health issues, and also causing Irish people to emigrate from Ireland.

The housing crisis demands innovative solutions, particularly those that empower individuals to plan for the future while providing politicians with insights and data. This project will use machine learning to address key questions about the housing market, aiming to predict future house prices and identify trends that can guide decisions for individuals and politicians.

Market inefficiencies and structural barriers have created a complex challenge requiring data-driven approaches and evidence-based policy making.

Key factors influencing house prices include:

* Supply and Demand, including the availability of new and second-hand homes.
* Economic conditions, such as inflation, unemployment, and wages.
* Government programs and their impact on first-time buyers and on developers
* Alternative uses of houses that were supposed to be in the rental/buying market, such as short-term rentals via platforms like Airbnb.
* Planning regulations and development timelines
* Construction sector capacity constraints
* Market sentiment and buyer confidence

By creating a predictive model, this project seeks to alleviate uncertainty surrounding the housing market and provide a pathway to informed decision-making for homebuyers and politicians alike.

The project follows the methodology of CRISP-DM (Cross-Industry Standard Process for Data Mining) and Agile practices to ensure that we are tracking deliverables but also ensuring we are flexible depending on the issues that arise during the project.

This capstone project will be completed over two semesters, with a focus on using historical housing data to predict future trends. The dataset, sourced from data.gov.ie, includes 589 rows detailing average house prices in key Irish regions: Dublin, Cork, Galway, Limerick, Waterford, and the rest of Ireland.

## **Problem Definition**

Ireland has been experiencing a housing crisis for at least a decade, if not longer. The root cause is complex, linked to:

* High demand and limited supply
* Wages and the current economy making it difficult for the general population to afford current house prices
* Immigration, which accounts for 15.5% of Ireland's population
* Platforms like Airbnb, where homeowners can earn more money by renting through the platform rather than selling or leasing their homes

This has a direct impact on homelessness, as many people cannot afford to buy or rent due to high prices. The crisis also negatively affects mental health and well-being, with many feeling hopeless and considering leaving the country as the situation shows no signs of resolving soon.

For some context and background, we can consider that the Irish house market has faced consistent pressure from the demand of new houses and high prices for buying and renting.

This issue also impacts several types of people, including first-time buyers, renters, developers and politicians.

## **Project Scope**

This capstone project spans two semesters and will analyze historical data from data.gov.ie, which has a dataset with information about the average house price on the main areas in Ireland: Dublin, Cork, Galway, Limerick, Waterford and the rest of Ireland, and it has a total of 589 rows.

While other EU nations and individual Irish counties may face similar housing challenges, this project will remain concentrated on these main areas to maintain clarity and relevance. By keeping the scope, we aim to produce more targeted insights that can be directly applied locally, also ensuring the findings are both practical and impactful for Ireland's housing market.

## **Key Phases of the Project:**

1. Data Understanding and Preparation:  
   * Analyze the dataset for trends and anomalies.
   * Ensure data quality by addressing missing values and inconsistencies.
2. Development of Predictive Models:  
   * Use machine learning techniques to predict housing prices for 2025 and 2026.
   * Evaluate multiple algorithms to ensure accuracy and interpretability.
3. Insights and Reporting:  
   * Identify counties with the highest and lowest predicted price increases.
   * Develop a user-friendly report summarizing findings for prospective homeowners and politicians.
4. Ethical Considerations:  
   * Address biases in the dataset to ensure equitable insights.
   * Evaluate the impact of predictions on vulnerable populations.

## **Timeline:**

* Semester 1: Focus on data understanding, preparation, and initial model development.
* Semester 2: Refine models, evaluate results, and document learnings and insights.

## **Success Criteria and Indicators**

The success of this project will be measured by checking if it achieves its main goals and creates useful results for different users. Below are the key criteria for success:

1. Model Accuracy  
   * The predictions for house prices in 2025 and 2026 should be close to real data when it becomes available.
   * The model should work well for all the main regions included in the project.
2. Clear Insights  
   * The project should identify which areas (accordingly to the dataset) in Ireland will see the highest and lowest increases in house prices.
   * The final report should be easy to understand and helpful for people planning to buy a house and for policymakers.
3. Positive Impact  
   * The report should help people make better decisions about buying houses by providing useful information.
   * Politicians should be able to use the results to improve housing programs and strategies.
4. Ethical and Social Awareness  
   * The project should not create unfair results or ignore certain groups of people or regions.
   * It should explain clearly how the model works and why certain predictions are made.
5. On-Time Delivery  
   * The work should follow the planned timeline and finish all parts of the project on time.
   * Each part of the project, such as data preparation and model building, should meet the quality expected by supervisors.
6. Long-Term Value  
   * The model and results should be easy to update with new data so they can stay useful in the future.
   * Policymakers or other organizations should find the results helpful enough to consider using them.

## **Project Plan: Implementation Approach**

The implementation of this project will follow a structured approach that incorporates regional analysis and socioeconomic factors to ensure comprehensive predictions.

Our analysis will include comparing price trends across Dublin, Cork, Galway, Limerick, and Waterford, while also considering employment rates, population changes, and economic indicators that influence housing markets.

We'll evaluate how different stakeholders are affected by the housing crisis and assess the effectiveness of existing housing policies like the Help-to-Buy scheme and planning regulations.

From a technical perspective, our machine learning approach will use feature engineering, time series analysis, and ensemble methods to improve prediction accuracy, while ensuring our models are interpretable for decision-making.

## **Data Understanding**

The dataset used in this analysis was selected from data.gov.ie, which contains historical house price information across the major regions in Ireland. It contains 588 records across multiple years since 1975, with eight rows providing detailed information about house prices in different areas.

Key Variables: Statistic Label: These variables categorize the house prices into two types:

* New House Prices
* Second Hand House Prices

Years Coverage: The years the house was bought

Area Coverage:

* National (country-wide statistics)
* Dublin
* Cork
* Galway
* Limerick
* Waterford
* Other areas

Price Information in euros

From an initial assessment, there's not too many empty values within the dataset. The dataset is also consistent with all values being in Euros.

## **Advanced Data Analysis Methods**

Our data analysis will use pattern recognition algorithms to identify trends in housing price fluctuations, including seasonal patterns and potential market correction signals. We'll also investigate correlations between housing prices and economic factors using regression analysis and principal component analysis to develop accurate predictive models.

## **Data Preparation**

The first steps taken for Data preparation were checking for missing values in our dataset. Our analysis revealed only 5 missing values in the VALUE column, which were handled by using the mean value for that.

I've also identified and assessed the relevance of each column, and during this process, I've removed the columns named: TLIST(A1), C02343V02817 and UNIT as is not adding anything to our analysis at the moment.

Additionally, I've considered removing data from the dataset as it contains values since 1975, but considering that we have only 588 entries, I decided to keep the dataset complete for now.

I've also added boxplots to analyze the data distribution for each region. These boxplots helped identify several outliers, particularly in the Dublin region. I have to examine these outliers to determine whether they represented market trends or errors in our dataset.

We've also introduced feature scaling using the StandardScaler library. This was important because the price values from our dataset ranged from €12,478 to €512,461, while our year values were between 1975 and the present. The scaling ensures all features contribute proportionally to the model's learning process.

**Hypothesis**

**General Goal**

## **Ethical Considerations**

Housing has long been recognized as a determinant of health, as highlighted in a 1934 report by Britten (R. Britten, The Relation Between Housing and Health). This is particularly relevant for vulnerable populations such as the homeless, refugees, and immigrants, who often face historical and systemic challenges related to housing. I will ensure that this project does not introduce any bias and that it is sensitive to the challenges faced by those striving to purchase a home, by ensuring the data analyzed doesn't contain any specific information such as nationality, gender and social situation. This approach ensures that we're maintaining objectivity by focusing only on variables related to housing, and so the insights will remain impartial.

## **Reference list**

www.linkedin.com. (n.d.). Understanding the Housing Crisis in Ireland: Causes, Consequences, and Solutions. [online] Available at: https://www.linkedin.com/pulse/understanding-housing-crisis-ireland-causes-solutions-tafura-khatun-lltme/.

Britten, R.H. (1934). The Relation between Housing and Health. Public Health Reports (1896-1970), [online] 49(44), pp.1301–1313. doi:https://doi.org/10.2307/4581354.

of the Taoiseach, D. (2024). Migration - the Facts. [online] Www.gov.ie. Available at: https://www.gov.ie/en/collection/aeea0-migration-the-facts/.

GeeksforGeeks. (2019). ML | Handle Missing Data with Simple Imputer. [online] Available at: <https://www.geeksforgeeks.org/ml-handle-missing-data-with-simple-imputer/>.

<https://realpython.com/python-matplotlib-guide/>

<https://realpython.com/linear-regression-in-python/>

# Introduction to Machine Learning with Python: A Guide for Data Scientists Paperback –

by [Andreas C Müller](https://www.amazon.co.uk/s/ref=dp_byline_sr_book_1?ie=UTF8&field-author=Andreas+C+M%C3%BCller&text=Andreas+C+M%C3%BCller&sort=relevancerank&search-alias=books-uk)

<https://www.w3schools.com/python/python_try_except.asp>

<https://www.evalacademy.com/articles/how-to-combine-data-from-multiple-sources-for-cleaning-and-analysis>

<https://saturncloud.io/blog/pandas-how-to-concatenate-dataframes-with-different-columns/#:~:text=To%20concatenate%20dataframes%20with%20different%20columns%2C%20we%20use%20the%20concat,new%20dataframe%20that%20combines%20them>.

<https://www.statology.org/pandas-factorize/>

**Github Repo**

Overall Notes:  
  
When combining the two datasets for data from 2010 to 2016  
Average absolute difference: 19.36%

The differences aren't too bad. We can probably combine them.

Include that I added a new dataset