Housing Price Analysis in Ireland

The real estate sector has always been an integral part of the economy, with property prices proving to be a strong indicator of a country's economic health. Both in Ireland and in many other countries, the real estate market has been influenced by political, economic, and social factors that have caused property prices to fluctuate over time.

# Importance of Analysis

The real estate market is complex and vast, containing a huge amount of available data. Therefore, traditional methods for accurately analyzing market patterns are often flawed and insufficient. Machine learning can bridge this gap, as it has tools with the ability to process large volumes of data and make accurate predictions and classifications.

An important capability to help investors and buyers understand future property price trends, both for buying and selling, and for making investments in the sector. Machine learning methods can also help policymakers and urban planners create better housing policies through detailed analysis of regional market conditions.

# Problems to be Addressed

## Determination of Future Prices

To assist investors and consumers in planning appropriately, it is crucial to make an effective forecast of property prices. Prediction algorithms can identify trends of increasing or decreasing prices based on historical data, reducing uncertainties in the market and developing better strategies in the real estate market.

## Regional Analysis

Ireland may present a heterogeneous real estate market, providing specific price patterns and needs for different properties depending on the county. Through classification algorithms, it is possible to divide the market and understand each region, thus helping to formulate public policies to develop the real estate market in the region.

# Justification for the Use of Prediction and Classification Algorithms

Machine learning algorithms can identify complex patterns in data that are often not visible through more traditional analysis methods. Regarding housing prices:

**Prediction algorithms** can model relationships between multiple variables and effectively predict future prices.

**Classification algorithms** help categorize data into groups or classes, such as price ranges or market categories (e.g., market price vs. non-market price).

# Project Objectives

1. Predict future property prices in Ireland, using historical data to train predictive models.
2. Conduct a detailed regional analysis to understand trends and peculiarities of the real estate market in different Irish counties.

These objectives aim to provide valuable insights that can benefit not only investors and planners but also ordinary individuals in their real estate decision-making.

# Characterization

## About Dataset

### Context

All residential properties sold in Ireland from 2010 to May 28th 2021.

Dataset retired from Kaggle on Adress: <https://www.kaggle.com/datasets/erinkhoo/property-price-register-ireland/data>

**Collaborator**: Erin Khoo (Owner)

### Content

**Rows**: 476,745

**Columns**: 9

**Memory usage**: 32.7MB

**Variables:**

* SALE\_DATE => Date of sale (dd/MM/yyyy) | datatime64[ns]
* ADDRESS => Address | string
* POSTAL\_CODE => Postal Code | string
* COUNTY => County | string
* SALE\_PRICE => Price (€) | float32
* IF\_MARKET\_PRICE => Not Full Market Price | int8
* IF\_VAT\_EXCLUDED => VAT Exclusive | int8
* PROPERTY\_DESC => Description of Property | string
* PROPERTY\_SIZE\_DESC => Property Size Description | string

### Acknowledgements

Data sourced from publicly available site: <https://propertypriceregister.ie>

### Source

National Property Price Registry

### License

[EU ODP Legal Notice](https://commission.europa.eu/legal-notice_en)

# Data Preprocessing

## Data Cleaning

To prepare the dataset for analysis, several actions were taken:

1. The columns POSTAL\_CODE and PROPERTY\_SIZE\_DESC were deleted because they contained a very large amount of null data (81.17% and 88.92% respectively).
2. Unrecognized characters were found in the PROPERTY\_DESC column. These data were changed to null and then deleted.
3. 770 records were deleted as they were considered duplicate data.

These actions were sufficient for data cleaning, deletion of null data, and keeping the data in good condition.

## Data preparation

The variable SALE\_DATE was split into SALE\_MONTH and SALE\_YEAR to provide flexibility and facilitate temporal analysis, allowing for the identification of seasonal trends over time or grouping data into specific time intervals.