A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Santhosh Shanmugasundaram |
| *Student Number* | Sba24100 |
| *Module Title* | Strategic Thinking |
| *Assessment Title* | Rent Predictor: A Machine Learning Approach to Forecast Dublin Home Rent |
| *Assessment Due Date* | 18th May 2025 |
| *Date of Submission* | 18th May 2025 |

**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.

Rent Predictor: A Machine Learning Approach to Forecast Dublin Home Rent

Contents

[1 Introduction 1](#_Toc198657503)

[2 Problem Definition 1](#_Toc198657504)

[3 Project Scope 1](#_Toc198657505)

[3.1 Business Understanding 2](#_Toc198657506)

[3.2 Data Understanding 3](#_Toc198657507)

[3.3 Data Preparation 3](#_Toc198657508)

[3.4 Modeling 4](#_Toc198657509)

[3.5 Evaluation 5](#_Toc198657510)

[3.6 Deployment 5](#_Toc198657511)

[4 Details of Dataset 6](#_Toc198657512)

[5 Key Challenges 6](#_Toc198657513)

[6 Machine Learning Algorithms 7](#_Toc198657514)

[7 References 7](#_Toc198657515)

[8 GitHub Link 7](#_Toc198657516)

# Introduction

Rental homes are an integral part of the housing ecosystem, providing an alternative to homeownership. A properly functioning housing market is an essential ingredient of a properly functioning economy and society. (IPOA and IPAV)

From an economic, social and political perspective, housing is currently the greatest challenge faced by Irish policymakers. The housing challenge is manifested in a lack of supply of owner-occupier and rental properties; prohibitively high house prices and rents; and a serious problem of homelessness. (IPOA and IPAV)

# Problem Definition

The inability to provide an adequate supply of suitable and affordable housing supply for those who want to rent has very negative economic and social consequences. These consequences include:

* High and rising rents take spending power out of the economy and render it very difficult for aspiring house buyers to build up a sufficient deposit.
* High and rising house rents put upward pressure on wages, and this undermines national competitiveness.
* The availability of an abundant supply of high-quality housing to rent or purchase at affordable prices is a necessary condition for labour mobility within a country and between countries. For Ireland, inward migration is an essential part of the economic model, and housing can act as a major impediment to such labour flows.  
  (IPOA and IPAV)

# Project Scope

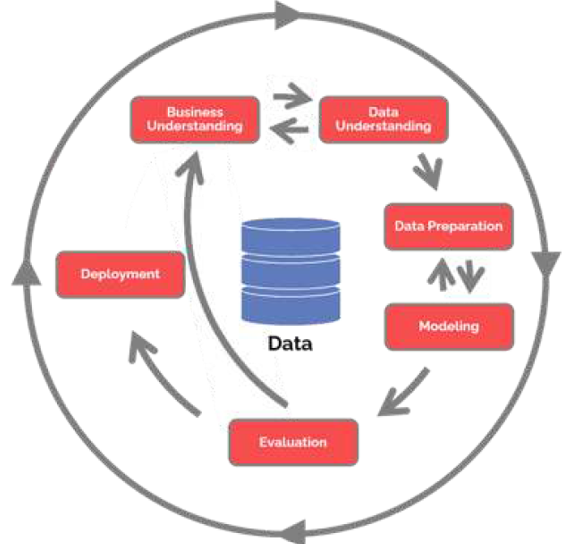
The scope of this project is to understand the **Project Management** strategy using key factors like increase in Dublin home rent price every quarter for three different types of homes namely apartment, terrace house and semi-detached house. The target of this study is to help the renters to decide on the future rent prices in Dublin so that expenses are planned.

Use the AI technology along with the Machine Learning models like ARIMA, Linear Regression and Random Forest algorithm to predict the future rent at Dublin based on the **CRoss Industry Standard Process for Data Mining** (CRISP-DM).

This is a process model that serves as the base for a data science process that includes 6 phases as below.

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modelling
5. Evaluation
6. Deployment

(Hotz)



## Business Understanding

Any good project starts with a deep understanding of the customer’s needs. Data mining projects are no exception and CRISP-DM recognizes this.

The Business Understanding phase focuses on understanding the objectives and requirements of the project. Aside from the third task, the three other tasks in this phase are foundational project management activities that are universal to most projects:

1. **Determine business objectives:** You should first “thoroughly understand, from a business perspective, what the customer really wants to accomplish.” ([CRISP-DM Guide](https://web.archive.org/web/20220401041957/https:/www.the-modeling-agency.com/crisp-dm.pdf)) and then define business success criteria.
2. **Assess situation:** Determine resources availability, project requirements, assess risks and contingencies, and conduct a cost-benefit analysis.
3. **Determine data mining goals:** In addition to defining the business objectives, you should also define what success looks like from a technical data mining perspective.
4. **Produce project plan:** Select technologies and tools and define detailed plans for each project phase.

While many teams hurry through this phase, establishing a strong business understanding is like building the foundation of a house – essential.

## Data Understanding

 Next is the *Data Understandin*g phase. Adding to the foundation of *Business Understanding*, it drives the focus to identify, collect, and analyze the data sets that can help you accomplish the project goals. This phase also has four tasks:

1. **Collect initial data:** Acquire the necessary data and (if necessary) load it into your analysis tool.
2. **Describe data:** Examine the data and document its surface properties like data format, number of records, or field identities.
3. **Explore data:** Dig deeper into the data. Query it, visualize it, and identify relationships among the data.
4. **Verify data quality:** How clean/dirty is the data? Document any quality issues.

## Data Preparation

A common rule of thumb is that 80% of the project is data preparation.

This phase, which is often referred to as “data munging”, prepares the final data set(s) for modeling. It has five tasks:

1. **Select data:** Determine which data sets will be used and document reasons for inclusion/exclusion.
2. **Clean data:** Often this is the lengthiest task. Without it, you’ll likely fall victim to garbage-in, garbage-out. A common practice during this task is to correct, impute, or remove erroneous values.
3. **Construct data:** Derive new attributes that will be helpful. For example, derive someone’s body mass index from height and weight fields.
4. **Integrate data:** Create new data sets by combining data from multiple sources.
5. **Format data:** Re-format data as necessary. For example, you might convert string values that store numbers to numeric values so that you can perform mathematical operations. (Data Science PM)

## Modeling

What is widely regarded as data science’s most exciting work is also often the shortest phase of the project. Here you’ll likely build and assess various models based on several different modeling techniques. This phase has four tasks:

1. **Select modeling techniques:** Determine which algorithms to try (e.g. regression, neural net).
2. **Generate test design:** Pending your modeling approach, you might need to split the data into training, test, and validation sets.
3. **Build model:** As glamorous as this might sound, this might just be executing a few lines of code like “reg = LinearRegression().fit(X, y)”.
4. **Assess model:** Generally, multiple models are competing against each other, and the data scientist needs to interpret the model results based on domain knowledge, the pre-defined success criteria, and the test design.

Although the [CRISP-DM Guide](https://web.archive.org/web/20220401041957/https:/www.the-modeling-agency.com/crisp-dm.pdf) suggests to “iterate model building and assessment until you strongly believe that you have found the best model(s)”, in practice teams should continue iterating until they find a “good enough” model, proceed through the CRISP-DM lifecycle, then further improve the model in future iterations.

## Evaluation

Whereas the *Assess Model* task of the *Modeling* phase focuses on technical model assessment, the *Evaluation* phase looks more broadly at which model best meets the business and what to do next. This phase has three tasks:

1. **Evaluate results:** Do the models meet the business success criteria? Which one(s) should we approve for the business?
2. **Review process:** Review the work accomplished. Was anything overlooked? Were all steps properly executed? Summarize findings and correct anything if needed.
3. **Determine next steps:** Based on the previous three tasks, determine whether to proceed to deployment, iterate further, or initiate new projects. (Data Science PM)

## Deployment

Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise.

A model is not particularly useful unless the customer can access its results. The complexity of this phase varies widely. This final phase has four tasks:

1. **Plan deployment:** Develop and document a plan for deploying the model
2. **Plan monitoring and maintenance:** Develop a thorough monitoring and maintenance plan to avoid issues during the operational phase (or post-project phase) of a model
3. **Produce final report:** The project team documents a summary of the project which might include a final presentation of data mining results.
4. **Review project:** Conduct a project retrospective about what went well, what could have been better, and how to improve in the future.

Your organization’s work might not end there. As a project framework, CRISP-DM does not outline what to do after the project (also known as “operations”). But if the model is going to production, be sure you maintain the model in production. Constant monitoring and occasional model tuning is often required. (Data Science PM)

# Details of Dataset

Residential Tenancies Board is an independent public body that regulates the private rented sector. It is responsible for multiple activities like tenancies registration, dispute resolution, investigation and sanction of landlords. It has also published many numbers of reports and data. (Citizensinformation.ie)

For the specific case study, I will use the dataset available in the below link.

<https://data.cso.ie/table/RIQ02>

# Key Challenges

Collecting valid, meaningful and up-to-date data from a trusted source, poses a significant challenge. Moreover, selecting the right dataset for analysis with correct filters and formatting the data before consumption takes significant amount of time.   
  
Exploratory Data Analysis (EDA) has been performed to make sure there is no duplicates entries nor null values is present. Also, scaling the data by eliminating unnecessary columns also had to be done.   
  
Hyperparameters in machine learning are extremely crucial and significant because they greatly affect the performance and efficiency of the selected model. I faced several typical difficulties with hyperparameters, which are:

Incorrect tuning of the parameters can result in either overfitting or underfitting. There are numerous hyperparameters for machine learning models, and selecting the appropriate one with the optimal configuration is quite challenging. Sometimes the models take longer to run during training, which complicates their use with larger datasets.

This dataset contains the average rental price at Dublin city based on number of bedrooms (1, 2, 3) and type of property (semi-detached house, terrace house, apartment) starting from fourth quarter 2020 till first quarter 2024.

# Machine Learning Algorithms

Three different types of algorithms were used to evaluate and predict the data namely,

1. Random Forest Regression
2. Linear Regression
3. ARIMA

# References

(IPOA and IPAV)

IPOA, and IPAV. *THE IRISH PRIVATE RENTAL MARKET*. IPAV.ie, June 2022.

Available at: <https://www.ipav.ie/sites/default/files/ipav_ipoa_jim_power_updated_report_june_2022.pdf>

(Hotz)

Hotz, Nick. “What Is CRISP DM?” *Data Science Project Management*, 2024, www.datascience-pm.com/crisp-dm-2/.

# GitHub Link

<https://github.com/santhosh-sba24100/SBA24100_CA3_Strategy_Thinking>