# **Project Proposal**

## **Predicting Perth Housing Prices**

**Introduction**

The objective of this project is to develop a predictive model for housing prices in Perth, Western Australia. By predicting housing prices, potential buyers and sellers can make more informed decisions about the housing market, this will enable them to navigate the market more effectively and optimize their transactions. In addition, government agencies can also use this model to modify government policies to make house ownership more accessible to people.

This proposal will outline the background, data description, analysis methods, and model prediction evaluation for developing such a predictive model.

**Background**

The cost of buying a house is steadily increasing in many parts of the world, making it more difficult to afford homes. This increase can be for various reasons such as economic growth, population density, and geographic location.

Hence, there can be several parameters in determining the values of house prices. Still, this analysis is limited to the size of the house, the actual area of the house, population data for each suburb, the distance of the house from the CBD, and several highly ranked schools. By leveraging these factors and using machine learning algorithms, models can predict specific area housing prices with varying degrees of accuracy.

**Description of Data:**

As part of this project, we will collect and utilize the following data:

1. House Pricing Dataset: A dataset containing information about the prices of housing in Western Australia, is part of this dataset. It contains 33,656 observations of 19 variables recorded between 1990 and 2020.
2. Census Data: This dataset contains the population of Western Australia. It contains 74888 observations of 12 variables recorded in 2021.

To make a prediction, this model takes input like size, location, number of bedrooms, number of bathrooms, land area, floor area, distance from CBD, age, nearest school, and its rank. It then combines these parameters with the variables such as the population of that suburb, Australian Citizens, Non-Australian Citizens, and age in the 10 years age group collected from Australian Bureau of Statistics [1].

The data used in this study comes from the Kaggle dataset [2]. The dataset contains historical housing prices in Perth, and it has been used for various analyses by other researchers and data scientists.

**Analysis methods**

A house price prediction model can be built by summing up the results from a training dataset, as well as the results from a test dataset, in this proposal, we will outline the steps for building such a model. Our model will be trained using the training dataset, while its performance will be evaluated by using the test dataset. We will experiment with the following model to train on the training dataset:

* Simple Linear Regression: Fitting a linear relationship between the house price and each predictor variable separately.
* Multiple Linear Regression: Fitting a linear relationship between the house price and all predictor variables simultaneously.
* Decision Tree and Random Forest

The performance of each model will be evaluated using cross-validation, so we will be able to select the one that performs the best.

**Model prediction evaluation**

Using the test dataset as a basis for evaluation, we will assess how well our model performs in the real world. The predictive analysis model will be validated by using 80/20 percent variation of dataset values and current actual values, as the currently obtained dataset ends by Dec-2020.A variety

of metrics will be used to evaluate the performance of the model, such as mean squared error, mean absolute error, and R-squared, to measure the accuracy of the model.

If models do not appear accurate or sufficient to make predictions, then other prediction models will be researched.

**Statement of Originality**

1. This assignment is my/our own original work, except where I/we have appropriately cited the original source (appropriate citation of original work will vary from discipline to discipline).

2. This assignment has not previously been submitted in any form for this or any other unit, degree or diploma at any university or other institute of tertiary education.

3. I/we acknowledge that it is my responsibility to check that the file I/we have submitted is: a) readable, b)the correct file and c) fully completed

# References

|  |  |
| --- | --- |
| [1] | Australian Bureau of Statistics, “Australian Bureau of Statistics,” Australian Bureau of Statistics, 20 April 2023. [Online]. Available: https://www.abs.gov.au. [Accessed 24 April 2023]. |
| [2] | M. S. ZAINAL, “Kaggle,” 2021. [Online]. Available: https://www.kaggle.com/datasets/syuzai/perth-house-prices/discussion?select=all\_perth\_310121.csv. [Accessed 24 April 2023]. |