

# Proposal

Garrick Ho  
Department of Statistics  
University of Connecticut

October 9, 2023

## 1 Introduction

With the never-ending demand for housing, real estate businesses are tasked with finding a system for predicting the housing market. Being able to predict the housing market will help multiple people with buying and selling houses are the right time. The reason why I wanted to choose to research the housing market is to be able to use what I find in this research and apply my own knowledge to the real market.

## 2 Specific Aims

This research aims to predict house prices based on this data set. I selected this question because I wondered what are the key characteristics of a house that can affect the cost of the house. This is important to understand because if this model works, we can use this with real live data and apply it to the real housing market.

## 3 Data Description

The data set that I have found comes from Kaggle. There are thirteen columns and no missing values in this data set. Price is the total cost of the house. Area refers to the total area of the house. Bedrooms is the total number of bedrooms in the house. Bathroom is the total number of bathrooms. Stories is the number of how many levels are in the house. Main road, guestroom, basement, hot water heating, and air conditioning are either the house does or does not have those qualities. Parking is the number of parking spots the house has. Prefarea is if the house is in a preferred area. Furnishing status is whether the house comes furnished, semi-furnished, or unfurnished when buying the house.

## 4 Research design

Because the data does have any missing values there is no need for dropping or replacing values. Then analyze the data to see how the variables are distributed and look for any patterns. After that, the data can go through some data manipulation and create the train and test data sets. While doing that, we can find the variables that are significant and drop the ones that are not significant. After that, we can run linear regression models to try to predict the prices of houses.

## 5 Discussion

Given the data set that is being used, I would expect to see that the linear regression model will be accurate in predicting the cost of houses. Because there are lots of variables that can correlate to the cost of the house, the model will be able to use them for the prediction. My work may help contribute to other works that have been done in the past. The potential impacts of this work would help the understanding of what causes the cost of a house to fluctuate. If the results of the investigation are not what I expected, then I could go back and try to use a different model for testing and see if it is better than the linear regression. Or I can go back to the linear regression model and conduct some hyper-parameters to see if it produces a better result.

## 6 Conclusion

With the housing market moving up and down every day, it is imperative that real estate businesses find a way to predict the cost of a house. The goal of this will be constructed in this research. The main focus of this research is to learn what causes the houses to fluctuate in cost. At the end of this, there will be a linear regression model that can predict the cost of a house.