

## Capstone Two: Project Proposal

Kemal Tezgin

**Project Name:** House price prediction in Australia

**Problem:** The goal of this capstone project is to estimate house price trends in Australia based on the prices of recent sales of houses in the country. Through the use of this dataset, we will be able to predict house prices based on historical data, allowing us to make smart decisions about whether or not to buy a house at a certain price.

**Clients:** The primary target market of this project is real estate companies that are looking to make their way into Australia.

**Data:** To achieve our goals, we aim to use the following dataset uploaded on Kaggle:

<https://www.kaggle.com/srikanthladda/house-price-prediction?select=train.csv>

The data has 1460 rows of data and 81 columns, including numerical and categorical variables.

**Approach to the Problem:** To resolve this problem, we will follow a series of steps:

1. The first step will be data wrangling. Several data fields are unfilled, which makes the model building more difficult. Data cleaning will be necessary to fix the problem.
2. The second step of this project will be exploratory data analysis. The cleaned data set will be visualized in terms of independent variables. We can then determine which factors to consider when determining the price of a house.
3. As a third step, several statistical models will be developed to estimate house prices. Initially, we will build baseline models, such as linear regression, Lasso regression, and Ridge regression, and collect appropriate performance metrics for comparison.
4. To improve performance, we will seek additional models besides the baseline ones. During this step, we envision using such techniques as k-nearest neighbors and random forest regression.
5. As a final step, we will perform interpretability analyses to determine how the variation of different features impacts the variation of house prices.

**Deliverables:** Our findings will be presented in the form of a slide deck, project report, and Jupyter notebook. Additionally, we will make the documents available to the public on a GitHub page.