Capstone Two: Project Proposal

Project Name: House price prediction in Australia

Problem: This capstone project attempts to predict house prices in Australia based on a set of data that represents houses that have recently been sold in the country. With the help of this dataset, we aim to predict the prices of sold houses to determine whether it is profitable to buy a house at a certain price.

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

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

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

the data has 1460 row data and 81 columns, including numerical, 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. Data cleaning will be necessary since several data fields are unfilled, which will make model building more difficult.
2. The cleaned data set will be visualized in terms of independent variables. This will help us to understand what variables have more weight in determining house sale prices.
3. Then, we will develop several models to estimate the house prices. There will be supervised learning models such as Lasso and Ridge linear regression models, as well as a k-nearest neighbors non-parametric supervised learning method.
4. Training and test sets will be produced for each model, and after the models have been trained on the train set, we will assess their accuracy in the test set. Specifically, we will analyze root mean squared errors of each model to determine the most accurate model for estimating house prices.

Deliverables: We intend to present our findings in the form of PowerPoint slides and the code will be delivered in the form of a Jupyter notebook. Additionally, we will create a special GitHub page where the documents will be made available.