

Project Proposal

Team Member: Yu Chen, Zifei Dong, Ning Pan, Sifan Tao, Jennifer Yao

Dataset: Kaggle Competition: House Price Prediction

Data source:

According to Kaggle, Ames Housing dataset was compiled by Dean De Cock. The House Prices dataset is from the Kaggle platform. It has two CSV files, "train.csv" and "test.csv" and has [79 explanatory variables](#) describing (almost) every aspect of residential homes and contains information on various features of houses sold in Ames, Iowa from 2006 to 2010.

Motivation and goals:

Our group's motivation is to build a model that accurately predicts the sale price of a house based on various features. This model can be used by various stakeholders such as real estate agents, homeowners, and property investors to make informed decisions about buying, selling, or investing in a property. The model should be able to provide accurate predictions of house prices, which will enable homeowners to price their homes competitively and real estate agents to provide informed advice to their clients. Additionally, the model can be used by property investors to identify undervalued properties and make informed decisions about buying or selling them.

The ultimate aim of our project is to create a model that performs well on new and unseen data, is interpretable, and provides insights into the most important factors that drive house prices. This will enable stakeholders to gain a better understanding of the housing market and make informed decisions based on the insights provided by the model. We expect our model to have the potential to provide valuable insights and assist stakeholders in making informed decisions related to the housing market.

Preliminary ideas on technique:

Sifan Tao:

Exploratory data analysis:

You may find our exploratory data analysis [here](#).