

Capstone 1 Proposal- Predict House Prices in Ames, Iowa

What problem do you want to solve?

Pricing a home can be very difficult and currently real estate agencies rely on manually comparing a home to several other comparable homes (also known as 'comps') sold recently in the same vicinity. This process yields less than ideal results since the agent will likely only choose a few homes and a few features causing the process to be prone to error.

Who is the client?

A real estate agency in Ames, Iowa is very interested in predicting the amount a home will sell for. This information will allow them to identify homes which are being sold under market value which could be an indicator of a good investment opportunity. Additionally, it would help them to quickly establish a sales price or recommend an offer price to their clients. This ensures that a home is correctly priced for sale without having to perform this research on their own, which is the conventional approach. They have decided to hire a data scientist to analyze data from home sales in Ames, Iowa and to develop a predictive model for home sales price.

What data is being used?

The dataset being used is from a Kaggle competition and contains 79 variables detailing almost every feature of residential homes in Ames, Iowa.

[House Prices Ames, Iowa](#) - The dataset contains about 1459 rows of data including the following features: Sales price, Building Class, Zoning, Lot Frontage, Lot Area, Street, Alley, LotShape, LandContour, Utilities, LotConfig, LandSlope, Neighborhood, Year Built.. Etc.

If needed, perhaps the data can be enriched with some data from Zillow or similar websites.

Proposed Solution/Approach:

To solve this problem, first the data will be analyzed for cleanliness and any issues with the core dataset corrected. Then via exploratory data analysis, the approach will be to identify some of the key features affecting home sales price. Finally, once these features are identified, a predictive model will be constructed using regression and other machine learning techniques.

Deliverables:

Project deliverables will include code for the project as well as a powerpoint presentation showing the work conducted and the results of the analysis. Finally, a dashboard will be created

for the real estate agents who would like to look up estimated sales prices by entering several key features of their choice.