Project Proposal

Problem Statement

How accurately can I predict the prices of homes, in Ames, Iowa, using up to 80 variables and factors?

Context

Ames Real Estate Co. (AREC) is on the fence to hire a permanent Data Scientist team. The technology and the science seems interesting, but they have yet to see how it could benefit them. Their new CTO Randy Cornell recommended that AREC hold a contest to see how accurately Data Scientists predict prices of homes using last year's data and home prices. Suffice it to say, not only would higher scores convince AREC of the efficacy of Data Science as a whole, but this could mean job opportunities for the most useful models. My task is to explore the data and accurately predict the prices of homes.

Criteria for Success

Create a model (or models) to predict home prices with at least 90% accuracy.

Scope of Solution Space

Scope of solution space includes explanatory analysis of 79 variables that can affect the prices of homes.

Constraints Within Solution Space

I am the mercy of the data. There could be missing values, mislabeled data, and even wrong information. This could highly influence my modeling and ultimately my results.

Stakeholders to provide key insights

1. Chief Technology Officer, Randy Cornell
2. Kenneth Gil-Pascual, Data Scientist Consultant

Key Data Sources

1. A single CSV file that contains 79 home variables for Ames, Iowa.
2. A TXT file describing the data coding and headers.
3. A PDF file from the Journal of Statistics Education comprehensively describing the Ames, Iowa Housing Dataset as originally published by Dean De Cock.

Deliverables

1. GitHub repository explaining my exploration process, analysis, and results.

2. A Project Report describing my process, conclusions, findings, and any other important details.
3. A PowerPoint presentation outlining and summarizing my work for AREC's Board of Directors.