House Prices Problem

from Kaggle

Ernestas Monkevicius
Isaac Dempsey
Yiming Chen
Xinyue Wang

A.k.a The Kagglers

Problem

We decided to solve a Kaggle competition problem for our project.

The problem we settled on involves predicting the *final price* of a home given 79 explanatory variables. These explanatory variables include detailed categorical information about each house.

Data

	81 Columns from *ID* to *sale price,* **79** explanatory variables.
	1459 rows of data
	1461 rows of data

Analysis

Firstly, we will clean our dataset, looking for missing data, outliers, duplicates etc.

Next, understanding. We will examine the dataset closely to try and find the shape of each feature. We will look for strong relationships between explanatory variables and the target feature - the final sales price.

Modeling/ Prediction

We'll investigate the performance of the following models:

- Linear Regression
- Log Linear Regression
- Random Forest

The aim is to ***predict the house prices***.

Result

The expected result is a strong relationship between a subset of the explanatory variables and the final house price.

We hope to uncover the nature of this relationship during our investigation.