**OBJECTIVES**

The goal of this project was to use EDA, visualization, data cleaning, preprocessing, and linear models to predict home prices given the features of the home, and interpret your linear models to find out what features add value to a home. The [data](https://www.kaggle.com/c/house-prices-advanced-regression-techniques" \t "https://towardsdatascience.com/_blank) was originally taken from Kaggle.

From Kaggle:

Ask a home buyer to describe their dream house, and they probably won’t begin with the height of the basement ceiling or the proximity to an east-west railroad. But this playground competition’s dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence.

With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, this competition challenges you to predict the final price of each home.

### **Exploratory Data Analysis (EDA)**

As with any data exercise, we began with some Exploratory Data Analysis.

****Numeric Variables****

There are 36 relevant numerical features. MSSubClass, which "identifies the type of dwelling involved in the sale", is encoded as numeric but is in reality a categorical variable.

There are 36 numerical features, of the following types:

* Square footage: Indicates the square footage of certain features, i.e. 1stFlrSF (First floor square footage) and GarageArea (Size of garage in square feet).
* Time: Time related variables like when the home was built or sold.
* Room and amenties: data that represent amenties like “How many bathrooms?”
* Condition and quality: Subjective variables rated from 1–10.

Most of the variables that deal with the actual physical space of the apartment are postively skewed — which makes sense, as people tend to live in smaller homes/apartments apart from the extremely wealthy.

Sale Price also has a similar positively skewed distribution — I hypothesize that the variables dealing with the actual dimensions of the apartment have a large impact on Sale Price.