# House Prices Advanced Regression Techniques

Overview

There are several factors that influence the price a buyer is willing to pay for a house. Some are apparent and obvious and some are not. Nevertheless, A large data set with 79 different features (like living area, number of rooms, location etc) along with their prices are provided for residential homes in Ames, Iowa. The challenge is to learn a relationship between the important features and the price and use it to predict the prices of a new set of houses.

01-eda: Exploratory data analysis

Plot distribution of the numerical features examine the skewness

Plot correlation matrix between the features

02-cleaning: Cleaning and preprocessing of data

remove skewenes of target features

handle missing values in categorical features

handle missing values in numerical features

feature selection

03-feature\_engineering: Engineering new features

Some examples:

A total area was created as a new feature by adding the basement area and living area.

The number of bathrooms were added together to create a new feature.

For numerical features with significant skewness, logarithms were taken to create new features.

Some features were dropped that did not contribute significantly in predicting the SalePrice.

04-modelling: Fitting different models on the cleaned data and predict the house price on test set

**Lastly, as a chronic over-achiever:**

* Find at least two unexpected phenomena in the data and provide a visualization and analysis to document their presence.

**Considerations**

Remember, the people reading your analysis will NOT be data analysts. Your audience will be general public . Your data and analysis need to be presented in a way that is focused, concise, easy-to-understand, and visually compelling. Your visualizations should be colourful enough to be included in press releases, and your analysis should be thoughtful enough for dictating programmatic changes.

**Assessment**

Your final product will be assessed on the following metrics:

* Analytic Rigor
* Readability
* Visual Attraction

**Hints**

* You may need to get creative in how you combine each of the CSV files. Don't just assume Tableau is the right tool for the job. At this point, you have a wealth of technical skills and research abilities. Dig for an approach that works and just go with it.
* Don't just assume the CSV format hasn't changed since 2013. Subtle changes to the formats in any of your columns can blockade your analysis. Ensure your data is consistent and clean throughout your analysis. (Hint: Start and End Time change at some point in the history logs).
* Consider building your dashboards with small extracts of the data (i.e. single files) before attempting to import the whole thing. What you will find is that importing all 20+ million records of data will create performance issues quickly. Welcome to "Big Data."
* While utilizing all of the data may seem like a nice power play, consider the time-course in making your analysis.
* Remember, data alone doesn't "answer" anything. You will need to accompany your data visualizations with clear and directed answers and analysis.
* As is often the case, your clients are asking for a LOT of answers. Be considerate about their need-to-know and the importance of not "cramming in everything". Of course, answer each question, but do so in a way that is organized and presentable.
* Keep a close eye for obvious outliers or false data. Not everyone who signs up for the program is answering honestly.
* In answering the question of "why" a phenomenon is occurring, consider adding other pieces of information on socioeconomic or other geographic data. Tableau has a map "layer" feature that you may find handy.
* Don't be afraid to manipulate your data and play with settings in Tableau. Tableau is meant to be explored. We haven't covered all that you need -- so you will need to keep an eye out for new tricks.
* The final "format" of your deliverable is up to you. It can be an embedded Tableau dashboard, a Tableau Story, a Tableau visualization + PDF -- you name it. The bottom line is: This is your story to tell. Use the medium you deem most effective. (But you should definitely be using Tableau in some way!)
* Treat this as a serious endeavour! This is an opportunity to show future employers that you have what it takes to be a top-notch analyst.
* Good luck!

### **REQUIREMENTS**

Submissions must meet the following requirements:

* Include a Project built with the required developer tools and meets the above Project Requirements.
* Include a text description that should explain the problem your Project is attempting to solve and its features and functionality.
* Include a demonstration video of your Project. The video portion of the submission:
* should be less than three (3) minutes
* should include footage that shows the Project functioning on the device for which it was built
* must be uploaded to and made publicly visible on YouTube and a link to the video must be provided.
* must not include third party trademarks, or copyrighted music or other material unless the Entrant has permission to use such material.
* Include a URL to a code repository on GitHub or another code repository platform. If the code repository is private, Entrant must provide access to the GitHub account
* Include a list of the APIs and Development tools used within the project.
* Include potential further improvements to your Project if more time were permitted.
* Be the original work of the submitter, be solely owned by the submitter, and not violate the IP rights of any other person or entity.
* Multiple Submissions: An Entrant may submit more than one Submission; however, each submission must be unique and substantially different from each of the Entrant’s other Submissions.