**Data Analysis Ideas Beyond Basic EDA**

**Central Question**: The simple linear model based on **GrLivArea** point to price/ft^2 as an important derived target. On the other hand the fan like price /ft^2 scatterplot implies that different houses have a wide range of price/ft^2. We would like to study the major driving factors for these **Ames** homes to have significantly different price/ft^2.

No matter which local housing market we inspect, the house prices are the on-going consensus among the home buyers and the home sellers depending on the supply and demand of the current house market. The house market supply and demand is influenced by the local labor market/economical condition, job creation, population density, interest rate, price momentum, etc. Among all the area related features (e.g. 1st floor area, 2nd floor area, basement finished area, etc.), do the local home buyers show some preference towards some features over the others? In other words, are the home buyers willing to spend an equal amount for one square-foot of 1st floor living space, 2nd floor living space, deck space, finished basement space, etc.?

Frame your study using multiple linear regression and interpret your findings in therms of the price discounts. Consider the following points:

* Which of the house styles, ranch or colonial, is more popular in terms of the house prices?
* What is the popularity interms of the housing inventory (based on the available data)?
* Does the house price sensitivity on 2ndFlrSF depend on the house style (1st story excluded)?
* How is GrLivArea engineered from the other area-related-feature? If you would design a gross living feature (to replace GrLivArea) fpr the purpose of descriptive modeling, what would you do? What are the pros and cons when aggregating several features together?
* Do home buyers value upstairs bathrooms and basement bathrooms equally? Do they price a half a bath as ½ of a full bath? Based on your finding, how would you design an aggregate bathroom feature?
* How much of a premium were the Ames home buyers (back in 200-2010) willing to pay for a swimming pool, an open deck, or an enclosed porch? Did the home buyers value the optional additions based on their sizes?
* How important is a garage in adding value to the house? How about the differences between an attached vs detached garage?
* Are there feature combinations which increase the house value? What are the feature combinations which drive the house value down (making the house less desirable)?
* House Quality Condition: The Ames dataset grades each house by the appearance, kitchen, basement, garage, heating fireplace,… and finally an overall quality/condition.
  + How are the overall quality/condition features enginerred from the individual quality/condition metrics?
  + How do these metrics’ importances affect local house prices?
  + To what extent does the curb appeal affect the house prices?
* Does home remodeling add value to the house? For example, how do the old houses which were remodeled in the 1990s compare to the houses which were built in the 1990s?

**Neighborhood Analysis**

Many home buyers purchase their homes based on their neighborhood preferences and the location. Study the statistics on all the **Ames** neighborhoods, including price, price/area, lot area, gross living area, year built, the ratio of one story homes, the lot utility rate (how much of the lot is utilized to build the house), various quality/condition metrics, house styles, etc. Please depict a story of the urban development of Ames and relate it to the Ames housing prices.

To get familiar with the various Ames neighborhoods and the ratings of the puvlic schools, visit the map tab of relator.com. While there are many housing info, web sites, realtor.com uses a neighborhood system which is more compatible with the one used in the dataset. Getting the long-lat coordinates of the 2,500 houses also helps to pin-point the relative geographic locations of the neighborhoods.

Consider the following questions:

* Does the price sensitivity on quality depend on the neighborhood?
* Given that Ames has about a 30% poverty rate, and that ISU is the largest employer of the city, identify the affluent and poor neighborhoods on your data analysis and the external information. Identify the neighborhoods which offer a convenient job commute.
* What types of dwellings are popular in Ames? How about their popularities in the different neighborhoods?
* Time Series Analysis
  + Between 2006-2010, does Ames housing market show any pricing trend? Is your observation consistent with the fact that that the recession occurred in 2008-2009
  + How would you describe the price trend of different types of dwellings in terms of time series (Hint: pivot/pivot\_table)?
  + In the US, there are more house-shopping activities during the warmer seasons (Spring/Summer). What is the seasonality of Ames house transactions?

**Descriptive Modeling Ideas**

The goal of a descriptive housing price modeling is to build a pricing model to describe the pricing rationale (the consensus of the home buyers/sellers), to describe the key driving factors that affect house prices, and to understand how they affect the local home prices. While optimizing