**Case Study**

**Zillow**

The website Zillow estimates the home prices for over 100,000,000 homes around the United States. (Well, actually they call them Zestimates.) In their own words,“We use proprietary automated valuation models that apply advanced algorithms to analyze our data to identify relationships within a specific geographic area, between this home-related data and actual sales prices. Home characteristics, such as square footage, location or the number of bathrooms, are given different weights according to their influence on home sale prices in each specific geography over a specific period of time, resulting in a set of valuation rules, or models that are applied to generate each home’s Zestimate. Specifically, some of the data we use in this algorithm include:

Physical attributes: Location, lot size, square footage, number of bedrooms and bathrooms and many other details.

Tax assessments: Property tax information, actual property taxes paid, exceptions to tax assessments and other information provided in the tax assessors’ records.

Prior and current transactions: Actual sale prices over time of the home itself and comparable recent sales of nearby homes

Currently, we have data on 110 million homes and Zestimates and Rent Zestimates on approximately 100 million U.S. homes. (Source: Zillow Internal, March 2013) " ———–

<https://www.zillow.com/>

The data for this Case Study is labeled as **“Real Estate Case Study Data.xls”**

The case study Questions::

1. What are the various types of variables that are present in the data?
2. How would you describe the distribution of house sizes?
3. Plot house sizes and prices of those houses. What do you conclude from the plot? Are there any outliers in the data?
4. Are houses with fireplaces generally more expensive than houses without? (Hypothesis Testing)
5. Compare the distribution of housing prices that have a water front and those houses that do not have water front. Test whether the prices of these two types of houses are same.