CAPSTONE PROJECT 2: **Zillow's Home Value Prediction (Zestimate)**

What is the problem you want to solve?

Zillow is an online real estate database company that was founded in 2006 since then their home valuation model 'Zestimates' is one of the most popular in the U.S. real estate industry.

In this competition, participants will develop an algorithm that makes predictions about the future sale prices of homes. The contest is structured into two rounds, the qualifying round which opens May 24, 2017 and the private round for the 100 top qualifying teams that opens on Feb 1st, 2018. The qualifying round requires to build a model to improve the Zestimate residual error. The final round is to build a home valuation algorithm from the ground up, using external data sources to help engineer new features that give your model an edge over the competition.

Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

A home is often the largest and most expensive purchase a person makes in his or her lifetime. Ensuring homeowners have a trusted way to monitor this asset is incredibly important. The Zestimate was created to give consumers as much information as possible about homes and the housing market, marking the first time consumers had access to this type of home value information at no cost.

In brief, outline your approach to solving this problem (knowing that this might change later).

First, data wrangling and exploratory data analysis will be applied to clean and understand the data better. This step will be beneficial before moving to the modeling part. My initial analysis has showed that the data has too many missing values, there are also too many redundant variables. After cleaning the data, machine learning algorithms like random forest, principal component analysis, boosting and more will be applied to make predictions.

For the primary analysis: link

What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

Deliverables for this project are the Jupyter Notebook with the analysis and a slide deck which summarizes the methodology and results.