Capstone Project Proposal

1. **What is the problem you want to solve?**

The problem I want to solve is how to calculate the total tax assessed value of a house in the Los Angeles region.

1. **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?**

My clients are builders who are looking to build new houses in the area, homeowners of new houses, and real estate agents looking at new houses. All three clients could use this information to determine what the market value of a newly constructed home in a given area could be.

Builders would be able to determine if the home is worth building or not. New homeowners could determine if the price that is being asked for the house is reasonable or not. And real estate agents could see if this is a particular home that would be of interest for them to work with given the expected selling price.

1. **What data are you going to use for this? How will you acquire this data?**

I am going to use 2017 property data from the ‘Zillow’s Home Value Prediction (Zestimate)’ Kaggle competition. I will not be participating in the actual competition and will not be trying to solve the same problem the competition is working to solve.

Statistics about the dataset:

Number of Rows (in main table) – 2.9 million

Number of Columns (in main table) - 58

Data Types in Columns (in main table) – numeric and categorical

Number of Tables - 8

Data Source: https://www.kaggle.com/c/zillow-prize-1/data

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

First, I will gather the data from Kaggle and load it into Python. Then I will analyze the data to determine the data quality. I will format and clean the data in this step. I will then explore the features of the dataset in order to determine which features are most important for my model. Next, I will split the data into my training and test data sets. After that, I will select the most appropriate algorithm to use. I will then create a model. Then I will evaluate that model and adjust it accordingly.

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

My deliverables will be a paper, a slide deck summarizing my paper, and the code associated with my project.