# Capstone 1 Project Proposal

**Problem:** San Francisco's real estate has been suffering a supply and demand issue for at least a decade now. This supply and demand issue has led to soaring property values, making it extremely difficult to purchase property in San Francisco, and consequently making it difficult to live in San Francisco because the cost of living is increasing drastically. Notable publications have written that delays in the issuance of building permits has had a major effect on San Francisco's real estate issues. I will be examining a dataset of San Francisco building permits and try to create a model that will help predict the issuing times for building permits in San Francisco.

**Client:** My client is a real estate developer in San Francisco. The client cares about the problem because he/she is looking to find out what the issuing times for building permits will be, what type of permits matter more to the San Francisco Department of Building Inspection, and insights that can be made about the real estate development in San Francisco. Based on my analysis, the client will be able to make a decision as to what type of projects should be focused on and then the client will be able to figure out when to file the permit.

**Questions To Explore:**

- Which features, produce the best model, for predicting the issuing time of a building permit?
- Do permit types affect the issuance of a building permit and the date it is issued?
- What conclusions can be drawn upon the real estate development in San Francisco?
- Can certain areas (i.e. zip code, districts, neighborhoods) be associated with long or short approval times?
- Suppose I submit a building permit file in the summer in a given area, what would be the approval time?

**Data:** The data I will be using originates from Kaggle. The URL is: https://www.kaggle.com/aparnashastry/building-permit-applications-data

Features to be potentially used:

- Permit Type - Type of permit
- Location - Longitude and Latitude
- Permit Type Definition - Description of permit type
- Description - Details about purpose of permit
- Filed Date - Date permit is filed
- Number of Existing Stories - Stories currently in building. (Not applicable for some permit types)

- Number of Propose Stories - Stories proposed. (Not applicable for some permit types)
- Estimated Cost - Estimated cost of project
- Revised Cost - Revised estimation of project cost
- Existing Use - Existing use of buildings
- Existing Units - Existing units in buildings
- Proposed Use - Proposed use of buildings
- Proposed Units - Proposed number of units
- TDIF compliance - TDIF compliant or not, new legal requirement
- Site Permit - Permit of site
- Supervisor District - Supervisor district to which building location belongs to
- Neighborhoods - Neighborhoods to which buildings belong to
- Zip code - Zipcode of building address

**Outline:** I will be using various data wrangling techniques to clean the data and analyze for any outliers in the data. Next, I will run an analysis to see which features will work best in the model, and then finally create a predictive model that will using features selected. The accuracy of the model will determine the best model.

**Deliverables:** The deliverables will be code, a machine learning model, a final report, and a slide deck.