Capstone Project 1 Proposal
Predicting Wind Farms Impact on Home Values

Problem:
There are many factors that influence a home's value such as the region's economy, condition of the home, proximity to shopping & other attractions, school district rating, etc…, but a wind farm may be a greater factor on home values than one, or even a combination of several of these factors.

Imagine you just opened your favorite news feed and the local news is that a wind farm is coming to your area. As a homeowner, should you be concerned about your home's value? As a town board member should you raise concerns about property tax revenue (due to a decrease in home values) at the next meeting?

We propose to use data from various sources on wind farms and home values to build a machine learning model for predicting the impact of windfarms on home values. The model will calculate the year-over-year changes in home value for zipcodes with a windfarm and compare it to the year over year home value change in a zipcode 25 miles away that does not have a windfarm.

Client:
Home owners, prospective home buyers, and local government officials interested in tax revenue can use this model to predict the likelihood that a wind farm in their area will impact home prices. They can then take steps, such as decide to not move to an area, petition their city to not go forward with a wind farm, or in the case of city officials insist on revenue from the wind farm that will compensate for the loss in tax revenue from homes.

Dataset:
The wind farm data will be acquired from the [US Wind Turbine database](#). This data contains numerous information about each wind mill in each wind farm; of interest for this study is the windmill Id, project id, location, installation date, and number of windmills in the project.

In addition, dataset/s containing information (location, year built, assessed value by year) about homes is needed for this study. The final datasets to use are being evaluated with leading candidates being [Zillow](#) and [Redfin](#).

Dataset parameters:
● The dataset timeframe for wind farm data is farms constructed in the years 2009 – 2018.
● Zipcodes with windfarms and zipcodes without windfarms 25 miles away will be evaluated

Approach:
Since the objective is to assess if percentage change in home values vary between zipcodes with farms and zipcodes without, a regression algorithm will be used to build the model. The algorithm will provide a prediction of the percentage change in home value.

The predictor variables will be…
● windfarm (windfarm id, location, etc...)
● houses in a given zipcode (location, percentage change in value in prior years, median income, population density, etc...)

Deliverables:
1. Milestone report on the capstone project

a. Data wrangling
b. Data story
c. Exploratory data analysis

2. Final report
   a. In-depth analysis
   b. Final paper
   c. Presentation/slide deck
   d. Online video or blog post

Appendix:

Request for housing data:

I am interested in API access (or download access) to historical housing data. Specifically, I need the ability to view average sales price, median sales price for residential housing by state, by county, by zip code, by neighborhood, and by latitude/longitude.

The history period I'm looking for is years 2000 - 2010, but I could make some other time frame work as well. Prior to signing up, I'd like confirmation from you that the data described above can be obtained by me once I am a member.