Capstone 2: Project Proposal

Category: Real Estate

Problem Statement

How can we automate the identification of profitable investment opportunities (for properties that can be flipped) based on images alone?

Use: I imagine this being used by an investor to get an automatically curated set of potential investment opportunities to browse through, where the program does the dirty work of sifting through numerous properties trying to determine which ones are good investment opportunities, i.e. which ones are outdated and need renovation

Problem: Currently, the best that can be done, short of hiring someone to manually do the sifting, is to set automated alerts based on basic filters of number of bedrooms, year built, etc. But there is no way to directly identify / filter for properties that are outdated in style. You can only do that by looking at the pictures. These sites don't have this filter built in and there is no field for "Outdated" with values of Yes or No. At best the original description or agent notes can be searched for keywords like "investment" opportunity, etc. But oftentimes the ones that are prime for flipping are NOT considered by the listing agent to be flippers! Like that's not the first thing that comes to mind, that's not what they're thinking of / that's not who they're thinking of / imagining will buy this place. They see it as a cute, cozy, sweet home sweet home

What We’ll Need

* A dataset of real estate listing images and their associated sale prices
  + <https://www.kaggle.com/amir22010/house-price-estimation-from-image-and-text-feature/data>
* A script to pull images of sold listings
* A way of projecting, based on ‘fixed up’ / ‘modernized’ homes, what the sale price of the unrenovated home in question *WOULD* be *once* it is fixed up

Of course, as is with any predictive model, ultimately a human would need to sift through the presented findings, at least visually, if not in person, to determine if there are any factors that aren’t being caught by the model, such as an oddly placed feature or lack of main floor bathroom or inability to remove a wall without compromising structural integrity

Thoughts:

* Lower weight/score for non-trendy colors - walls gotta be gray/

Other stuff:

* Kaggle competition on predicting housing prices
* May be tricky to train data set with mixing pictures and text - try doing only one or the other for now / at first / at a time
* Dataset found only has 4 pictures for each house – would like to make a model that utilizes more pictures, especially since may not be using *ANY* textual data!
* Could do that by writing a script / using selenium web driver to mimic a person using a website
* Need to learn about Convolution Neural Networks & Deep Learning for dataset with images – get briefed by Raghu & get training resources