Capstone 2 – Final Project Report

* needs to be in PDF
* explain problem, approach & findings
* ideas for further research
* 3 concrete recommendations on how your client can use your findings
* give the document an appropriate title, such as Capstone\_Final\_Report or

Capstone\_Project

* keep final code in notebook
* ‘Model Metrics file’ – a separate text file of the final model features, params, hyperparams, & performance metrics! make as easy to read as possible

\*\*\*don’t know if there’s a diff b/w wrangling & preprocessing?? look at the SB dox??

* use Final Notebook – **WORKING COPY** to scrape ‘cute’ comments, notes, insights
* try moving threshold up or down! don’t know which makes more sense
* some images were probably grainy – don’t know if made a diff
* See if I need to further simplify / adjust training & test images!!
  + ***MAYBE WE SHOULD PROBABLY FOCUS ON \*JUST\* ONE ROOM, LIKE THE KITCHEN, SINCE THAT’S USUALLY THE \*MOST TELLING\* OF THE STATUS/UPDATEDNESS OF THE HOUSE, THE SINGLE BEST INDICATOR/ PREDICTOR OF THE ‘HEALTH’ OF THE HOME!!!! THE \*SUMMARIZER\* OF THE REST OF THE HOME!!!!***
    - *I’m sure it could be done where you combine, I mean that’s what Ahmed/Moustafa did! but yeah, maybe there's a way to like basically like do each room SEPARATE,* ***BUT, \*ASSIGN WEIGHTS TO EACH\* ROOM AND PUT EM TOGETHER FOR A FINAL PREDICTION 0->1!!!!***
    - *or we can simply even say like we’re ONLY in like the ‘home renovation business’! and like we look at listings for sale, and then approach these sellers to see if they’d wanna update to make their house more sellable! / up the value of their home! CUZ THAT’S AMAZING DATA FOR THEM!!!! they’re getting \*FREE\* LOOKS INTO PEOPLE’S HOUSES! so they can see what’s out there to see what services they could offer them!*
    - *now, this is different, cuz, the CLIENT/SELLER is getting more money for the home, iA,* ***YOU’RE*** *not getting like that increased profit – unless THIS IS WHAT YOU MAKE THE DEAL/YOUR MODEL!- that the bulk of your pay, or even* ***ALL*** *of it comes from getting a* ***PERCENTAGE*** *of the sale of the home!!!!* ***but what if they decide to NOT sell it cuz they’re like OH WOW THIS IS SOWWW NAAIIIIYYYCEEE!*** *well, you can put in contract that they’ll then be charged/billed at the regular rate etc!*
    - *but even if kitchen remodeler isn’t getting* ***ANY*** *piece of the house sale, which would solve the no-longer-selling problem*
    - *oh also it might be weird if ppl see like oh, this was just up, now it looks completely different and the price is WAY higher!*
    - *TOO LATE TO DO CONSTRUCTION WHEN PPL WANNA \*SEE\* THE PLACE!!!/IT’S* ***ALREADY*** *LISTED!! and some ppl might just take it as is! might look bad to DELIST/RELIST! #shady triggers! plus the construction downtime could cause significant delays in the timeline! and that’ll make the sellers nervous about like WHEN IS THIS GONNA BE DONE! so would be stressful/high for remodelers unless like super confident/efficient! then that’ll be why they get paid the BIG BUCKS! will show from testimonials/WORD OF MOUTH IS BEST!!!!*
    - *would be smart if kitchen modeler was teamed/partnered with, or* ***WAS*** *him/herself a real estate agent, so that BEFORE THE HOUSE IS EVEN LISTED, then do the remodeling at THAT point!!!*
    - ***OH SHOOT THIS ALREADY \*IS\* JUST KITCHENS!!!!***
      * *but we can further trim it down to make more consistent! like start off w/ making as* ***EASY AS POSSIBLE!!!!*** *like make sure they all have islands, and SQUARE islands at that!!!! (for not flips)*
    - *and using default params, we’re at 72% Accuracy – so either we gotta get better TRAINING IMAGES, or we gotta get better PARAMETERS,…* ***OR BOTH!?!***
* go back and change ‘flip/not flip’ to ‘buy/don’t buy’!!!!
* shorten ‘training’/‘testing’ to train/test
* how do you know the appropriate image pixel size to make an image??
* maybe try to ensure more even number of examples of flip & non-flip for both training & test to avoid class imbalance/bias!!!!
* maybe I shouldn’t make the images SQUARES since they’re all rectangles, and at that, they’re always in LANDSCAPE orientation?! so by resizing to square, it’s SMUSHING them?? or, I guess it’s all relative anyway so doesn’t MATTER to machine?! all the same to it?!
* okay so right now in training I have a MIX of kitchens with SQUARED/RECTANGULAR islands and then *LONG/ANGLED/BENT islands!!! this might be throwing things off! if looking at shape! that might be TOO MUCH for it to process!! like it’s very likely it’ll pick upon the WRONG features!!! esp if it has such few examples as this!!!*
* Remem, in theory we COULD have multiple labels, as like SUB-TYPES of flips, and then GROUP together the ones that fall under broadest FLIP!!! like classifier could be for say *DIFF KINDS OF KITCHENS!!!* And so it’s like, if any of THESE, then FLIP, if any of THESE categories, then *DON’T FLIP!!!*
* *it’s much harder to tell w/ IMAGES for us HUMANS, cuz like subjective from OUR view, but* ***MAY WELL HAVE HAD ‘OUTLIERS’ IMAGES IN DATASET THAT THREW EVERYTHING OFF!? but like here may be ones like that DIDN’T HAVE AN ISLAND!!! OR HAD LONG/SKINNY/ANGLED ISLAND!!!***
  + *sA I think initially I might’ve thought the opposite – like that machine was like a smart adult – that like you could show them just a ‘handful of’ images of a ‘diff type’ / example class of non-flippers, for example, and that’d be enough for it to pick it up! but NO!!! one-offs can actually be what COMPLETELY \*RUIN\* IT/THE MODEL!!!! DON’T TAKE RISKS!!! STICK TO BASICS AT FIRST/IN EARLY STAGES!!!! ESP IF ONLY HAVE A FEW TRAINING EXAMPLES!!!! but yeah so I threw in/sprinkled a few ‘diff’ ones thinking THIS WILL HELP “ROUND IT OUT” AND TEACH IT/SHOW IT THERE’S SOME VARIATION SO WHEN IT SEES EXAMPLES LIKE THESE ON (THE) TEST, IT’LL BE PREPARED/KNOW WHAT TO DO!*
  + *but the better simple strategy would be to also trim the TEST set to only include examples/be consistent, for BOTH labels, with the labels in the TRAINING!!!!!*
* Also I think I should probably **CROP MORE!!!** but let’s start w/ working on the training set first! Cuz right now, at 28x28 of 32x32 is 76% of the picture
  + **but actually, more than simply the *percentage* of the picture is like *WHICH 75% OF THE PICTURE IS IT?! Like LET’S NOT ASHHYUME! LET’S ACTUALLY LOOK AT OUR PICTURES! I think we can especially trim from the TOP, cuz framing composition tends to be very similar, and in that composition, the top is typically a lot of CEILING!!! which isn’t really the crux/is NOISE!***
* *at this point do our ANGLES of like where the picture was taken ALSO need to be the same?!* Cuz, usually it’s the CABINETS that are surrounding the island / in the background…
* Also, if there’s CHAIRS at the island, that could throw it off?
* LOL! could even go as so far as to basically JUST zoom into the island!!! cuz if there’s an island, that’s usually in the center of the kitchen/center of the kitchen *shot!*
  + *lol in which case I can even JUST get pictures of islands!!!*
  + *LOL – so like the kitchen is the summary of the updatedness of the house, and the ISLAND is the summary of the updatedness of the KITCHEN, and* ***THUS****, from* ***JUST*** *looking at the island, YOU CAN TELL TO PRETTY GOOD CONFIDENCE/FOR NEARLY ALL CASES/PPL-* ***HOW UPDATED THE \*HOUSE\* AS A \*WHOLE\* IS!!!!!!*** *cuz usually ppl won’t have like an updated OTHER AREAS and* ***NOT*** *an updated KITCHEN!!! that’s usually the FIRST place! floors & kitchen!counters&cabinets! lol – that’s the other one* ***– FLOORS & CABINETS!!!!!*** *so we could like create* ***FEATURES*** *if we could figure out how to reliably tell it like what parts are what!!!!!! like how in that one type of modeling, maybe PCA/like? where it breaks image down into ‘PRINCIPAL COMPONENTS’!!!!!*
* *but short of that, the next best/easiest, or even easier thing we can do, to get good, CONSISTENT, images, which is what we want at this point, is just take the few good ones we have and* ***REVERSE GOOGLE IMAGE SEARCH THEM!!!! CUZ IT’S VERY GOOD AT GETTING LIKE IMAGES!!!!***
* *but remem – we still got 70-75%, when only like 8/70 aka ≈10% of images were consistent w/ this, and maybe another 10% consistent on/to ANOTHER proto/type!*
* *so it should probably be enough that I only made sure the DON’T FLIPS were very consistent, so that if something doesn’t fit that mold, even if the FLIPS are more variegated, it’ll just call those FLIPS!?*
* dang, so for some reason, it’s a lot hard to find many consistent good images for DON’T flip than I thought! whether older/contemporary OR modern*!!*
  + *and yeah, I went ahead and tried it where I had WAY MORE \*DON’TFLIPS\* THAN FLIPS\* AND IT IDENTIFIED \*ALL\* AS FLIPS!!!*
  + *maybe try TRIMMING DOWN flip set to only a few / near same number!!!*
  + ***LOL! NOPE! REALLY BAD! THEN MADE THEM ALL DON’T FLIPS!***
  + *just gonna have to do compromise I guess between getting more DON’T FLIPS and CUTTING DOWN THE FLIPS!!*
* does it matter what size we resize the image to??
* should I have a ‘random seed/state’?

possible params to adjust?:

* img = cv2.resize(img, dsize = (32,32), interpolation=cv2.INTER\_CUBIC)
* *try BICUBIC??*