

Data Science Project Class

KEVIN MCALISTER

Administrative Stuff

Introductions!

Administrative Stuff

We'll meet each Monday and Wednesday in MSC 307C from 4:00 – 5:15 PM

- Attendance at these meetings is expected
- Unexcused absences will result in a 1% deduction from overall final grade
- A lot of life is just showing up, so just show up!

Administrative Stuff

If you work hard and show up, you will get a good grade for this class!

- This is an absolute banger of a project that has a real chance to have wide reaching impact.
- I'm just asking for you to apply your killer data science skills to the cause.

Project Summary

In conjunction with Neighborhood Nexus, we're going to be working to create a new data source for affordable housing research!

- A little different than the original project proposal, but a better project

Project Summary

A big problem in the affordable housing community is that homeowners can fall behind on property upkeep

- Why do you think that this is a problem?
- What are some reasons that you think homes fall into disrepair?

Project Summary

Property upkeep can be expensive!

- It's tough to pay for roof repairs, lawn maintenance, foundation repairs, energy efficient windows, etc.
- Where are owners who are struggling financially going to get that money?

Project Summary

Two options:

- Sell the home to an investment firm who buys it for less than it would be worth with investment in getting it up to date
- Let the property deteriorate to the point that the city needs to step in. Ends up with homeowners paying large legal fees and still needing to pay for repairs.

Project Summary

Reasons this happens:

- Low disposable income to mortgage ratio
- High property taxes
- Heir's properties
- Poor knowledge of resources that exist to assist owners in repairs

Project Summary

As you might expect, this problem disproportionately impacts marginalized communities:

- The elderly
- Minority home owners
- Historically disadvantaged communities and neighborhoods

Project Summary

There are resources out there to help those struggling!

- InvestATL 0% loans
- City of Atlanta homeowner services

Need to figure out how to target those in need.

Project Summary

Current methods:

- Wait for people to come on their own
- Drive around and see if there are homes in specific neighborhoods that look like they are in code violation and need repairs

Project Summary

Similar problem:

- Certain homes are abandoned. Either rentals that aren't currently filled or owned homes where no one lives.
- Abandoned properties arise for a number of different reasons.
- Poorly maintained abandoned properties are ripe for purchase by affordable housing orgs to provide affordable housing options.

Same story – can only know if people directly observe the property!

Project Summary

Question:

Given limited nonprofit resources to find homes in need of repair, is it possible to do so **algorithmically**?

Project Summary

Some work has been done in this area:

- Tax records
- Code violation complaints

Largely observational data that is only partially related to house condition.

Project Summary

Proposal: Could we use **overhead imaging of homes** to build a predictive model as to whether or not the home is in need of repairs?

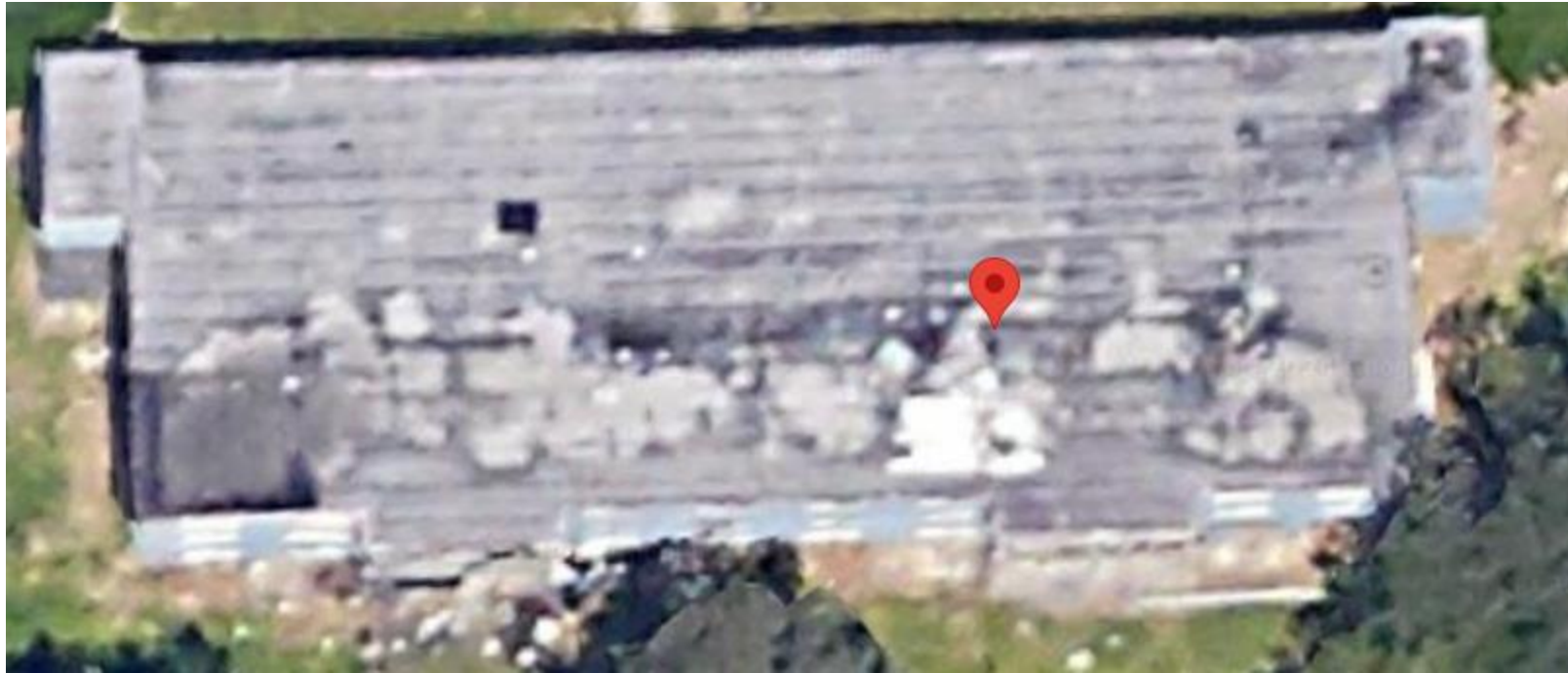
In theory, an image of a home could tell us if the home is in need of repairs or not.

- At least the roof, though roof condition is probably highly correlated with other problems...

Project Summary



Project Summary



Project Summary

Snapshots of overhead images taken 30m above the ground for the entire U.S. is available in yearly snapshots from a number of different organizations:

- Landsat via USGS and NASA
- Sentinel via the ESA

Project Summary

Higher Res imaging is done almost yearly by the National Agricultural Imagery Program

- Higher resolution overhead imaging
- Most recently available data from 2022

Project Summary

Easiest access to these images I've found is through **Google Earth Engine**

https://developers.google.com/earth-engine/datasets/catalog/USDA_NAIP_DOQQ

Can be accessed via Google's ee library in Python!

Project Summary

Free and licensed access for non-commercial research.

Can:

- Store derivative images
- Cache images
- Create databases with images
- Perform large scale analysis on images

Project Summary

Google Maps API has similar overhead images

- CANNOT STORE LOCALLY AND PERFORM IMAGE ANALYSIS USING THIS SOURCE!!!
- We don't want to fight Google...

Project Summary

In this class, you'll split into two groups that will work independently on these problems.

Merge work at a certain point.

Then diverge again after the initial database is created.

Project Summary

Step 1: Create a database of highest resolution images possible for all tax parcels in Fulton County, GA

- Start by working with tax parcel info to find **minimum bounding box** that includes each tax parcel (A rectangle around each parcel)

Project Summary

Step 1: Create a database of highest resolution images possible for all tax parcels in Fulton County, GA

- Develop a Python function that takes in a bounding box and returns the highest resolution image from an open source.
- Should theoretically work for counties beyond Fulton county

Project Summary

Step 1: Create a database of highest resolution images possible for all tax parcels in Fulton County, GA

- Use this function to get an image for each tax parcel in Fulton County, GA
- Create some sort of searchable database that includes:
 - Image
 - Address/Tax Parcel ID
 - Date image was taken
 - Residential/Nonresidential Zoned Parcel

Project Summary

Step 1: Create a database of highest resolution images possible for all tax parcels in Fulton County, GA

- Storage of this information will be a challenge. I'm exploring options to use QTM server as a central store for this info.
- At some point before the big scrape, the groups will merge their code into one super function and we'll just work with one image database

Project Summary

Step 2: Use images to perform analysis

- Actual question is up to the group.
- Some ideas

Project Summary

Step 2: Use images to perform analysis

- Create a predictive model that determines if a roof is in good or bad condition
- Use different data sources as labels for some of the images and develop classifier
- Use unsupervised methods to cluster roofs and see if we can locate bad condition roofs

Project Summary

Step 2: Use images to perform analysis

- Determine if property actually has a building!
- It is actually kinda tough to know which parcels are open and which aren't
- Train a classifier on roof/not roof to determine if we can see which ones are residentially zoned and may not actually have a building on it...

Project Summary

Step 2: Use images to perform analysis

- Use tax payment information and images to determine if a home has been abandoned
- No tax payment + Poor Upkeep could mean it isn't being used

Project Summary

Step 2: Use images to perform analysis

- Create better roof recognition pipeline
- Images contain roofs + other stuff
- Use image segmentation methods to extract roofs from images

Project Summary

Step 2: Use images to perform analysis

- There is **a lot** you could do here.
- All use cutting edge machine learning techniques with a large image data set
- To assist with computational burden in prototyping, I have Google Cloud credits for each of you. \$50 to start.

Class Schedule

Data Viz Assignment:

Before next Monday, try to use Python to get an image of the roof of your home here in Atlanta!

- Work together to figure this out on Wednesday and over the weekend

Class Schedule

Communication Slides:

Next Wednesday, each team should present 5-7 slides on

- Where collaborative code will be stored
- How you're going to communicate
- First tasks
- Division of labor

Class Schedule

Team Check-In presentations:

5-7 slides that talk about where you're at and any roadblocks you are hitting

Class Schedule

Before Spring Break:

- Minimum bounding boxes for Fulton County tax parcels
- A function that takes in a box and returns an image of the tax parcel **with documentation**
- A plan for organizing images in a searchable way

Class Schedule

During Spring Break:

- I'll try to run this function on the QTM server to get images ready and searchable after the break
- Depending on timing, may happen after Spring Break

Class Schedule

After Spring Break:

- Groups use images to address question of interest.
- Just try your best here
- Anything you produce here will be better than what the interested nonprofits can and have done!

Class Schedule

Deliverables from Each Group:

- 25 minute presentation covering work and potential impact on urban studies community
- A whitepaper that covers work and findings. No more than 12 pages (not including images and citations)
- A replicable code repository

Class Schedule

This is an ambitious project with potentially wide impact!

- Just do your best here.
- That's all we're asking from you.
- You're all really smart and well trained. Now, it's just putting it all together into a real project.

Class Schedule

Let's construct the groups.

- Paul and Lorna will be together.
- Who here has previous database construction experience?
- Who here has previous image analysis experience?
- Who here is comfortable with deep learning methods?

Class Schedule

Questions?