I. Predicting Wildfires

A. Questions

1. Can you predict wildfire locations in the future?
   a) Feed the model past wildfire locations and give it older satellite data so that it can recognize patterns in the variables that led the region to become a wildfire.
2. Predict age of wildfires in the past?
   a) Can you give the model a previously burnt area and can it make an educated guess on when the last wildfire took place? This would require knowing when past wildfires occurred.
3. Classify at risk species from satellite data?
   a) This will be harder to do, but in theory you can feed the model images (i.e., visual spectrum for the built environment, infrared for plants) along with other data, like elevation, to make a good estimate on what species are in a photo. Ground-truthing the data will ultimately be required, so this question is likely out of the scope of this course.

B. Process Used:

1. Classify satellite photos for dry forests and correlate weather data to teach the computer to make future (or past) predictions on wildfires.

C. Data Needed:

1. *Satellite photos*: Need to make sure that the photos are of high-enough resolution. This could limit what questions we can pursue. See here: https://earthexplorer.usgs.gov/
2. *LiDAR and elevation maps*: This could potentially help teach the model to answer some of the questions, as elevation or even point clouds are another data point to differentiate purely RGB orthophotos. See here: https://www.opentopography.org/
3. *Wildfire historical extents and dates*: This is key to predicting wildfires. We need to have a record of previous wildfires and where they started.
4. *Climate data*: Having a record of weather patterns, including temperature, wind, and precipitation, could potentially help the model understand what weather conditions help give rise to wildfires. See here: https://www.ncei.noaa.gov/cdo-web/datasets
5. *Powerline data*: This could be extracted from the images themselves, or from a separate layer. Old powerlines are known to cause wildfires.

D. Business Use Case/Potential Value:

1. Wildfires cause damage to property and to lives. In a warming world, I think that wildfires will increase. It's important to understand the interactions between the various data sources discussed above to see if we can help communities understand where wildfires are likely to occur next. There is a business case here, as a company can run these predictions and help communities prepare or escape. I want to focus on a GIS-based project. GIS (geographic information system) work can involve spatial, satellite-derived data and this project should be able to satisfy my interest in this topic.

II. Citibike Station Availability Predictor

A. Questions

1. Can you predict when stations will be full and when they will be empty?
    a) Users of Citibike are able to see if there are open bikes available to use. But what about knowing if they are, say, 30' away from their destination, will there be an opening for them to dock their bike?

B. Process Used:

1. Analyze usage data and then use machine learning to predict when stations are full and empty.

C. Data Needed:

1. *Citibike Dataset*: Citibike has available for download tens of millions of rows of usage data for us to feed into a model. The data is freely available, see here: https://citibikenyc.com/system-data
2. *Weather data, etc.*: Other datasets that could be pretty easily correlated to usage data, like weather, economy, health — things that might inform the model on when stations are empty or full.

D. Business Use Case/Potential Value:

1. This would be a very good service to have for users of Citibike. I also believe that it can also be applied to other realms, like rental cars, stores — places that have a stock of something that increases or decreases based on time. This also is a spatial analysis-focused project, since you have to factor in road distance from the station, not simply the shortest line between two points.

III. Global/Regional/Local Conflict Model

A. Questions

1. Do depressed economies, hotter temperatures, lower rainfall, and decreased health cause an increase in violence?
    a) By correlating various data sources together, along with historical conflicts that occurred within the timeframe of the datasources, we can train a model to predict the likelihood of future events.

B. Process Used:

1. Machine learning.

C. Data Needed:

1. *Satellite photos*: This might be helpful to put into context the size of the conflict.
2. *Economic data*: This is a key part of the model, to see how exactly depressed economies may influence the outbreak of conflicts. See here: https://tradingeconomics.com/
3. *Climate data*: Deeper problems may be dictating the outbreak of conflicts. Do hotter temperatures and lower rainfall contribute to famines or droughts, which cause unrest in a population?
4. *Health index*: Knowing the relative health of a population can be useful to round out the model and understand when conflicts are most likely to arise. See here: https://www.healthdata.org/institute-health-metrics-and-evaluation

D. Business Use Case/Potential Value:

1. As we have tragically come to learn with Putin's war in Ukraine, the world does not always enjoy peace all the time. Understanding when conflicts are likely to flare can be helpful for businesses and governments.

IV. Flooding and Property Values/Tax Base Impact in NYC/Region

A. Questions

1. How does flooding affect property values and the tax base?
    a) Could we predict where flooding will most likely occur, and then also have a model tell us what the economic impacts of the flooding will have on the community?

B. Process Used:

1. Spatial analysis and machine learning.

C. Data Needed:

1. *Satellite photos*: In order to understand where flooding is likely to occur, you need to have the ground elevation and the sea level.
2. *Building data*: We need to have a layer that shows where all the buildings are, so that we can see which buildings will get affected by rising sea levels. See here: https://www.nyc.gov/site/buildings/dob/find-building-data.page
3. *Economic data*: This may help to put the local economic trends in a wider context.
4. *Climate data*:Instead or in addition to sea level rise, we can see the effect that heavy rainfall has on the city. Where is water most likely to accumulate?

D. Business Use Case/Potential Value:

1. Flooding will increasingly become a problem for coastal cities the world over. Constructing a model of the economic impact of the flooding will help policymakers and businesses plan for a warmer and wetter world in the future.