Important Notes:

- OccuSpytial looks to be a python package for occupancy modeling. They say 'bayesian inference of single-season site occupancy models,' and occupancy modeling with 'spatial autocorrelation between neighboring sites for the occupancy covariates' (I guess that's where the bayesian inference comes in). Could be a good option for a more sophisticated model.
- Link here https://masonfidino.com/bayesian_integrated_model/ shows a demo for a model described as follows:
    - *"The Koshkina et al. (2017) integrated SDM, which was based on the model in Dorazio (2014), uses an inhomogeneous Poisson process model for the latent state. A Poisson process is just a mathematical object (e.g., a square with an x and y axis) in which points are randomly located. An inhomogeneous Poisson process, however, means the density of points on this object depends on your location on the object. Ecologically, this means that there is some region in geographic space (e.g., the city of Chicago, Illinois) and the modeled species abundance in that region varies across space."*
    - I guess this is spatial variation of a different kind than the first bullet point. The first is spatial autocorrelation of the covariates whereas this is a variation of species abundance itself. Maybe another option we could incorporate (unless we don't care about species abundance but rather simple presence/absence). Doesn't look to be available in a python package. The link gives sort of a working demo on how to hard code it (in R), but I don't think it would be too bad to translate to python.
- Another link with some explanatory background and demos https://kevintshoemaker.github.io/NRES-746/Occupancy.html#Occupancy_Models. The examples are in R but maybe could still prove useful, especially if there are things we have to hard code in Python.
-

Occupancy models - useful because they take into account both the observation of a species in a particular location and the probability that we detect the species there. That is, they accommodate imperfect detection and so give a more reliable estimate of species presence, numbers, etc (I'm seeing the 'occupancy' sometimes referred to as 'latent state').

Questions that occupancy models allow us to answer (taken from this nice source: https://kevintshoemaker.github.io/NRES-746/Occupancy.html)

- Where does the species occur (species distribution)?
- What influences where a species is found?
    - These influencing factors can be incorporated into the occupancy model as 'covariates.'

Assumptions of Occupancy models:

- Area being studied is closed to changes in species occupancy between sampling times

- Surveys and units are independent (i.e., whether or not a species was previously identified at a site has no impact on the current survey detection at that site).
- No false detections (observer believes they saw the species of interest, but it was actually a different one)
- Probability of species detection, given the species was present, is constant across sites (can be relaxed if incorporated into the model)
- Probability of occupancy of the species is constant across sites (also can be relaxed if incorporated into the model).

Occupancy Model Statement:
Need Bernoulli distributions (binomial distributions with only one trial, i.e. only two outcomes for a single event - success with probability p (and n = 1) and failure with probability 1-p (and n = 0)),

We can write occupancy model as: $y_i|z_i \sim \text{Bernoulli}(p * z_i)$

$z_i \sim \text{Bernoulli}(\psi)$

$\text{logit}(p) = \alpha_0 + \alpha_1 * \text{covariate}_1$

$\text{logit}(\psi) = \beta_0 + \beta_1 * \text{covariate}_1$

Where:
y_i = data at site
p = detection probability
z = true occupancy state at site i
ψ = occupancy probability
α = parameters to estimate for the detection probability p
β = parameters to estimate for the occupancy probability ψ

Gonna write this out in words so it helps me:

Data at site i | true occupancy state at site i ~ a Bernoulli distribution with probability of success given by (detection probability * true occupancy state)
While true occupancy state ~ a Bernoulli distribution with probability of success given by occupancy probability

Logit function is defined as the inverse of the standard logistic function. We define $logit(p) = \ln\frac{p}{1-p}$ for $p \in (0,1)$. Logit is also called the log odds function (where the odds are $\frac{p}{q} = \frac{p}{1-p}$). So the statements

$\text{logit}(p) = \alpha_0 + \alpha_1 * \text{covariate}_1$

$\text{logit}(\psi) = \beta_0 + \beta_1 * \text{covariate}_1$

Say that we're modeling the odds of the detection probability with some linear model and modeling the odds of the occupancy probability with some other linear model.

We call occupancy models *hierarchical* since the state we want to know about (latent state, z) is present both in the data model (the conditional distribution y|z) and in the latent process model (the Bernoulli distributions for the z_i themselves). So we need to know about both the data and the detection probability in order to learn about the occupancy state and occupancy probability. The logits are *link* functions - they link the deterministic and stochastic processes for the detection and occupancy probabilities.

Here's the tree of events unfolding in an occupancy model/data collection process: