Section 3

Video: Suitability modeling in data science

Print

Time	Caption
0:00	♪ [music] ♪
0:15	- [Shannon] Hi, everyone, Shannon and Lauren here.
0:18	- [Lauren] Hi. - [Shannon] Welcome back
0:19	to another week in the Spatial Data Science MOOC,
0:22	and this week is all about suitability.
0:24	- [Lauren] Now suitability analysis is a key approach to a foundational
0:29	spatial problem that people have been trying to solve forever.
0:31	Where's the best, the most suitable location for a new store,
0:36	for a conservation area, for a solar farm?
0:39	Really anything and everything you can think of,
0:41	which makes it a key aspect of spatial data science.
0:45	- [Shannon] So let's dig into what we mean by suitability, and I think
0:47	probably the best way for us to do that is with some examples.
0:51	- [Lauren] Definitely. Also, though, we should remember
0:53	that, like most complex problems,
0:55	there are a lot of different ways to solve this problem.
0:58	- [Shannon] Now one approach to solving this type of analysis
1:00	is to evaluate suitability based on a set of criteria.
1:04	So, for example, if we were all a real estate development company
1:08	looking for the perfect site for our next project,
1:11	some of the criteria we might consider would be based on
1:14	the surface of the earth, the topography.
1:15	What does the slope of the site look like?

1:18	Is it relatively flat, or are we building on the side of a mountain?
1:21	I mean, that would impact the ultimate cost of our development.
1:23	- [Lauren] It would. - [Shannon] Or we might look at
1:26	some environmental factors, such as what does
1:28	the land cover on a site look like?
1:30	Or are there any endangered species that we need to be aware of,
1:34	because that would impact whether or not we can develop at all.
1:37	- [Lauren] Now we'd also definitely want to
1:39	take into account some human factors, right?
1:41	So things like how far do people have to travel from work?
1:44	How many entertainment or retail venues are there in proximity?
1:48	- [Shannon] Now, despite how different these datasets might be,
1:51	we want to use them all in the suitability model.
1:53	And how we do that depends on how each dataset relates to our objective.
1:56	So, for instance, let's take that distance to retail.
2:00	For this particular project, let's assume the closer we are to retail,
2:03	the better the site, and the further we are,
2:05	the less attractive the site is for us.
2:08	We want to capture that relationship within our model.
2:10	And this part of the process is called reclassification,
2:13	because we're translating that raw data,
2:15	whether it's a distance, a slope, a land cover, you name it,
2:19	and we're converting that into a suitability that's based on
2:23	a scale that's shared across all of the datasets in our analysis.
2:27	- [Lauren] Now that reclassification step is really
2:30	one of the most important steps in the suitability analysis process.
2:33	That's really where we are deciding
2:35	how each criteria fits into our model.

2:38	Once we've reclassified, we also have the ability
2:41	to weight each of those criteria.
2:43	So we can take a criterion that's particularly important to us,
2:45	give it a higher weight, and it'll have a bigger impact
2:48	on our resulting suitability surface.
2:51	- [Shannon] Once we have reclassified and weighted
2:53	all of the datasets in our analysis, we combine them together to create
2:56	a surface where we can calculate the overall suitability
3:00	of all of the locations in our study.
3:02	And this weighted overlay approach is
3:05	a really common spatial analysis technique.
3:07	- [Lauren] Now another approach is more probability-based.
3:10	So let's take habitat modeling as an example, where we've collected
3:14	data on where a particular species has been seen.
3:18	We can model where we expect to see that species
3:22	using machine learning or regression analysis
3:25	or some other modeling technique to predict
3:28	where we expect to see that species
3:31	based on a series of underlying variables,
3:33	things like temperature, elevation,
3:35	distance from water sources, and others.
3:38	Now, in this approach, we really let the data speak for itself
3:42	and hope that it's representative.
3:44	- [Shannon] Now, whether we're doing a more weighted overlay approach
3:47	or a more data-driven probability-based one,
3:49	our end goal is the same.
3:51	We want to prioritize all possible locations
3:53	and then use that prioritization or ranking to be able to help us

3:57	make more informed decisions.
3:59	- [Lauren] Now, this week, we'll be focused on that
4:01	weighted overlay approach, not only because it's really
4:03	bread-and-butter spatial analysis, but because
4:06	it's such a key aspect of spatial data science that
4:09	every spatial data scientist needs it in their toolbox.
4:13	