Intro to Spatial Data Analytics

**Dr. Dave Wanik - University of Connecticut**

Simply stated - if your data has a spatial component (i.e. an address or a lat/lon coordinate) and you aren’t leveraging US Census data into your models, you are leaving money on the table! Come to this workshop and leverage the power of ‘where’ in your models. Topics include how to geocode addresses (from street address to lat/lon); how to do a ‘spatial join’ of points (lat/lon) to polygons (Census tracts and zip code boundaries) and extract hundreds of socioeconomic variables; how to identify the nearest features to a point of interest (nearest airport to all houses in Connecticut); and how to combine ALL OF THIS information into an end-to-end machine learning model that predicts house prices as a function of the newly-derived features you have created.

All Python codes will be provided to students so they can follow along and practice on their own. No prior Python experience needed – ***just listen, watch and enjoy*** the concepts we will demonstrate in this workshop.

You need to use your Google Drive and Colaboratory in order to run these scripts.

<https://colab.research.google.com>

Scripts

1. Basic EDA notebook (using Boston Housing - pre-req EDA materials)

<https://drive.google.com/file/d/1XtMx10gmMDpY-mqECuhmeP0_w-6OUke3/view?usp=sharing>

1. Intro to Geospatial Data Types: <https://colab.research.google.com/drive/1ptyAIUXtqjf5DuPXlT1chd5wQXZbFDyb?usp=sharing>
2. Geocoding and Spatial Join: <https://colab.research.google.com/drive/1F8nx5I3ezEooTzf-UTrkVmqOW_-4e54k?usp=sharing>
3. Nearest Neighbor: <https://colab.research.google.com/drive/1xrYQlvjnP1FZwAvo-WnzNGScfCIN0_FR?usp=sharing>
4. Data Enrichment: <https://colab.research.google.com/drive/19liyOvZt6ZtOb9mnrx1S5Bvm5sDYEsda?usp=sharing>
5. Build a Model: <https://colab.research.google.com/drive/1X3gO3jUcVy3CHXTFDneTidViRE8cCJSu?usp=sharing>

Data

* Real Estate (Real\_Estate\_Sales\_2014-2016.csv)
  + <https://drive.google.com/file/d/1AqU0lrw0DupBcG8poZPWN7UGFFoR4IUJ/view?usp=sharing>
* Zip Code Shapefile (tl\_2017\_us\_zcta510.zip)
  + <https://drive.google.com/file/d/1V4nDKJRK-7d5gg_MrMJ4-8XQckSCZiKz/view?usp=sharing>
* ZCTA5 Census Data (ZCTA5\_censusData.csv)
  + <https://drive.google.com/file/d/1e9BTlsm4BszynWrzFMoI07qCGQImnZVS/view?usp=sharing>

See you soon! And remember… it’s all about layers...

