School of Computer Science
Uwe Roehm

DATA2001/2901: Data Science, Big Data, and Data Diversity **Semester 1, 2021**

Tutorial Week 9: GeoData

In this week's tutorial, we will look at how to do integrate geo-location data into an existing data set in order to do a spatial join between two datasets. As part of this weeks tutorial we will also look at some requirements for the assignment

System Requirements You may use the following tools for this tutorial:

- The University's Cisco VPN client to access our internal Jupyter notebook server and the FortiClient VPN if you are accessing from China.
- A browser to access the Jupyter notebook server:
  - https://ucpu0.ug.it.usyd.edu.au/
- We are using two new python libraries today:
  - geopandas
  - shapely
- we will also be using the PostGreSQL server so make sure you have access to that from the previous weeks. If you are using your own PostGreSQL server - make sure you have the geodata extension PostGIS v3.0+ installed and enabled (http://postgis.net/install/).
  - Note: We will be using our credentials file to connect to PostGreSQL - you may reuse your credentials file from previous weeks. Remember to place it into the ./data folder of your working directory

Common Errors The certificates on the Jupyter notebook servers are acting up so you may get some 'unsafe certificate' warnings. You can bypass these by clicking on the 'Advanced' or 'More information' option that pops up and selecting the 'continue anyway' option.

Make sure to close your database connections!

Exercise 1. Finalise Groups for the Project

As part of the assessment for this course, you will have a group project. Since group-work is a core part of the assessment, you can NOT do this assessment alone. You must form groups of 2-3 students for this assessment. You can form groups within the combined tutorial timeslot that you have been allocated.

Advanced students should only form groups with other advanced students (as you have a different assessment rubric and an extended assignment tasklist)

By now you should have a group - If you do not have a group please mention it in class and the tutors will attempt to pair you up.

Please note that group participation will be measured in the assignment demo in week 12/13. Those that have not contributed adequately may find their mark adjusted.

Exercise 2. Demo Jupyter notebook explaining Geometry objects and GeoPandas

The first portion of the tutorial will involve going through the demo file which looks at how Geometry objects work to represent shapes and geography and how spatial joins are carried

Exercise 3. Tutorial jupyter notebook on GeoPandas and PostGIS

This notebook will work through using geopandas and PostGIS to interact with geometry data and answer a few straight forward questions.

Exercise 4. Homework

Try to complete the associated assignment based tasks with your group.

We will be releasing the datasets and assignment specs soon.