



תרגיל בית 1 – סדנת פיתוח בממג

Python & GIS

מטרת התרגיל

פיתוח ותרגול שאילתות GIS וניתוח גאוגרפי בסביבת Python

Section A (20 points)

Goal

Practice how to create geometric objects using Shapely module and how to find out different useful attributes from those geometries

Exercise

Our dataset **travel_times.csv** consist of travel times between specific locations in Helsinki Region.

Thus, we have many columns of data, but the few important ones are:

Column	Description
from_x, to_x	x-coordinate of the origin/ destination location (longitude)
from_y, to_y	y-coordinate of the origin/ destination location (latitude)
total_route_time	Travel time with public transportation at the route

Tasks

1. Read 4 columns, i.e. 'from_x', 'from_y', 'to_x', 'to_y' from the data into Python using Pandas.
2. Create two lists called `orig_points` and `dest_points`
3. Iterate over the rows of your DataFrame and add Shapely Point - objects into the `orig_points` -list and `dest_point` -list representing the origin locations and destination locations accordingly.
4. Compute Movements:
 1. Create a list called `lines`
 2. Iterate over the origin and destination lists and create a Shapely



LineString -object between the origin and destination point

Column	Description
Lat	y-coordinate of the post
Lon	x-coordinate of the post

3. Add that line into the lines -list. Find out what is the average distance of all the origin-destination LineStrings that we just created and print it out.

Section B (45 points)

Goal

how to create geometries in Geopandas and how-to re-project data and do some basic geometric calculations.

Exercise

The problem is creating geometric point -objects and putting them into a map. Here our aim is to plot a set of x and y coordinates that we should read from a some_posts.csv comma separated file that contains following kind of data:

```
lat,lon,timestamp,userid
-24.980792492,31.484633302,2015-07-07 03:02,66487960
-25.499224667,31.508905612,2015-07-07 03:18,65281761
-24.342578456,30.930866066,2015-03-07 03:38,90916112
-24.85461393,31.519718439,2015-10-07 05:04,37959089
```

The data has 81379 rows and consists of locations and times of social media posts inside Kruger national park in South Africa:

Note: although the data is based on real social media data, it is heavily anonymized. Userids and timestamps have been randomized, i.e. they do not match with real ones, also spatial accuracy of the data have been lowered.

Tasks

Part 1:

- Read the data into memory using Pandas
- Create an empty column called geometry where you will store shapely Point objects
- Iterate over the rows of the DataFrame and insert Point objects into column geometry (you might use .loc indexer to update the row).
- Convert that DataFrame into a GeoDataFrame, [see hints](#)



- Update the CRS for coordinate system as WGS84 (i.e. epsg code: 4326)
- Output: Save the data into a Shapefile called Kruger_posts.shp
- Create a simple map of those points using a GIS software and use `.plot()` -function in Python.
 - Recommendation: please check and verify that your plot is on the right place on the map. (please add a screen shot of your results overlay on a map)

Part 2: How long-distance individuals have travelled?

In this problem the aim is to calculate the distance in meters that the individuals have travelled according the social media posts (Euclidian distances between points).

In your code you should:

- Reproject the data from WGS84 projection into EPSG:32735 -projection which stands for UTM Zone 35S (UTM zone for South Africa) to transform the data into metric system.
- Group the data by userid
- Create an empty GeoDataFrame called movements
- For each user:
 - [sort](#) the rows by timestamp
 - create LineString objects based on the points
 - [add](#) the geometry and the userid into the GeoDataFrame you created in the last step
- Determine the CRS of the movements GeoDataFrame to EPSG:32735 (epsg code: 32735)
- Calculate the lengths of the lines into a new column called distance in movements GeoDataFrame.
- Save the movements of into a Shapefile called Some_movements.shp
- Create a simple map of those points using a GIS software and use `.plot()` -function in Python.

Questions

Write your answers below the questions. You should also print in your code the answers to the questions.

- What was the shortest distance travelled in meters?
- What was the average distance travelled in meters?
- What was the maximum distance travelled in meters?



Section C (35 points)

In this section we will study how to do geocoding, spatial joins and nearest neighbour analysis in Geopandas.

Tasks

What are the closest bus stops from your home& work?

- Create a txt-file called activity_locations.txt (use the same formatting as in Problem 1) with two columns:
 - id --> unique number (e.g. 0 and 1)
 - addr --> address of your work and home (you don't need to reveal your home/work if you don't want to, these can be whatever two addresses from Israel!)
- Read those addresses into Pandas and convert the addresses to Point objects using the geocoding functionalities of Geopandas
- Find out the five nearest bus stops (Bus_Stops.shp) to these points using the techniques shown during the lesson.
- What are the locations names of the bus stops? (use geocoding)- please add it to the GeoDataFrame of the bus stops.
- What are the average **Euclidean** distances from bus stops to your home & work?
- What is the closest bus station to the address? What is the distance between the address to the bus station?
 - Join the information from the closest bus stops to your address
 - Save and export the address & bus station information into a Shapefile called My_address_bus_stops.shp



הערות ואופן ההגשה:

- מועד ההגשה הוא 5.11.2020
- ההגשה היא **בזוגות**.
- יש להגיש דו"ח בפורמט מחברת (הכולל הסבר על הקוד, קלטים, פלטים, תהליך פיתרון)
- יש להוסיף לתשובות בכל אחת מהשאלות הסבר ומסקנות.
- תיעוד פנימי: יש לתעד את הקוד באנגלית: יש לתעד כל מחלקה, מהן המתודות והפרמטרים שלה. בנוסף, יש לתעד כל פונקציה.
- במקרים בהם על המשתמש להכניס נתונים: נתוב ללכתיבה\קריאה קובץ, נדרש להסביר טרם ההרצה. (הרצות שיכשלו, לא יבדקו)
- סיבוכיות מינימלית עבור כל אלגוריתם.
- אם נעשה שימוש במודול שלא למדנו בשיעורים יש להזכיר בתיעוד החיצוני ולהסביר אותו.
- קוד שאינו מתועד/ רץ לא ייבדק.
- שימו לב כי חשד להעתקה יוביל לציון אפס.
- יש להגיש למודל קובץ zip עם השמות של המגישים המכיל את קובץ התיעוד עם קבצי הקודים.

בהצלחה!