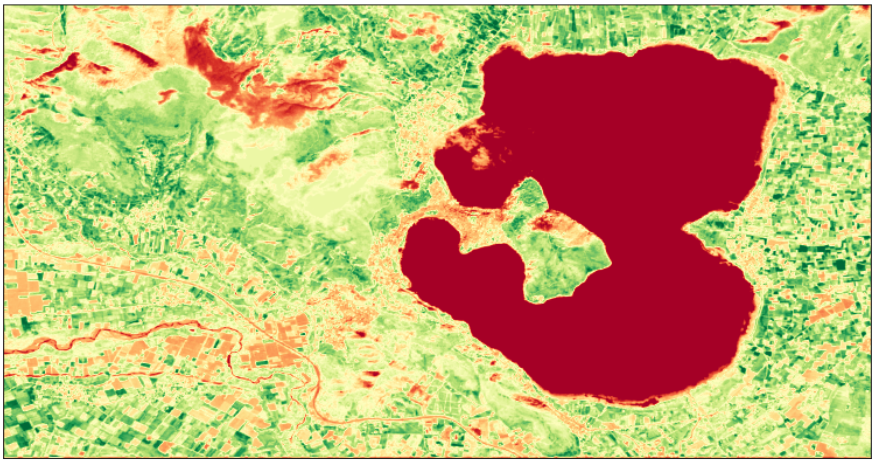
NATIONAL TECHNICAL UNIVERSITY OF ATHENS

MSC DATA SCIENCE AND MACHINE LEARNING

Geospatial Big Data Analytics
Project 2



Pavlos Kalfantis
MSc Student
pavloskalfantis@mail.ntua.gr
Student ID: 03400134

April 2022

Abstract

The scope of the second semester project of the graduate course 'Geospatial Big Data Analytics' was getting familiar with the different types of geospatial data and the different ways of getting and analyzing them with the Python programming language. Different types of raster and vector data, as well as time series data were collected through web services of the Open Geospatial Consortium (OGC) and other sources. The differences of Web Map Service (WMS) and Web Feature Service (WFS) were studied. Many different python libraries that are used for analyzing and visualizing geospatial data were explored, such as Geopandas, folium, rasterio and leafmap and both static and interactive maps were created on a Jupyter notebook environment. The notebook containing the python code for this project is included on the submitted documents.

Question 1

For the first question of the project, raster, vector and time series geospatial data was collected through various web services. Firstly, Sentinel-2 raster images of a large geographical area around Kastoria, Greece were loaded in the notebook from an open dataset. This dataset contained 10 different bands or indices and 24 different dates in 2016, for a total of 240 images. The first of those images, showing the Blue (B2) band for the first date, is shown on figure 1. The dataset also contained a file with land usage classes of the dataset, shown on figure 2.

Next, a geoJSON vector file with a polygon within the larger geographical area was created on https://geojson.io/. This vector file would be later used to mask the raster images on the next question of the project. The area that was selected is shown on figure 3 and contains the city of Kastoria, the adjacent lake, mountain and certain nearby suburbs. Furthermore, through this NASA website, meteorological data for a certain point for years 2016 to 2020 were created in csv format. More specifically, temperature, humidity and surface pressure time series data were collected, for the point shown on figure 4.

Finally, we were asked to use the Web Feature Service (WFS) to collect data from Corine Land Cover 2018. CORINE Land Cover (CLC) provides information on the biophysical characteristics of the Earth's surface. However, this service could not be located on the website and we use the Web Map Service (WMS) instead, which includes both raster and vector data. The WFS and WMS have been defined by the Open Geospatial Consortium (OGC) as methods for accessing geographic data. The WFS is concerned with direct access to the data – reading, writing, and updating your features. The WMS is concerned with transforming your data into a map (image).

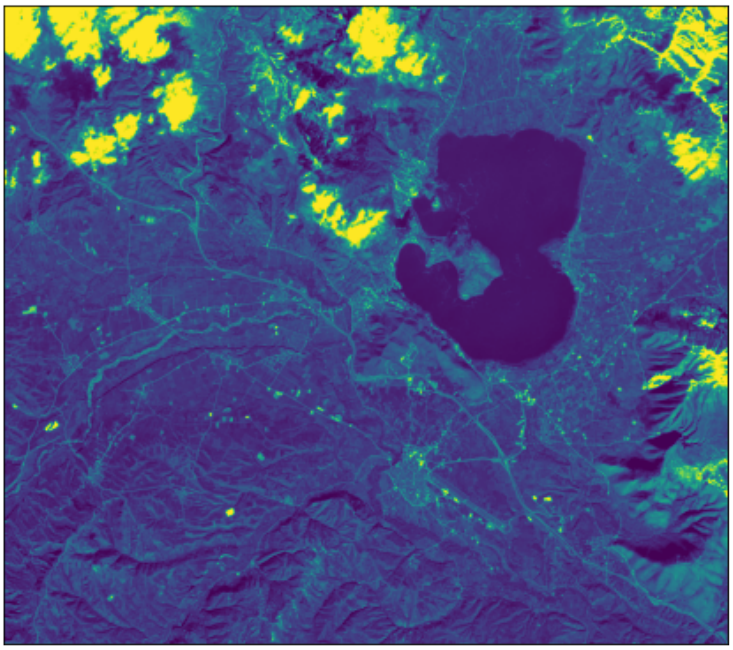


Figure 1: Plot of First Band of Kastoria Dataset

Question 2

For the second question of the project, raster images were plotted for different bands or indices. Firstly, the red, green and blue bands (B4, B3 and B2 respectively) for the first available date of our image dataset (2016-01-25) is presented on figure 5. Secondly, we were asked to mask the original dataset with the vector file we created and showed on figure 3. In order to do that, the original dataset had to be reprojected on the same Coordinate Reference System (crs). After our raster and vector data share the same crs, the raster data can be masked using the rasterio python library.

Next, for the clipped dataset, we are showing the Normalized Difference Vegetation Index (NDVI) for each available date on figure 7, shown on a Red-Yellow-Green colormap, from the smallest to the largest NDVI value. In addition, for each of those available dates, the mean NDVI of all pixels of our image is calculated and plotted as a time series plot, shown on figure 8.

Question 3

For the third question of the project, the geospatial python library Geopandas was used to answer geographical queries on our dataset, as well as build static maps. We get the vector data that is mainly used on this question from the Greek Government's geospatial data website, using their Web Feature Service. Then, different layers of this dataset

Reference Data

Kast_RefData_26Classes.tif

Class code	Class Description
1	DENSE URBAN FABRIC
2	SPARCE URBAN FABRIC
3	INDUSTRIAL, COMMERCIAL, TRANSPORT UNITS
4	Road/Asphalt Networks
5	PHOTOVOLTAIC UNITS
6	BROADLEAVED FOREST
7	CONIFEROUS FOREST
8	NATURAL FRASSLANDS
9	SPARCE SCLEROPHYLLOUS VEGETATION
10	Sparcely vegetated areas
11	VINEYARDS
12	FRUIT TREES
13	WHEAT
14	OTHER CEREALS
15	MAIZE
16	CHICKPEA
17	LENTIL
18	POTATO
19	BEANS
20	OTHER VEGETABLES (
21	CLOVER
22	ALFALFA
23	Beaches, dunes, sands
24	Bare rocks
25	INLAND MARSHES
26	WATER BODIES

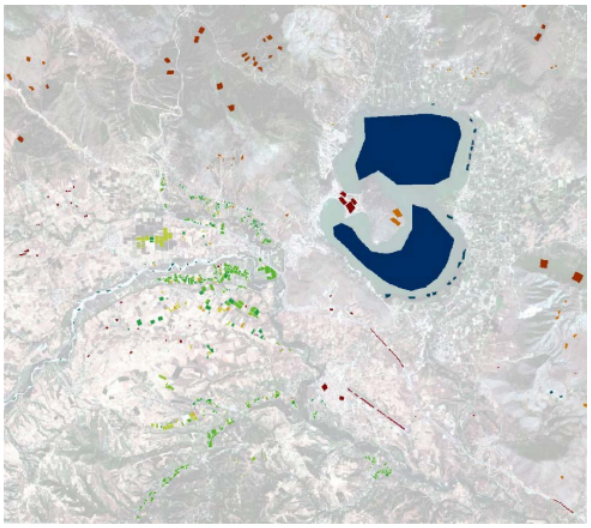


Figure 2: Classes of Kastoria Dataset

are joined using Geopandas' Spatial Join to answer questions, for example how many airports are located within each region of Greece. Examples of maps that have created by data from the different Greek government's geospatial data layers are shown on figure 9.

Next, by using the Web Map Service of Corine Land Cover 2018, static maps for Greece and Kastoria were created, showing the land usage of this dataset. The raster images as shown on figures 10 and 11 respectively, while the mapping of colors to land usage is shown on 12. Next step was to create a static map that shows the point for which we downloaded time series data, shown on figure 13. Finally, we plot the time series data that were downloaded (as shown on figure 4), with the resulting plots shown on figures 15, 16 and 17.

Question 4

The last question of this project was about creating interactive geographical maps, using the data that has been used so far. So both raster and vector data that was are plotted in maps that provide some type of functionality to the user and convey useful information. The python library that is most suitable for this type of analysis is leafmap, and

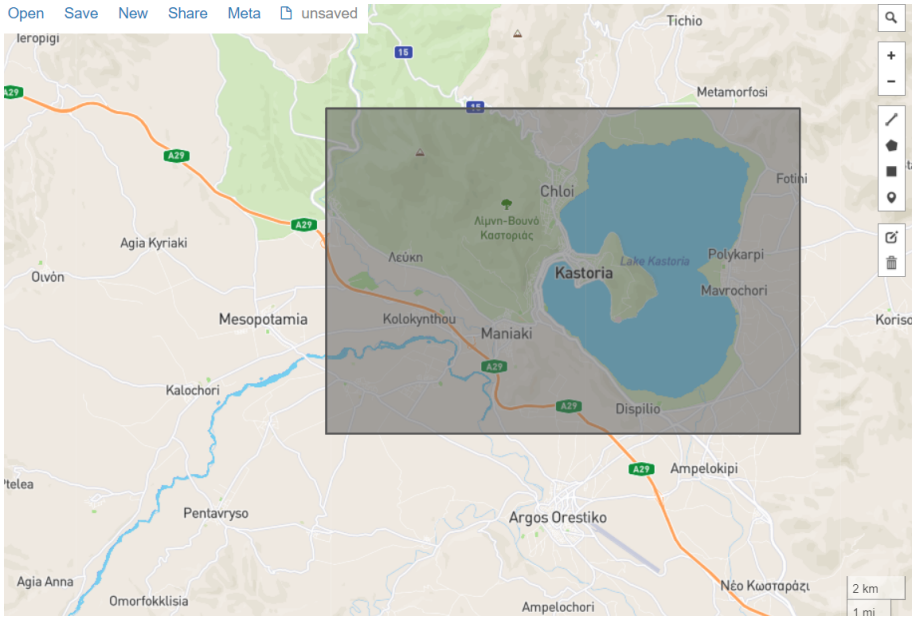


Figure 3: Creating Vector file within the Region of Interest

the resulting maps can be found on the Jupyter notebook, where the live functionality of the maps can be fully experienced.

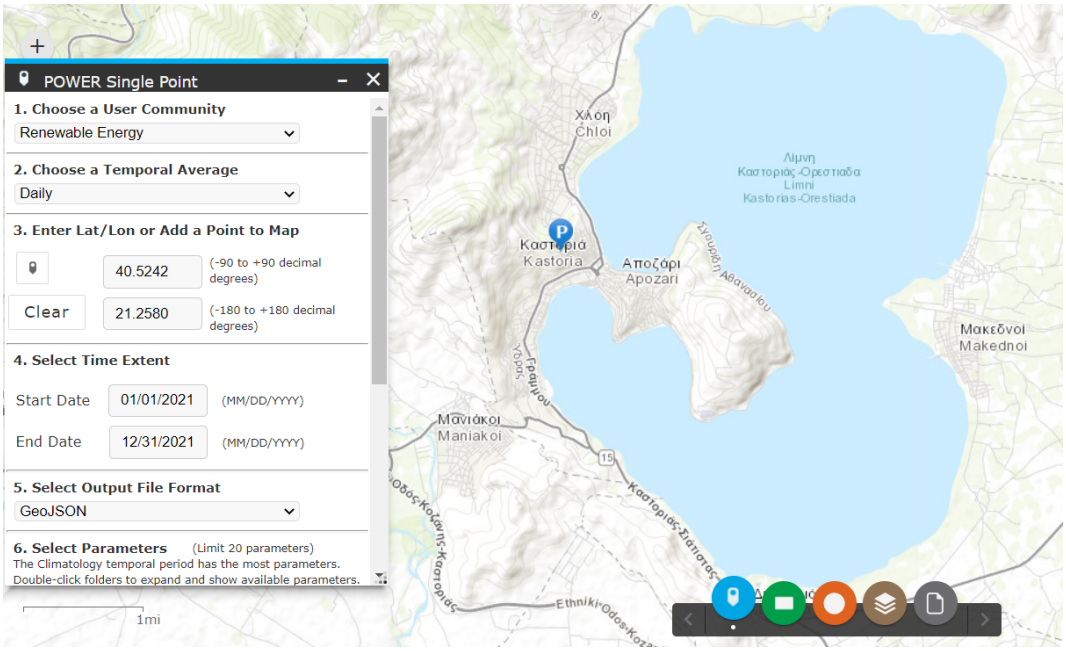


Figure 4: Getting Meteorological Data

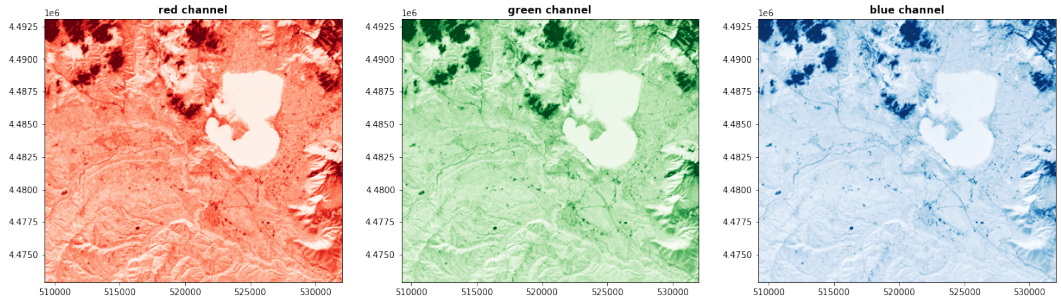


Figure 5: Plotting of Reg, Green and Blue Bands on first available date

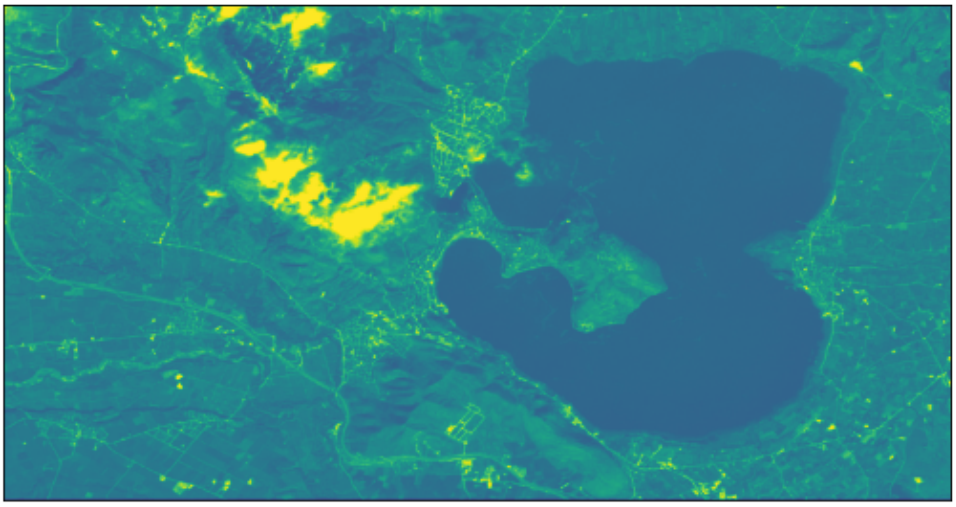


Figure 6: First Band of Clipped Image

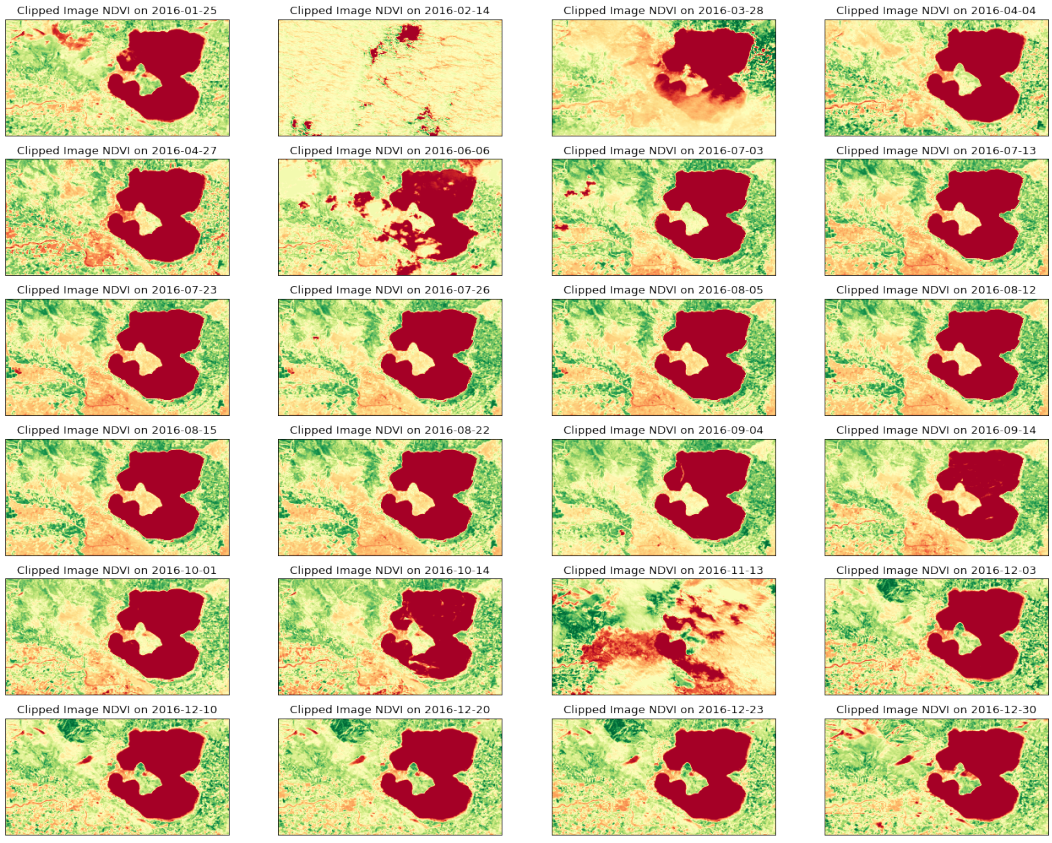


Figure 7: Images showing NDVI on available dates

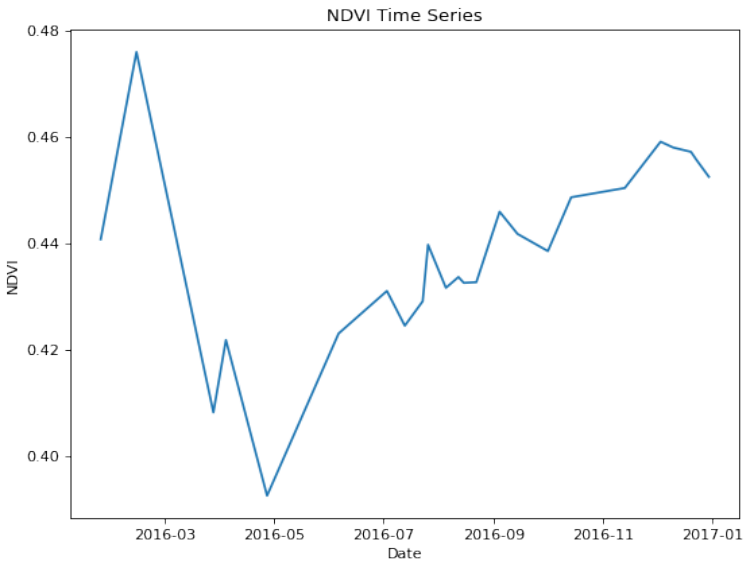


Figure 8: Mean NDVI Time Series

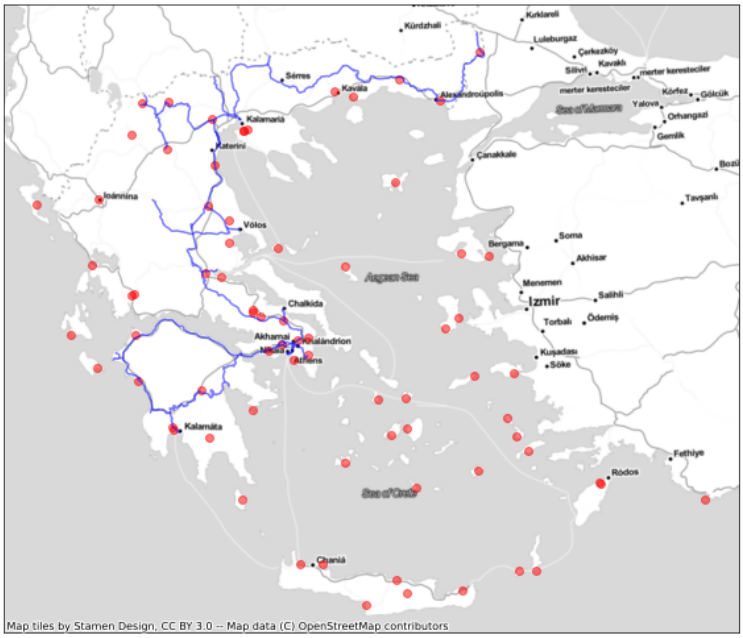


Figure 9: Railway Network and Airports of Greece

Figure 10: Corine Land Cover 2018 raster image for Greece

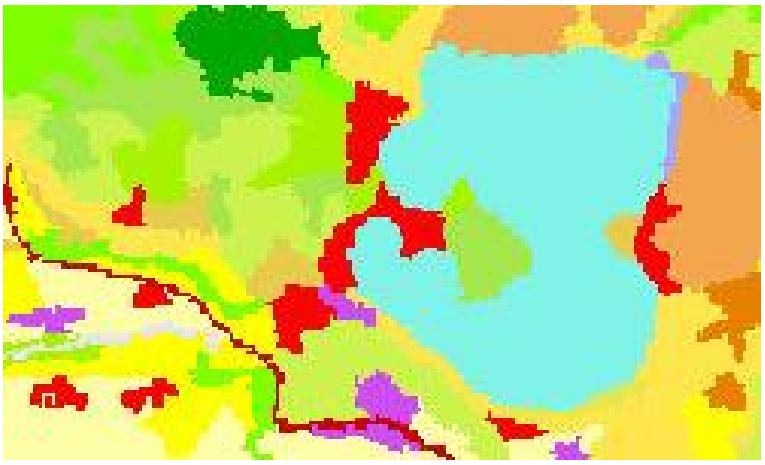


Figure 11: Corine Land Cover 2018 raster image for Kastoria

Figure 12: Corine Land Cover 2018 Classes

Figure 13: Clipped Kastoria Image and Selected Point

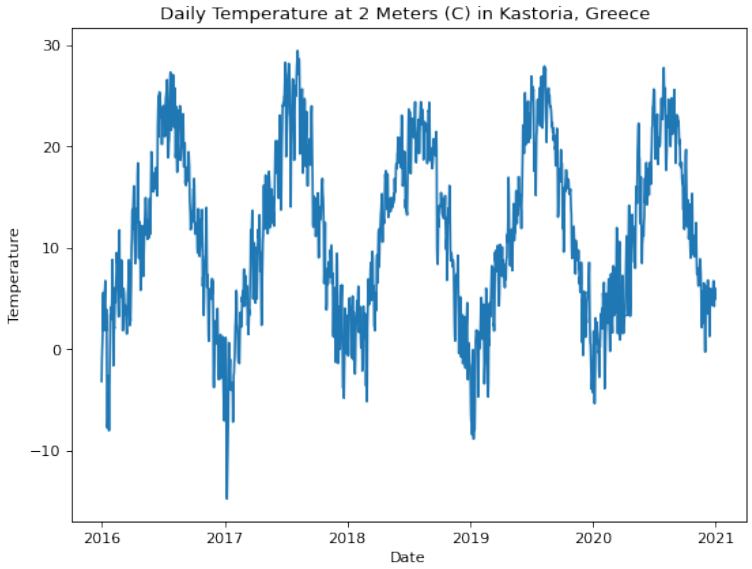


Figure 14: Time Series Temperature Data

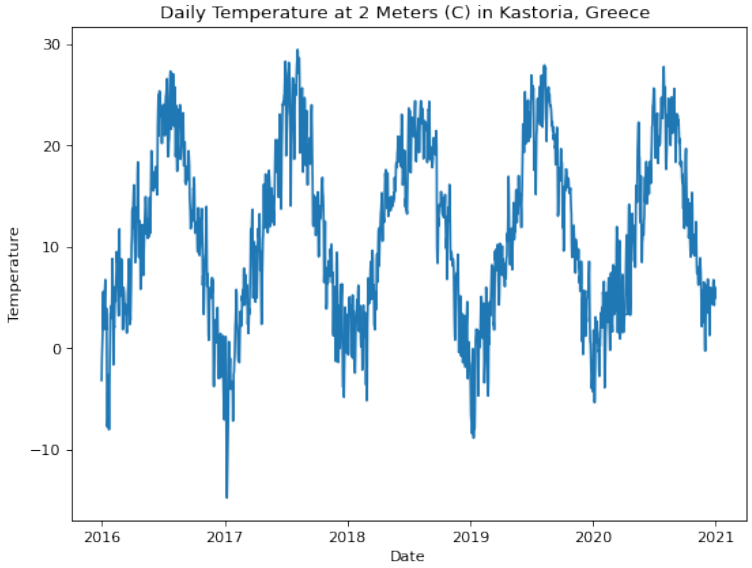


Figure 15: Time Series Temperature Data

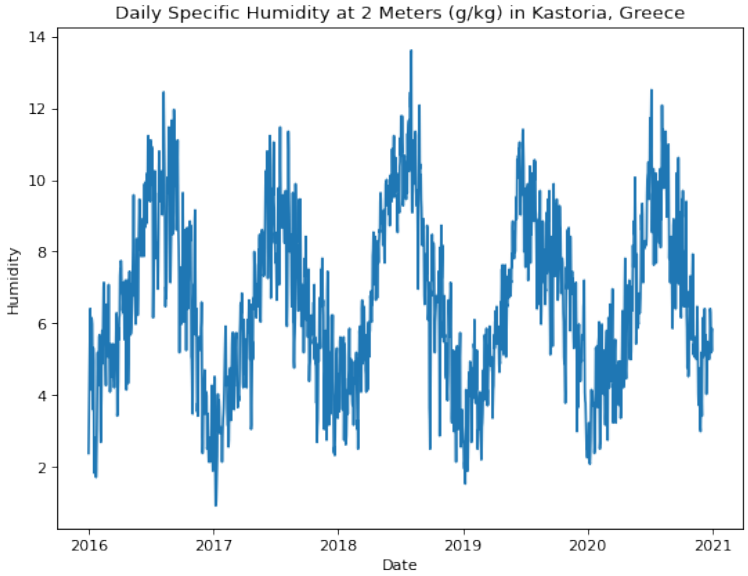


Figure 16: Time Series Humidity Data

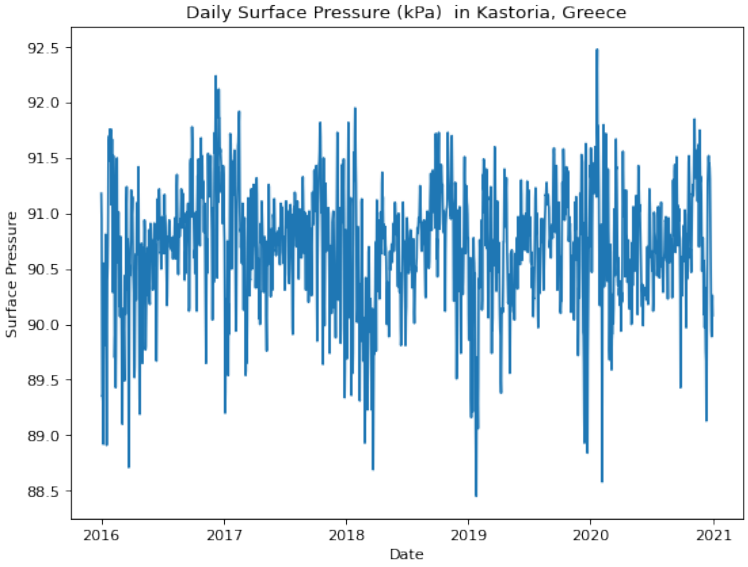


Figure 17: Time Series Pressure Data