

Time Series Data

1 Introduction

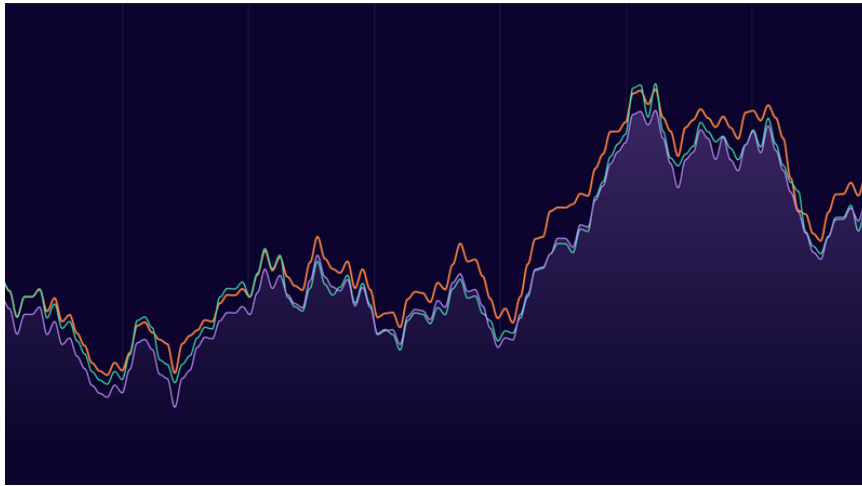


Figure 1: webflow.com

Introduction

In this lab session, you will work with electricity data from the [Open Power System Data \(OPSD\)](#). The data set includes daily totals of power consumption and production (solar and wind) in gigawatt hours (GWh) from 2006-2017 for Germany.

You can find more information about the data set as well as an in-depth tutorial on how to work with the data set in a blog by Jennifer Walker: [Time Series Analysis with Pandas \(Tutorial\)](#). Also have a look at [Working with Time Series in Python \(Tutorial\)](#) if you're interested in further reading on the topic.

Getting Started

Pandas offers the `TimeStamp` class as well as the `to_datetime()` function to handle temporal data. Dive into working with time series in pandas by using the example data set containing energy data of Germany. To begin, complete the following steps:

- Read the `opsd_german_daily.csv` into a `DataFrame` and call its `head()` and `describe()` methods.
- Check the column's datatypes. With what you just learned about `TimeStamps`, can you identify the steps you need to take to make the `Date` column usable in time series plots?

2 Lab Assignments

Important: Before you start with the actual assignments, please ensure that you set the index of your DataFrame to the Date column. Also, add new columns containing the year, month (January-December) and weekday (Monday-Sunday) for each row in the dataframe. The quickest way to do this correctly is by relying on the methods provided by pandas' [DateTimeIndex](#). Also, check out [time-based indexing](#) to get to know a quick way to select the data you need!

Germany's Daily Energy Consumption 2006-2017 ✓

Use `df.plot(...)` to plot a time series of Germany's daily energy Consumption from 2006-2017. Since pandas' DataFrame plotting methods rely on matplotlib, you can use `matplotlib.pyplot` to style your plots. Set a custom marker, linestyle, title as well as labels for the x-, and y-axis.

Subplots Showing Daily Consumption and Solar-, Wind-Generation ✓

Next, generate **one** figure with **three subplots** containing time series for Wind, Solar and energy Consumption. Ensure that you properly represent the different magnitudes of renewable energy generation and power consumption by **scaling the y-axis accordingly!**

Can you Spot Patterns? ✓

Select a single year from the dataset using pandas' timebased indexing (`df.loc[...]`) for further examination. Plot this years Consumption in a time series plot. Be prepared to comment on what you see next session!

Next, Find a way to plot a boxplot showing the monthly energy Consumption in Germany from 2006-2017. You may e.g. use [seaborn](#) for this. What do you see?

Calculate and Visualize Monthly Consumption and Generation ✓

First, resample the data to change the resolution of our dataset to reflect the monthly sums of energy Consumption and generation (split into Wind and Solar). To do so fairly easily, make yourself familiar with `df.resample(...)`.

Finally, create **a single time series plot** containing monthly sums of Consumption, Wind and Solar from 2006-2017. Try out different colors and styles (e.g. area plots, line plots, ...) within this single plot and make sure to **add a legend!**

Use Your Own Data! ✓

Go to [Wiki Pageviews](#) and download a dataset of the pageviews of a Wikipedia page of your choice.

1. Enter the page name(s) you want to find out more about.

2. Click into the Dates field and choose All time, then download the .csv file and load it into your notebook.
3. Change the index to make use of timebased indexing (`DateTimeIndex`).
4. Inspect the dataset (missing values?)
5. Check for patterns in your data (e.g. selecting just one year/month, grouping monthly/weekly, ...)
6. Finally, show **at least one plot** with a weekly aggregate of numbers.

3 Homework

At the beginning of the next lab session, you will get a chance to indicate which of the **lab assignments** (✓) you completed and some of you will be asked to present their solution in class. Each student will be asked to present their solutions **at least twice** over the course of the semester. Other than that, no additional upload is needed. Please see the [Course Syllabus](#) for details on how this marking of assignments and their presentation affects your grade.