

Hameed Abdul  
Artificial Neural Networks  
10/1/2018  
Homework Assignment 1

We were tasked with three objectives for this assignment. I will briefly outline the requirements and give detail for my approach to solving.

1. Write a Python program containing a function `generate_site_files` which accepts a .csv file of the above format and splits it into separate .csv files, one for each station. The files produced should be named `2017-StationID.csv`, where StationID is the numerical ID for that station.

This was a very straight forward problem to solve. I wanted to leverage recently developed tools in python to achieve this. I made use of the Path class from the `pathlib` class. This library enables python to easily join strings together to represent the path that will be used. Another perk is that the joining is cross platform, so when the path is built it uses system information to determine if forward/backward slashes are needed.

Once I had the main csv loaded in, I utilized pandas to store the csv into a dataframe. Pandas allows for quick data querying. So I used the StationID column and search for unique elements then stored them in an array. Essentially getting the value of each station once. With that done I looped through my array and pair each matching entry to its own file.

2. Also write a function `station_stats_by_month` which accepts a station .csv file and a month (1-12) and returns a list [min, max, mean], where each of these lists the minimum (maximum, mean) value for each attribute during that month.

So this was not as trivial as the first problem. One has to use an int ( between the value 1 - 12) to search for a given month and return max, mean and min stats. This is rather simple but given the fact that the `local_eastern_time` column is the only column with month information and it is in a comprehensive format (h-m-s D-M-YYYY) . Not to mention the month is stored between numeric data and is a string.

To solve this issue, I first took this column and 1 by 1 grabbed the sub string value that store the month for each entry. Afterwards I created a dictionary and mapped three character string values to their appropriate integer value. I looped through the sub string array and grabbed the mapped integer value for each entry and stored it into an array. Leveraging pandas I stored this `Month_Index` array as a column in the csv, that way this computation only is performed once and we can use for future monthly stat queries.

Pandas/Numpy also has built in mean, max and min functions. These were used as they could easily be passed back into a sub dataframe for monthly stats.

3. Write a routine print out the minimum, maximum, and mean solar radiation values for site 410.  
Solar radiation is the last column in the .csv file. It is measured in W/m<sup>2</sup>

Leveraged Pandas ran the mean, max, and min functions on the last column(solar radiation) of the 410 station.