Internet of Things Data Analytics

**Name: Rohit Mishra**

In this exercise I have tried to visualize seasonal activity of Tellus (study area in university of Oulu) by looking the PIR sensor readings, with dataset given for the task. I have used Pandas for reading data and Matplotlib for plotting interactive visualization.

**Questions:**  
  
**2.1. Visualization:**Through video link below you can get a visualization of PIR readings over a chosen period Aug 2017- Nov 2018.

<https://youtu.be/RYM_Jr88Vak>

**2.2. Seasonal variations:**

I have taken the variation of the PIR reading which for different sensors varies between minimum value 0 to maximum value 144 for each day and then took the time scale from Aug 2017 to Nov 2018. Even in this case combining different months and taking the variation of PIR reading with unit of analysis of a single day to week, affects the representation of data. Therefore, I decided to take unit of analysis as a single day and visualize how the sensor variation happened for the day and took it further for consecutive days for the time period. It gives following insights of the data which I have noted below:

1. The time period of sep to November for both year 2017 and 2018 have been the most active period.
2. November to Jan 2017 and May to Aug 2018 have shown little variation.
3. Sep 15 2017 and nov 08 have been the clear outliers in the data set. (Although I haven’t removed considering that I might be missing some point here.)
4. The morning time has generally been less active than the evening time slots. Particularly for Nov- Jan 2017 and Oct to Dec 2018.
5. In the month of November and December the clear change in use of the place can be seen in the time scale most of the activities siftes to 11:00 am to 09:00 pm with very less variation compare to other months.

To check this further, I tried to cluster the sensors based on their GPS coordinates (nearer one in one cluster). For doing this I have taken the mean of maximum and minimum differences of GPS coordinate. But, visualization in this case get affected by the different clock time in use of the different sensors, this further requires the time lag-analysis of to synchronize the clocks. I failed to come up with any meaningful way to do the same.

**2.3. Future implications:**

Breaking down the data in locations of tell us area and displays the change in visits to places could help in maintaining the air quality, light and in optimizing other environmental. Tell us being a collaborative place to study, such data set could also point out what areas are more frequently occupied, which could help in designing such more features in the other parts as well. Nonetheless, it is important to note that some part are designed for induvial study place and other for group activities, some part are time restricted in terms of booking while other are 24\*7 open. This could facilitate the clustering of location based on the research question we want to ask out of the data. Long and Nelson, 2013; identified 7 classes of such events for the analysis: (1) time geography, (2) path descriptors, (3) similarity indices, (4) pattern and cluster methods, (5) individual–group dynamics, (6) spatial field methods, and (7) spatial range methods.

Additional Reference:

[1] J. A. Long, "A review of quantitative methods for movement data," Int. J. Geogr. Inf. Sci., vol. 27, (2), pp. 292-318, 2013. DOI: 10.1080/13658816.2012.682578.