Homework 0

Data preprocessing and exploring the dataset
Due Date: 23:59, October 13 Friday, 2017
TA: 蕭立品 vul1688@gmail.com

In this homework, you need to do some data preprocessing, and then get some basic information about the dataset via tools.

Dataset:

Taipower -
https://drive.google.com/file/d/0B8L9BXtauLRvMFRTSy03aXVVeDA/view?usp=sharing

Weather -
https://drive.google.com/file/d/0B8L9BXtauLRvWW1IaTZVTkRVNm8/view?usp=sharing

Schema:

Taipower - Power supply and usage data.
2016/09/27 - 2017/09/02 : power.json
2017/09/02 - 2017/09/23 : *.csv
*.csv : each column is for Time , NorthSupply , NorthUsage , CenterSupply , CenterUsage , SouthSupply , SouthUsage , EastSupply , EastUsage.

Weather -
Weather record from 2016.07.03 to 2017.07.03. There are some features which may have relationship with the power data.

Tasks:

1. Extract the useful features and import the data into database.
2. After importing the data , try to do the folowing tasks in SQL.(I prefer you to do these query in SQL, but if you don't have any idea about how to do it , you can try other language.)
    a. Find the maximum power supply value and power usage value for each day and each area(North , Center , South , East) from 2016/10/01 to 2017/06/30
    b. List all the locationName(測站) in Weather data.
    c. Find the maximum temperature value for each day and each area(North , Center , South , East) from 2016/10/01 to 2017/06/30. (hint : you can easily choose the location you like to represent the

temperature of the whole area. For example , I choose 板橋 to represent the temperature of North.
   d. Find the highest and the lowest temperature value in Taiwan from 2016/10/01 to 2017/06/30.
3. For the task 2.a and 2.c , please draw three curve(power supply , usage and temperature) in one figure for each area(it means that you will have four figures) and try to explain what you found.
4. About time series data transformation , please choose two taipower data(Q and C) and try to compare the distance between Q and C, D(Q,C) , before and after the transformation below. For more information , please see the slide.
   a. Offset Translation
      $$Q = Q - mean(Q)$$
      $$C = C - mean(C)$$
   b. Amplitude Scaling
      $$Q = (Q - mean(Q)) / std(Q)$$
      $$C = (C - mean(C)) / std(C)$$
   c. Linear Trend Removal
      $$Q = detrend(Q)$$
      $$C = detrend(C)$$
   d. Noise Reduction
      $$Q = smooth(Q)$$
      $$C = smooth(C)$$
5. Do more interesting tasks to get much more score !

Report:

Your report should contain the result(answer , screenshot , figure...) , explaination, SQL. Tell me more detail of your work and I will give you more score :)

Please hand in your report in **.pdf** , **.html** or **.ipynb** format.

Also, you need to hand in the source code.

If you have anything to talk to me , feel free to contact me.