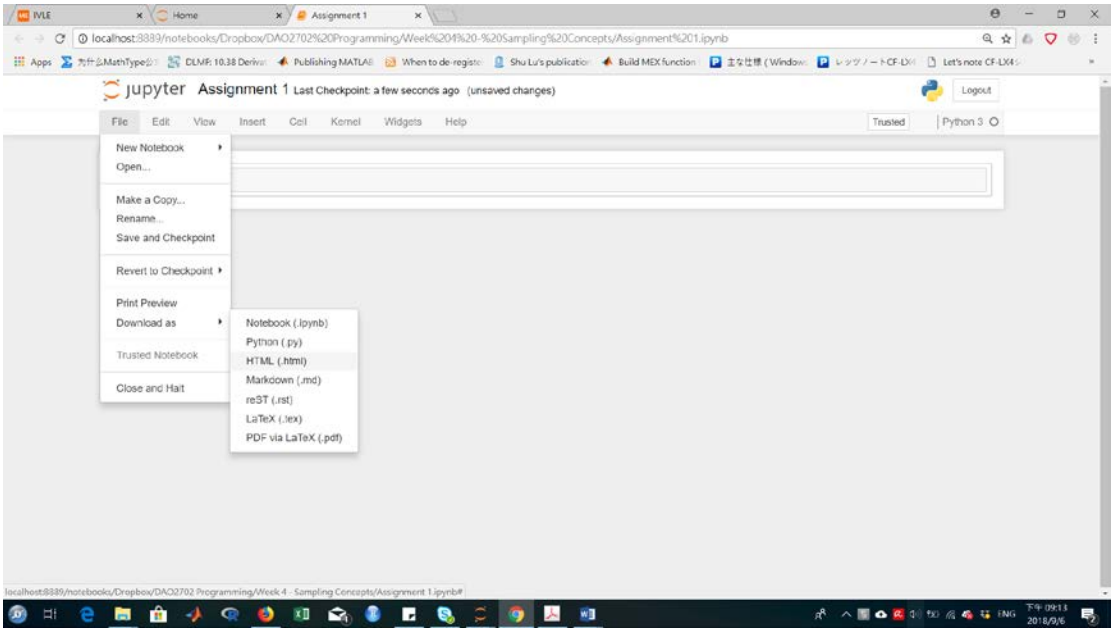
DAO2702 Individual Assignment 1

Due Date: September 21, 11:59PM

Special Notes:

1. **HONOR CODE FOR SUBMISSIONS: assignments cannot not be discussed with anyone. I'm sure that you understand.**
2. Assignment is due to IVLE (upload to student submission folder) before 11:59pm, September 21. There is a penalty for late submission. You will suffer from **25% deduction** if the assignment is submitted within 24 hours after the deadline. **100% deduction** will be imposed if your submission is later than 24 hours after the deadline.
3. E-mail submission is not accepted!
4. Submissions can be a Word file or HTML file. Please also submit your Python code and zip them as a single zip file. Hence, I will suggest you to do the assignment using Jupyter directly and create your HTML file using Jupyter.



5. For correctly identifying the student submitting the assignment, please name your file following the format as in this example:
   Assignment_1_G1_MatricNumber
   (Assignment 1 from tutorial Group 1)
6. **Wrong file name = non-submission!**

Q1.

Understanding demand is always a key issue in business operations. The Hospital.csv contains Singaporeans' arrivals at the major public hospitals' emergency departments (EDs) in 2011. To make a staffing plan, which decides the number of nurses and doctors to serve patients, the intra- and inter-day variations in patient arrivals are important descriptive statistics. Please answer the following questions:

(a) **What is the patient arrival pattern within a day?** Please draw a proper chart to present your findings. To answer this question, you should divide each day into 24 time intervals (each interval is an hour) and then find out the average number of patient arrivals in an interval. Moreover, you should assume there is no inter-day variability to find the average in part (a).

Specifically, you should first create a 31 by 24 table to store the counting results. The first row of the table will be patient arrivals on Oct 01, 2011, in different time periods, say 12 midnight to 1am, 1am to 2am and so on. Then, collapse your rows to have 24 averages and each average represents the average hourly patient arrivals in that time period and visualize your findings.

(b) **Is there any inter-day variation in patient arrivals?** Please draw a proper chart to present your findings. However, now you should assume there is inter-day variability in a week ***(That is, different days in a week can have different arrival patterns of patients)*** and find out the inter-day-specific average number of patient arrivals. Then, visualize the arrival patterns of patients by different days in a week (from Monday to Sunday).

Q2.

The data source is from Taiwan's Taxation Bureau, which consists of 5 local branches all over Taiwan. In the tables below, we can have a basic understanding of the monthly salary distribution of full-time Taiwanese employees in 2015. Although the total population in Taiwan is estimated over 23.5 million people in 2015, only 5.11 million people's salary information can be accurately ascertained after considering multiple demographic criteria and checking the accuracy of income records. All of the following statistics in the tables are calculated by Ministry of Finance, Taiwan, using the income data from Taxation Bureau.

Age	<=20	21-30	31-40	41-50	51-60	61-65	>=66	Total
Number of people	105,770	1,170,176	1,797,410	1,256,255	692,903	88,260	7,362	5,118,136

Age	<=20	21-30	31-40	41-50	51-60	61-65	>=66
Mean salary	17,487	32,481	47,044	56,863	59,514	76,572	114,779
S.d.	7,459	18,207	37,225	65,919	103,232	137,916	360,839

From the above tables, you can assume the population data of salaries of full-time Taiwanese employees are available for us to carry out more detailed analytics.

(a) Although the average monthly salary can be an indicator of central tendency to represent the typical earnings of a Taiwanese, it may not be a good indicator sometimes, due to the impact of extreme income levels in the population. Thus, we also want to know a Taiwanese's median monthly salary to get a full picture of the salary distribution. However, this piece of information is lacking. **Please find out the possible estimate of the median monthly salary by using Python simulation.** That is, please use simulation to **generate a random sample of size 200** using the population information provided in the above tables. Moreover, we make an explicit distributional assumption on the age-specific salary subpopulations. For example, we assume that for those Taiwanese under 20 years of age, their monthly salary population will be a normal distribution with mean 17,487 and variance $7{,}459^2$. Similarly, the salary population of Taiwanese with age between 21 and 30 follows another normal distribution with mean 32,481 and variance $18{,}207^2$. All other age-specific salary subpopulations can be defined accordingly. Please use Python to obtain a random sample of size 200 and estimate the sample median monthly salary.

Note: In your sample, the age-group composition of the sample must be the same (or very close to) the age-group composition of the population. For example, the 21-30 age group constitutes 22.86% (1,170,176/5,118,136) of the population. Hence, in your sample, 46 persons should come from 21-30 age group.

(b) Please perform repeated random sampling of size 200 from Taiwanese population and figure out the sampling distribution of sample median monthly salary. **(The sampling distribution must be constructed with 5000 repeated sampling outcomes and you need to visualize the sampling distribution simulated as your answers).**

(c) Please repeat your repeated sampling in (b), but now you need to increase the sample size from 200 to 500 and 1000. **Comment on the shape of the**

sampling distribution of the median monthly salary. When the sample size increases, is the sampling distribution of the median monthly salary approaching a normal distribution? **(Please draw a panel of histograms to present your visualization).**