

For this exercise, you will be using **activity.csv** to generate and display the needed information.

It is suggested that you look into the following modules (they might come in handy i.e. plotting bar charts, time series, histogram, formatting dates and computing ranges, mean, median, etc):

1. Numpy
2. Pandas
3. matplotlib.pyplot
4. datetime
5. statistics

Your task is to create a python script that would be able to display the information being asked for:

A. What is mean total number of steps taken per day?

For this part of the task, we ignore the missing values (NA) in the dataset.

1. Calculate the total number of steps taken per day
2. Make a histogram of the total number of steps taken each day
3. Calculate and report the mean and median of the total number of steps taken per day

B. What is the average daily activity pattern?

1. Make a time series plot of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all days (y-axis)
2. Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

C. Inputting missing values

1. Calculate and report the total number of missing values in the dataset (i.e. the total number of rows with NAs)
2. Devise a strategy for filling in all of the missing values in the dataset
3. Create a new dataset that is equal to the original dataset but with the missing data filled in.
4. Make a histogram of the total number of steps taken each day and Calculate and report the mean and median total number of steps taken per day.

D. Are there differences in activity patterns between weekdays and weekends?

1. Create a new factor variable in the dataset with two levels - "weekday" and "weekend" indicating whether a given date is a weekday or weekend day.
2. Make a plot containing a time series plot of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all weekdays or weekend days (y-axis).

All the files (.py, .csv, and figures must be saved in your github repository), make sure to include a readme.md file which contains the description of this exercise (you could copy paste whatever is written here actually)

Please send the link github to my email : imanuaba@binus.edu with subject "Machine Problem: CSVFileTask" before Thursday, 17 Dec 2020 at 9:00 AM

Good luck!