Laboratory Activities: Basic Pandas

CP020001 Computer Programming
Khon Kaen University

(10 Points) Basic Dataframe

Objective

This assignment aims to help students gain hands-on experience with data manipulation using the Pandas library in Python. By analyzing a dataset featuring 20 Thai politicians, students will practice the following key skills:

- **Creating and manipulating DataFrames.**
- **Performing data aggregation, sorting, and filtering operations.**
- **Using logical conditions and groupby operations to answer data-related questions.**

Disclaimer!!!

This lab is a mockup dataset created for academic purposes only. The data *does not* represent actual information about Thai politicians. Any resemblance to real individuals or political figures is purely coincidental. This dataset is intended solely for learning and practice with Python and data manipulation techniques and should not be used for any official or real-world representation.

Task Overview

You are provided with a DataFrame containing data about 20 Thai politicians and their respective details. The dataset consists of the following columns:

- **Name: Name of the politician**
- **Age: Age of the politician**
- **Salary_Million_THB: Annual salary in millions of Thai Baht (THB)**
- **Education: Country where the politician studied (US, UK, or Thailand)**
- **Party: Political party of the politician**

The dataset will be stored in the mockup_thai_politicians.csv file, which you can download from the GitHub repository. Please analyze the dataset by answering the following questions. Do not use the solutions provided in the document to answer the questions. You must write your own code to solve them.

Dataset:

The dataset is available in the file mockup_thai_politicians.csv on GitHub. You can download it and load it into a Pandas DataFrame for analysis.

[Download Thai Politicians DataFrame - github link]

```
import pandas as pd
# Load the dataset
df =
pd.read_csv('https://github.com/kaopanboonyuen/CP020001_ComputerPrograming_2026s1/raw/main/dataset/mockup_thai_politicians.csv')
df
```

Questions to Answer:

1. **Who are the top 5 highest-paid politicians?**
2. **How many politicians have studied abroad (US or UK)?**
3. **Find the average salary of politicians from each political party.**
4. **Which politician has the lowest salary but studied in the US?**
5. **Find all politicians aged between 50 and 70 years old and earning more than 30M THB.**
6. **Find the percentage of politicians from each political party.**
7. **Which political party has the highest total salary sum?**
8. **Create a new column ‘Salary_per_Age’ and find the top 3 politicians with the highest value.**
9. **Find all politicians whose name starts with ‘ส’ or ‘พ’.**
10. **Find the politician who has the closest salary to the median salary.**

Instructions

- **Download the dataset from the link provided above.**
- **Load the dataset into a Pandas DataFrame.**
- **Write your Python code to solve each of the questions listed above.**
- **Submit your Python script or Jupyter Notebook containing the code and outputs.**
- **Do not write the solutions directly in the document; instead, analyze the dataset and answer the questions programmatically.**

Submission Details:

- **Submit a Python script (`.py`) or Jupyter Notebook (`.ipynb`) file via the course portal.**
- **Ensure your code is well-commented to explain your thought process for each question.**
- **The file should be named as follows: `CP020001_YourName_BasicPandas.py` or `CP020001_YourName_BasicPandas.ipynb`.**

Evaluation Criteria:

- **Correctness of the answers.**
- **Quality and readability of the code.**
- **Proper use of Pandas operations such as `sort_values`, `groupby`, logical indexing, and other relevant techniques.**
- **Clear presentation of the output results.**

Deadline:

Submit your assignment by 2025 March.