EBA3400 Programming, data extraction and visualisation

Final exam

- Start date: **26/11/2021 09:00**
- Finish date: **26/11/2021 11:15**
- Total no. of pages: 3 incl. front page
- To be answered: **Individually**

Instruction:

- Download the file attached to this exam and place it in the same folder as your jupyter notebook.
- Use headings such as "Part 2(b)" to clearly indicate which part of the assignment your code is referring to.
- You can only use the following packages: pandas, seaborn, matplotlib.
- We encourage you to write some comments to explain your code. But whether you write comments or not will not affect your score.
- Code efficiency will not be part of the scoring criteria.

Allowed aids:

- You are allowed to use all course material and your notes.
- You are allowed to use resources on the internet.
- You are **NOT** allowed to get help from others.

Submission:

- All the tasks in this assignment must be answered and delivered as a jupyter notebook (.ipynb). Please submit the **ipynb file** to **WISEflow** before the deadline.

Points:

- Part 1: (a) 5 points (b) 20 points (c) 5 points
- Part 2: (a) 15 points (b) 5 points
- Part 3: (a) 5 points (b) 10 points (c) 5 points (c) 5 points
- Part 4: (a) 5 points (b) 5 points (d) 5 points (d) 10 points

1. Grading

(a) Student A took three exams with the following scores: exam1 = 60, exam2 = 85, exam3 = 80. Create a three-element tuple containing these scores.

(b) Write a function that takes **a three-element tuple** as the argument and use the following weights to calculate the final score: exam1 (35%), exam2 (35%), exam3 (30%). **Return the final grade** based on the following table.

Final score	Final grade
$75 \leq$ score ≤ 100	A
$65 \leq$ score < 75	B
$55 \leq$ score < 65	C
$45 \leq$ score < 55	D
$35 \leq$ score < 45	E
$0 \leq$ score < 35	F

For example, the final score of student A is 60×0.35+85×0.35+80×0.3=74.75 and final grade is B.

(c) Use the function in (b) to get the grades of student B and student C. Print out the returned grades.

- Student B: exam1 = 70, exam2 = 50, exam3 = 65.
- Student C: exam1 = 80, exam2 = 75, exam3 = 90.

2. Annual balance

(a) Suppose you deposit 10000 in the bank and the annual interest rate is 5%. By the end of the year, your annual balance will be 10000 plus 500 interest. (Hint: interest = (current balance) × (annual interest rate).)

Write a function with two arguments (deposit and annual interest rate), calculate the annual balance for each of the next five years and print out the result. Round the number to two decimal places. For example, if deposit = 10000 and annual interest rate is 5%, the function should print out the result as follow:

Year-1: 10500.00

Year-2: 11025.00

Year-3: 11576.25

Year-4: 12155.06

Year-5: 12762.82

(b) Test your function by passing deposit = 18000 and annual interest rate is 4.5%.

3. Customer profile

A store collected some information from 10 customers as shown in the following tables.

CustomerID	Gender
C01	F
C02	M
C03	M
C05	F
C06	M
C08	F
C10	F

CustomerID	Age
C01	20
C03	45
C04	25
C06	35
C07	40
C08	30
C09	55

(a) Create two pandas DataFrames based on the tables shown above.

(b) Combine these two DataFrames created in (a) into one DataFrame. The resulting DataFrame should contain all customer IDs, which means that the number of rows should be 10.

(c) Replace all missing values with the text "Unknown".

(d) Use a bar chart to show the number of customers in each gender group (F, M, Unknown).

4. Oslo city bike

The dataset **citybike.csv** contains anonymous trip data from Oslo City Bike in August 2021. The following table includes the column descriptions required for this task. Other column descriptions can be found here https://oslobysykkel.no/en/open-data/historical.

Column	Description
started_at	Timestamp of when the trip started
duration	Duration of trip in seconds
start_station_name	Name of start station

(a) Import the data (citybike.csv) and set the column "started_at" as DatetimeIndex.

(b) Select the 10-day data from August 11th to August 20th (inclusive) and use this subset to complete the **task (c) and (d)**.

(c) How many trips started at "BI Nydalen" station on these days?

(d) Calculate the total riding time for each day and display the result with a line chart. Note that the riding time of any trip is included in the total riding time for the day when the trip started, even if the trip ended next day. (Hint: Use the column "duration")