Final Exam Computer science (ETL and visualization)

UCA-2022-2023 fall semester

!9th of December 2022

This exam is a take home exam with a time Span of 24 hours.

Note 1- You should submit 2 documents in the moodle.

1- Your Python codes.
2- 2- A link to your final production

This exam is about retrieving the necessary data from different database and create a dynamic dashboard for visualization of the blended data.

You may use Python programming language and all libraries that you consider useful, you can use sqlite relational databases.

IMPORTANT 1: If you find difficulties with deploying your final results, please explain the issue and provide the correct way of running your code on localhost.

IMPORTANT 2: Although the data queries in this exam are very simple. if you have trouble writing the sql queries, you should mainly search for groupby solutions. There are no joins involved. Some sample queries and the HR database diagram can be found in https://www.w3resource.com/sqlite-exercises/

Each QUESTION HAVE 20 MARKS FOR ATOTAL OF 100

Exercise 1-

Use the HR sqlite database available on https://www.w3resource.com/sqlite-exercises/, search for a solution to produce the diagram of the database using python. Do not invent solutions, search for existing solutions and try to present a fairly readable database diagram.

Exercise 2-

Present a bar chart with the number of employees with the same job. Use Plotly-Dash to make your bar chart dynamic and allow the user to choose the job titles to be seen in the chart.

Exercise 3-

Present a horizontal bar chart with the difference between the maximum and minimum of each job salaries. Make it dynamic by using a scale over salary. The scale can be visual or can be as 2 input boxes as min and max

Exercise 4-

Find the average salary. Then automatically go to https://www.itjobswatch.co.uk/jobs/uk/sqlite.do site and read 3 camps of 10^{th} Percentile, 20^{th} Percentile, 75^{th} Percentile, 90^{th} Percentile. This should be done by using scrapping methods (beatifulsoup, selenium, etc)

Plot a scatter chart where the average salary of your database is shown in black, and the 12 numbers read from the 3 columns are shown in green. Make it dynamic in function of year.

Exercise 5-

Put all the above exercises into one dynamic dashboard and deploy. If you get any issues while deploying, make it run on local host but you need to explain clearly why you could not deploy your dashboard and the ways your tried to solve it without success.