Final Examination

Date: 02/02/2024; Duration: 100min
Open book, no internet connection
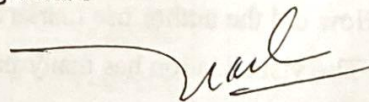
SUBJECT: Data Science and Data Visualization (IT138IU)	
Approval *by the ScSE* Signature	Lecturer: Signature
Full name: Assoc.Prof. Nguyen Van Sinh	Full name: Trần Thanh Tùng
Proctor 1 Signature	Proctor 2 Signature
Full name:	Full name:

STUDENT INFO	
Student name: Student ID:	

INSTRUCTIONS: the total point is 100 (equivalent to 40% of the course)

1. *Purpose:*
   - Test your knowledge of data science and data visualization in the following topics:
     - visualization: storytelling, design, and evaluation visualization (CLO2, CLO3)
   - Examine your skills in
     - Analyzing and designing visualization (CLO2)
2. *Requirement:*
   - Write the answers and draw models CLEAN and TIDY.

Student Name:............................
Student ID:............................

1. (30pts) List, and describe 2 most important interactive features for visual analysis on **BIG DATASETS.** Explain why those features are the most important.

2. (40pts) The following chart shows the quantity of exported fruits each month. However, it is a bad example of data visualization. Analyze and redesign it.

   a. (10pts) What is its data? types of data?

   b. (5pts) How did the author use marks and channels to encode the data?

   c. (10pts) The visualization has many problems, describe them.

   d. (15pts) Redesign it.
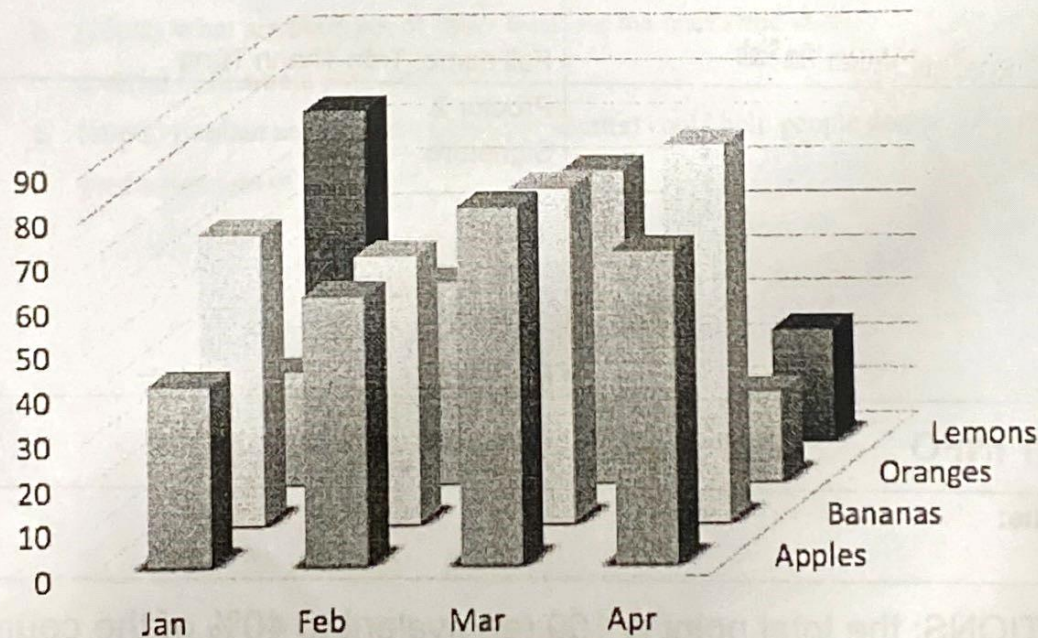
3. (30pts) At University X, the late graduation rate is high meaning that most students cannot finish the program within the designated timeline of the curriculum.

   The reasons could be that

   - Students cannot pass courses after the first attempt,

   - Students did not register for courses in the designated semester,

   - Students intentionally lengthen the learning path by registering for fewer courses than the suggested courses in the curriculum,

   - Or other reasons

   Hence, people would like to analyze the learning progress of students to make early warnings to students and the dean of the school regarding the potential late graduation problem.

- The data for the learning progress of students is grouped by enrolment batch.
- Each batch has an enrolment year and a major i.e. Batch 2020-Computer Science, Batch 2021-Data Science, and Batch 2019-Information Technology
- Each batch has its curriculum including a list of courses students of the batch should take each semester like the following table

Batch	Semester	Course	
2020-CS	1	Introduction to Computing	(3 credits)
		C/C++ programming	(4 credits)
2020-CS	2	Object-Oriented Programming	(4 credits)
		Computer Network	(4 credits)
		Discrete Mathematics	(4 credits)
2020-CS	3	Data Structures and Algorithms	(4 credits)
		Linear Algebra	(3 credits)
		Database	(4 credits)
...	4	...	

- Each batch has a summarized of grades of students by semester like the following table

	Batch	Semester-Year	Course	Number of registrations /Total students of the batch	Number of students who passed the course (score > 50)
1	2020-CS	Sem 1-2020 (Sep/2020)	Introduction to Computing	60/60	60
2	2020-CS	Sem 1-2020 (Sep/2020)	C/C++ programming	50/60	40
3	2020-CS	Sem 2-2020 (Feb/2020)	C/C++ programming	5/60	5
4	2020-CS	Sem 2-2020 (Feb/2020)	Object-Oriented Programming	40/60	30
5	2020-CS	Sem 2-2020 (Feb/2020)	Computer Network	60/60	50
6	2020-CS	Sem 2-2020 (Feb/2020)	Discrete Mathematics	30/60	27
7	2020-CS	Sem 1-2021 (Sep/2021)	Database	40/60	35
8		...	...	...	...

Student Name:... ... ...
Student ID:... ... ...

In the table above, row 2 shows that 50 students of 60 students of batch 2020-Computer Science registered for "C/C++ Programming" in semester 1 of 2020. And only 40 of those 50 students passed the course.

Row 3 shows that in semester 2 of 2020, 5 students of batch 2020-Computer Science registered for "C/C++ Programming" and all those 5 students passed the course.

Your tasks are to design chart(s) to analyze the learning progress of students and highlight potential late graduation rate

a. (10pts) State the problem described above in terms of abstract tasks in visualization

b. (10pts) What are the types of data? What are the marks and channels to use? Draw a sample chart(s) to illustrate your design,

c. (10pts) Explain and describe how your chart(s) could help people detect the potential high late graduation rate of a batch.

-- The end --