DIGITAL TECHNOLOGY

Academic Year 2023-24
Prof. Barbara PERNICI

Exam 17/6/2024
Total time 1.30 h

Last name	
First name	
Matricola/Person code	
Signature	

Please remember that:
- *The exam is closed books.*
- *The use of cellular phones or any other electronic devices during the exam is forbidden.*

It is necessary to answer at least partially all the questions for a positive evaluation.
Given answers should be explained, lists of bullet items are insufficient to answer a question.

Please write the answer for each question on a separate piece of paper.

Question 1 [11 points]

Assume you have a database table with the following schema:

Visit(RecordID, PersonID, Name, Institution, Town, StartingDate, Duration)

The visit records are derived from different sources of data.

Discuss the following:
- Which quality problems could be found in such a table?
- Assume that you want to design a data warehouse for these data, to analyze trends in time in the starting dates (Time analysis dimension) and in the visited places at different granularities for the visited locations (Location analysis dimension):
  - Identify the facts for the data warehouse
  - Design the Dimensional Fact Model for the requested data warehouse
  - Which data is it necessary to clean and transform to be able to load the data in the data warehouse?
  - In the Star schema corresponding to the Dimensional Fact model, how would you represent the Time analysis dimension?

Question 2 [11 points]

Explain how continuous delivery of value is implemented in agile project management. Compare the advantages and disadvantages of this approach to traditional project management, focusing on stakeholder engagement, risk management, and project outcomes.

[please turn over]

Question 3 [11 points]

Consider the following fragment of Python code:

```
import pandas as pd

diz1 = {'house_id': [1, 2, 3, 4], 'rooms': [2, 3, 1, 2], 'street_name': ['A', 'A', 'B', 'B']}
diz2 = {'zip_code': [2020, 2021, 2020, 2020], 'house_id': [1, 2, 3, 4],  'price': []}
diz2['price'] = [10 * (i+1) if i % 2 == 0 else 50 for i in range(len(diz2['zip_code'])) ]

df1 = pd.DataFrame(diz1)

df2 = pd.DataFrame(diz2)

df3 = df1.merge(df2, on='house_id')

for var, obj in df3.groupby(['zip_code', 'rooms']):
   print(obj['price'].max())   # Change this line
```

Questions:

1. Qualitatively explain what the program does and describe the expected output.

2. Change the last line of the code, without using the mean function of pandas, such that it computes the average price.

3. Explain the difference between a set, a list, and a tuple.