INTRODUCTION TO DATA SCIENCE

MIDTERM MAKE-UP EXAM

(This is a 90-minute exam)

STUDENT ID:	FULLNAME:	SIGNATURE:

Q.1:	Q.2:	Q.3:	Q.4:	TOTAL:

Q.1) (20) Data can be measured on different scales. Explain **the four levels** of measurement and provide examples for each. For each level, make sure to define **the key characteristics** of the measurement scale, and provide **at least two clear examples** of attributes that can be measured on that scale.

INTRODUCTION TO DATA SCIENCE
MIDTERM MAKE-UP EXAM
(This is a 90-minute exam)

Q.2) (20) Imagine you're building a social media platform. Users can follow each other, creating connections, and post updates that appear in a feed based on the followed users.
What data structures would you use to represent:

- The order of posts in a user's feed: This should ensure the user sees the most recent updates first.
- The follower relationships between users: This needs to efficiently track who follows whom for personalized feed generation.

Explain your choices, highlighting why each data structure is suitable for its respective purpose.

INTRODUCTION TO DATA SCIENCE

MIDTERM MAKE-UP EXAM

(This is a 90-minute exam)

Q.3) A company is launching a new marketing campaign that relies heavily on customer data for targeting and personalization. The data comes from various sources, including website forms, CRM systems, and loyalty programs. Considering the importance of data quality for successful marketing campaigns, answer the following questions.

(15) (a) Identify three potential data quality issues that could arise when integrating data from multiple sources. Explain how these issues might impact the marketing campaign.

(b) (15) Describe two data quality checks you would recommend performing on the customer data before launching the campaign. Explain what each check aims to achieve.

INTRODUCTION TO DATA SCIENCE
MIDTERM MAKE-UP EXAM
(This is a 90-minute exam)

Q.4)

a) (15) What is the output of the code snippet given below.

```
import pandas as pd

df = pd.DataFrame({
    'Category': ['A', 'B', 'A', 'B'],
    'Values': [10, 20, 30, 40]
})

grouped_df = df.groupby('Category').sum()
print(grouped_df)
```

b) (15) What is the output of the code snippet given below.

```
import pandas as pd

data = {'Age': [25, 30, 22, 38],
        'City': ['New York', 'Los Angeles', 'Chicago', 'Miami']}
df = pd.DataFrame(data, index=['Alice', 'Bob', 'Charlie', 'David'])

x = df.iloc[1, :]
print(x)
```