# United International University
## B.Sc. in Data Science

DS 1115: Object-Oriented Programming for Data Science

Final Exam: Fall 2024 Time: 2 Hours Marks: 40

**Answer all of the following questions.**

1. (a) Write Python code for a function named *processSensorData* that takes a list of temperature readings (*temperatures*) in Celcius and an integer (*threshold*). The function should return a filtered list of temperatures that is greater than the threshold. However, the function must enforce the following constraints:

- The *temperatures* list must not be empty.
- All elements in the list must be numbers (int or float).
- The threshold must be an integer.
- Temperature values must be within the range of -100 to 100.

Create custom exceptions for each of these constraints and raise them appropriately from the function. [7]

(b) Find the output of the following program. [3]

```
def f(x):
    try:
        print("Start")
        if x == 1:
            raise ValueError("Invalid value!")
        elif x == 2:
            return "Returning from try"
        elif x == 3:
            raise TypeError("Wrong type!")
        return "End of try"
    except ValueError as e:
        print("Caught ValueError: ", e)
        return "Returning from except"
    except TypeError as e:
        print("Caught TypeError: ", e)
    finally:
        print("Executing finally")
```

```
        return "Returning from finally"


f(1)
f(2)
f(3)
f(4)
```

2. (a) Write a decorator named *time_limit(n)* that limits the execution time of a function. If the function takes longer than $n$ seconds to execute, it should just print "Time limit exceeded." [5]

   (b) In a distant futuristic city named Numera, a group of scientists is working on a robotic assistant called E.V.E (Even Value Extractor). The robot is designed to generate even numbers for various scientific experiments. Your task is to write a generator function that takes an integer $n$ and yields the first n even numbers, starting from 0. The scientists expect you to write Python code. [5]

3. You are given a dataset containing information about customer transactions at an online retail store. The dataset has the following columns: **Order_ID** (unique identifier for each order), **Customer_Name** (name of the customer), **Product_Category** (category of the purchased product), **Price** (price of the purchased item), **Quantity** (number of items purchased), **Purchase_Date** (date of purchase in *dd/mm/yyyy* format), **City** (city where the order was placed), and **Payment_Method** (payment method used like Credit Card or PayPal). Based on this, answer the following questions: [2x5=10]

   (a) How many unique product categories are there in the dataset?

   (b) Which city had the highest total sales?

   (c) What is the name of the customer who made the highest total purchase?

   (d) Find the most frequently used payment method among customers who bought products from the "Electronics" category.

(e) Determine the month with the highest total sales.

4. (a) A data analyst at a retail company wants to visualize different aspects of their sales data. For each of the following scenarios, identify which type of graph would be the most appropriate and **justify your choice**. For each of the following scenarios, draw an example graph using values of your choice labeling the x and y axes. [3x2=6]

   i. The analyst wants to show the total sales revenue for five different product categories to compare their performance.

   ii. The analyst needs to display how the company's total sales have changed over the past 12 months.

   iii. The analyst wants to represent the proportion of total sales contributed by each of the five product categories.

(b) Carefully observe the following table and answer the questions.

| Date | Category | Total Sales ($) | Units Sold | Price Per Unit ($) |
|---|---|---|---|---|
| 2024-01-01 | Electronics | 1500 | 10 | 150 |
| 2024-01-02 | Clothing | 800 | 15 | NaN |
| 2024-01-03 | Electronics | NaN | 8 | 160 |
| 2024-01-04 | Groceries | NaN | 12 | 30 |
| 2024-01-05 | Groceries | 500 | NaN | 25 |
| 2024-01-06 | Clothing | 1200 | 20 | 60 |

   i. How would you fill-up the missing values in the table? Provide explanations and show calculations. You **do not** need to write any code. [3]

   ii. What will the missing values in the *Total Sales* column be replaced with if you were to run **interpolation** on that column? Show all calculations. [1]