United International University
B.Sc. in Data Science
DS 1115: Object-Oriented Programming for Data Science
Final Exam: Spring 2024   Time: 2 Hours   Marks: 40

Answer all of the following questions.

1. (a) Write Python code for a function named `calculateInterest` that will take two parameters as input: `balance`, `rate` and return interest by multiplying balance by interest rate. However, this function can not operate on negative balance and interest rates that are not in the range $[0, 1]$. Your code should check these conditions and raise exceptions as necessary. Create custom exceptions for these conditions and raise them from the function. [3+3=6]

   (b) Find the output of the following program. [4]

```python
def process_input(value):
    try:
        if not isinstance(value, int):
            raise TypeError("Input must be an integer")
        elif value < 0:
            raise ValueError("Input must be a positive integer")
        elif value % 2 != 0:
            raise ValueError("Input must be an even number")
        else:
            print("Input is valid")

    except TypeError as te:
        print("TypeError occurred:", te)
    except ValueError as ve:
        print("ValueError occurred:", ve)
    except Exception as e:
        print("An exception occurred:", e)


process_input(10)
process_input("hello")
process_input(-5)
process_input(7)
```

2. (a) Find the output of the following program. [5]

```python
def even_generator():
    num = 2
    while True:
        yield num
        num += 2

def odd_generator():
    num = 1
    while True:
        yield num
        num += 2

def alternate_generators():
    even_gen = even_generator()
    odd_gen = odd_generator()

    while True:
        try:
            yield next(even_gen)
            yield next(odd_gen)
        except StopIteration:
            break

gen = alternate_generators()

for _ in range(10):
    print(next(gen))
```

(b) Write Python code for a decorator that allows a function to be run only 5 times. Name your decorator StopAtSixth. [5]

[hints: use a dictionary to register each function and keep count or alternatively use a inner variable to keep count]

3. Suppose you work for a company that analyzes weather data and tries to extract useful information from the data. One day you are given a CSV file and upon reading it you see that the file looks like the table given below. Now answer the following set of questions based on this data.

day	temperature	windspeed	event
1/4/2024	41	6	Sunny
2/4/2024	42	9	Sunny
3/4/2024	40	NaN	NaN
4/4/2024	33	7	Rain
5/4/2024	46	NaN	Sunny
6/4/2024	44	NaN	Sunny
7/4/2024	NaN	NaN	Rain
8/4/2024	43	8	Sunny
9/4/2024	32	12	Rain

(a) If the name of the file you were given is 'weather.csv', write code to read the contents of the file into a Pandas DataFrame. [2]

(b) You want to see just the average value for all the columns in the data. How would you do that? [2]

(c) If you notice the data, you see some cells have the value "NaN". What does NaN mean and how can it cause problems when extracting information from the data? [2]

(d) The temperature value for the row '7/4/2024' is missing. You want to fill this missing value with the most logical replacement. You noticed that the event for this day is 'Rain'. Based on this observation, what should be the new value for temperature on that day? [2]

(e) In the 'windspeed' column, there are 3 consecutive missing values for the days 5/4/2024 to 7/4/2024. One popular method of dealing with this chain of missing values is to interpolate the missing values in between. The idea of interpolation is that the gaps are filled with uniformly spaced values. For example, if one missing value is between 6 and 10, the interpolating point will be 8. What should be the value of the 3 missing points in the "windspeed" column for 5/4/2024 to 7/4/2024 if you use interpolation to fill the data? [2]

4. You are applying for a job in two very renowned companies, **Company X** and **Y**. You are trying to evaluate which company will be better for your career. You are looking into the

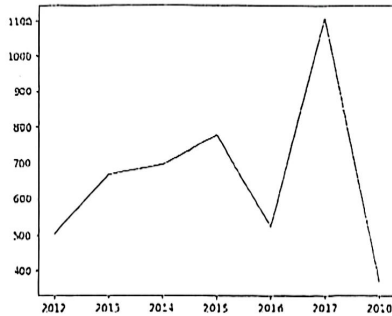number of people joined in these companies from year **2012 to 2018.**
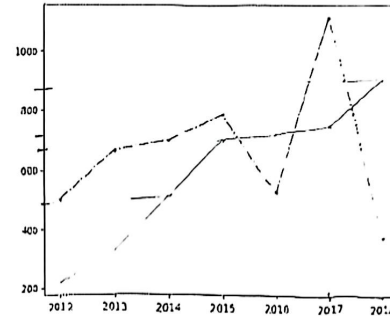

Figure 1: X Employee Turnover

Figure 2: Turnover Comparison

(a) Observe Figures 1 and 2. Although the plots are alright, it is a bit difficult to understand the meaning of the plots. What should be the labels of the axes of these graphs?   [2]

(b) Which company do you think will be a better choice? Based on these plots, Find out the years in which the maximum number of people joined in both of these companies.[4]

(c) Consider the following dataset.

Product line	Quantity
Health and beauty	854
Fashion accessories	902
Home and lifestyle	911
Sports and travel	920
Food and beverages	952
Electronic accessories	971

Company X has these **Product Categories** and **Quantity** sold in a month. Which type of graph plot will be a better choice in this scenario so that it will be easier to understand which Category needs better marketing? Write down the necessary codes to replicate the type of graph you mentioned.   [4]