Foundation of Data Science (Practical) - Lab 2 (Introduction to Data science Tools: Numpy and Pandas) Assignments

1. Numpy assignment: Perform edge detection in a photo.
Steps:

- Import a photo using PIL library (install if necessary) and convert it to numpy (*PIL.Image.open()*).
- Covert it to grayscale (For each pixel take mean of R,G,B values which will be the grayscale value)
- Use edge detection logic (using Sobel Filter):
    - Add padding pixels around the boundary of the image (use *numpy.pad()*)
    - For each pixel at position (i, j) perform following:

Image[i, j] = absolute_value_of(*Gradx*) + absolute_value_of(*Grady*)
where,
Gradx = Sum(*3x3 neighbourhood of Image[i, j] * Kx*) i.e. element-wise multiplication of two 3x3 matrices followed by sum of all elements

Grady = Sum(*3x3 neighbourhood of Image[i, j] * Ky*)

3x3 neighbourhood of Pix[i, j] = Image[i-1:i+2, j-1:j+2]

Kx = [[-1, -2, -1],
[0, 0, 0],
[1, 2, 1]]

Ky = [[-1, 0, 1],
[-2, 0, 2],
[-1, 0, 1]]

- Convert above obtained numpy array to PIL image (*Image.fromarray()*)
- Show/save the image using PIL

2. Pandas assignment: On the attached salaries dataset, perform following:

- What is the job title and base pay of “David Shinn”
- What is the average base pay
- Fill the missing numeric value in each column using the column's average
- Select and display rows where the JobTitle is "CAPTAIN III (POLICE DEPARTMENT)".
- Find employees whose BasePay is greater than $200,000.
- Identify the employees who received the lowest Benefits.
- Find the top 5 employees based on their TotalPay, and display their EmployeeName, JobTitle, and TotalPay.
- Calculate the average BasePay for each unique JobTitle. Find the job title with the highest average BasePay.