Assignment 1 (Python programming exercise)

Python is a popular programming language (https://www.zdnet.com/article/programming-language-pythons-popularity-ahead-of-java-for-first-time-but-still-trailing-c/) and is popular for data mining application. This exercise is aimed at getting some simple Python programming done.

Write a Python program which input two n-dimensional vectors **x** and **y** and compute the following similarity measures:

1. Euclidean
2. cosine
3. Jaccard
4. correlation

Input requirement
The user enters n and then two strings of n numbers. Each number is separated by spaces. The number may be integers or floating point numbers. The program checks the input and asks the user to re-input if there are less than n numbers.

Output requirement
Print the values of the 4 measures.

Notes
The measures should be implemented from scratch. Do not use existing functions, but basic functions (e.g. sqrt ( )) can be used.

Marking Scheme

1. The measures are correctly implemented. Test it with Appendix. (80%)
2. The program is adequately commented. Marks will be deducted if there are no or few comments. (20%)

3. You'll get 0 mark if your program cannot run.

What you should hand in

One Jupyter notebook containing your code, output, and answers.

Appendix

For the following vectors, $\mathbf{x}$ and $\mathbf{y}$, calculate the indicated similarity or distance measures.

(a) $\mathbf{x} = (1, 1, 1, 1)$, $\mathbf{y} = (2, 2, 2, 2)$ cosine, correlation, Euclidean

$\cos(\mathbf{x}, \mathbf{y}) = 1$, $\text{corr}(\mathbf{x}, \mathbf{y}) = 0/0$ (undefined), $\text{Euclidean}(\mathbf{x}, \mathbf{y}) = 2$

(b) $\mathbf{x} = (0, 1, 0, 1)$, $\mathbf{y} = (1, 0, 1, 0)$ cosine, correlation, Euclidean, Jaccard

$\cos(\mathbf{x}, \mathbf{y}) = 0$, $\text{corr}(\mathbf{x}, \mathbf{y}) = -1$, $\text{Euclidean}(\mathbf{x}, \mathbf{y}) = 2$, $\text{Jaccard}(\mathbf{x}, \mathbf{y}) = 0$

(c) $\mathbf{x} = (0, -1, 0, 1)$, $\mathbf{y} = (1, 0, -1, 0)$ cosine, correlation, Euclidean

$\cos(\mathbf{x}, \mathbf{y}) = 0$, $\text{corr}(\mathbf{x}, \mathbf{y}) = 0$, $\text{Euclidean}(\mathbf{x}, \mathbf{y}) = 2$

(d) $\mathbf{x} = (1, 1, 0, 1, 0, 1)$, $\mathbf{y} = (1, 1, 1, 0, 0, 1)$ cosine, correlation, Jaccard

$\cos(\mathbf{x}, \mathbf{y}) = 0.75$, $\text{corr}(\mathbf{x}, \mathbf{y}) = 0.25$, $\text{Jaccard}(\mathbf{x}, \mathbf{y}) = 0.6$

(e) $\mathbf{x} = (2, -1, 0, 2, 0, -3)$, $\mathbf{y} = (-1, 1, -1, 0, 0, -1)$ cosine, correlation

$\cos(\mathbf{x}, \mathbf{y}) = 0$, $\text{corr}(\mathbf{x}, \mathbf{y}) = 0$