HW2

First, select the correct dataset to which your PIN belongs.
e.g., If your PIN is 5, then select 1-10.csv
If your pin is 20, then select 11-20.csv
In the remaining assignments, we address 3 different prediction problems using ML:

- Predicting 2022 citation numbers using the university rank and 2017-2021 citation numbers.
- Predicting the h-index using the university rank and all (2017-2022) citation numbers.
- Predicting the i10-index using the university rank and all (2017-2022) citation numbers.

For HW2, use Principal Components Analysis on the data set provided to you consisting of 100 data points from 10 universities, for one of the three prediction problems mentioned above.
(Use the intended inputs to ML, not the variable whose values are to be predicted!)
Divide the data into 4 quarters based on the variable to be predicted: red for the lowest quarter, orange/pink for the 26th-50th percentile, green for the 51st-75th percentile, and blue for the top percentile. Generate a scatter plot for all data using these 4 colors, whose axes are the two most important principal components.
What are your conclusions regarding the usefulness of principal components for this prediction task?

Your submission should include:
1. code
2. output (you can add it in the report.pdf file)
3. report.pdf (explaining your results and your conclusions)