Proposed idea:

Take data from baseball reference of all of the pitchers over a mlb season
(We have data from baseball reference downloaded, we are just having trouble downloading it as a csv)

- Assign some sort of value to the winner of the cy young that differs them from the rest of the pitcher
- Use data from 1990(?)-2010(?)
    - Train it to learn which statistics and their values tend to have the highest correlation to the CY young winner
    - Have it learn which is more valuable (through linear regression?)
- Then give it the data from 2011-2023 and have it try and guess which pitchers won the cy young award that year
- 13 years, 26 award winners, see how many it guesses right?

If the model is able to predict the cy young winners very accurately, maybe try to do the same with MVP since MVP is more complicated (pitchers and batters can win, and a lot more batting/fielding statistics exist)