Vimeo has hundreds of millions of videos on our platform. Our curation team hand selects "Staff Picks" and once a month the whole company watches and votes on a staff pick of the month which is featured on the site.

A product manager wants to create a system to suggest "similar" staff pick videos, but they aren't quite sure what they mean by "similar" yet. Your mission, should you choose to accept it is to write a program that takes a clip_id as input and outputs "similar" clips using your best judgement after exploring the dataset. This could be visual, text based, some other features or all of these features combined, it's up to you! We've included metadata of around 4,000 staff picks in a dataset along with their categories they belong to and links to their thumbnails which you can download here: https://drive.google.com/open?id=1zbCF5jVmc2Gy5L1_JqJICGvbhRexwWaX

To complete the exercise, your program must:

- Return a list of the 10 most similar videos in the dataset in order of similarity given a single clip_id (also from the dataset.)
- Please output the results for the following clip_ids in your submission:
    - 14434107, 249393804, 71964690, 78106175, 228236677, 11374425, 93951774, 35616659, 112360862, 116368488

Before you create the program itself, you should evaluate the dataset in an ipython/jupyter notebook with pandas detailing:

- How you explored, cleaned, and combined the data.
- Which features you explored and chose for your final model and why.
- Algorithms or external models you used and why.

Things we care about:

- Clean, readable and well documented code
- README / jupyter notebook explaining your approach and design
- You have fun while building this.

Bonus points for:

- Implementing a web interface where you can submit a clip_id and get the similar clips listed.
- Using graphs and charts in your writeup.
- Describing how you would scale your solution to millions of videos.
- Detailing how you would evaluate your model in production against other options.
- Python 3