1 Understand the World

In the real world, applying data science can be much harder than analyzing a data-set provided by a tutorial or a kaggle challenge.

In this exercise you'll be tested on your capability to solve something a bit more realistic than any educational example.

If you accept it, you'll be extracting and analyzing features from a 42 minute video.

For any of these exercises you can use Python and any library you want. No need to implement very complex algorithms.

You also don't need to analyze every single frame, but instead one frame per second. For example take only the first frame and assume that it represents that 1 second time interval. Assuming the video is 42 minutes, your loop should iterate 2,520 times, one for each frame.

1.1 Face detection

1. What's the percentage of time a face is shown in the video?
2. What is the average amount of faces in a frame?
3. What is the average amount of faces in a frame with at least one face?
4. How does the number of faces evolve during the show? (Plot)
5. Any extra analysis you would like to do.

1.2 OCR

SubRip is a well know subtitle format. In a text editor it looks like this:

```
1
00:00:20,000 --> 00:00:24,400
Howdy!
2
00:00:24,600 --> 00:00:27,800
Why, hello there!
What's your name?

Where 00:00:20,200 represents hours:minutes:seconds,milliseconds
```

1. Extract the text from the footer of each frame, when the background of the footer is blue.
2. Write the subtitles in file with the SubRip format.

1.3 Extra

How would you apply Artificial Intelligence to the features extracted during the previous exercises to build a use case?

2 Theory

1. How do you deal with an imbalanced data-set, assuming SMOTE is not an option?
2. Explain in your words what model calibration is.
3. When is explainability important?
4. How do you define the right threshold for a binary decision?
5. What's the difference between a score and a probability in a classification model?

3 Logic

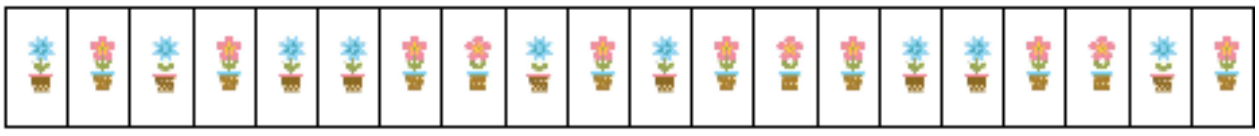
You're a florist, and a spy. Your boss is a criminal mastermind.

It's your mission to know when he is meeting with an arms dealer the next day. You and your colleagues know it's somewhere between noon and 7pm.

You have to give this information to your fellow spies. However, you have no direct communication with them.

Everyday, at your flower shop, your boss spreads out 20 flowers in 20 different pots, numbered and perfectly aligned with the storefront.

Below is an example of one of these days:



Flowers are either red or blue, and they are randomly selected out of a huge pot.

Even though the Boss is very strict with rules he does allow you to change one, and just one of the flowers in the pot, for a different colour.

It is your job to arrange an algorithm with your fellow colleagues that will allow them to know at what time, from 12pm to 7pm your boss is meeting the arms dealer the next day, and the only thing you can do is change (or not change) the color of a single flower, out of the 20.

Good luck!