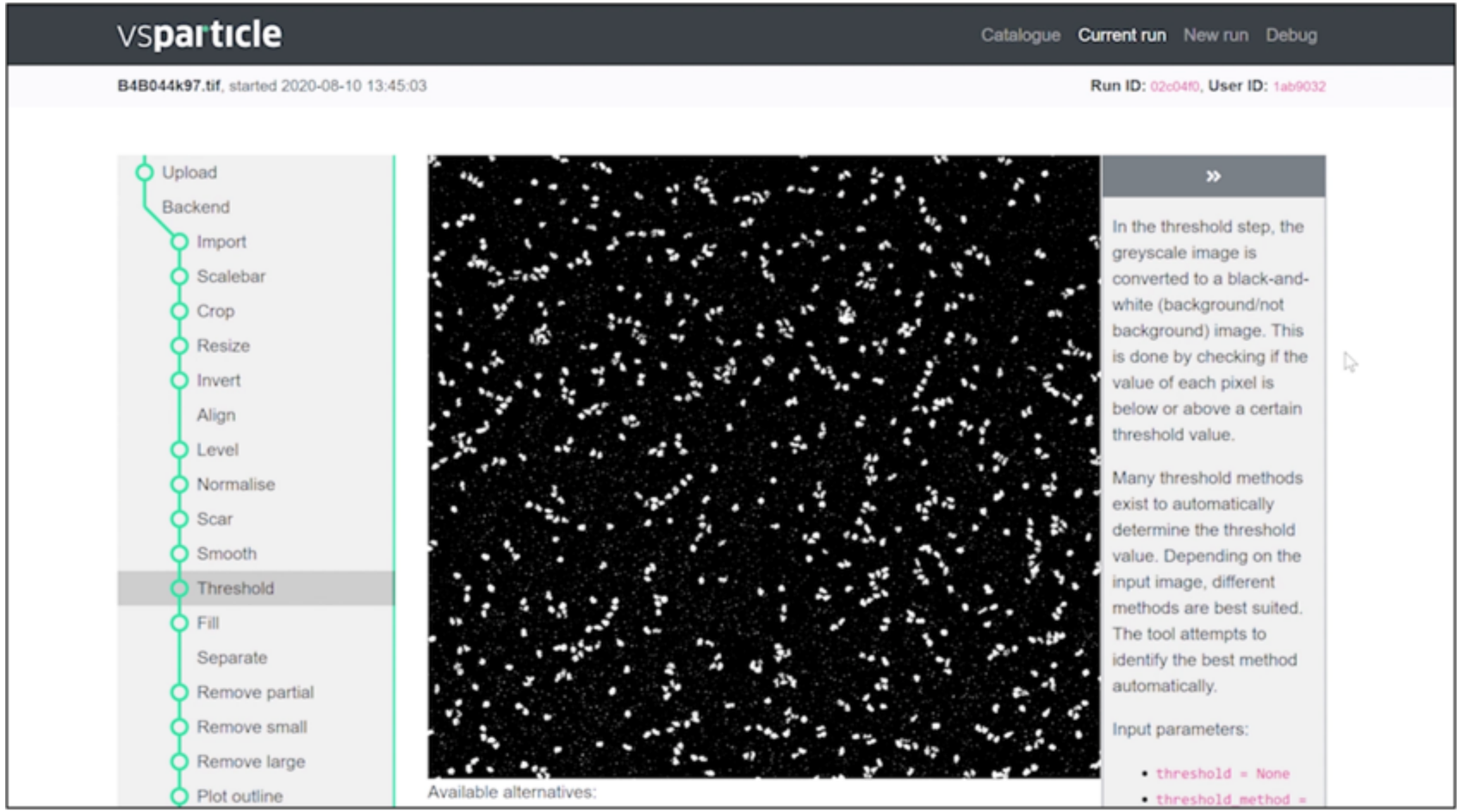
12/10 – 23/10

DANI, KLARA, YORAN, OSCAR

NANO — WEEK 7

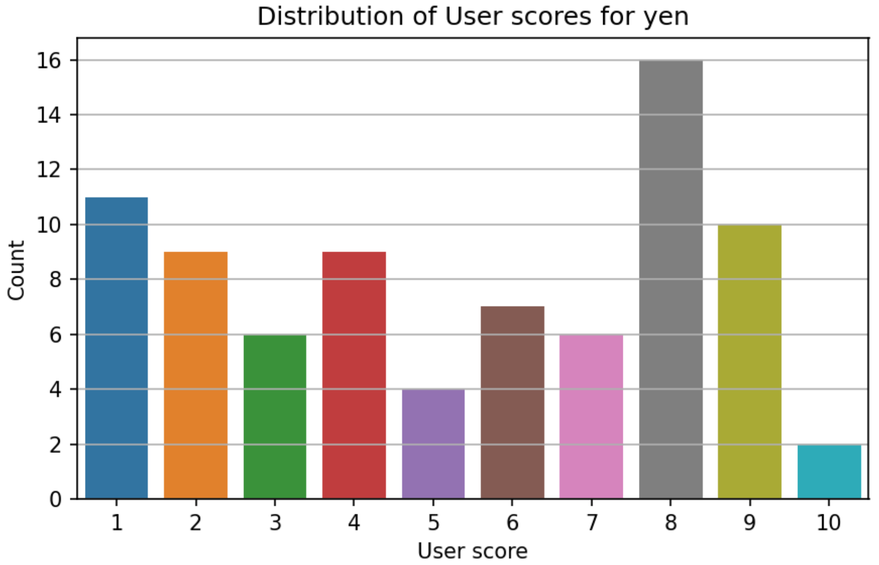
OBJECTIVE

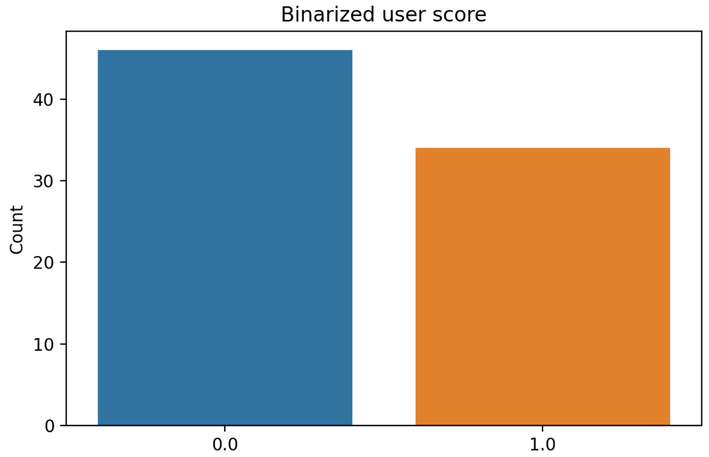
- VSParticle's image analysis software
- Create a model that predicts the thresholding method

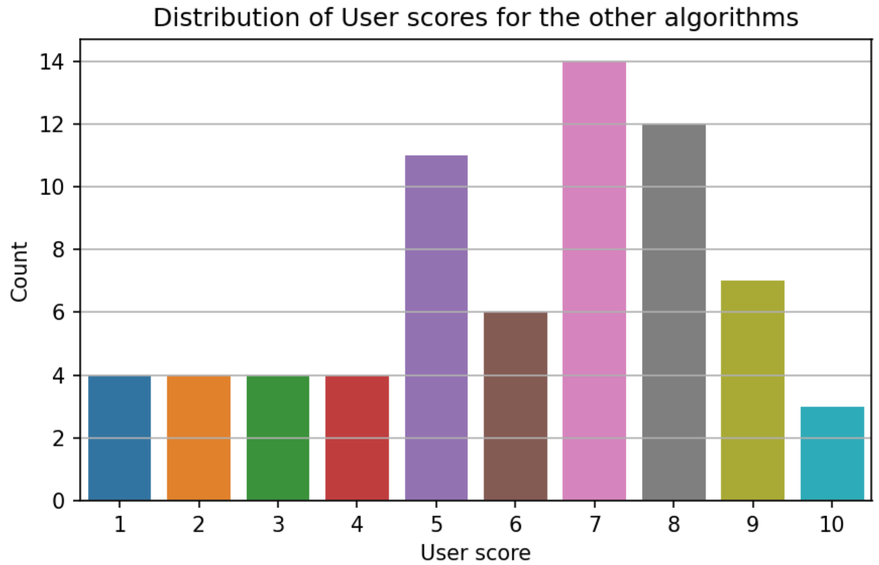


EXPERIMENTAL DATASET

- Dataset containing:
  - Used threshold algorithm
  - Used algorithm feature scores
  - User score
- Only use yen
- Make user score binary
- File contained some faulty values
  - Fixed!







CORRELATION MATRIX

- Find out which scores are closest related to the user score

Pearson method

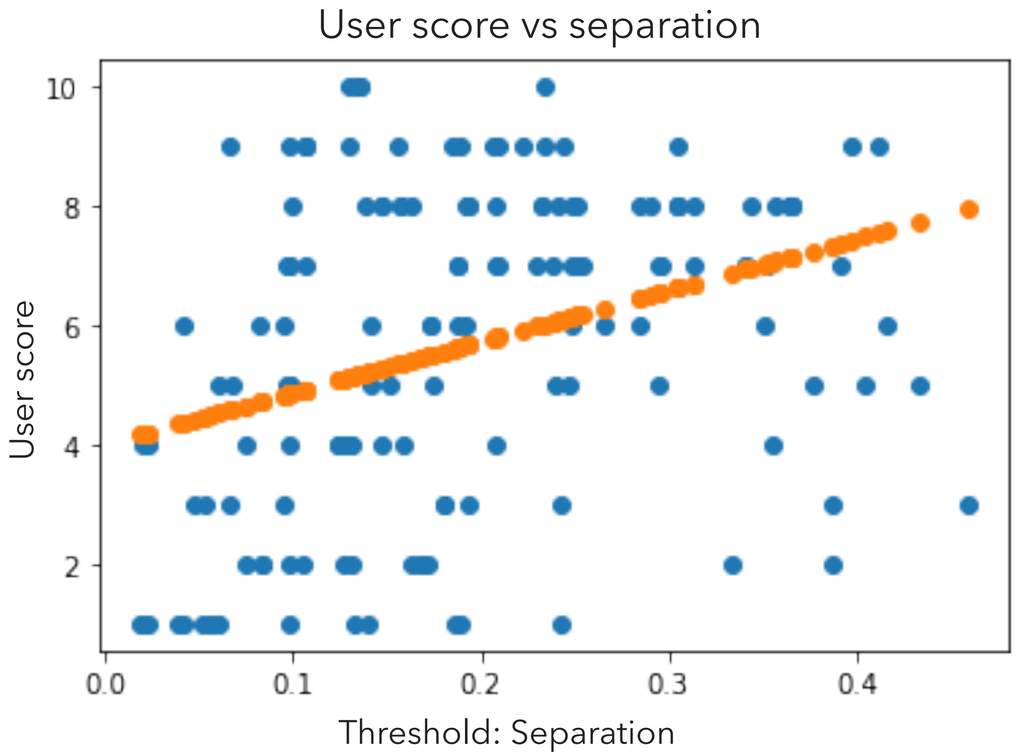
	Threshold: area spread	Threshold: border	Threshold: count	Threshold: fill	Threshold: intensity	Threshold: separation	User score
Threshold: area spread	1.000000	-0.383638	0.161977	0.807861	0.309858	-0.471778	-0.118352
Threshold: border	-0.383638	1.000000	-0.082287	-0.300276	0.462293	0.727519	0.205863
Threshold: count	0.161977	-0.082287	1.000000	0.070393	-0.233837	-0.061663	0.032932
Threshold: fill	0.807861	-0.300276	0.070393	1.000000	0.417174	-0.363062	-0.102722
Threshold: intensity	0.309858	0.462293	-0.233837	0.417174	1.000000	0.171531	0.070804
Threshold: separation	-0.471778	0.727519	-0.061663	-0.363062	0.171531	1.000000	0.438307
User score	-0.118352	0.205863	0.032932	-0.102722	0.070804	0.438307	1.000000

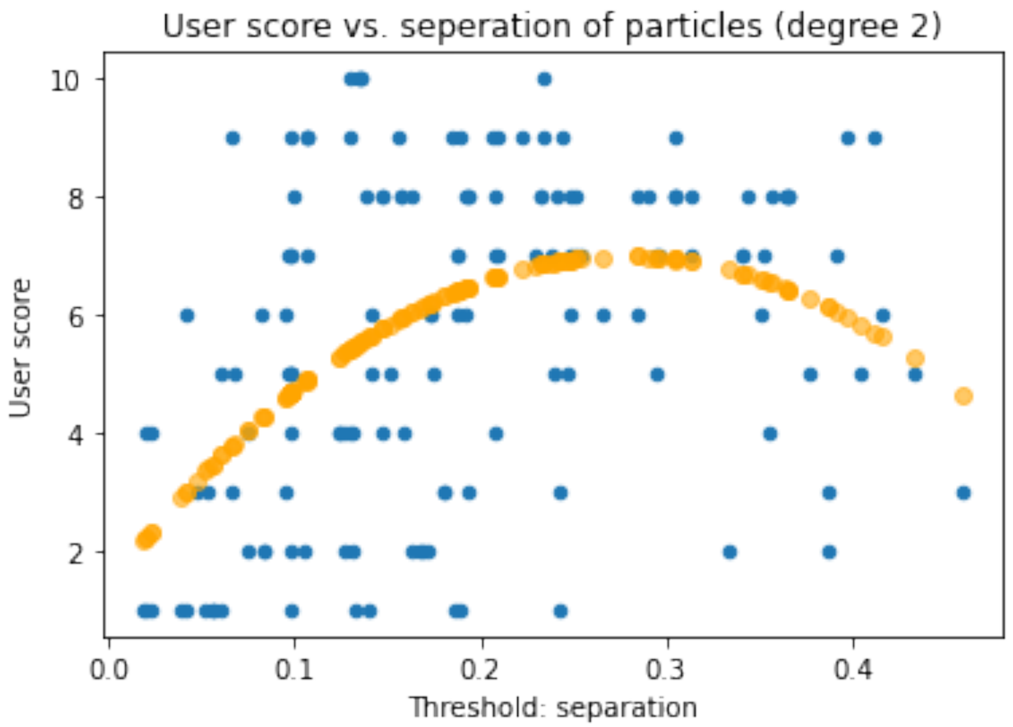
Spearman method

	Threshold: area spread	Threshold: border	Threshold: count	Threshold: fill	Threshold: intensity	Threshold: separation	User score
Threshold: area spread	1.000000	-0.431829	0.430281	0.793319	0.278076	-0.509923	-0.083765
Threshold: border	-0.431829	1.000000	-0.050624	-0.259820	0.292144	0.839448	0.335396
Threshold: count	0.430281	-0.050624	1.000000	0.530522	-0.026856	-0.084076	0.154634
Threshold: fill	0.793319	-0.259820	0.530522	1.000000	0.494403	-0.279252	-0.026909
Threshold: intensity	0.278076	0.292144	-0.026856	0.494403	1.000000	0.199913	0.117228
Threshold: separation	-0.509923	0.839448	-0.084076	-0.279252	0.199913	1.000000	0.451202
User score	-0.083765	0.335396	0.154634	-0.026909	0.117228	0.451202	1.000000

MORE ML MODELS

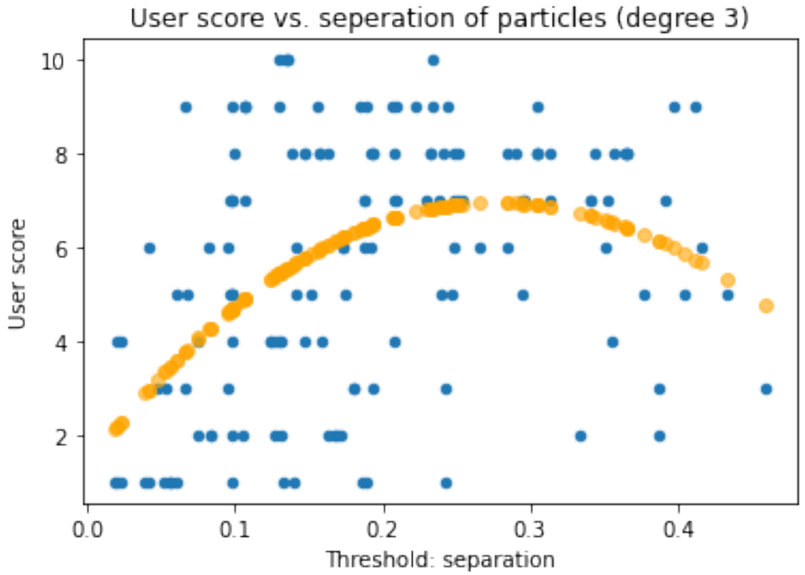
- Use 1 of the feature scores as input, predict user score
  - Linear regression

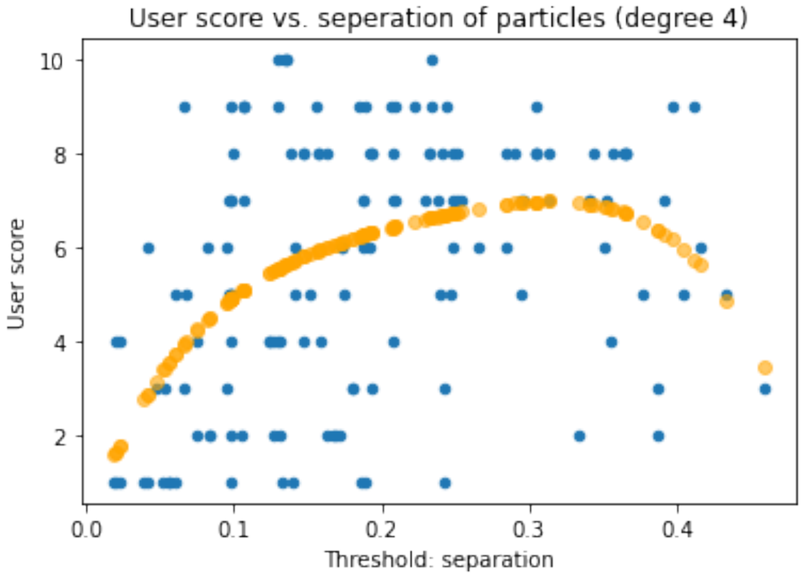


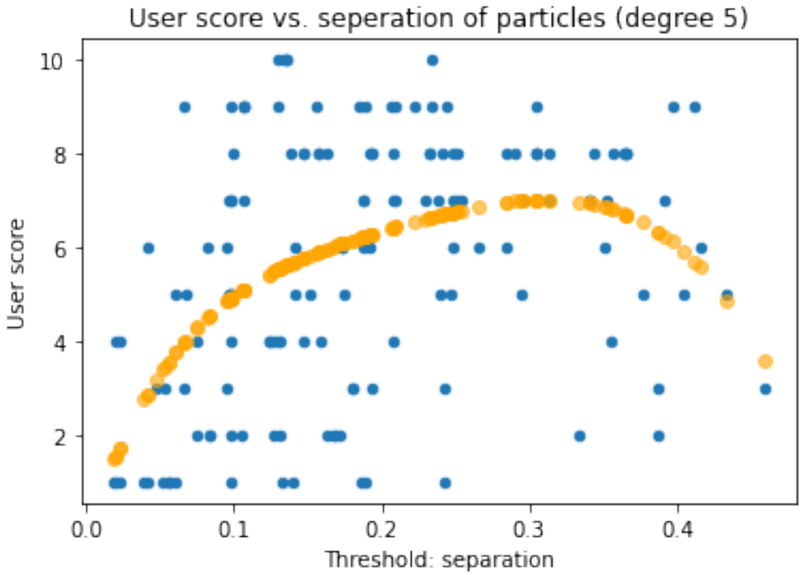


MORE ML MODELS

- Use 1 of the feature scores as input, predict user score
  - Linear regression



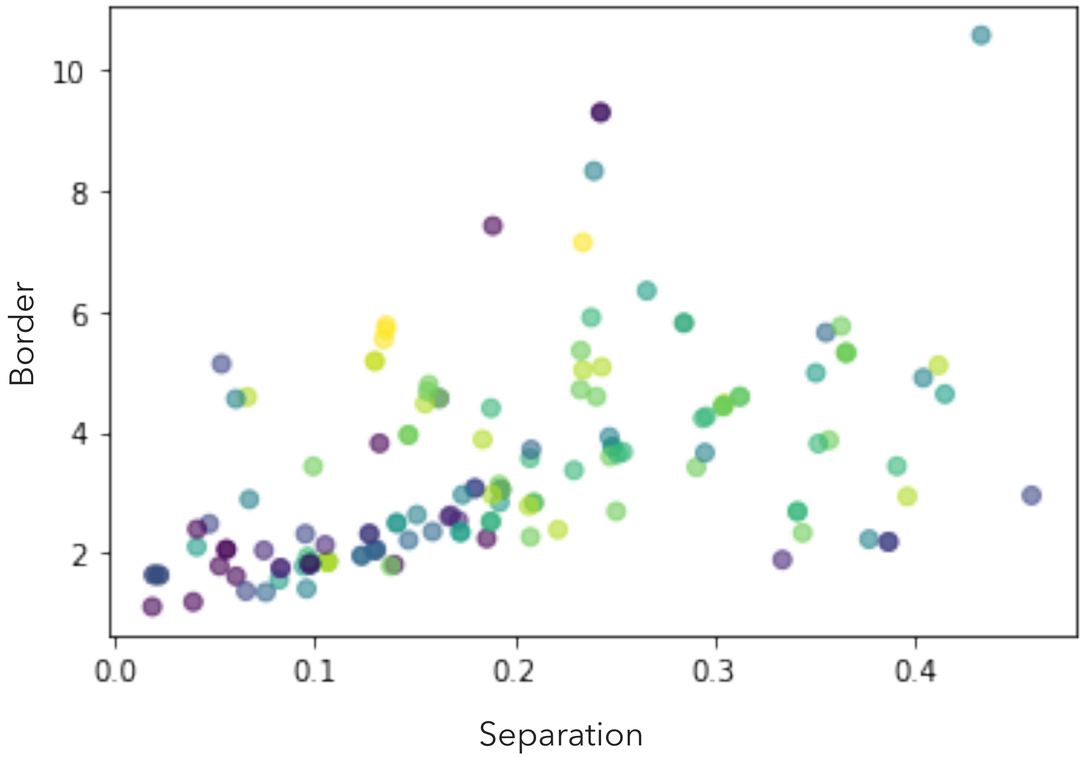




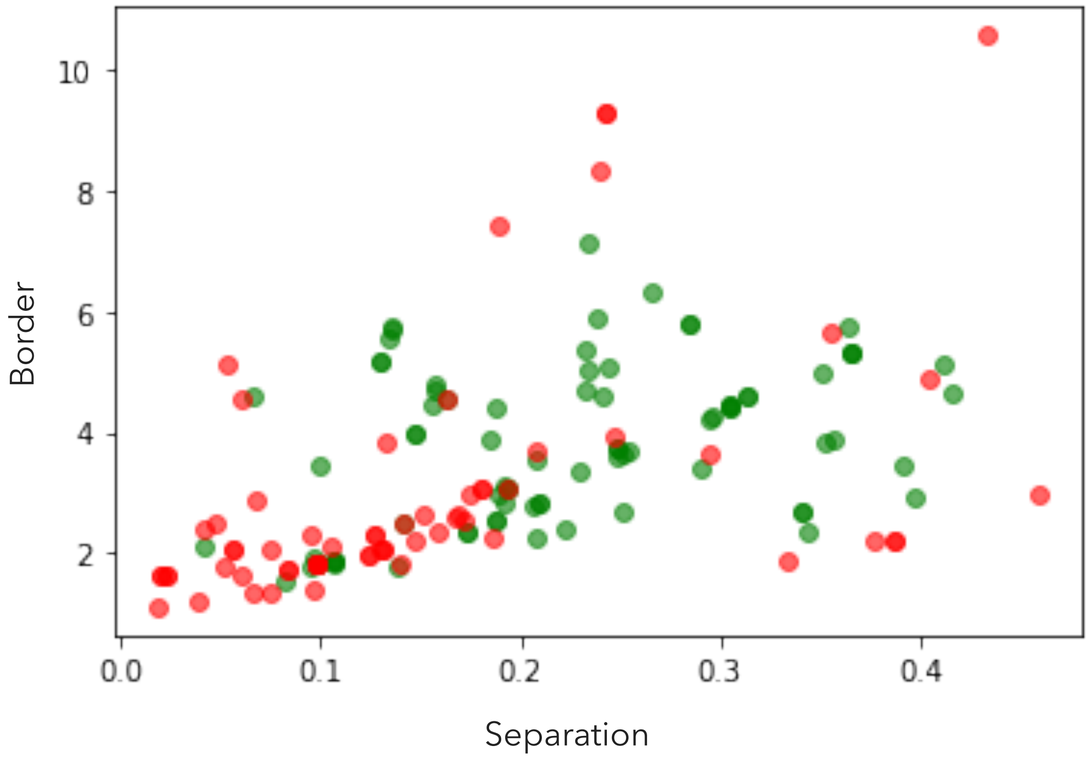
VISUALIZATIONS

- Separation vs. Border score

All user scores



Binary user scores



MEETING WITH PRODUCT OWNER

- NDA: no access to the actual images
- See images of specific runs (outliers)
    - Reasons quickly became clear

COMING WEEK

- Continue experimenting with models and feature combinations
  - Model with multiple features (separation, border and more)
  - Analyse features
- Implement corrected CSV file
- Look into non-linear scaling
- Work on individual portfolios

THANK YOU!

ANY QUESTIONS OR FEEDBACK?