Original Image #1:

With Random Initialization, seed = 1342

K	2	3	10	20	40
MSE	1580.27186392 04127	1007.9519 383571846	167.47270 702309882	82.782904 02001679	48.247059 40029903
Iterations until convergence	15	38	17	25	26

With Random Initialization, seed = 4132

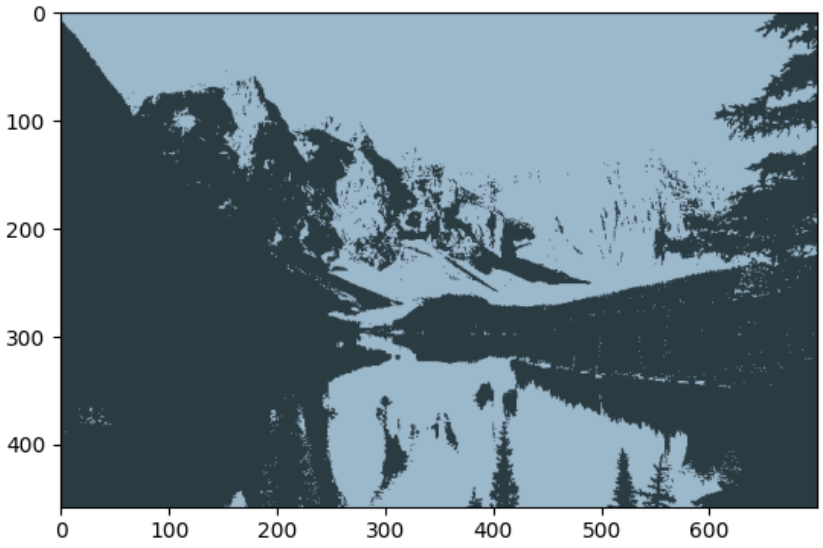
K	2	3	10	20	40
MSE	1594.51 217646 38618	827.78173 95618559	189.87726 73633638	70.252444 4772527	38.222893 76605777
Iterations until convergence	21	36	28	25	56

With Random Initialization and minimum distance, seed = 1342

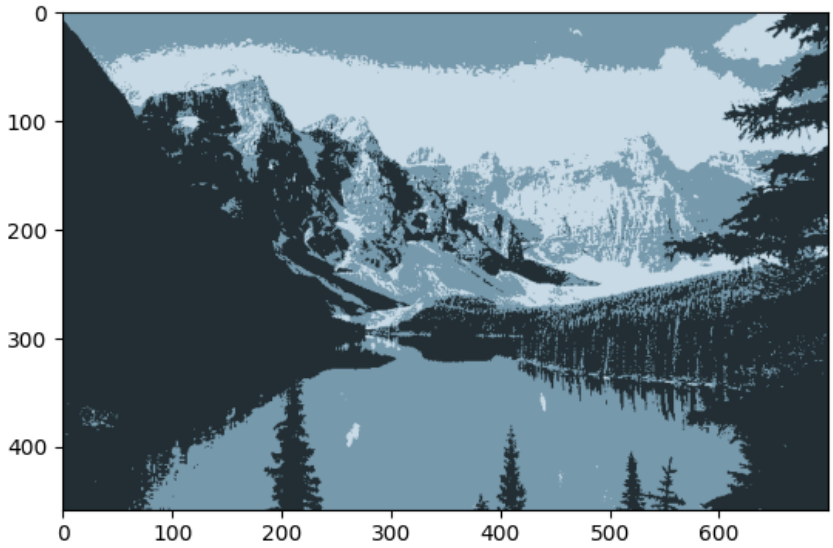
K	2	3	10	20	40
MSE	1560.88072355 41128	705.6168808755369	149.8118676316325	71.256210626326	37.55616888449374
Iterations until convergence	10	10	16	24	29

Pictures with Random Initialization, seed = 1342

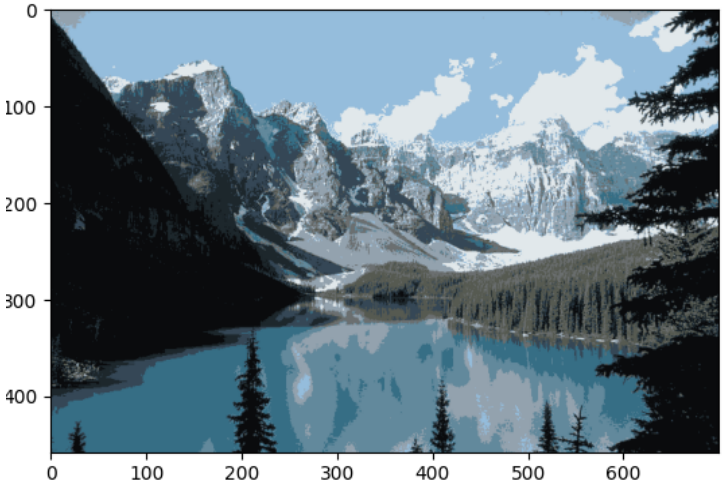
K = 2:



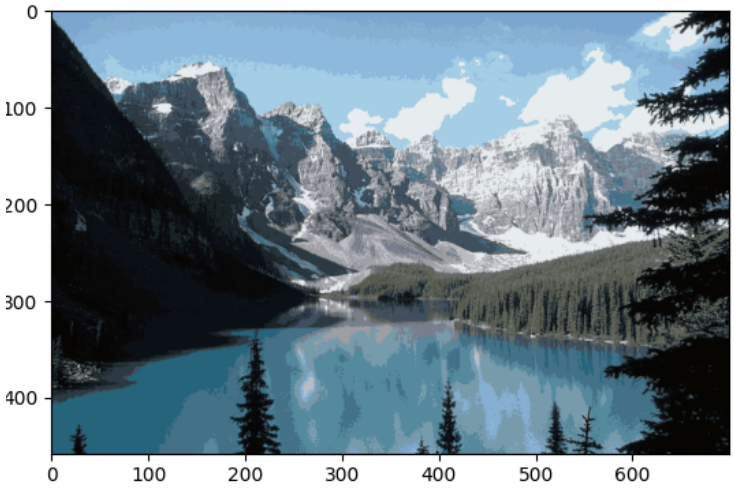
K = 3:



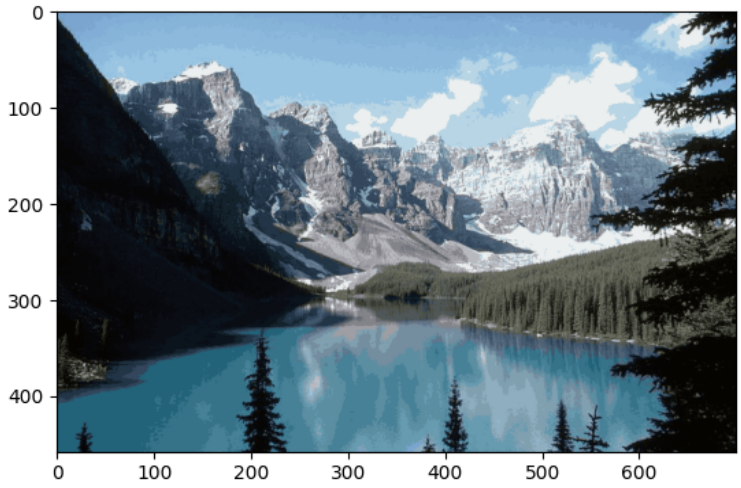
K = 10:


0
100
200
300
400
0
100
200
300
400
500
600

K = 20:


0
100
200
300
400
0
100
200
300
400
500
600

K = 40:


0
100
200
300
400
0
100
200
300
400
500
600

Original Image #2:

With Random Initialization, seed = 1342

K	2	3	10	20	40
MSE	934.555780540 2003	706.5678230140982	196.99872995135826	109.11216621570146	58.33125310342356
Iterations until convergence	6	13	18	21	42

With Random Initialization, seed = 4132

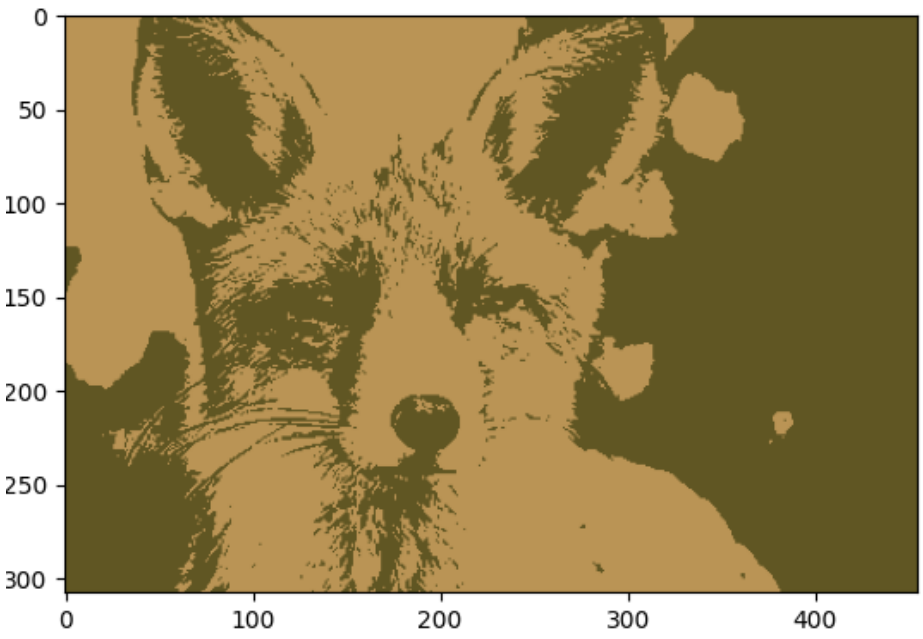
K	2	3	10	20	40
MSE	1011.29456417 50075	709.6135298054027	198.47918582713694	109.97069691397137	63.64481957568519
Iterations until convergence	14	23	20	28	46

With Random Initialization and minimum distance, seed = 1342

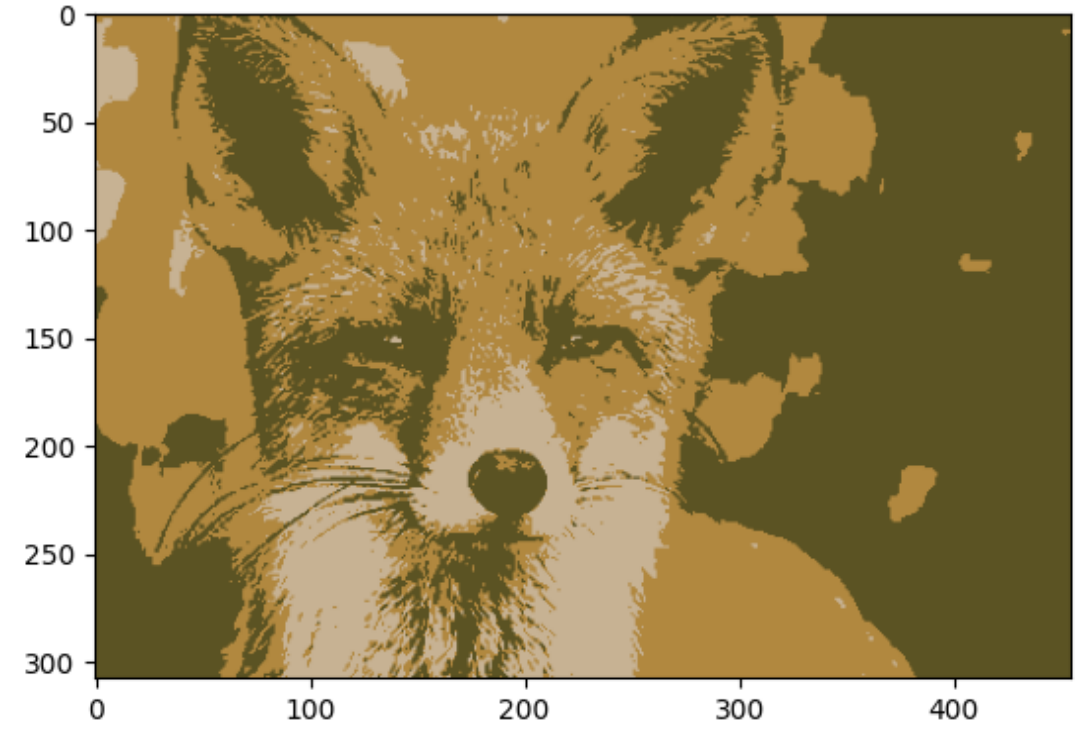
K	2	3	10	20	40
MSE	936.696483818 2356	656.3143349639489	210.5847195596923	105.32414030279718	53.4997952075581
Iterations until convergence	9	5	20	36	32

Pictures with Random Initialization, seed = 1342

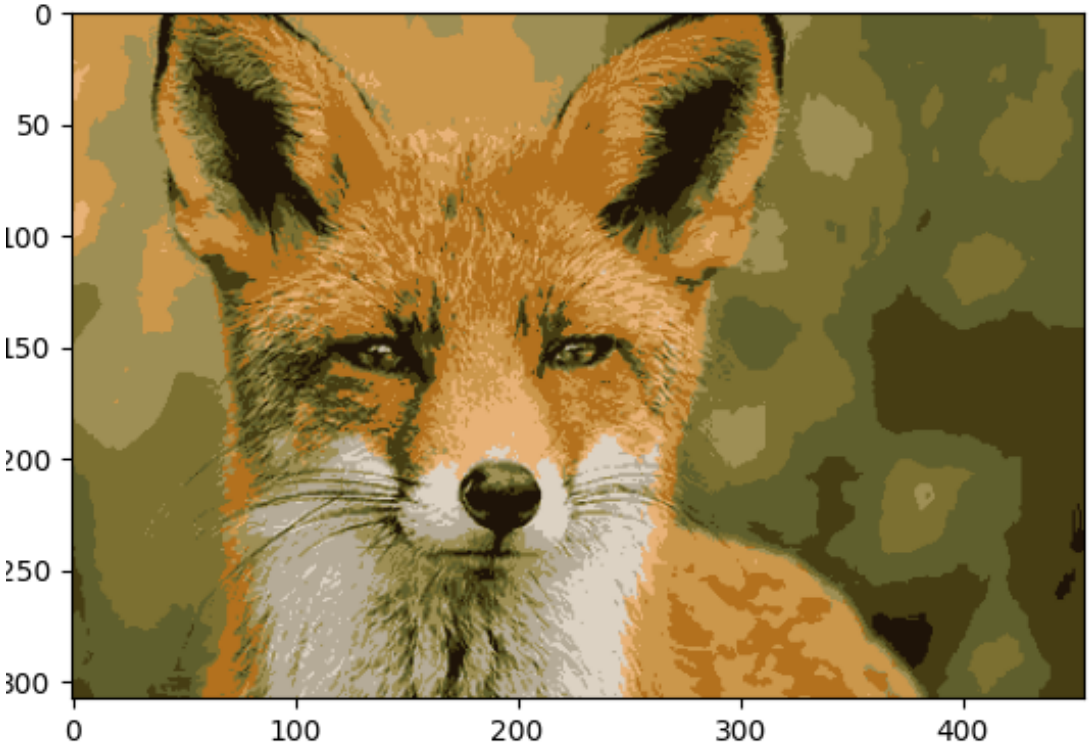
K = 2:



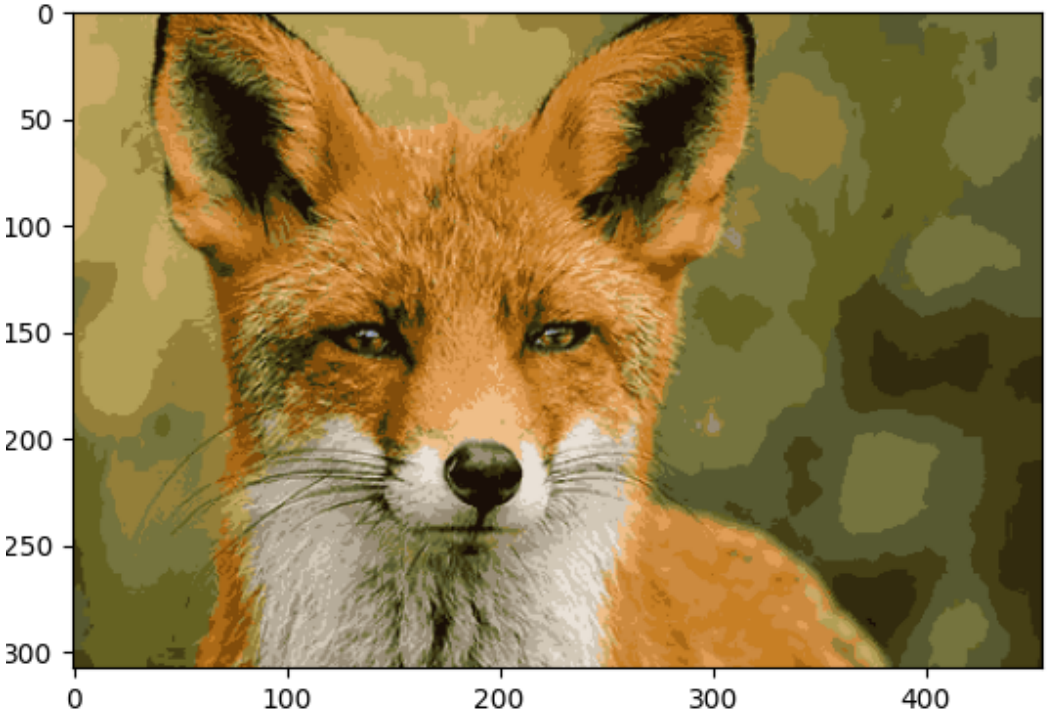
K = 3:


0
50
100
150
200
250
300
0
100
200
300
400

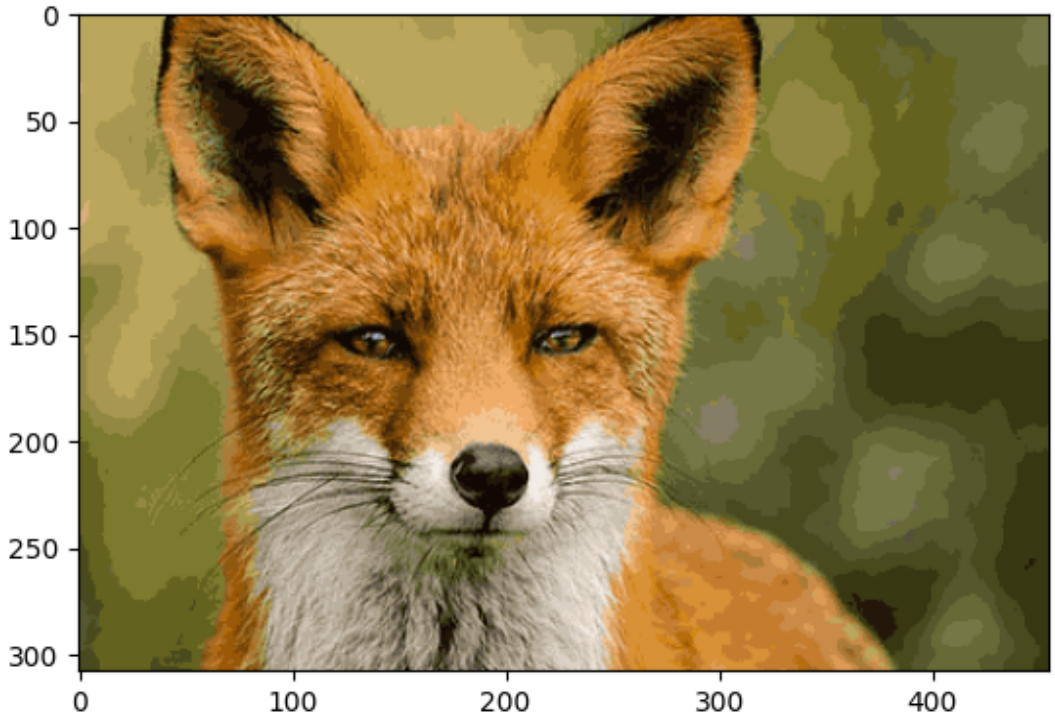
K = 10:


0
50
100
150
200
250
300
0
100
200
300
400

K = 20



K = 40:



Initialization Strategies:

Firstly, a shuffled copy of the pixel array is created. For the first initialization strategy, the first k points from the shuffled copy are used as the initial centres. For the second strategy, a minimum distance is used as a threshold for selecting our centres. The first point in the shuffled copy will be our first centre. The distance of the next point from each centre currently in the list is then computed and checked to be greater than the defined threshold. If so, it's appended as another centre, and the algorithm repeats until ass centres are found. In terms of selecting the minimum threshold, it will depend on the value of k. At lower k values, this threshold can be quite large, since we don't have as many centres to account for, however this initialization strategy does not make much of a difference here anyways. As we increase k, the threshold must decrease to be able to find enough centres. This value was roughly estimated to be around 30 for k = 40, 50 for k = 20, and 80 for k = 10.

Discussion:

When comparing the different initialization strategies, we can see a notable improvement in the time it takes to converge. At lower k values, the minimum distance makes almost no difference due to there not being enough random centres. However, at larger k values the algorithm is able to converge faster, and to a lower mean squared error. Additionally, we can see that k is directly inversely proportional to the mean squared error. As k increases, MSE decreases quickly, which is expected. However, as mentioned above, the initialization strategy makes little difference on the MSE, even on higher values of k (could be due to only testing it with one seed). In terms of the image quality, we can see that it is also proportional to k, and MSE.