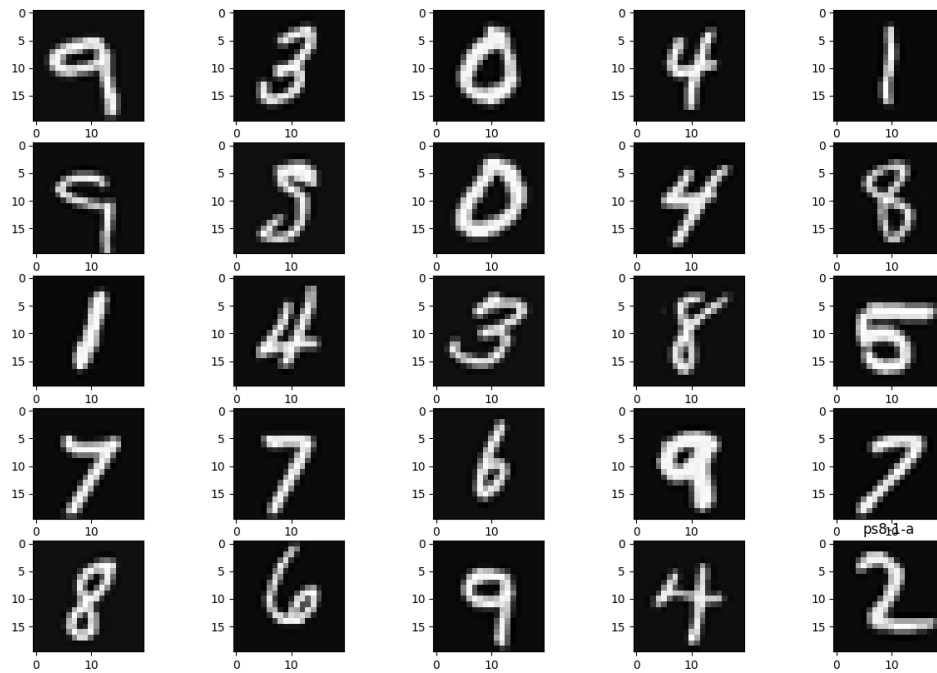Name: Timothy Horrell
ID: 4475990

Questions:
1) .
   a. *Image output*

ps8-1-a.png

   b. .
   c. .
   d. **One-vs-One SVM**
      i. Error on X1: 2.89%

      ii. Error on X2: 12.56%
          Error on X3: 11.56%
          Error on X4: 11.00%
          Error on X5: 11.33%

      iii. Error on Test: 11.2%

e. **KNN (k=3)**
   i. Error on X2: 4.78%

   ii. Error on X1: 12.22%
       Error on X3: 10.78%
       Error on X4: 11.11%
       Error on X5: 10.33%

   iii. Error on Test: 10.60%
f. **Logistic Regression**
   i. Error on X3: 0.67%

   ii. Error on X1: 12.67%
       Error on X2: 13.44%
       Error on X4: 12.33%
       Error on X5: 12.78%

   iii. Error on Test: 11.00%
g. **Decision Tree**
   i. Error on X4: 0.00%

   ii. Error on X1: 34.11%
       Error on X2: 33.00%
       Error on X3: 31.33%
       Error on X5: 33.44%

   iii. Error on Test: 30.40%
h. **Random Forest**
   i. Error on X5: 0.00%

   ii. Error on X1: 12.11%
       Error on X2: 10.33%
       Error on X3: 9.67%
       Error on X4: 9.56%

   iii. Error on Test: 9.80%
i. **Majority Voting Rule**
   i. Error on Test: 7.00%
j. **Discussion:**
   i. Each classifier performs very well on the training set it was trained with. There is around 0-4% error on the training set it was trained on, and anywhere from 10-30% testing error. The majority rule reaches a 7% testing error.
   ii. The SVM, KNN, and logistic regression classifiers perform similarly, with a low 0-4% training error, but 10-12% testing error. Both trees achieve 0% training error.

The decision tree has around 30% testing error however, which is indicative of an overfit on the data. The random forest is more reasonable and has testing error like the first three classifiers which is around 10%. Overall, Random Forest, SVM, KNN, and Logistic Regression seem to achieve the same performance on their own, with the Decision Tree performing poorly in comparison. When bagging these together and taking a vote, a test error of 7% is achieved, which is lower than any classifier individually. A roughly 3% drop in error is significant enough to show that taking a vote between classifiers is helpful in producing a more accurate model. I believe it could be even more helpful if the decision tree was removed, or improved, as it could confound some votes given its inaccuracy.

2) .

**I did not get the third image to output cleanly. I believe it had something to do with being size M, N, 4, rather than 3. Below is the best I was able to get, which shows the attributes of the city skyline, but doesn't represent the image correctly. I will be looking at the panda and cardinal both and ignoring this third image when evaluating effects of parameters.
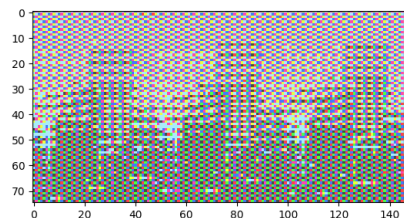

When comparing all the parameters, it seems K has the largest visual effect of any of them. By adding more clusters, one can see a sharper distinction of colors because areas where colors would previously merge can become two different clusters. The panda image below shows max iterations, max restarts, for 3, 5, and 7 clusters (left to right). The leaf detail increases with each cluster increase, and eyes, paws, and shadows show up on the panda.
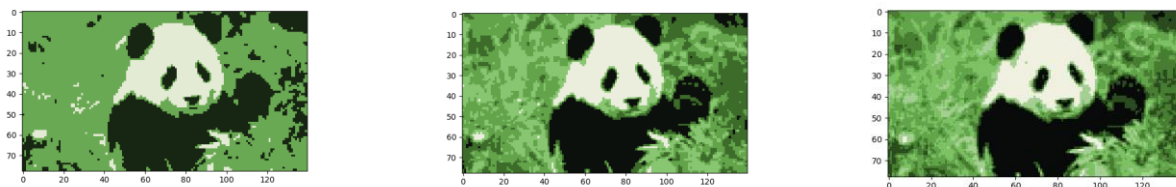

Next, the number of iterations seemed to have minimal effect on the images. The assumed effect of iterations is to get the closest and refined clusters, and therefore the most accurate representation of

the image by colors. The birds below are all 3 clusters, 25 randomization birds, with 7, 13, and 20 iterations. There is minimal differences between them.
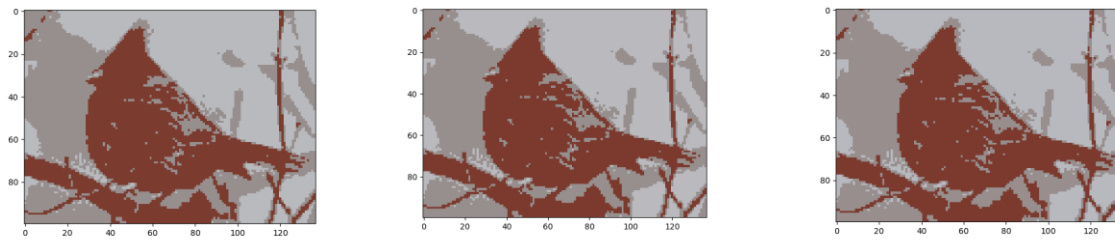
Finally, the number of randomizations does play a large role in consistent image output. The best example of this is in the cardinal set, specifically the 3-cluster portion. Out of the nine images, only one captured the bright red of the cardinal's body as shown below (3 cluster, 20 iterations, 5 randomizations):
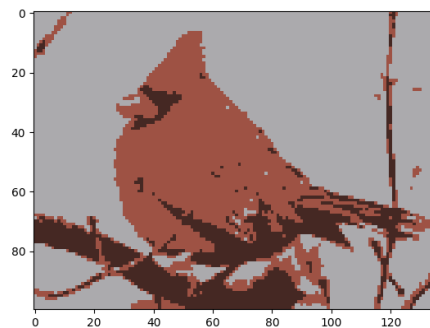
It makes sense this would be an outlier at only 5 random starting points. When dialed in to 25 randomizations, the image output appears much more standard. It is safe to assume that with greater randomization, the SSD is minimized more and more. The next images are cardinals with 25 randomizations and 7 clusters, that change the number of iterations. You can see the middle bird lacks texture on its back; however, all birds and backgrounds are closely the same. That is the effect that randomization has.
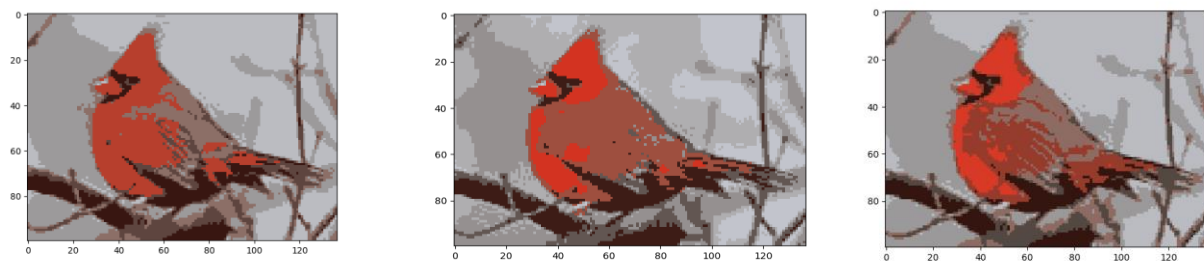
Overall, the number of clusters has the greatest effect, followed by randomizations, and lastly iterations. It seems with a reasonable number of iterations; the clusters will converge. Visually, 7 to 20 iterations make a small difference. Larger random instances help get consistent output and avoid outliers, and a larger number of clusters adds a lot of detail.