Experiment 1:

Due to many failures of creating a 8x8 pixel that is clear and good enough to seen clearly, I have decided to omit them in the report, and compensate with the associated digits.

Experiment 1 run 1:

Associated digits: 8, 0, 1, 5, 7, 6, 3, 4, 4, 2

Average MSE: 0.00

MSS: 1293.44

Mean entropy: 1.00

Accuracy: 68.84%

Confusion matrix:

	0	1	2	3	4	5	6	7	8	9
0	90	0	0	0	88	0	0	0	0	0
1	0	58	21	1	0	0	4	0	98	0
2	1	2	149	9	0	0	0	3	13	0
3	0	0	0	165	0	1	0	9	8	0
4	0	5	0	0	162	0	0	6	8	0
5	1	1	0	30	1	148	1	0	0	0
6	1	0	0	0	1	0	176	0	3	0
7	0	6	0	0	1	2	0	167	3	0
8	1	8	1	34	0	4	2	2	122	0
9	0	23	0	145	0	6	0	5	1	0

From what I can gather from the confusion matrix, it is more accurate than I thought it would be. Due to classifying the clusters by most frequent digit, it happens to lead a clustering bias away from the value 9. It is possible the attributes in each cluster are the same for specific values, leading to a confusion with other values that have similar attributes.Also, the value 1 is being confused with the value 8, which possibly means the the closest centroid, for some of the one values is the centroid associated with the value 8.

Experiment 1 run 2:

Associated digits: 4, 2, 0, 6, 7, 1, 6, 3, 5, 8

Average MSE: 1.00

MSS: 1310.17

Mean entropy: 1.00

Accuracy: 73.79%

Confusion matrix:

	0	1	2	3	4	5	6	7	8	9
0	176	0	0	0	2	0	0	0	0	0
1	0	57	18	1	0	0	8	0	98	0
2	1	2	149	9	0	0	0	3	13	0
3	0	0	0	165	0	1	0	9	8	0
4	0	5	0	0	162	0	0	6	8	0
5	0	1	0	31	1	148	1	0	0	0
6	0	0	0	0	0	0	180	0	0	0
7	0	7	0	0	1	1	0	167	3	0
8	1	8	1	34	0	4	2	2	122	0
9	0	23	0	145	0	6	0	5	1	0

Looking at the confusion matrix, again, the value 9 does not seem to be predicted correctly at all, leading to another case of confusion. This time, the value 1 is getting confused with the value 8, meaning the closest centroid is associated with the value 8, leading to consistency. When it comes to the other variables, they are pretty accurate.

Experiment 1 run 3

Associated digits: 2, 6, 6, 5, 7, 3, 8, 1, 4, 7

Average MSE: 1.00

MSS: 1311.91

Mean entropy: 1.00

Accuracy: 64.22%

Confusion matrix:

	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	2	0	0	176	0	0
1	0	58	18	1	0	0	8	0	97	0
2	0	2	148	10	0	0	0	4	13	0
3	0	0	0	165	0	1	0	9	8	0
4	0	5	0	0	162	0	0	6	8	0
5	0	1	0	26	1	153	1	0	0	0
6	0	0	0	0	0	1	179	0	1	0
7	0	5	0	0	1	4	0	166	3	0
8	0	8	1	34	0	4	2	2	123	0
9	0	23	0	145	0	6	0	5	1	0

In this run, the values 0 and 9 are being confused, being confused with 7 and 3 respectively. The value 9 seems to be a big problem in this run and the previous runs, which is puzzling. The value 1 is being confused as well, however, it one of the associated digits, which leads to just an attribute problem with the value 8, which is also puzzling. Overall, there is consistency in the value 9 being not in the associated digits.

Experiment 1 run 4:

Associated digits: 3, 6, 7, 0, 2, 4, 5, 1, 8, 6

Average MSE: 4.00

MSS: 1311.61

Mean entropy: 0.99

Accuracy: 74.01%

Confusion matrix:

	0	1	2	3	4	5	6	7	8	9
0	176	0	0	0	2	0	0	0	0	0
1	0	58	18	1	0	0	8	0	97	0
2	1	2	148	10	0	0	0	3	13	0
3	0	0	0	165	0	1	0	9	8	0
4	0	5	0	0	162	0	0	6	8	0
5	0	1	0	26	1	153	1	0	0	0
6	0	0	0	0	0	1	179	0	1	0
7	0	6	0	0	1	3	0	166	3	0
8	1	8	1	34	0	4	2	1	123	0
9	0	23	0	145	0	6	0	5	1	0

Looking at the confusion matrix, the value 9 consistently is not being part of the associated digits, there leading to confusion with another value, in this case, 3. The value 1 is again being confused with the value 8. This is happening to be a correlating pattern, where 9 is not being associated to any of the clusters and 1 is being confused with 8. The only hypothesis I can come

up with is 1s' attributes are closer to the centroid associated with the value 8, which the following runs are beginning to confirm due to its consistency of leading to this hypothesis.

Experiment 1 run 5

Associated digits: 7, 2, 0, 3, 2, 4, 1, 8, 5, 6

Average MSE: 1.00

MSS: 1300.14

Mean entropy: 0.97

Accuracy: 74.23%

Confusion matrix:

	0	1	2	3	4	5	6	7	8	9
0	176	0	0	0	2	0	0	0	0	0
1	0	56	23	1	0	0	2	0	100	0
2	0	1	163	0	0	0	0	2	11	0
3	0	0	2	163	0	1	0	9	8	0
4	0	5	0	0	162	0	0	6	8	0
5	0	1	0	31	1	148	1	0	0	0
6	1	0	0	0	1	0	176	0	3	0
7	0	7	0	0	1	1	0	168	2	0
8	1	8	1	34	0	4	2	2	122	0
9	0	23	0	145	0	6	0	5	1	0

Looking at the confusion matrix, the value 9, again, is being confused with the value 3 and is not along the associated digits, confirming my hypothesis, again. The value 1 is being confused with the value 8 again, which confirms my hypothesis even more. Overall, the k-means clustering is a

simple algorithm for supervised learning, but it does lead to confusion if the associating is to the wrong digit. If each cluster was for the each corresponding digit, then the k-means might work,but there is no guarantee.