Emily Shepherd
CS654

What I Did and What I Learned:
I already had some experience with KMeans Clustering through a Udemy course I have taken. While I understood the theory behind the method, I had only worked with using kmeans imported from sklearn.cluser. For this assignment, I went through the hardcoding. First, I went through the demonstration with two clusters. Then I modified the code to use three. I checked my results using sklearn. The two methods had similar outcomes.
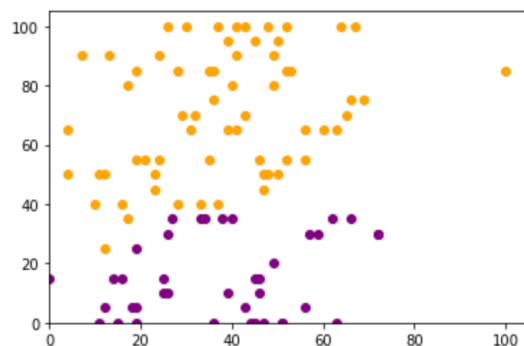
Initial Clustering: (k =2)

Initial Centers: (11, 50), (26, 10)

Cluster_1: [19, 55] [48, 100] [49, 80] [67, 100] [100, 85] [10, 40] [11, 50] [48, 50] [52, 85] [46, 55] [40, 80] [52, 100] [56, 65] [12, 25] [28, 85] [4, 65] [41, 65] [17, 80] [60, 65] [24, 90] [64, 100] [65, 70] [31, 65] [39, 65] [52, 55] [29, 70] [47, 45] [21, 55] [33, 40] [36, 85] [37, 40] [47, 50] [4, 50] [7, 90] [12, 50] [23, 45] [37, 100] [45, 95] [28, 40] [49, 90] [13, 90] [23, 50] [69, 75] [36, 75] [41, 100] [35, 55] [41, 90] [56, 55] [16, 40] [32, 70] [19, 85] [35, 85] [43, 100] [66, 75] [53, 85] [26, 100] [50, 95] [50, 50] [63, 65] [24, 55] [30, 100] [17, 35] [39, 95] [43, 70]

Cluster_2: [38, 35] [46, 10] [47, 0] [57, 30] [72, 30] [12, 5] [49, 20] [19, 25] [72, 30] [40, 35] [11, 0] [15, 0] [56, 5] [59, 30] [25, 15] [62, 35] [39, 10] [63, 0] [0, 15] [14, 15] [18, 5] [19, 5] [25, 10] [27, 35] [19, 0] [51, 0] [34, 35] [66, 35] [36, 0] [44, 0] [33, 35] [45, 15] [46, 15] [26, 30] [45, 0] [43, 5] [26, 10] [16, 15] [51, 0]

Clustering Visualization:

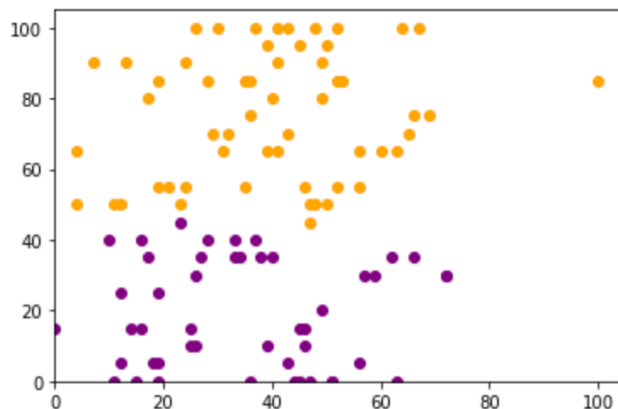SSE of Cluster_1: 213315
SSE of Cluster_2: 44385

Final Clustering:

Final Centers: (39, 75), (34, 19)

Final Cluster_1: [19, 55] [48, 100] [49, 80] [67, 100] [100, 85] [11, 50] [48, 50] [52, 85]
[46, 55] [40, 80] [52, 100] [56, 65] [28, 85] [4, 65] [41, 65] [17, 80] [60, 65] [24, 90] [64, 100]
[65, 70] [31, 65] [39, 65] [52, 55] [29, 70] [47, 45] [21, 55] [36, 85] [47, 50] [4, 50] [7, 90]
[12, 50] [37, 100] [45, 95] [49, 90] [13, 90] [23, 50] [69, 75] [36, 75] [41, 100] [35, 55] [41, 90]
[56, 55] [32, 70] [19, 85] [35, 85] [43, 100] [66, 75] [53, 85] [26, 100] [50, 95] [50, 50] [63, 65]
[24, 55] [30, 100] [39, 95] [43, 70]

Final Cluster_2: [38, 35] [46, 10] [47, 0] [57, 30] [72, 30] [10, 40] [12, 5] [49, 20] [19, 25]
[12, 25] [72, 30] [40, 35] [11, 0] [15, 0] [56, 5] [59, 30] [25, 15] [62, 35] [39, 10] [63, 0] [0, 15]
[14, 15] [18, 5] [19, 5] [25, 10] [27, 35] [19, 0] [33, 40] [37, 40] [51, 0] [23, 45] [34, 35][66, 35]
[36, 0] [28, 40] [44, 0] [33, 35] [45, 15] [46, 15] [16, 40] [26, 30] [45, 0] [43, 5] [26, 10] [16, 15]
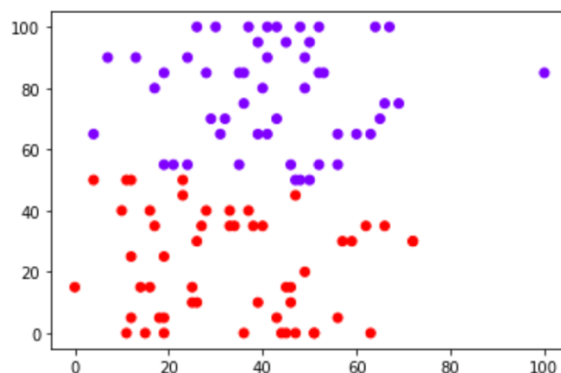[17, 35] [51, 0]

Final Clustering Visualization:

SSE of Cluster_1: 208514   **SSE of Cluster_2:** 55616

I also completed the clustering by importing KMeans from sklearn.cluster. I obtained near identical results. Here is the visualization:
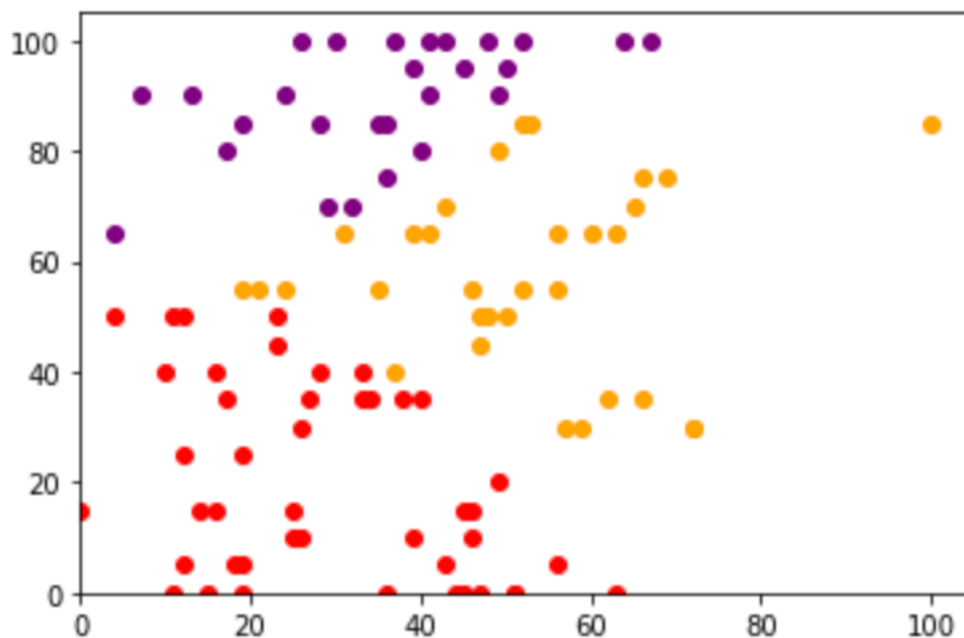
Initial Clustering (k = 3):

Initial Centers: (56, 55) (26, 100) (25, 15)

Cluster_ 1:[19, 55] [49, 80] [57, 30] [72, 30] [100, 85] [48, 50] [52, 85] [46, 55] [56, 65] [72, 30] [41, 65] [59, 30] [60, 65] [62, 35] [65, 70] [31, 65] [39, 65] [52, 55] [47, 45] [21, 55] [37, 40] [47, 50] [66, 35] [69, 75] [35, 55] [56, 55] [66, 75] [53, 85] [50, 50] [63, 65][24, 55] [43, 70]

Cluster_2: [48, 100] [67, 100] [40, 80] [52, 100] [28, 85] [4, 65] [17, 80] [24, 90] [64, 100] [29, 70] [36, 85] [7, 90] [37, 100] [45, 95] [49, 90] [13, 90] [36, 75] [41, 100] [41, 90] [32, 70] [19, 85] [35, 85] [43, 100] [26, 100] [50, 95] [30, 100] [39, 95]

Cluster_3: [38, 35] [46, 10] [47, 0] [10, 40] [11, 50] [12, 5] [49, 20] [19, 25] [12, 25] [40, 35] [11, 0] [15, 0] [56, 5][25, 15] [39, 10] [63, 0] [0, 15] [14, 15] [18, 5] [19, 5] [25, 10] [27, 35] [19, 0] [33, 40] [51, 0] [4, 50] [12, 50] [23, 45] [34, 35] [36, 0] [28, 40] [44, 0] [23, 50] [33, 35] [45, 15] [46, 15] [16, 40] [26, 30] [45, 0] [43, 5] [26, 10] [16, 15] [17, 35] [51, 0]

SSE of Cluster_1: 27834
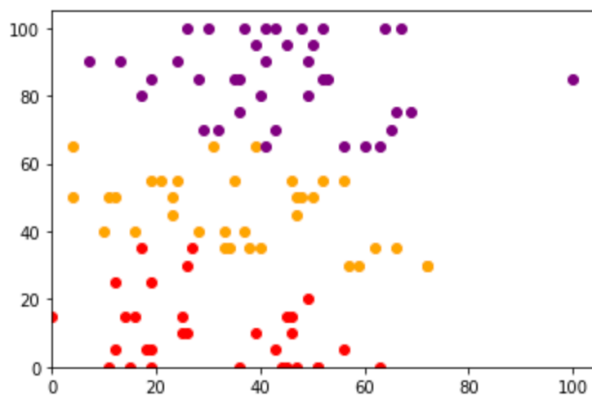SSE of Cluster_2: 20735
SSE of Cluster_3: 40761

Final Clustering (k = 3):

Final Centers: (36, 45) (43, 85) (31, 10)

Cluster_1: [19, 55] [38, 35] [57, 30] [72, 30] [10, 40] [11, 50] [48, 50] [46, 55] [72, 30] [4, 65] [40, 35] [59, 30] [62, 35] [31, 65] [39, 65] [52, 55] [47, 45] [21, 55] [33, 40] [37, 40] [47, 50] [4, 50] [12, 50] [23, 45] [34, 35] [66, 35] [28, 40] [23, 50] [33, 35] [35, 55] [56, 55] [16, 40] [50, 50] [24, 55]

Cluster_2: [48, 100] [49, 80] [67, 100][100, 85][52, 85] [40, 80] [52, 100][56, 65][28, 85][41, 65][17, 80] [60, 65][24, 90][64, 100][65, 70][29, 70][36, 85][7, 90][37, 100][45, 95][49, 90][13, 90][69, 75][36, 75] [41, 100][41, 90][32, 70][19, 85][35, 85][43, 100][66, 75][53, 85][26, 100][50, 95][63, 65][30, 100][39, 95] [43, 70]

Cluster_3:[46, 10][47, 0][12, 5][49, 20][19, 25][12, 25][11, 0][15, 0][56, 5][25, 15][39, 10][63, 0][0, 15] [14, 15][18, 5][19, 5][25, 10][27, 35][19, 0][51, 0][36, 0][44, 0][45, 15][46, 15][26, 30][45, 0][43, 5][26, 10] [16, 15][17, 35][51, 0]

SSE of Cluster_1:46209. **SSE of Cluster_2:**63190   **SSE of Cluster_3:** 21959

I also completed the clustering by importing KMeans from sklearn.cluster. I obtained near identical results. Here is the visualization: