K Means Algorithm using Euclidean distance

**Team Members**

1. **Jahnavi Mesa**
2. **Mohammed Ather Ahmed Shareef**

**Goal:** Implementing the k-means algorithm using Euclidean distance on a dataset with two attributes

**Language Used:** Python

**Summary of program result:**

Cluster Points

0 [ 0 1 7 9 16 52 56 60 72 84 88 93 97 98]

1 [ 6 8 14 22 29 34 42 43 44 45 46 48 61 65 66 75 80 85 86 89]

2 [27 35 36 47 71 81 82 94 95]

3 [ 5 11 12 18 19 23 26 28 33 38 40 50 53 54 57 62 64 74 76 77 87 92 96]

4 [13 17 20 32 39 49 68 73 79 83 90]

5 [ 3 10 21 31 59 67 69 70 91]

6 [15 55]

7 [37 51 58]

8 [ 2 24 25 41 63 99]

9 [ 4 30 78]

Value of Sum of Squared Error (SSE): 0.8848185752431142

**Assumptions:**

Note that data.csv is indexed differently from what is given by Sir. As python uses 0 based indexing, I have renamed ID column starting from 0. Please refer data.csv. Please use data.csv for running the code.

Clusters are named from 0 to K-1

**Run: Blue points are data points, Red cross are initial centroids and Red Square is the final position of centroids**

**python KMeans.py 10 "C:/Users/ather/Desktop/data.csv" "C:/Users/ather/Desktop/result.txt"**

Cluster Points

0 [ 2 3 10 15 31 51 55 58 59 63 69]

1 [ 5 11 12 19 23 37 40 74 76 92 96]

2 [ 8 43 48 65 75 85 89]

3 [ 6 14 61 80 86]

4 [ 0 1 7 16 52 56 60 64 93 97 98]

5 [13 17 20 24 25 32 39 41 49 68 73 79 83 90 99]

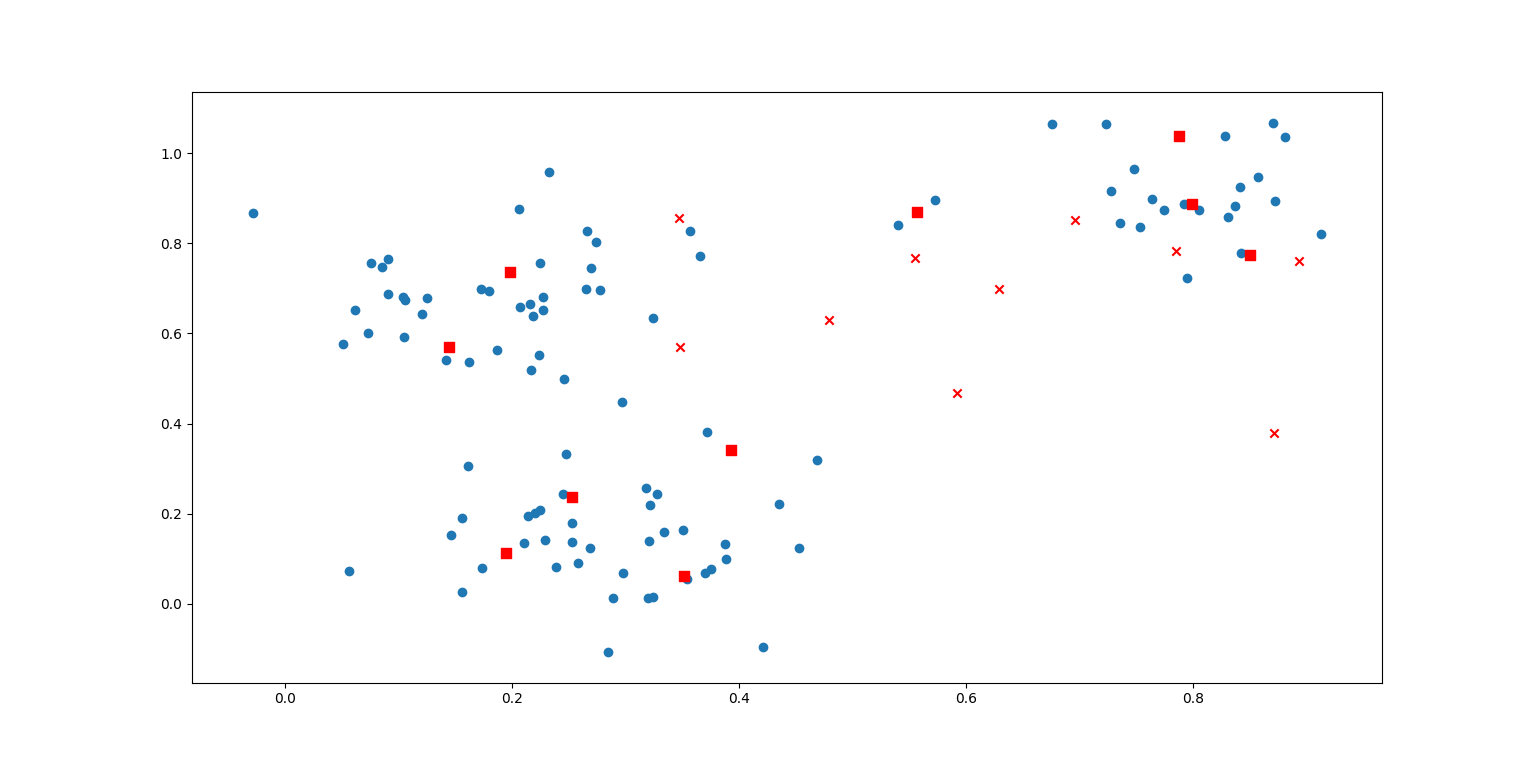
6 [ 9 18 26 28 33 38 50 53 54 57 62 72 77 84 87 88]

7 [21 27 35 36 47 67 70 71 81 82 91 94 95]

8 [ 4 30 42 66]

9 [22 29 34 44 45 46 78]

Value of Sum of Squared Error (SSE): 0.775870512037962



**K=6**

Cluster Points

0 [13 17 20 21 27 32 35 36 39 47 49 67 71 73 81 82 83 90 94 95]

1 [ 3 10 31 51 58 59 63 69 70 91]

2 [ 4 6 8 14 22 29 30 34 42 43 44 45 46 48 61 65 66 75 80 85 86 89]

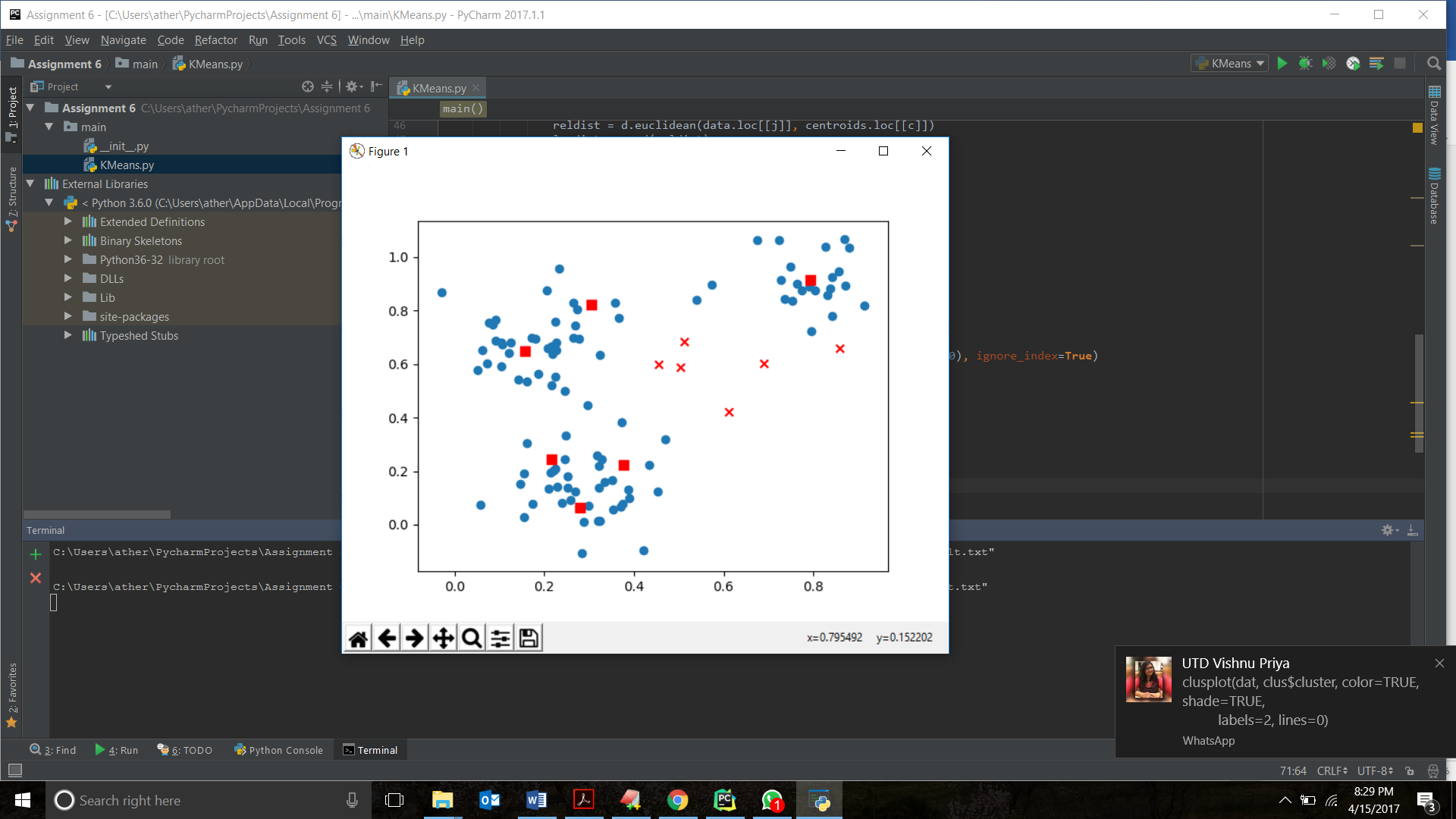
3 [ 2 15 24 25 37 41 55 68 79 99]

4 [ 5 9 11 12 18 19 23 26 28 33 38 40 50 52 53 54 56 57 62 64 72 74 76 77 84

87 88 92 96]

5 [ 0 1 7 16 60 78 93 97 98]

Value of Sum of Squared Error (SSE): 1.2318090079240682



Cluster Points

**K=15**

Cluster Points

0 [14 22 29 34 44 45 46 65 75 80 85 89]

1 [30 78]

2 [35 36 47 81 82 94 95]

3 [13 17 20 32 39 49 68 73 83 90]

4 [ 9 18 19 23 26 28 33 38 40 50 53 54 57 62 72 76 77 84 87 88]

5 [51 58]

6 [ 2 24 25 41 79 99]

7 [ 0 1 7 16 52 56 60 93 97 98]

8 [ 5 11 12 37 64 74 92 96]

9 [ 4 6 42 61 66 86]

10 [10 21 27 31 67 69 70 71 91]

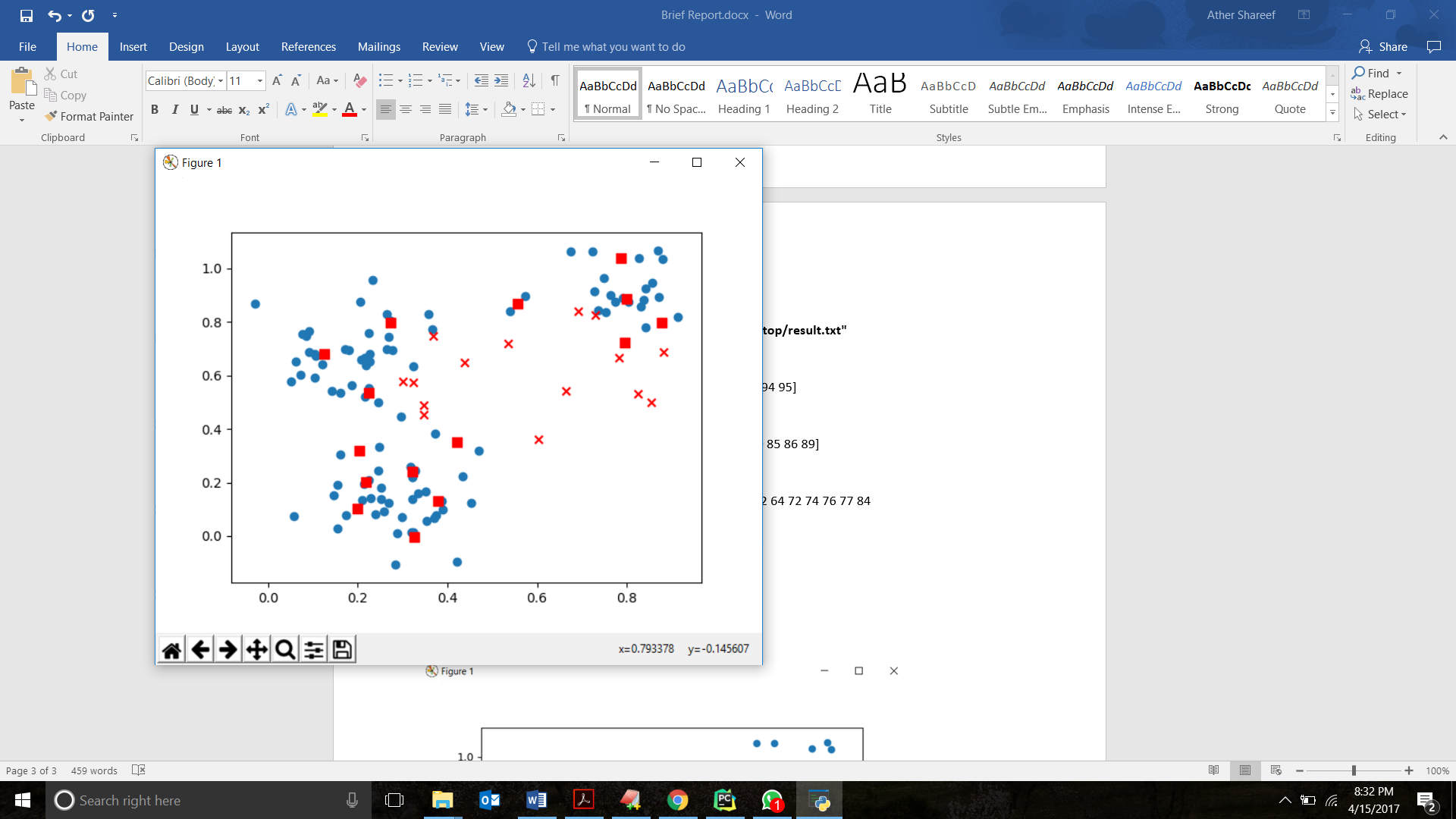
11 [43]

12 [ 3 59 63]

13 [ 8 48]

14 [15 55]

Value of Sum of Squared Error (SSE): 0.5426507682539682



**K=3**

Cluster Points

0 [ 2 3 10 13 15 17 20 21 24 25 27 31 32 35 36 39 41 47 49 51 55 58 59 63 67

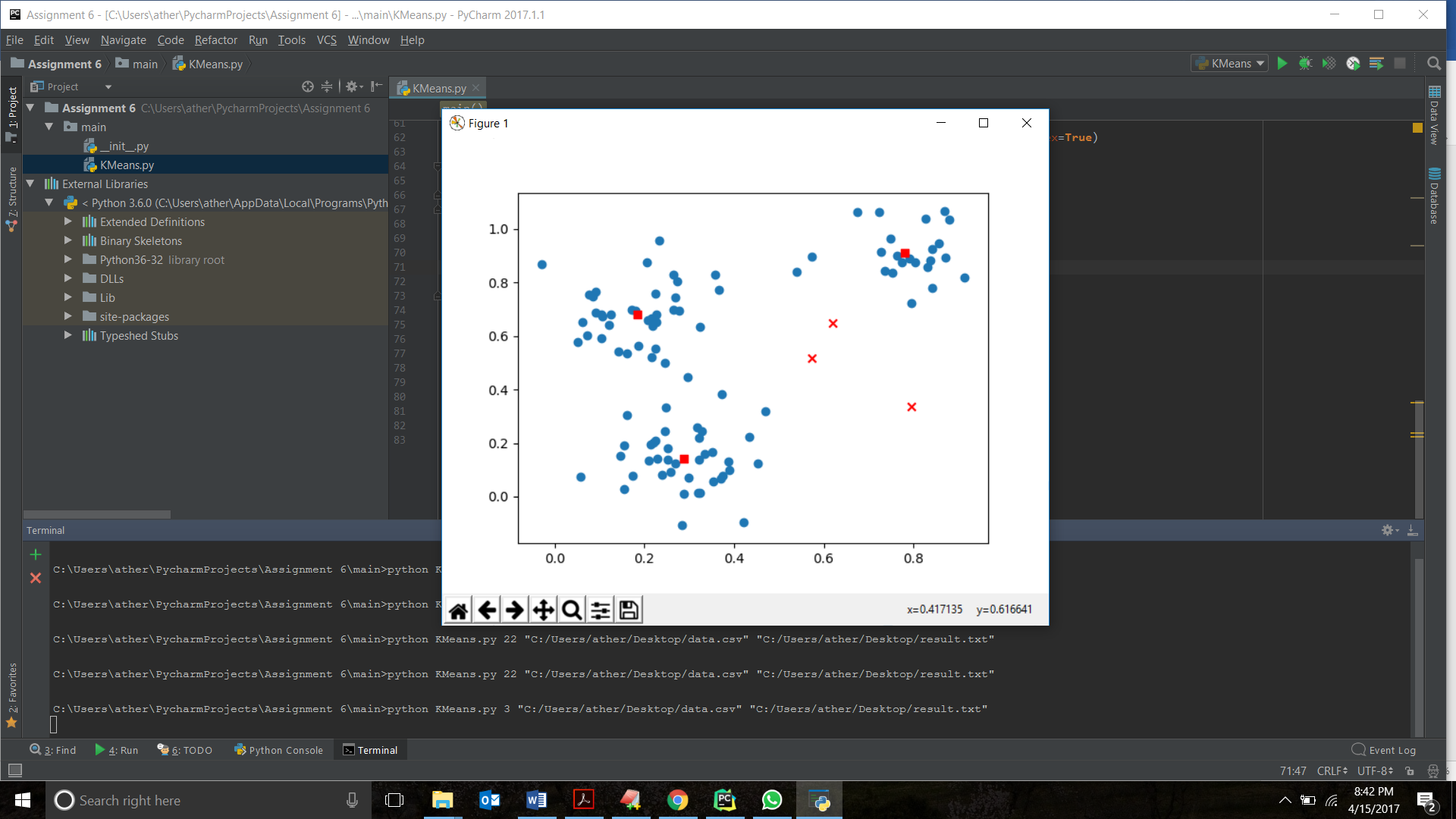
68 69 70 71 73 79 81 82 83 90 91 94 95 99]

1 [ 0 1 5 7 9 11 12 16 18 19 23 26 28 33 37 38 40 50 52 53 54 56 57 60 62

64 72 74 76 77 84 87 88 92 93 96 97 98]

2 [ 4 6 8 14 22 29 30 34 42 43 44 45 46 48 61 65 66 75 78 80 85 86 89]

Value of Sum of Squared Error (SSE): 1.9106917060670068



**K=22**

Cluster Points

0 [10 21 27 31 67 69 70 71 91]

1 [ 9 18 19 26 40 53 54 72 76 77 84 87 88]

2 [43]

3 [13 17 20 24 25 32 39 41 49 68 73 79 83 90 99]

4 [ 1 23 28 33 38 50 52 56 57 60 62 64]

5 [ 5 11 12 37 74 92 96]

6 [ 2 3 59 63]

7 [15 55]

8 [ 8 48]

9 [51 58]

10 []

11 []

12 []

13 [14 22 29 44 45 46 61 65 75 80 85 86 89]

14 []

15 [ 0 7 16 78 93 97 98]

16 []

17 []

18 [ 4 6 30 34 42 66]

19 [35 36 47 81 82 94 95]

20 []

21 []

Value of Sum of Squared Error (SSE): 0.5927122852869353

