CS 6375

ASSIGNMENT 2

Names of students in your group:

Maviya Yaseen (mxy200016)

Khoa Diep (ktd170030)

Farrel Raja (fjr190000)

Number of free late days used: 0   
Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

|  |  |  |
| --- | --- | --- |
| K Value | SSE | Size of Each Cluster |
| 5 | 2586.96 | 1: 661 tweets  2: 487 tweets  3: 882 tweets  4: 927 tweets  5: 282 tweets |
| 10 | 2467.61 | 1: 748 tweets  2: 487 tweets  3: 735 tweets  4: 160 tweets  5: 285 tweets  6: 493 tweets  7: 102 tweets  8: 89 tweets  9: 6 tweets  10: 134 tweets |
| 15 | 2360.58 | 1: 154 tweets  2: 138 tweets  3: 97 tweets  4: 484 tweets  5: 266 tweets  6: 114 tweets  7: 293 tweets  8: 496 tweets  9: 99 tweets  10: 255 tweets  11: 321 tweets  12: 96 tweets  13: 249 tweets  14: 74 tweets  15: 103 tweets |
| 20 | 2330.80 | 1: 309 tweets  2: 566 tweets  3: 32 tweets  4: 136 tweets  5: 127 tweets  6: 59 tweets  7: 107 tweets  8: 93 tweets  9: 171 tweets  10: 253 tweets  11: 175 tweets  12: 161 tweets  13: 43 tweets  14: 62 tweets  15: 172 tweets  16: 100 tweets  17: 277 tweets  18: 41 tweets  19: 213 tweets  20: 142 tweets |
| 30 | 2246.69 | 1: 275 tweets  2: 37 tweets  3: 56 tweets  4: 65 tweets  5: 269 tweets  6: 81 tweets  7: 93 tweets  8: 53 tweets  9: 173 tweets  10: 151 tweets  11: 41 tweets  12: 35 tweets  13: 120 tweets  14: 240 tweets  15: 16 tweets  16: 51 tweets  17: 161 tweets  18: 65 tweets  19: 17 tweets  20: 114 tweets  21: 75 tweets  22: 54 tweets  23: 44 tweets  24: 77 tweets  25: 124 tweets  26: 177 tweets  27: 58 tweets  28: 23 tweets  29: 173 tweets  30: 321 tweets |
| 50 | 2131.07 | 1: 117 tweets  2: 28 tweets  3: 77 tweets  4: 41 tweets  5: 66 tweets  6: 114 tweets  7: 6 tweets  8: 31 tweets  9: 12 tweets  10: 19 tweets  11: 48 tweets  12: 75 tweets  13: 17 tweets  14: 34 tweets  15: 88 tweets  16: 24 tweets  17: 83 tweets  18: 37 tweets  19: 240 tweets  20: 77 tweets  21: 30 tweets  22: 14 tweets  23: 114 tweets  24: 110 tweets  25: 9 tweets  26: 64 tweets  27: 93 tweets  28: 6 tweets  29: 51 tweets  30: 56 tweets  31: 72 tweets  32: 57 tweets  33: 269 tweets  34: 174 tweets  35: 61 tweets  36: 39 tweets  37: 55 tweets  38: 17 tweets  39: 40 tweets  40: 95 tweets  41: 74 tweets  42: 62 tweets  43: 93 tweets  44: 60 tweets  45: 4 tweets  46: 133 tweets  47: 27 tweets  48: 53 tweets  49: 25 tweets  50: 48 tweets |
| 100 | 1907.47 | 1: 104 tweets  2: 4 tweets  3: 26 tweets  4: 27 tweets  5: 126 tweets  6: 22 tweets  7: 19 tweets  8: 21 tweets  9: 11 tweets  10: 66 tweets  11: 15 tweets  12: 43 tweets  13: 33 tweets  14: 31 tweets  15: 49 tweets  16: 38 tweets  17: 15 tweets  18: 54 tweets  19: 83 tweets  20: 20 tweets  21: 85 tweets  22: 34 tweets  23: 20 tweets  24: 29 tweets  25: 10 tweets  26: 15 tweets  27: 52 tweets  28: 59 tweets  29: 110 tweets  30: 25 tweets  31: 10 tweets  32: 18 tweets  33: 8 tweets  34: 45 tweets  35: 25 tweets  36: 25 tweets  37: 26 tweets  38: 18 tweets  39: 14 tweets  40: 41 tweets  41: 11 tweets  42: 28 tweets  43: 1 tweets  44: 34 tweets  45: 23 tweets  46: 32 tweets  47: 29 tweets  48: 13 tweets  49: 21 tweets  50: 68 tweets  51: 47 tweets  52: 3 tweets  53: 14 tweets  54: 56 tweets  55: 16 tweets  56: 8 tweets  57: 30 tweets  58: 7 tweets  59: 11 tweets  60: 54 tweets  61: 99 tweets  62: 23 tweets  63: 31 tweets  64: 3 tweets  65: 36 tweets  66: 58 tweets  67: 17 tweets  68: 26 tweets  69: 23 tweets  70: 66 tweets  71: 17 tweets  72: 5 tweets  73: 42 tweets  74: 21 tweets  75: 53 tweets  76: 16 tweets  77: 14 tweets  78: 8 tweets  79: 49 tweets  80: 38 tweets  81: 64 tweets  82: 19 tweets  83: 49 tweets  84: 7 tweets  85: 42 tweets  86: 34 tweets  87: 22 tweets  88: 35 tweets  89: 20 tweets  90: 13 tweets  91: 2 tweets  92: 1 tweets  93: 27 tweets  94: 32 tweets  95: 43 tweets  96: 3 tweets  97: 88 tweets  98: 70 tweets  99: 15 tweets  100: 26 tweets |

Through our testing, we found that as the K value increases, the SSE, or the Sum of Squares Due to Error, decreases. This means that as there are more clusters being formed, those individual clusters become more tight knit around the centers.