CS 6375

ASSIGNMENT 3

Names of students in your group:

**Anjali Sundareswaran Varier AXV180027**

**Aparna Cleetus AXC190011**

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.

* + We have used the cnnhealth.txt dataset from the below link

<https://archive.ics.uci.edu/ml/datasets/Health+News+in+Twitter>

**PART B – Tweets classification using K-means**

**Output:**

|  |  |  |
| --- | --- | --- |
| **Value of K** | **SSE** | **Size of each cluster** |
| 5 | 3758.1519 | 1: 1611 tweets  2: 364 tweets  3: 533 tweets  4: 821 tweets  5: 728 tweets |
| 10 | 3529.0107 | 1: 835 tweets  2: 187 tweets  3: 726 tweets  4: 208 tweets  5: 262 tweets  6: 90 tweets  7: 170 tweets  8: 1219 tweets  9: 213 tweets  10: 147 tweets |
| 20 | 3414.8576 | 1 : 334 tweets  2 : 145 tweets  3 : 82 tweets  4 : 486 tweets  5 : 54 tweets  6 : 70 tweets  7 : 440 tweets  8 : 93 tweets  9 : 139 tweets  10 : 97 tweets  11 : 122 tweets  12 : 343 tweets  13 : 212 tweets  14 : 246 tweets  15 : 106 tweets  16 : 642 tweets  17 : 199 tweets  18 : 59 tweets  19 : 107 tweets  20 : 81 tweets |
| 50 | 3243.6787 | 1 : 388 tweets  2 : 11 tweets  3 : 92 tweets  4 : 41 tweets  5 : 18 tweets  6 : 205 tweets  7 : 108 tweets  8 : 41 tweets  9 : 60 tweets  10 : 87 tweets  11 : 58 tweets  12 : 12 tweets  13 : 24 tweets  14 : 136 tweets  15 : 121 tweets  16 : 62 tweets  17 : 71 tweets  18 : 146 tweets  19 : 50 tweets  20 : 55 tweets  21 : 16 tweets  22 : 16 tweets  23 : 112 tweets  24 : 72 tweets  25 : 6 tweets  26 : 122 tweets  27 : 30 tweets  28 : 23 tweets  29 : 52 tweets  30 : 57 tweets  31 : 132 tweets  32 : 13 tweets  33 : 9 tweets  34 : 79 tweets  35 : 49 tweets  36 : 72 tweets  37 : 25 tweets  38 : 39 tweets  39 : 326 tweets  40 : 22 tweets  41 : 14 tweets  42 : 55 tweets  43 : 8 tweets  44 : 461 tweets  45 : 13 tweets  46 : 39 tweets  47 : 144 tweets  48 : 57 tweets  49 : 147 tweets  50 : 61 tweets |
| 100 | 3109.04507 | 1 : 152 tweets  2 : 73 tweets  3 : 2 tweets  4 : 14 tweets  5 : 12 tweets  6 : 35 tweets  7 : 27 tweets  8 : 421 tweets  9 : 10 tweets  10 : 20 tweets  11 : 19 tweets  12 : 46 tweets  13 : 22 tweets  14 : 31 tweets  15 : 70 tweets  16 : 5 tweets  17 : 45 tweets  18 : 37 tweets  19 : 13 tweets  20 : 57 tweets  21 : 18 tweets  22 : 13 tweets  23 : 9 tweets  24 : 26 tweets  25 : 18 tweets  26 : 93 tweets  27 : 2 tweets  28 : 28 tweets  29 : 41 tweets  30 : 68 tweets  31 : 63 tweets  32 : 24 tweets  33 : 18 tweets  34 : 120 tweets  35 : 7 tweets  36 : 18 tweets  37 : 18 tweets  38 : 7 tweets  39 : 54 tweets  40 : 6 tweets  41 : 6 tweets  42 : 203 tweets  43 : 94 tweets  44 : 41 tweets  45 : 12 tweets  46 : 13 tweets  47 : 9 tweets  48 : 2 tweets  49 : 7 tweets  50 : 7 tweets  51 : 50 tweets  52 : 98 tweets  53 : 15 tweets  54 : 48 tweets  55 : 28 tweets  56 : 74 tweets  57 : 83 tweets  58 : 38 tweets  59 : 45 tweets  60 : 54 tweets  61 : 106 tweets  62 : 31 tweets  63 : 29 tweets  64 : 10 tweets  65 : 10 tweets  66 : 40 tweets  67 : 29 tweets  68 : 6 tweets  69 : 83 tweets  70 : 29 tweets  71 : 23 tweets  72 : 21 tweets  73 : 12 tweets  74 : 35 tweets  75 : 24 tweets  76 : 58 tweets  77 : 16 tweets  78 : 42 tweets  79 : 26 tweets  80 : 54 tweets  81 : 19 tweets  82 : 88 tweets  83 : 32 tweets  84 : 55 tweets  85 : 149 tweets  86 : 13 tweets  87 : 6 tweets  88 : 17 tweets  89 : 10 tweets  90 : 13 tweets  91 : 29 tweets  92 : 3 tweets  93 : 20 tweets  94 : 7 tweets  95 : 172 tweets  96 : 8 tweets  97 : 21 tweets  98 : 9 tweets  99 : 2 tweets  100 : 11 tweets |