K-Means Clustering Algorithm using Jaccard Distance

|  |  |  |
| --- | --- | --- |
| Value of K | SSE | Size of each cluster |
| 5 | 1173.087972022422 | 1 : 462  2 : 263  3 : 285  4 : 279  5 : 116 |
| 10 | 1124.1548024615427 | 1 : 97  2 : 156  3 : 190  4 : 108  5 : 280  6 : 100  7 : 84  8 : 48  9 : 244  10 : 103 |
| 15 | 1092.3619385824613 | 1 : 46  2 : 104  3 : 145  4 : 88  5 : 159  6 : 130  7 : 56  8 : 33  9 : 20  10 : 149  11 : 68  12 : 169  13 : 41  14 : 73  15 : 134 |
| 20 | 1075.502412464582 | 1 : 27  2 : 76  3 : 120  4 : 64  5 : 84  6 : 149  7 : 47  8 : 28  9 : 94  10 : 99  11 : 31  12 : 170  13 : 23  14 : 69  15 : 90  16 : 90  17 : 57  18 : 14  19 : 49  20 : 39 |
| 25 | 1064.0917608835816 | 1 : 79  2 : 48  3 : 98  4 : 30  5 : 69  6 : 99  7 : 32  8 : 25  9 : 78  10 : 91  11 : 18  12 : 141  13 : 23  14 : 57  15 : 85  16 : 73  17 : 117  18 : 36  19 : 48  20 : 46  21 : 41  22 : 19  23 : 15  24 : 10  25 : 47 |
| 30 | 1045.586363515455 | 1 : 70  2 : 52  3 : 82  4 : 47  5 : 62  6 : 105  7 : 29  8 : 21  9 : 15  10 : 73  11 : 49  12 : 125  13 : 30  14 : 23  15 : 76  16 : 29  17 : 109  18 : 23  19 : 51  20 : 37  21 : 28  22 : 53  23 : 13  24 : 6  25 : 40  26 : 17  27 : 20  28 : 41  29 : 26  30 : 78 |

With the increase in the value of K, the value of SSE kept decreasing.

K = 100, SSE = ﻿915.3508385857389

K = 250, SSE = ﻿736.687871461663

K = 500, SSE = ﻿518.5039638595792

K = 1000, SSE = ﻿180.5138428655096

K = 1300, SSE = ﻿34.77136016935829