**Report : Part II**

**Tweets Clustering using k-means**

**Dataset used : gdnhealthcare**

|  |  |  |
| --- | --- | --- |
| **Value of K** | **SSE** | **Size of each cluster** |
| 5 | 2284 | 1 : 171  2 : 358  3 : 1260  4 : 433  5 : 773 |
| 10 | 2186 | 1 : 93  2 : 168  3 : 296  4 : 106  5 : 354  6 : 257  7 : 795  8 : 231  9 : 213  10 : 482 |
| 15 | 2181 | 1 : 85  2 : 75  3 : 38  4 : 201  5 : 52  6 : 390  7 : 349  8 : 497  9 : 271  10 : 108  11 : 108  12 : 280  13 : 59  14 : 385  15 : 97 |
| 20 | 2146 | 1 : 32  2 : 159  3 : 38  4 : 116  5 : 203  6 : 123  7 : 29  8 : 299  9 : 214  10 : 148  11 : 199  12 : 71  13 : 64  14 : 182  15 : 168  16 : 136  17 : 307  18 : 309  19 : 153  20 : 45 |
| 25 | 2057 | 1 : 198  2 : 38  3 : 258  4 : 28  5 : 232  6 : 159  7 : 146  8 : 29  9 : 83  10 : 152  11 : 230  12 : 19  13 : 69  14 : 35  15 : 256  16 : 73  17 : 36  18 : 136  19 : 22  20 : 281  21 : 183  22 : 83  23 : 48  24 : 31  25 : 170 |