Assignment5

Information Retrieval
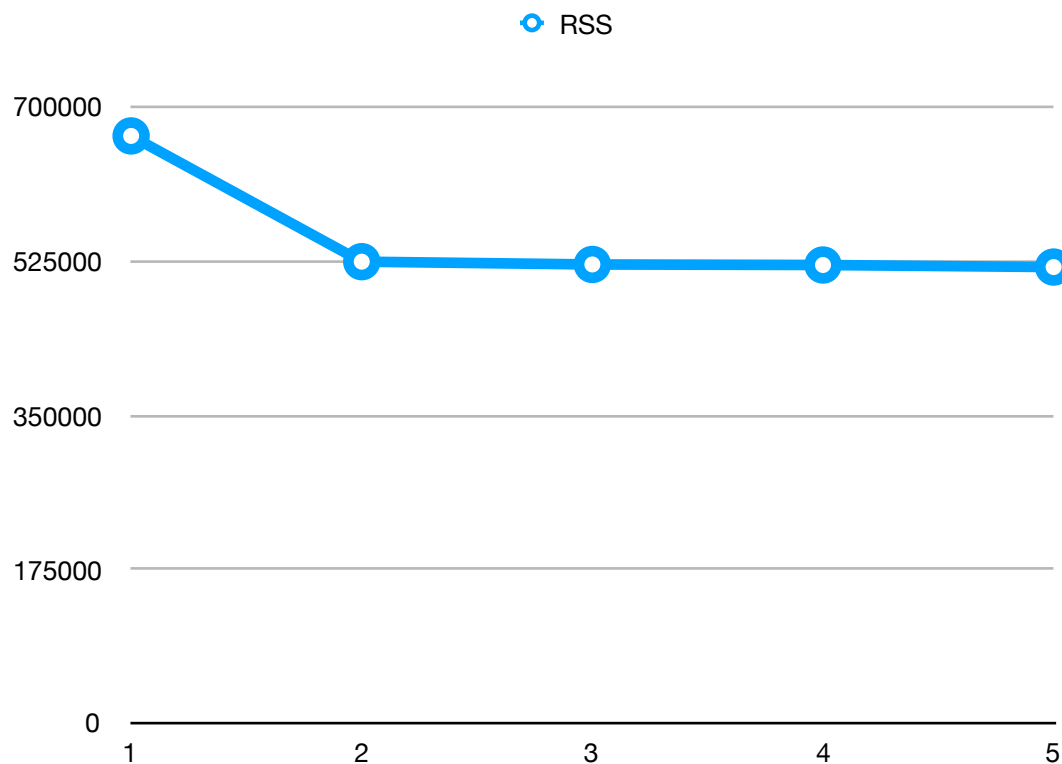
Submitted by: Himanshu Aggarwal
MT17015

Kmeans has been evaluated using RSS value.

Plots are as below for increasing values of K.
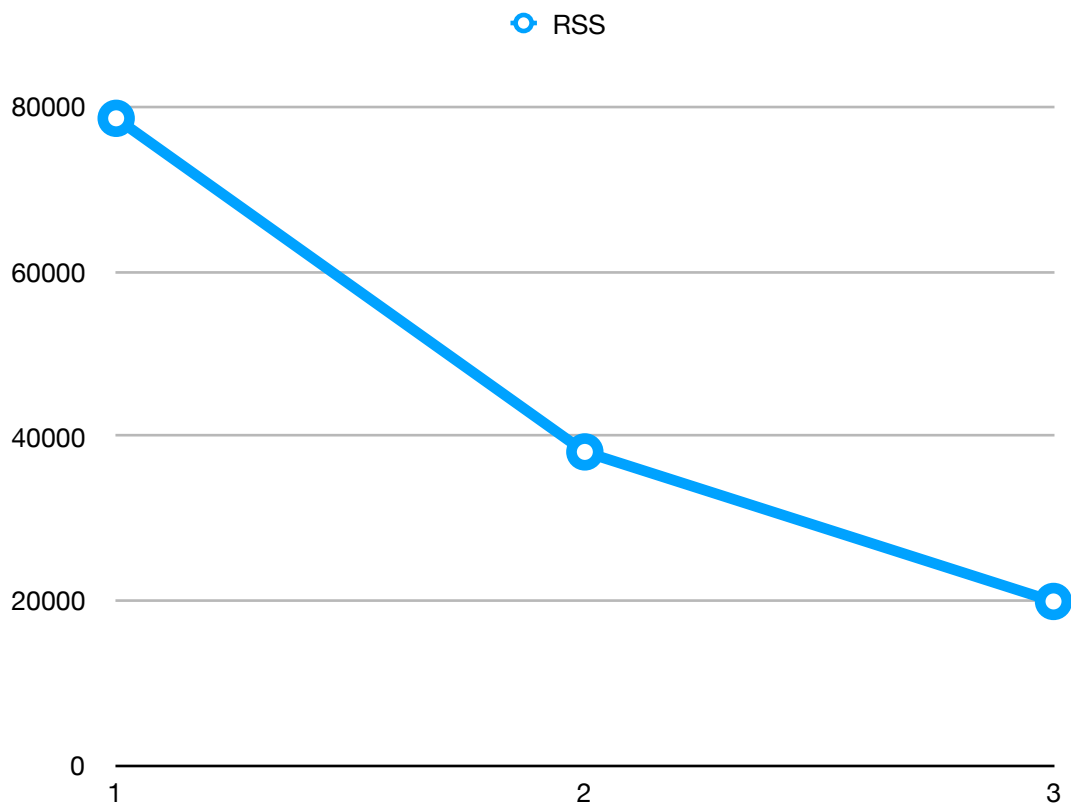
Bag of Words

K	RSS
1	666784
2	524148
3	520991
4	520381
5	518046

RSS

700000					
525000					
350000					
175000					
0					
	1	2	3	4	5

Word2Vec

K	RSS
1	78619.7605133
2	38095.7499648
3	19929.431972

○ RSS

Analysis
- There is significant difference in the results of bag of words and word2vec approach.
- Word2Vec model converged in 3 iterations, where as other took 34 iterations to fully converge. (not all shown)
- This is due to the quality of the vectors formed, which is better in word2vec.