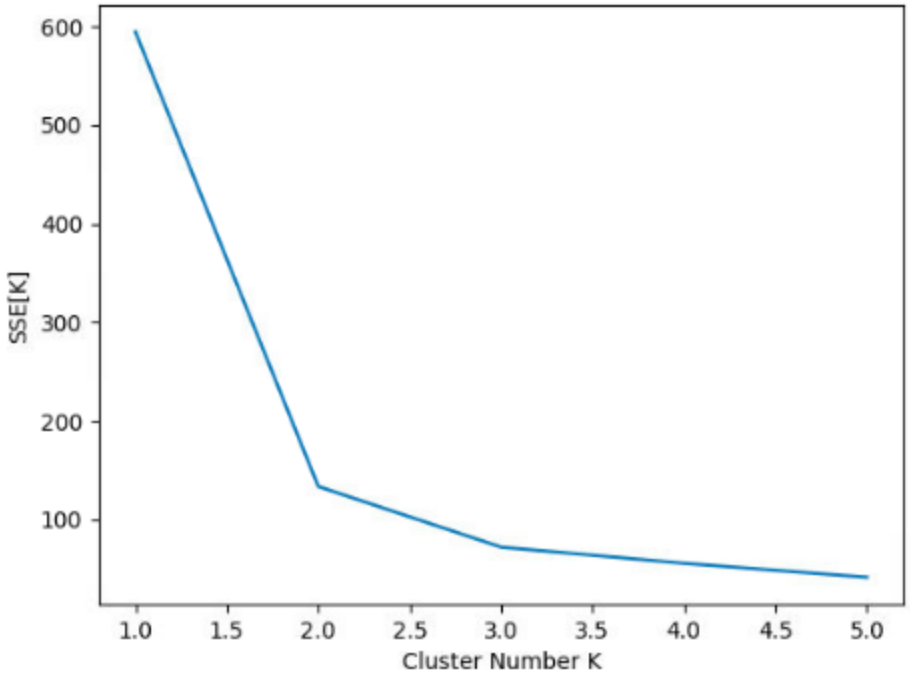
PART 1: Clustering using K-Means

Dataset: Iris - located on UCI website, but also included with sklearn

Experiment Number	Value of K	SSE
1	1	594.8006666666665
2	2	133.46431822602608
3	3	71.94591674975074
4	4	54.38195714285716
5	5	40.40045714285715

3 is the best value of k for the Iris dataset, by the Elbow Method. In the Elbow method, you are selecting the smallest k value that is associated with a significantly minimized SSE. SSE trends to 0, so a higher k value could be selected to see a further minimized SSE. But no significant reduction in SSE happens after k = 3, and a smaller value of k is preferable.

I am satisfied with this result because there are 3 classes in the Iris dataset so it makes sense that 3 clusters would be best, since the Iris datasat is a suitable dataset for clustering.



Part 2 - Color Quantization

Language: Python
Tools/Packages: numpy, matplotlib, sklearn, skimage
Code based off of sklearn documentation:
https://scikit-learn.org/stable/auto_examples/cluster/plot_color_quantization.html

How to Run
Include 3 images in files section of Colaboratory notebook, then run all notebook cells

Results

Image	Original File Size	Value of k	Image Quality	Image Size
Image 1	364 KB	64	Looks slightly grainier/lower quality	327 KB
Image 2	549 KB	64	Looks slightly grainier/lower quality	292 KB
Image 3	613 KB	64	Looks slightly grainier/lower quality	423 KB
Image 1	364 KB	16	Looks grainier/lower quality	342 KB
Image 2	549 KB	16	Looks grainier/lower quality	300 KB
Image 3	613 KB	16	Looks grainier/lower quality	412 KB
Image 1	364 KB	256	Looks similar to original	317 KB
Image 2	549 KB	256	Looks similar to original	287 KB
Image 3	613 KB	256	Looks similar to original	422 KB