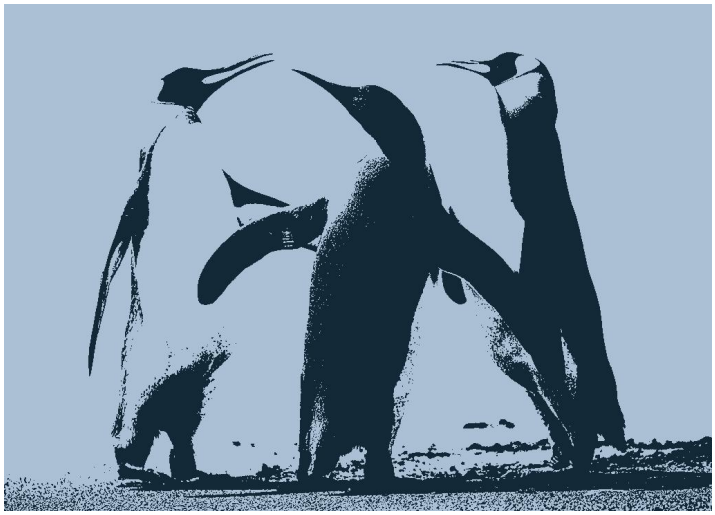
Report

KMEANS

1. Display the images after data compression using K-means clustering for different values of K (2, 5, 10, 15, 20)

Penguin



K = 2

K = 5

K = 10

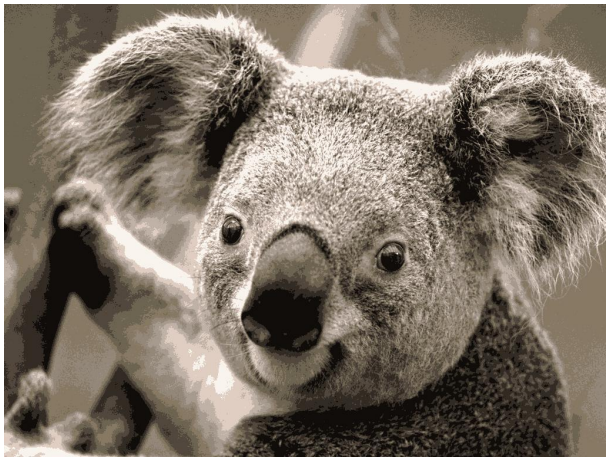
K = 15

K = 20

Koala

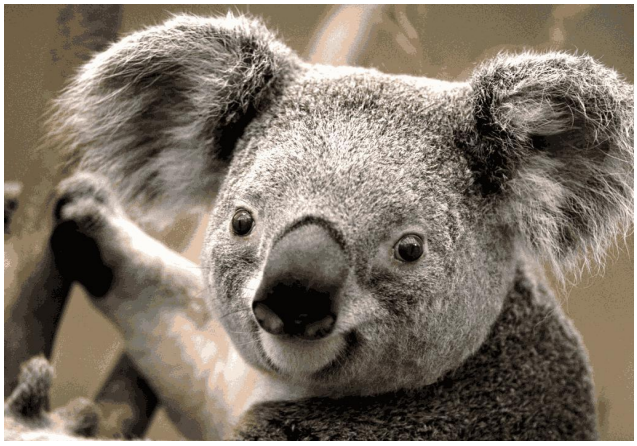
K = 2

K = 5



K = 10

K = 15



K = 20

Penguin	Compression Ratio

Initialization	K=2	K=5	K=10	K=15	K=20
0	0.10946	0.136213	0.15027	0.148079	0.149597
1	0.109581	0.136196	0.149191	0.147529	0.150739
2	0.109649	0.136181	0.150556	0.149899	0.150666
3	0.109481	0.136913	0.152962	0.148004	0.147342
4	0.109562	0.137677	0.150241	0.14815	0.15009
5	0.10955	0.132594	0.149298	0.148907	0.147499
6	0.109467	0.13789	0.150252	0.155317	0.151536
7	0.109649	0.136802	0.15043	0.148969	0.150514
8	0.109581	0.138275	0.149672	0.148876	0.151535
9	0.109487	0.132641	0.150511	0.1528	0.149137
Mean Compression Ratio	**0.1095467548**	**0.136138127**	**0.1503383**	**0.149652947**	**0.1498654**
Variance of Compression Ratio	**4.54E-09**	**3.58E-06**	**9.83E-07**	**5.56E-06**	**2.00E-06**

Koala	Compression Ratio				
Initialization	**K=2**	**K=5**	**K=10**	**K=15**	**K=20**
0	0.174506	0.224846	0.209321	0.205204	0.199813
1	0.174515	0.226674	0.21049	0.204532	0.202112
2	0.167502	0.222904	0.207871	0.204811	0.201124
3	0.166606	0.221655	0.208377	0.206640	0.201324
4	0.167149	0.225493	0.210103	0.205768	0.199041
5	0.174527	0.224347	0.21263	0.208013	0.199844
6	0.167054	0.226267	0.211777	0.203476	0.200509
7	0.166963	0.225161	0.21045	0.204401	0.202346
8	0.167098	0.225138	0.210827	0.203286	0.199646
9	0.166962	0.226957	0.210006	0.204929	0.199175
Mean Compression Ratio	**0.1692882**	**0.2249441**	**0.21018504**	**0.2051060**	**0.2004933**
Variance of Compression Ratio	**1.18E-05**	**2.45E-06**	**1.85E-06**	**1.82E-06**	**1.26E-06**

2. What are the compression ratios for different values of K? Note that you have to repeat the experiment multiple times with different initializations and report the average as well as variance in the compression ratio.

Refer to table above

3. Is there a tradeoff between image quality and degree of compression? What would be a good value of K for each of the two images?

Yes, there is a tradeoff between the image quality and degree of compression. The more compressed the image is (less size) the lesser the quality is. Since with the K value, we are just compressing the image to have only K pixels. In other terms the less the compression ratio the more compressed the image is, but lesser-quality.
There are instances where we can't see any increase in compression ratio because the original image itself contains few pixel variations.

The good value of K for image 1 (penguin) would be K = 10 or K =15 where we are unable to tell the differences between the images and once we reduce the value of K the quality of the image drops. Hence K = 10 and K = 15 provides our best option. We can narrow down the choice using the compression ratio, which in this case is K = 15

Note:

Compression formula used: $\frac{Compressed\ Image\ Size}{Original\ Image\ Size}$