K-Means Clustering

ByeongKyu Park

(byeonggyu.park)

# The programming language used for the project.

: **Python**

from skimage import io

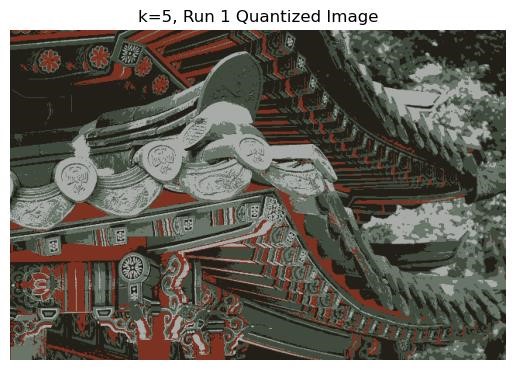
import ‘os,’ ‘numpy,’ and ‘matplotlib.pyplot’

# <Image 1>

original K=3 k=4



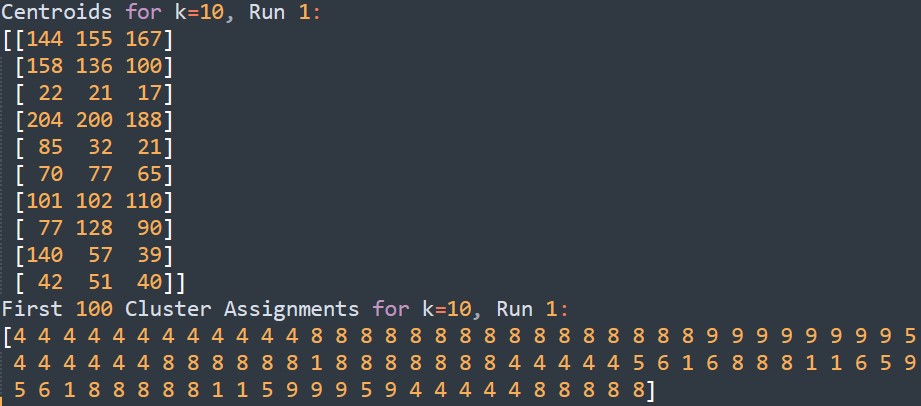
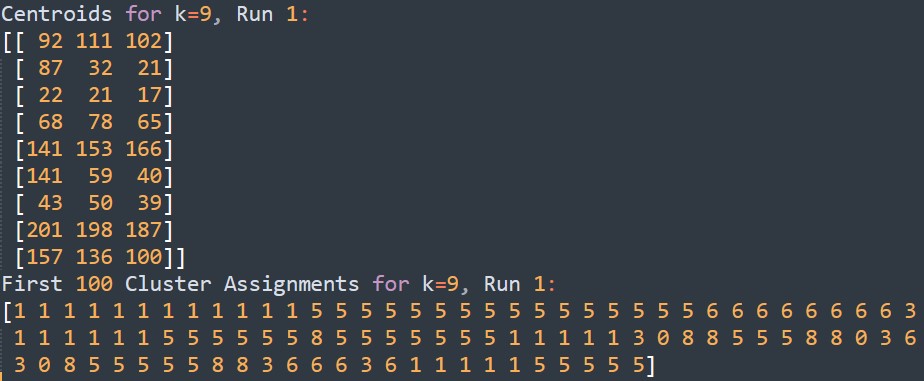
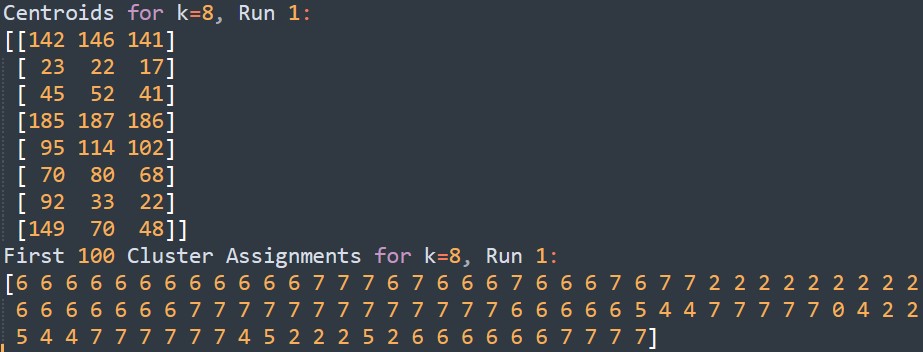
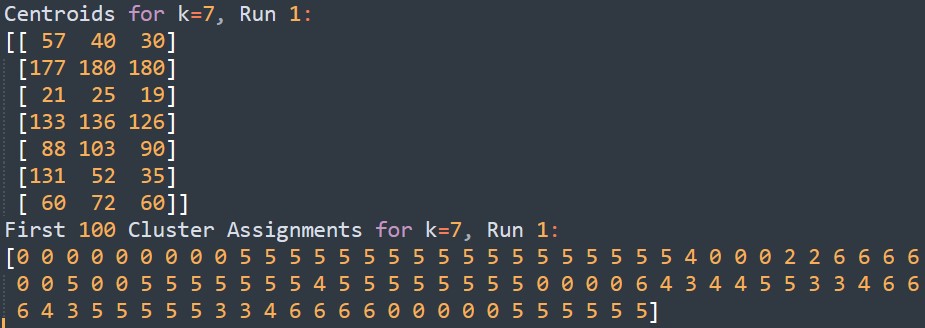
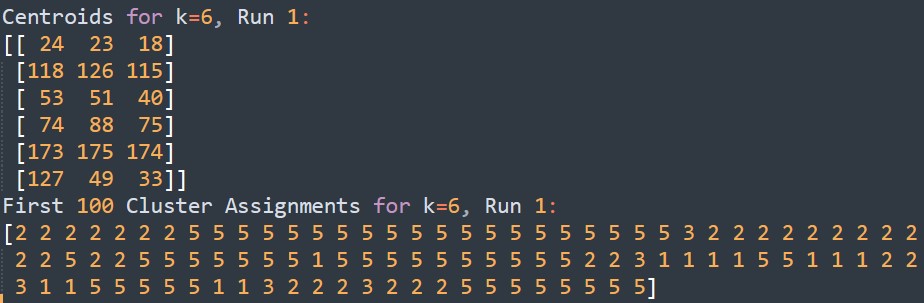
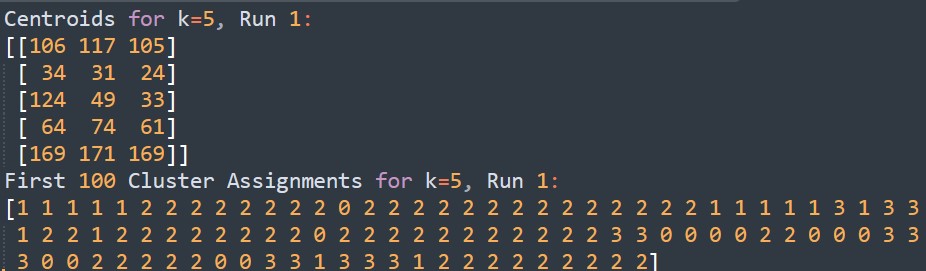
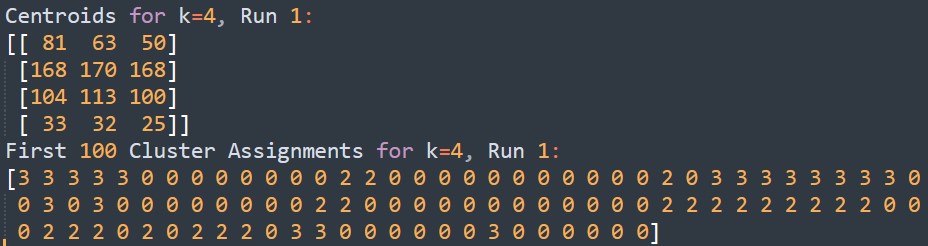
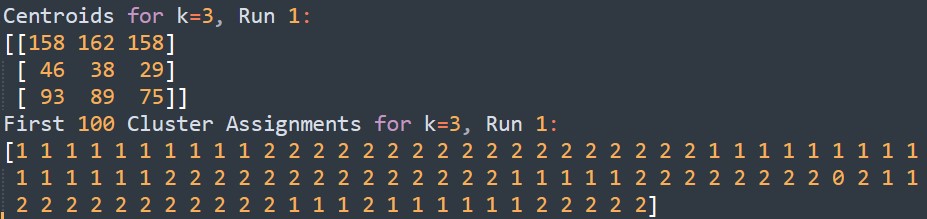
K=5 k=6 k=7



K=8 k=9 k=10



- Clustering for each k;



best choice of k and explanation why

Optimal k: At **k = 10**, the algorithm was able to detect subtle shades and the greenish tint that were missed at lower k values.

(explanation)

At k=3: The output image appeared almost black and white. This indicates that with only three colors to represent the image, the algorithm could not capture the complexity of the original colors and instead simplified the image to its most basic form.

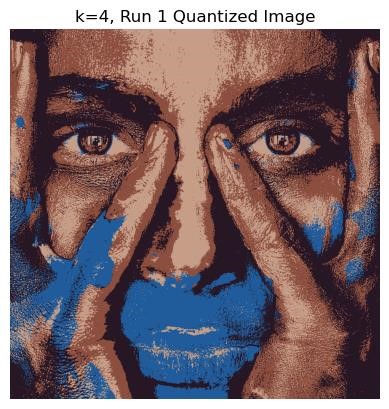
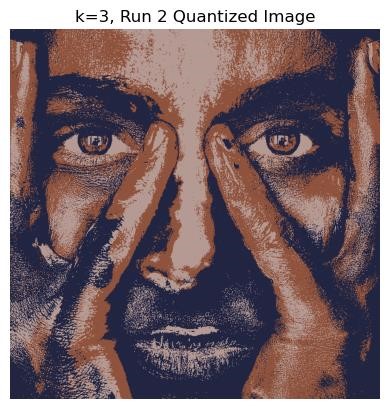
At k=4: The output began to show the two most prominent colors in the image, identified as shades of grey and brown. This shows that increasing the cluster count allows for a slightly more complex color representation.

From k=5 to k=9: A gradual increase in the distinction of more color shades was observed. This progression suggests that as the number of clusters increases, the algorithm can capture more subtle differences between colors, leading to a more detailed and nuanced color palette in the output image.

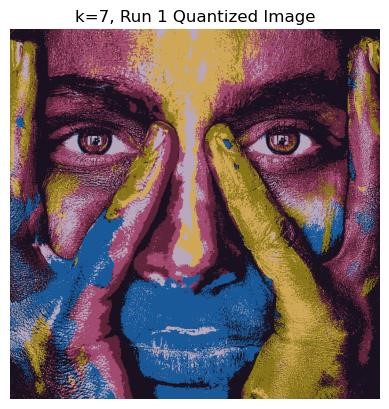
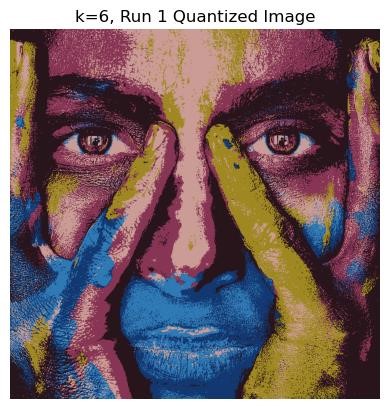
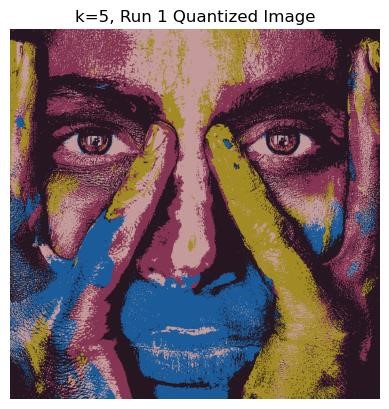
At k=10: The algorithm was able to distinguish between grey and light green, which were previously treated the same at lower cluster counts. This separation indicates that at this cluster level, the algorithm has sufficient granularity to differentiate colors that are closely related but not identical.

# <Image 2>

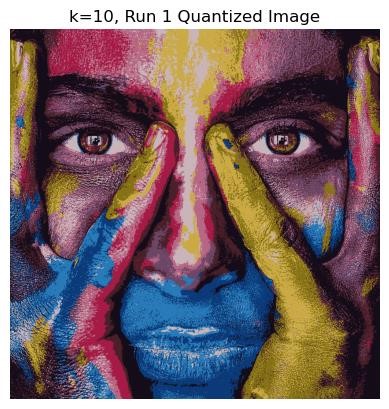
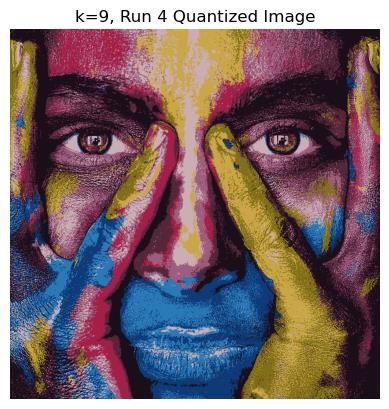
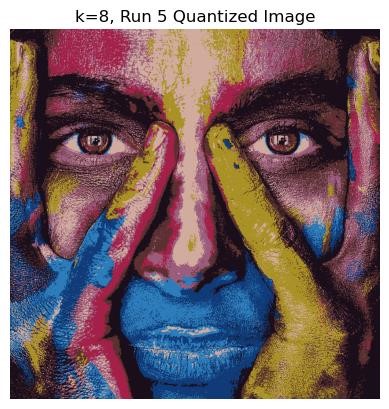
original k=3 k=4



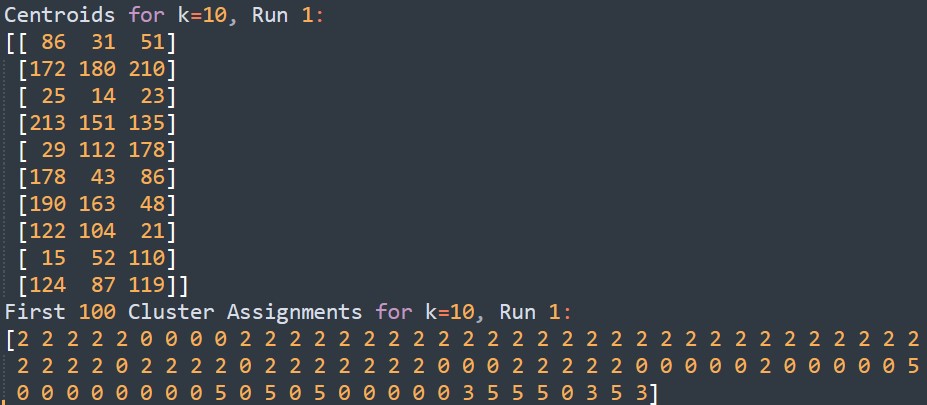
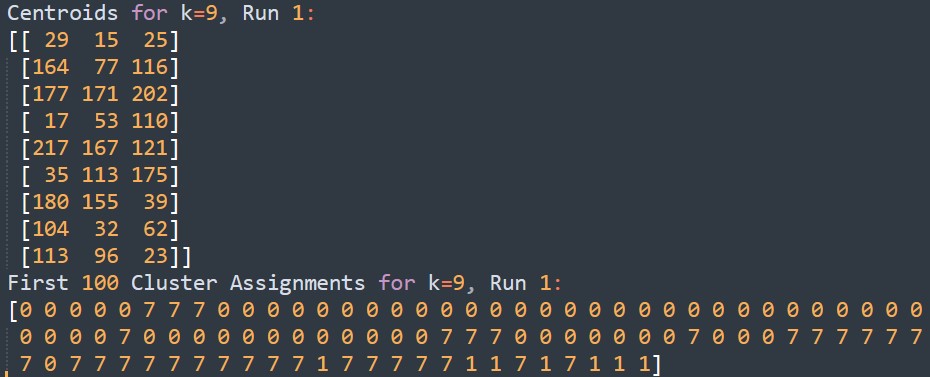
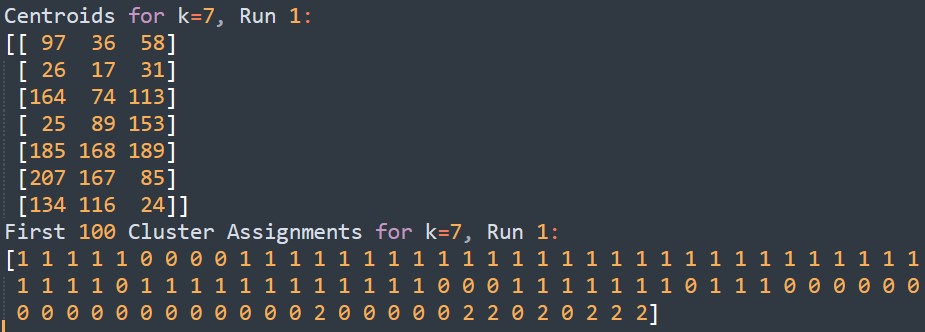
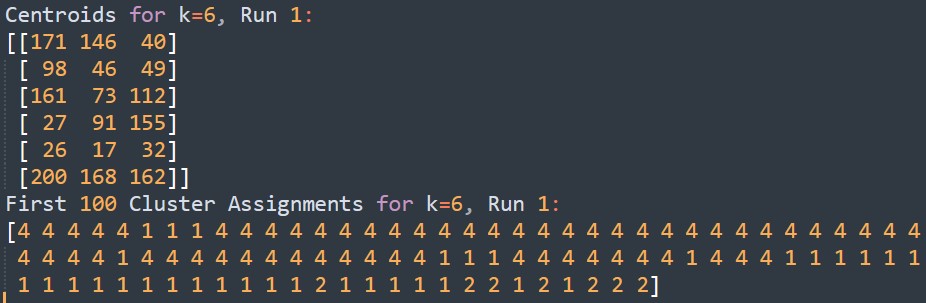
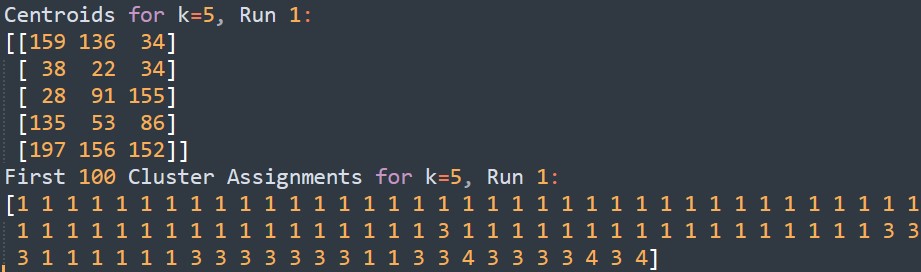
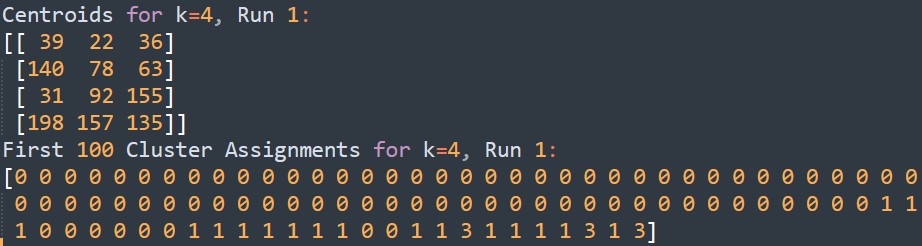
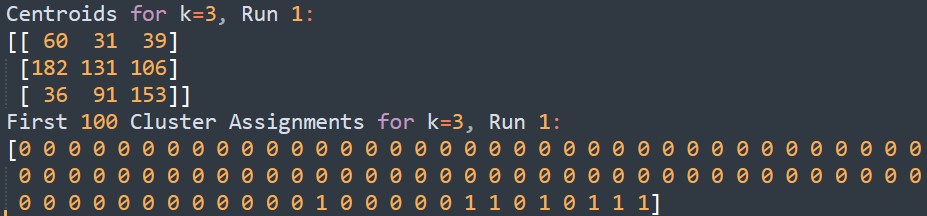
K=5 k=6 k=7



K=8 k=9 k=10



## Clustering for each k;



## best choice of k and explanation why

Optimal **k=8**, the red shades were correctly represented. It was observed that beyond k=8, the output images were identical. This suggests that there is a threshold at which increasing the number of clusters leads to a negligible improvement in capturing additional detail.

(explanation)

* **At k=4:** The algorithm failed to detect the blue paint around the mouth. This suggests that with only four clusters, the algorithm prioritizes the more dominant colors in the image and may overlook colors that do not occupy large areas or are not as prominent.
* **At k=5:** The yellow color on the fingers started to be detected. It implies that increasing the number of clusters allows the algorithm to recognize more varied and less dominant colors.
* **At k=8:** The red shades were correctly represented. This level provides a good balance between the color variety and fidelity to the original image, indicating that eight clusters are sufficient to capture this image's complexity without oversimplifying it.
* **For k=8, 9, 10:** The output images showed minimal differences, indicating that increasing k beyond 8 does not improve the color representation.

<Image 3>

original k=3 k=4 k=5

z



k=6 k=7 K=8 k=9

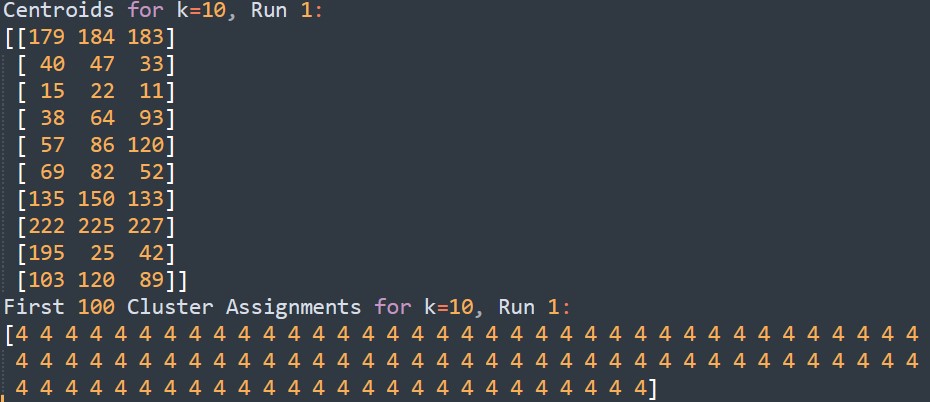
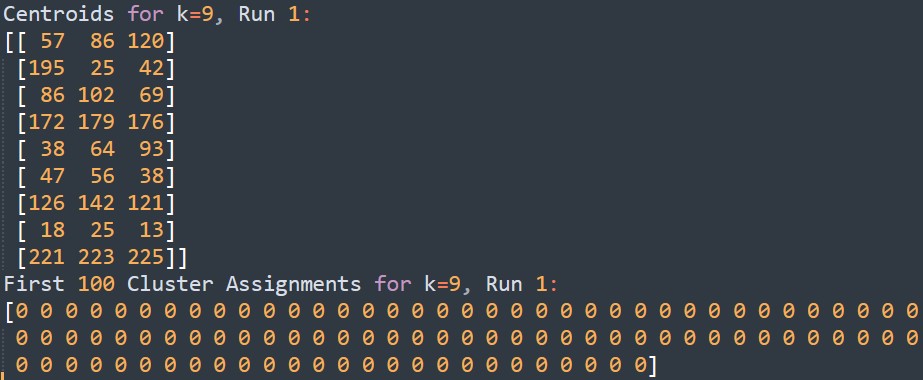
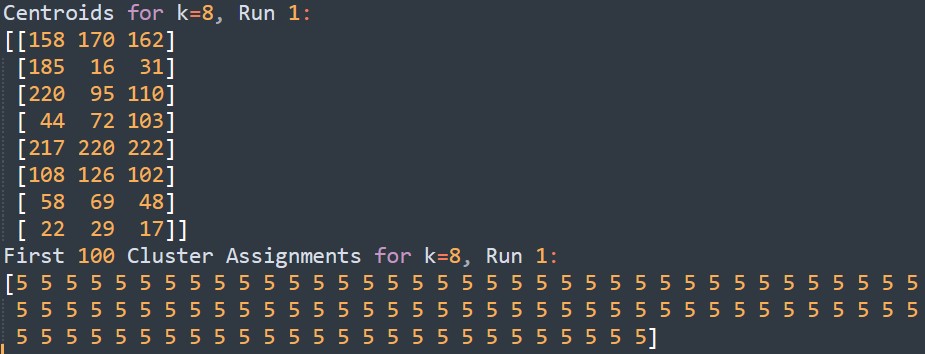
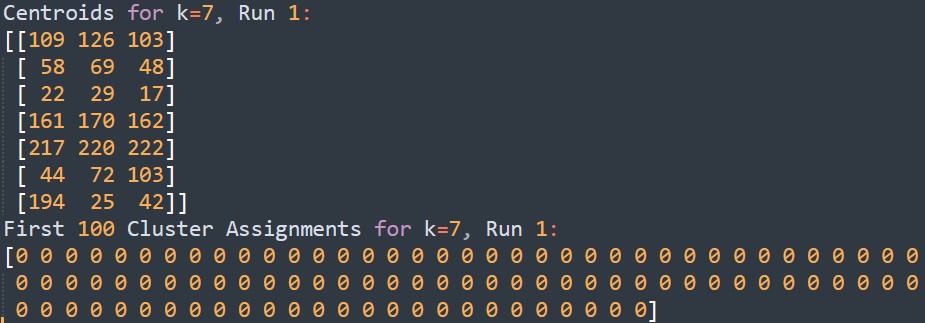
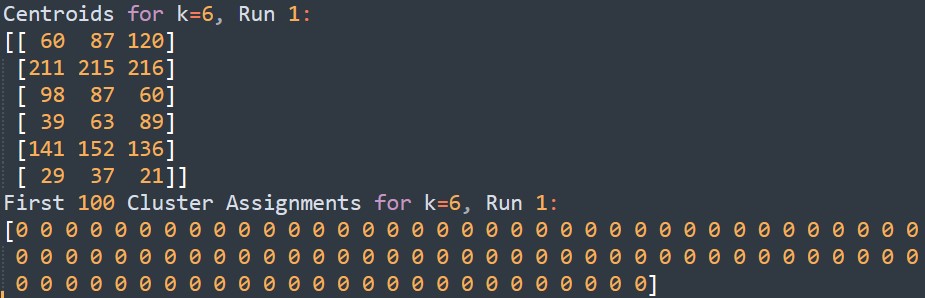
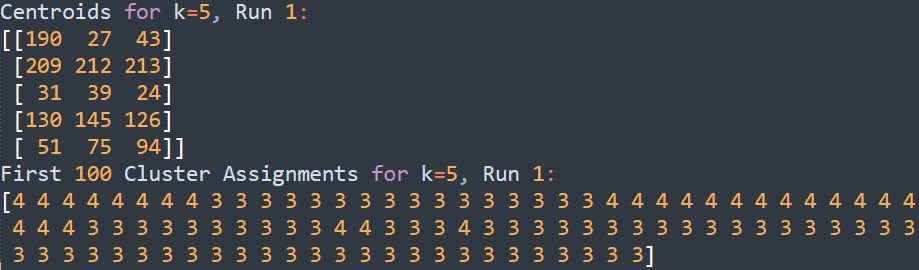
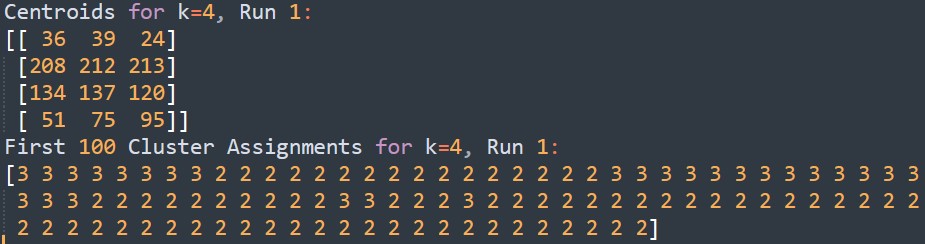
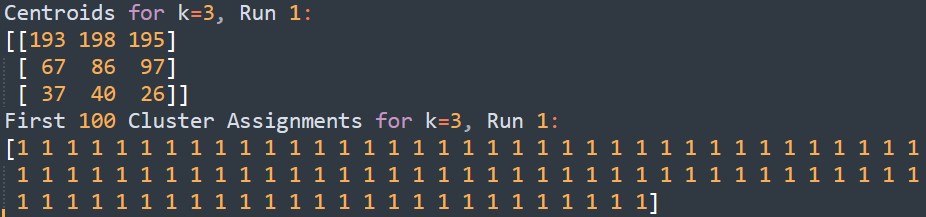


K=10



## Clustering for each k;

The top rows of pixel data represent the color of the sky, resulting in consistent values across these rows as anticipated.



## best choice of k and explanation why

Best **K=6**: the k value of 6 is the most efficient choice for this image. It allows for a representation that captures essential colors, such as the reds on the football, while maintaining c computational efficiency. The findings suggest that there is a threshold of diminishing returns when incrementing the number of clusters in K-means clustering for color quantization of images. Choosing a k value of 6 for this image provides a good balance, as it captures the primary colors and significant details like the red pattern on the football and the contrast with the green grass without becoming too complex.

(explanation)

* **At k=6:** The algorithm successfully detected the red color on the football. This indicates that six clusters are sufficient to distinguish significant color features in the image.
* **At k=8:** There was a minimal detection of the clouds in the sky, suggesting that more clusters can capture more subtle details.
* **For k=6, 7, 8, 9, 10:** The outputs were similar, which implies that increasing the number of clusters beyond 6 does not significantly enhance the image's detail for the human eye.

The original images are in the 'images/original' directory, while the processed images are in the images/output directory. Please note that due to the high resolution and large file size of the images, processing each one can take several hours. The resulting clusters for each k-value are saved in the 'clustering' directories.