Answer Sheet
Connor Deakin
MAT 345
Project 4

Programming Language

Python

Image Coloring

 I decided to run the k means clustering algorithm on two different images. These images and the output for every value of k specified can be found in the directories blood_eagle/ and urban_night/. Based on the output of these two images, there isn't really a best value of k for all images. The best value for k is dependent on the image itself.

 First, we need to define what best even means in this context. To me, the best would be the value of k such that all dominant colors and shades of those colors are clearly represented in the final output. Looking at output_k_3.png in blood_eagle/ (*Figure B*), we can see that the dominant color from the original image (*Figure A*) is well preserved because there is only one, red. Since the image does not have significant changes in color, only the changes in the shade of the color are clearly represented. This low variation allows small values of k to produce decent outputs.

Figure A (input.png)

Figure B (output_k_3.png)

The images in urban_night/ show some very different results. Unlike the input image in blood_eagle/, the urban_night image (*Figure C*) has more colors and more shades of those colors. In output_k_3.png (*Figure D*) colors are removed from the image when the k means clustering algorithm is used to process it. Instead, the clustering algorithm pulled out the different shades in the image rather than the colors. Higher values of k give the algorithm more centroids to work with. This allows the algorithm to pull out more dominant colors and shades of those colors. This effect shows up immediately in output_k_4.png (*Figure E*) and continues to improve with higher values of k. Due to the high contrast of both color and shades in urban_night, only high values of k will produce an image that is more representative of the original.

Figure C (input.png)
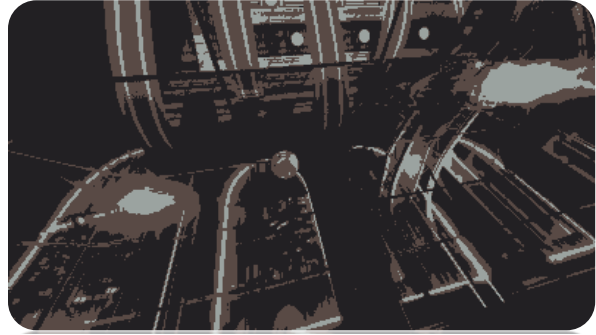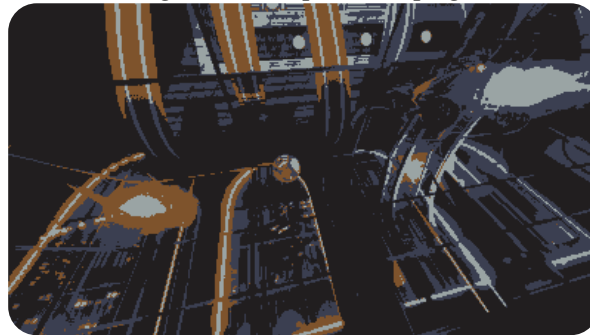

Figure D (output_k_3.png)

Figure E (output_k_4.png)

Both urban_night and blood_eagle make it clear that images with higher variation in color and shade require higher values of k. If we have an image with low variation, low values of k with be sufficient, but as the amount of variation in an image increases, we need higher values of k to represent the higher number of dominant colors and shades.