

# COMPRESSION OF IMAGE OR VIDEOS USING KMEANS

## Description of data

The dataset consists of an image with dimensions 1105 pixels in height, 736 pixels in width, and 3 color channels (RGB). This results in a total of 813,280 pixels, each represented as a point in a 3-dimensional color space.

Each pixel in the image is characterized by its color intensity across three channels: red, green, and blue. The values for each channel range from 0 to 255, indicating the intensity of that color component at that pixel.

Reshaping the image into a one-dimensional array with 813,280 elements allows us to treat each pixel as a data point in a high-dimensional feature space. This flattened representation enables us to apply clustering algorithms like k-means or hierarchical clustering to group similar pixels together based on their color similarity.

By clustering the pixel data, we aim to find representative colors (centroids) that effectively summarize the color distribution in the image. These centroids will then be used to compress the image, where each pixel in the compressed image will be represented by the centroid that it is closest to in color space.

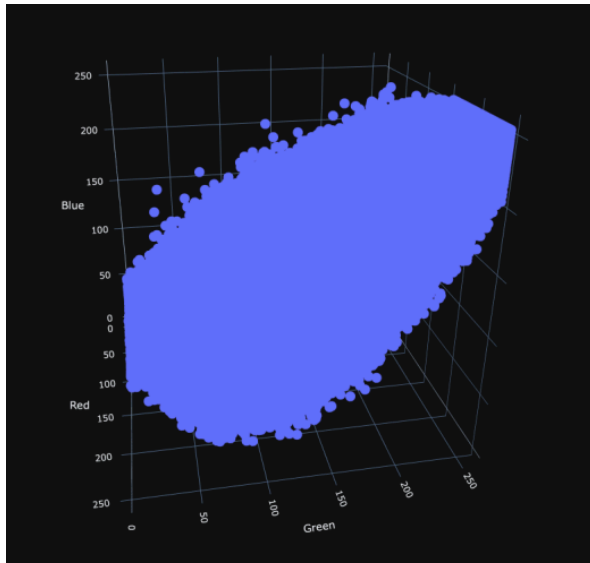
In summary, the dataset represents an image with rich color information, and our goal is to compress this image by finding a reduced set of representative colors using clustering techniques.

## Main objective of the analysis

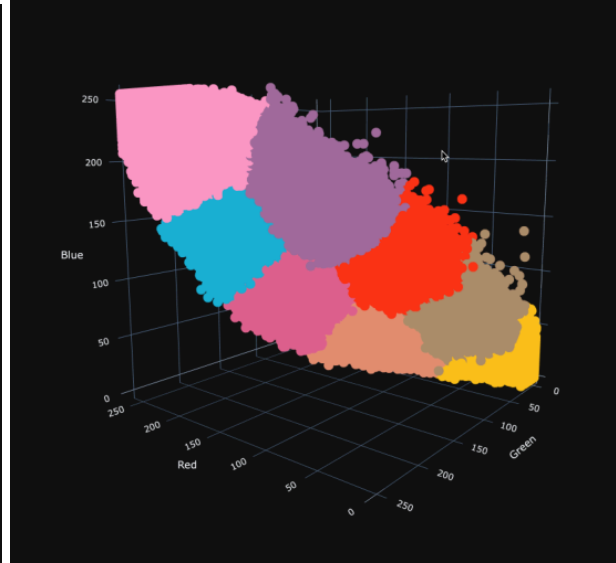
To encode the image in such a way that similar colors are matched together. We will use less color variation in the compressed image. This technique applied to videos will help to reduce the burden on the servers to send the data. The decompression burden will be increased at the client side. The videos encoded in this way are very small in size.

## Variations of the unsupervised models. And which one suits the scenario well.

We have tried kmeans clustering as well as heirarchical clustering, but the kmeans clustering works the best because the centroid is good for the color than the mean of the points.



This was the original space to cluster



After the kmeans clustering

## Key findings related to the main objective of the analysis

The image can be reduced more than 80% by using this technique.



Uncompressed

Compressed.

Possible flaws in the model and the plan of action to revisit in future.

The sharpness of the image gets reduced. The color are limited in the image, so if the image contains too many different colors, than it would be difficult to compress it.