E0-270: Assignment 2

K-Means Clustering :

- The dataset required to be clustered is all the pixel values of size (3,1)(RGB values) in the given image.
- If the image size is M*N*3, then the dataset size is (M*N,3), where each datapoint is a point in 3-Dimensional space.
- Initially randomly initialised the K clusters with K random pixels in the input image.
- Each cluster centre is of size (3,1).
- In each iteration, performed the alternate optimisation of cluster assignment and updated the cluster centres.
- **Cluster Assignment :**
  - Assigned each data point X_i to a cluster whose cluster centre is closer to the X_i.
- **Update Cluster centres:**
  - K Cluster centres are updated as mean of all data points assigned to that cluster during **Cluster Assignment**.
- Iteratively performed Cluster Assignment and Updating cluster centres for 100 iterations.

Reconstruction of image :

- Image is reconstructed from the K cluster centres by changing each pixel(RGB value) in the image to the assigned cluster centre.
- Mean Squared error between the original and reconstructed is calculated as the mean of squares of the difference between each pixel value in the original and reconstructed image.

Original image

Reconstructed images by running K-Means clustering on the original image :

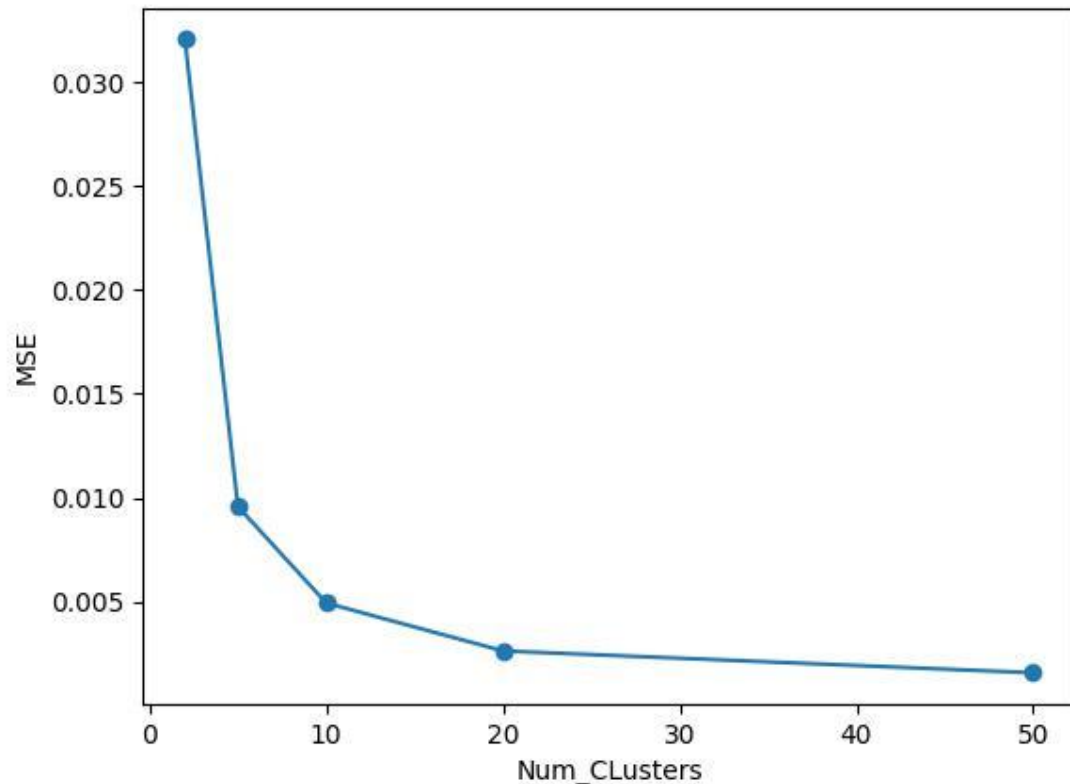**K = 2**   **K = 5**

K = 10

K = 50

K = 50

The plot of the number of clusters on the X-axis and MSE between original and reconstructed on the Y-axis

- Applying K-Means clustering on the image can be seen as finding K different pixel(RGB) values in the image and assigning each pixel to its closest RGB value.
- As we increase the number of clusters, more RGB values get captured, that result in the decrease of MSE between the original and reconstructed image.
- As we keep increasing the number of clusters, the reconstructed image is perceived better.