**Clustering Image Data**

DESCRIPTION

Analyzing images with codes can be difficult. Therefore, it's necessary for you to make your code understand the context of an image. In doing so, the first step will be identifying the  
dominant colors.

[Hint: Refer the following url for image processing documentation: <http://omz-software.com/pythonista/docs/ios/PIL.html>]

**Objective**: To identify the dominant color in the image.

**Action to Perform:**

1. Open and display the image “dog.jpeg.”
2. Use K-means clustering for image segmentation, which will include the following steps:
3. Find out the dimensions of the image and convert it to a two-dimensional array.
4. Use K-means clustering with k set to 3 and cluster the image.
5. [Hint: Refer to the K-means module of scikit learn]
6. Predict the cluster label of every pixel in the image and plot it back as an image.
7. Find out the three dominant colors in the image.
8. [Hint: The cluster centers should correspond to three dominant colors]

Solution:

Steps:

1. Import all the necessary Libraries.
2. Open and display the image dogs.jpeg

A picture containing text, dog

Description automatically generated

1. Convert the image to array.
2. Check the shape of the array and reshape it to 2d array

Table

Description automatically generated

1. Use K-means clustering with k set to 3 and cluster the image. K means clustering is an unsupervised learning algorithm. That is the data is not labeled. It deals with partitioning a data set into K distinct, nonoverlapping clusters. To perform K-means clustering, we must first specify the desired number of clusters K; then, the algorithm will assign each observation to exactly one of the K clusters.

Graphical user interface, text, application, email

Description automatically generated

1. Plot it back as an image. Note that 3 clusters of 3 different colors are formed.

A picture containing graphical user interface

Description automatically generated

1. Find out the three dominant colors in the image. The cluster centers should correspond to three dominant colors.

Chart

Description automatically generated

Chart

Description automatically generated

A picture containing graphical user interface

Description automatically generated

1. How to find the k in K meeans clustering? We use elbow plot for this. The below code calculates sum of squared errors (SSE) for k values 1 to 10. We plot it in a plot. The best k value is the one that is found in the elbow of the plot. From the below plot, we find that the best k value is 3.

Text

Description automatically generated

Chart, line chart

Description automatically generated