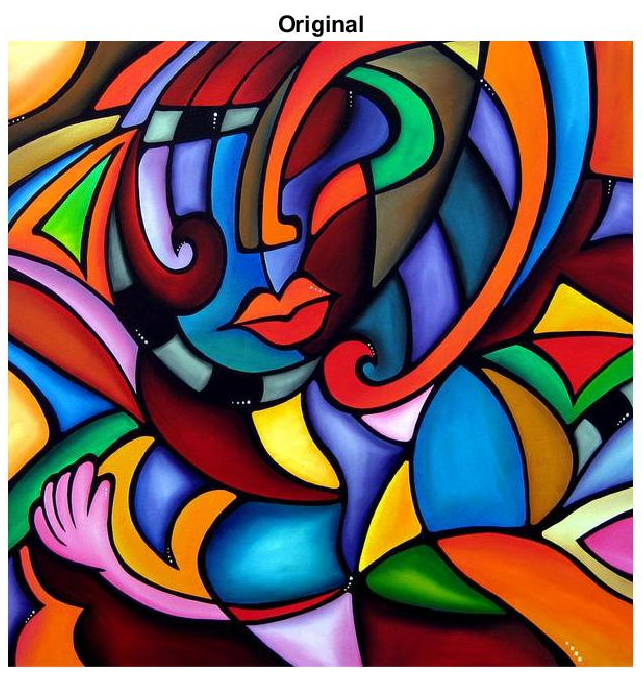
**Segmentation Report**

The following image was used for the comparison between K means and mean shift algorithm, while varying the parameters.



The code tries both K means and mean shift algorithms to segment the above image based on colour.

**K Means**

The following was the result obtained from running the algorithm for both Euclidean and Manhattan distance metric, across various k parameter values.



The K means clustering is simple, fast and has a time complexity of *O(knT)* where *k* is the number of cluster, *n* the number of points and *T* is the number of iterations.

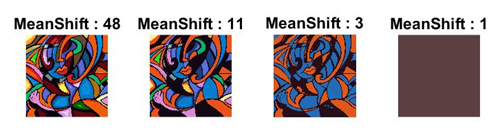
The assumption that K means considers are:

1. The number of clusters are already known.
2. The clusters are shaped spherically or elliptically. (It tries to find centres with spheres around them, which is not always required.)

K means is also very sensitive to initializations, which is why we have to run the clustering multiple times to find the most stable cluster configuration.

**Mean Shift**

The following was the output (cluster number sizes 48, 11, 3 and 1) obtained at bandwidth values for the Gaussian kernels at 0.1, 0.2, 0.4 & 0.8 respectively.



The assumption that the mean shift algorithm considers is:

1. The set of points taken as input are from a probability distribution.

Mean shift algorithm in clustering tries to automatically find all the local maxima or the local modes in the input data probability distribution. Hence, we don’t need to know the number of clusters present in the input (unlike k means). Also, mean shift algorithm does not make any assumptions on the shape of the cluster and it’s not sensitive to outliers (unlike k means).

For each input data point, mean shift defines a window (a.k.a region of interest) around it and computes the centroid of the data points in that window. The vector pointing from initial centre of the window to the new centroid is called mean shift vector. Then it shifts the centre of the window to the centroid and repeats the algorithm till it converges (mean shift vector decrease during these iterations and finally vanishes). After each iteration, we can consider that the window shifts to a denser region of the dataset. The stationary data points obtained after convergence represent the mode of the density function. All points associated with the same stationary point belong to the same cluster.

Mean shift algorithm is a non-parametric algorithm, although bandwidth parameter of the gaussian kernel used can be changed. The choice of bandwidth in influences convergence rate and the number of clusters. Choice of bandwidth parameter is critical. A large bandwidth (i.e. short fat kernel) might result in incorrect clustering and might merge distinct clusters, since all data points go for one mode. A very small bandwidth (i.e. tall skinny kernel) might result in too many clusters, since one mode will be found for each few data points.

Means shift algorithm is very time intensive with time complexity given by *O(n2T)* where *T* is the number of iterations and *n* the number of input data points.

Mean shift requires reasonable number of input points within the region of interest to shift the centroid. Hence, the algorithm might not work in high dimensions.