Question 1:

1. Variance in the data-set explained by Principal Component0 = 54.18%
   Variance in the data-set explained by Principal Component1 = 45.82%
2. Since we have calculated the Principal components by the eigendecomposition of covariance matrix which actually does centering automatically as variance actually means the average of the square of the deviations of data points from mean. Therefore Centered and noncentered dataset have the same covariance matrix and hence centering doesn't make any difference in PCA.

4. For the given dataset, the RBF kernel is best suited because as per the dataset plot the features are non-linearly related. So, standard PCA doesn't help. Using polynomial Kernel function with degree (2 and 3) also couldn't linearly separate the dataset in Voronoi regions (as observed). When we use Radial Basis Function, as seen by the plots, the projections of the data points in this kernel can be linearly separated, hence this is best suited. However, choosing the best sigma can be difficult. For the given values of sigma to experiment with I found that sigma 0.9 or 1 would be the best option.

Question 2:

3. RBF kernel performs better than polynomial kernel as per experiments done. Out of Interest, I have also performed Spectral K-means Clustering using Laplacian which is found to outperform RBF kernel. However, as we increase the radius greater than 10, the clustering results in poor performance.

4. For the given dataset, this mapping performs close to RBF (for some values of sigma) and degree2 polynomial kernel.This might be because in higher dimensional space, for a datapoint, the maximum component of eigenvector indicates that the point is closer to that particular eigenvector and hence it makes sense to assign all the points having maximum same eigenvector component identical cluster. Polynomial kernel of higher degrees are not performing good for the given dataset.

References:

1. https://xavierbourretsicotte.github.io/Kernel_feature_map.html (Kernel PCA)
2. https://towardsdatascience.com/spectral-clustering-aba2640c0d5b (Kernel K-means)
3. https://www.kaggle.com/code/vipulgandhi/spectral-clustering-detailed-explanation/notebook (Kernel K-means)