Task 3 for Machine Learning
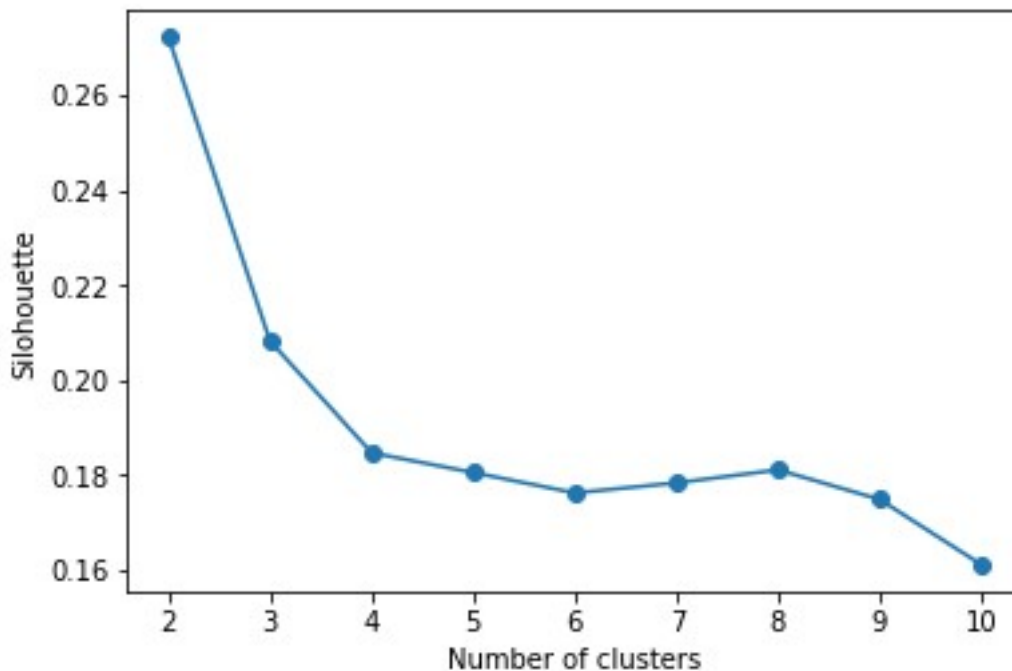# by
Eduardo Sánchez López
and
José Alejandro Libreros Montaño

The dataset is from Iquitos.

The number of elements left in our dataset is 200, after deleting some outliers that we talked about in the last task.
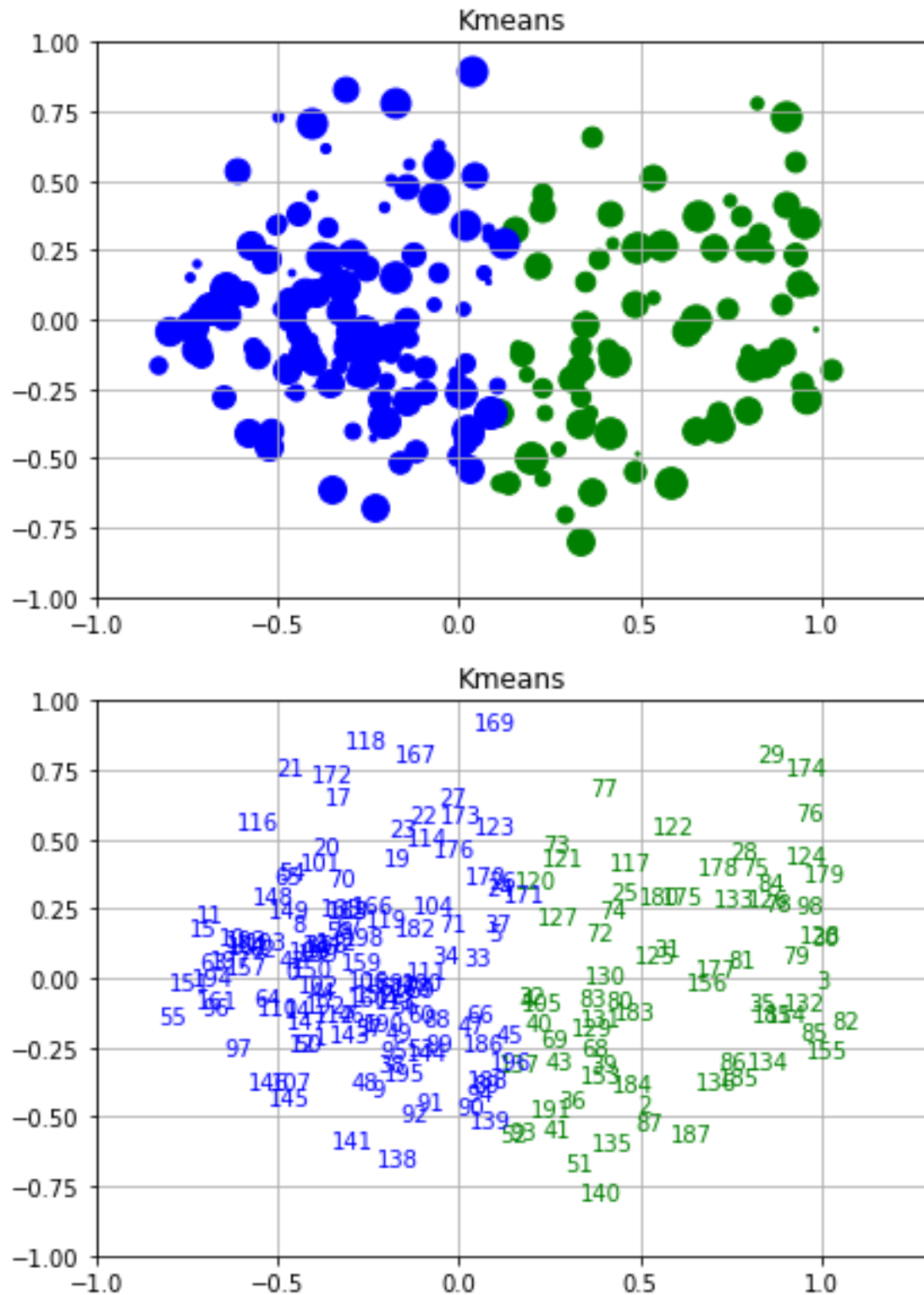
We choose k (our number of clusters) to be 2, since the silhouette is at max when we get that value.

Also we are using k-means++ since it provides us the most consistent results. The variation with random is not too big, so exchanging one with another shouldn't make the results much different.

Silhouette Coefficient (Kmeans++): 0.275
[0.4130468 0.20305457]

The PCA resulting of this is:

The main differences between these two groups are the precipitations, the green group has much more precipitations than the blue one.