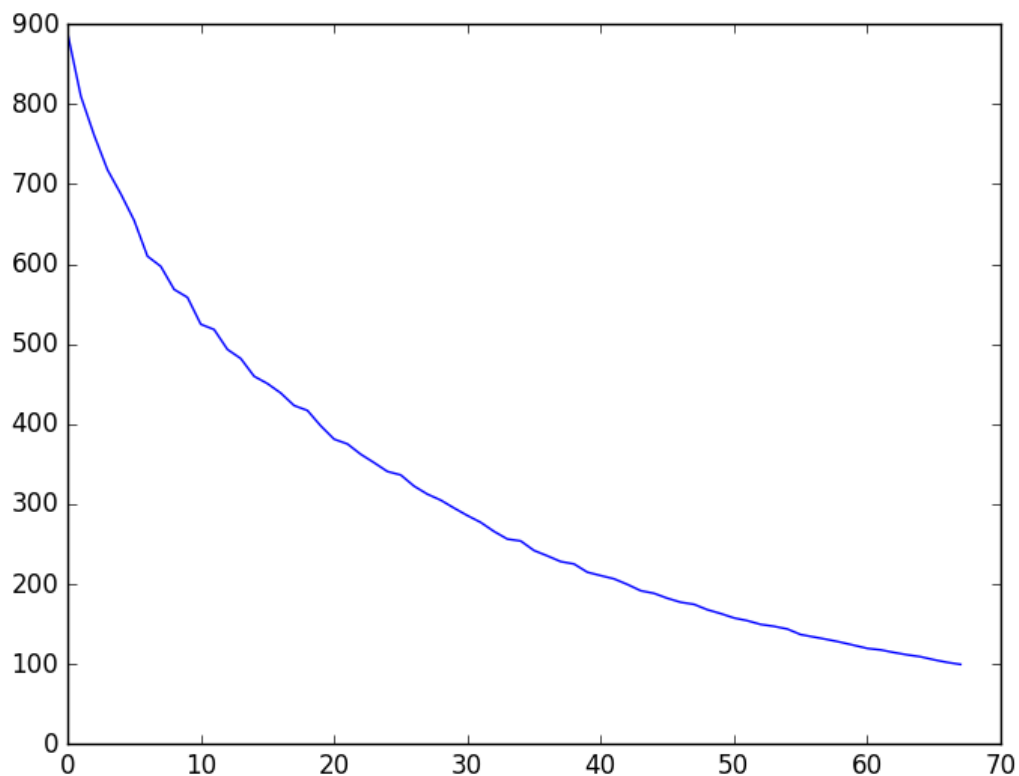


HW8 Q4

For this question I used SSE which is defined as the sum of the squared distance between each member of the cluster and its centroid.

I used `km.inertia_` to find the SSE values (Which is usually returns a large float number) for my data set. I generated a for loop between the range (2,70) and initial centroids=100.



y-axis = SSE values

x-axis = number of clusters

As we can see from the picture, The line starts to become horizontal for the value 65.

Therefor I have chosen number of cluster for my data set to be 65.