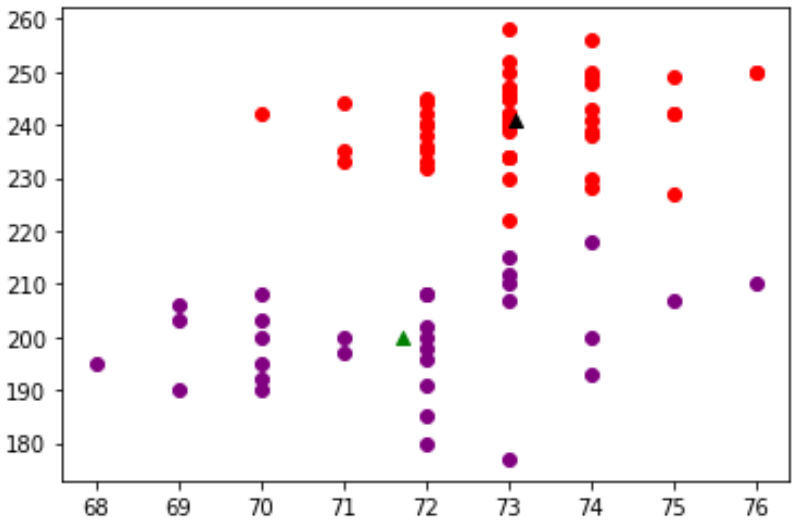
PROJECT 5

Eric Cai

In this project, we separated data into two clusters using K-means clustering. This was repeated for a total of five runs. The initial centroids, final centroids, and final J values of all five runs are shown below.

#	Initial Centroid 1		Initial Centroid 2		Final Centroid 1		Final Centroid 2		J Value
	x1	x2	x1	x2	x1	x2	x1	x2	
1	74	192	69	218	71.70967742	199.87096774	73.07843137	241.05882353	71.07781428857932
2	75	204	78	261	71.70967742	199.87096774	73.07843137	241.05882353	71.07781428857932
3	76	188	67	214	71.70967742	199.87096774	73.07843137	241.05882353	71.07781428857932
4	74	223	68	241	71.70967742	199.87096774	73.07843137	241.05882353	71.07781428857932
5	66	240	71	168	73.07843137	241.05882353	71.70967742	199.87096774	71.07781428857932

This is a plot of the best result. The best centroids are [73.07843137, 241.05882353] and [71.70967742, 199.87096774]. They give an error term of 71.07781428857932.



The green triangle is the final centroid 1. It is the centroid of the purple cluster. The black triangle is the final centroid 2. It is the centroid of the red cluster.