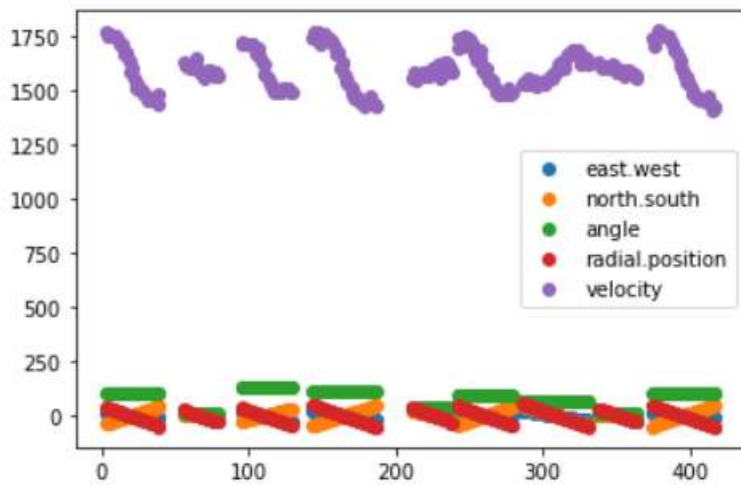
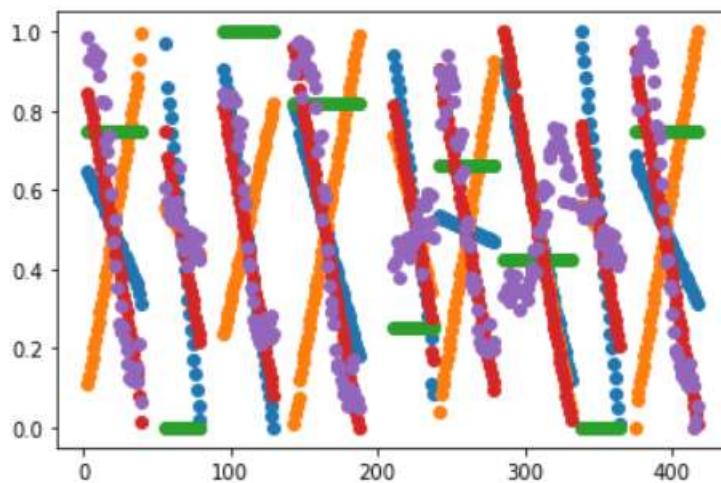


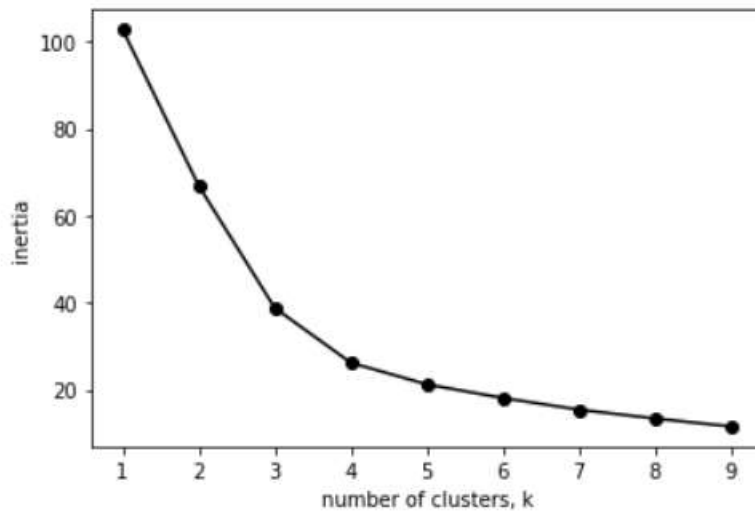
Radial GALAXY Data Set before scaling:

Observation: Needs scaling for a better view.

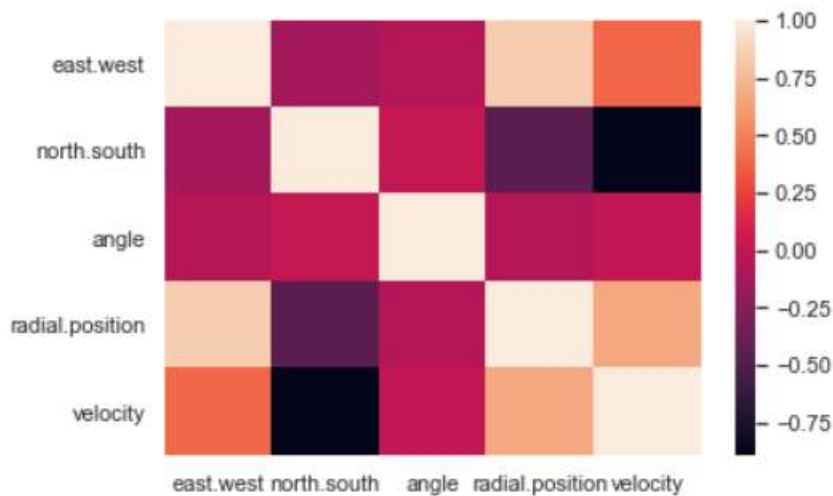
Galaxy Data after Scaling:

Does not look seperable at all.

Definitely needs dimension reduction before clustering.

Elbow method to choose k Clusters:

Action:- I choose 3 clusters

Correlation Map:

Observation:- a) Implies feature “angle” is not correlated with any other feature

b) Implies feature “north.south” is also not correlated with any other feature
Hypothesis: These features won't matter during classifying data points.

Let's do PCA to see which features are actually contributing to the data:-

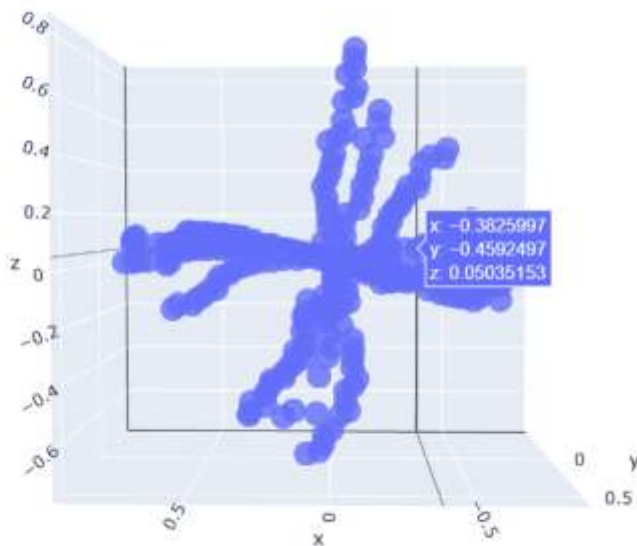
Rotation Matrix:

	PC1	PC2	PC3	PC4	PC5
east.west	-0.43035405	-0.6187657	-0.15420177	0.621137517	0.14944715
north.south	0.45838027	-0.5787659	-0.14423245	-0.139860809	-0.64385827
angle	0.03219063	0.2237573	-0.97322249	-0.006307842	0.04116579
radial.position	-0.54683545	-0.3347213	-0.08611298	-0.756604819	0.09521720
velocity	-0.55192180	0.3464871	0.02900598	0.148783393	-0.74320329

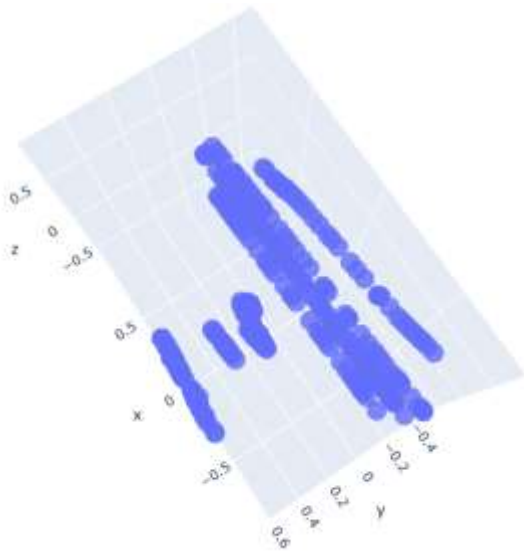
Initial hypothesis proved :- Feature 'angle' does not contribute much.

Kernal Pca (kernel = Linear)

Reduced the whole galaxy dataset to 3 dimensions:

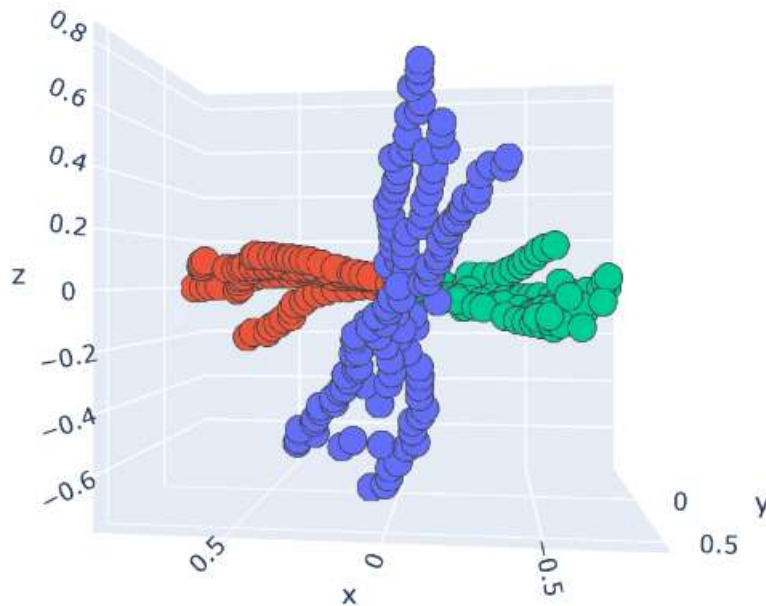


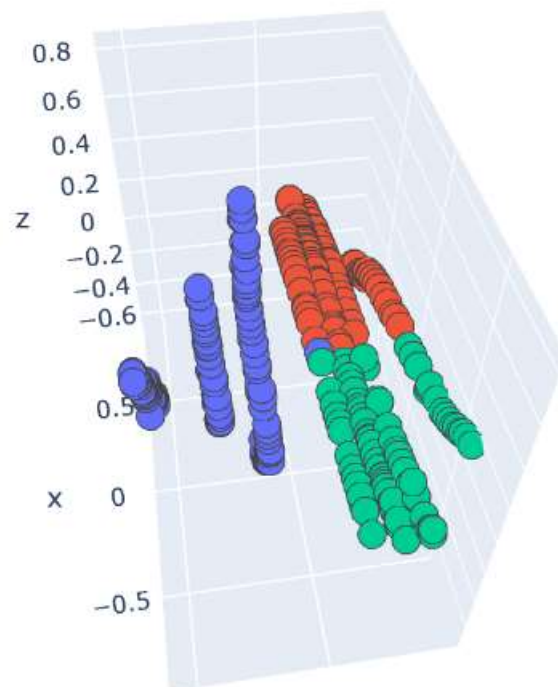
3D-Figure



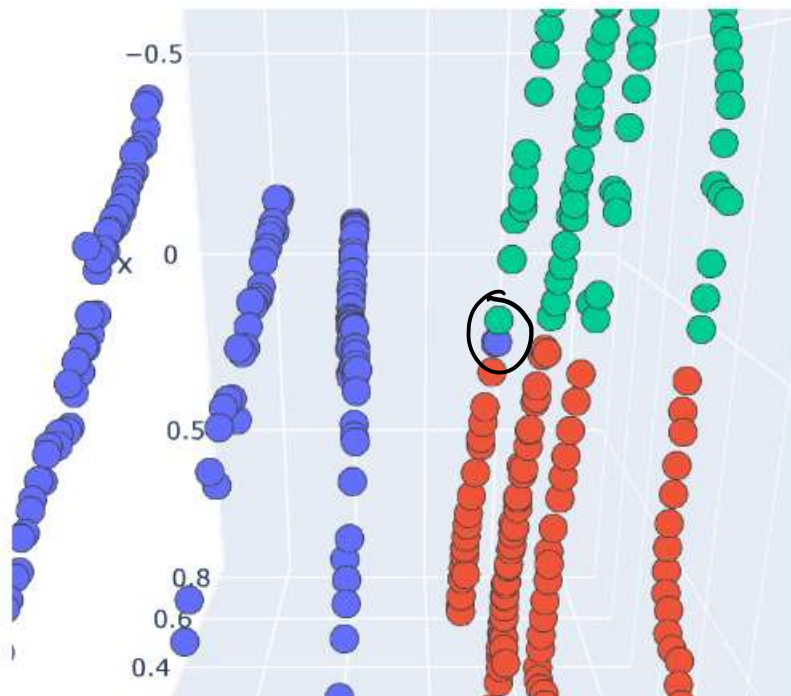
(rotated view)

Performing Kmeans Clustering:





(Rotated View)

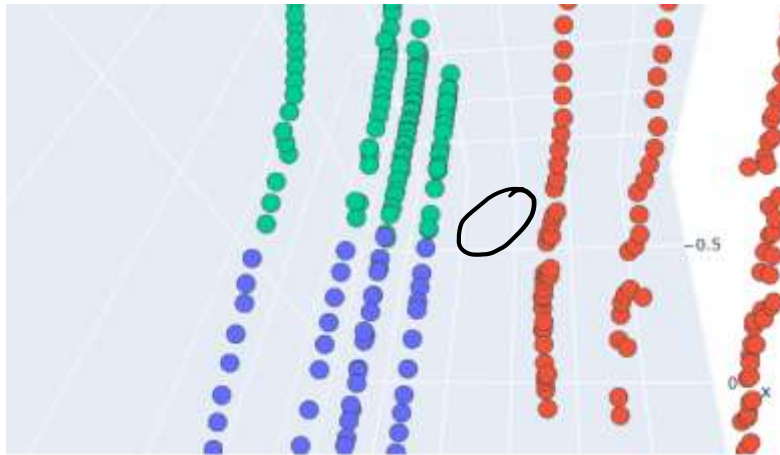
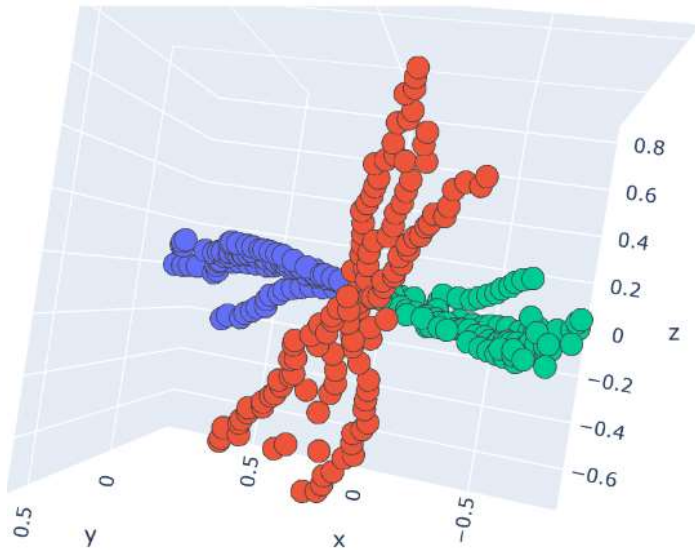


Observations: Two points from cluster 1 is closer to cluster 2

Hypothesis: hoping this would change with spectral clustering.

Spectral Clustering:

Performed Spectral Clustering on the data set



Look how perfect this clustering is; no misclassified points / No points too far from clusters1

Hypothesis proved. Hence, Spectral clustering > Kmeans !

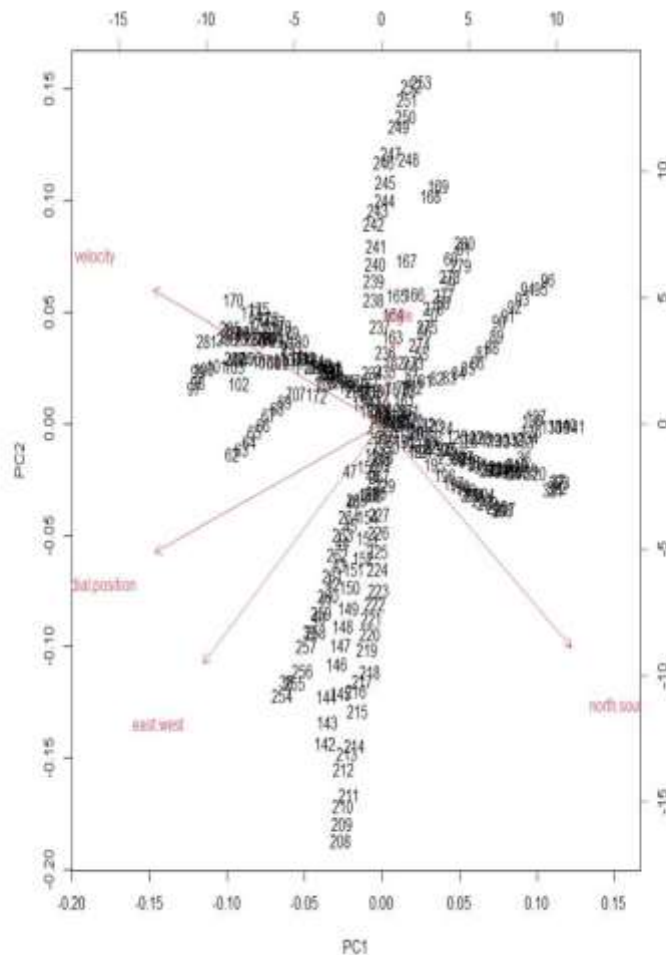
ISOMAP and LLE:

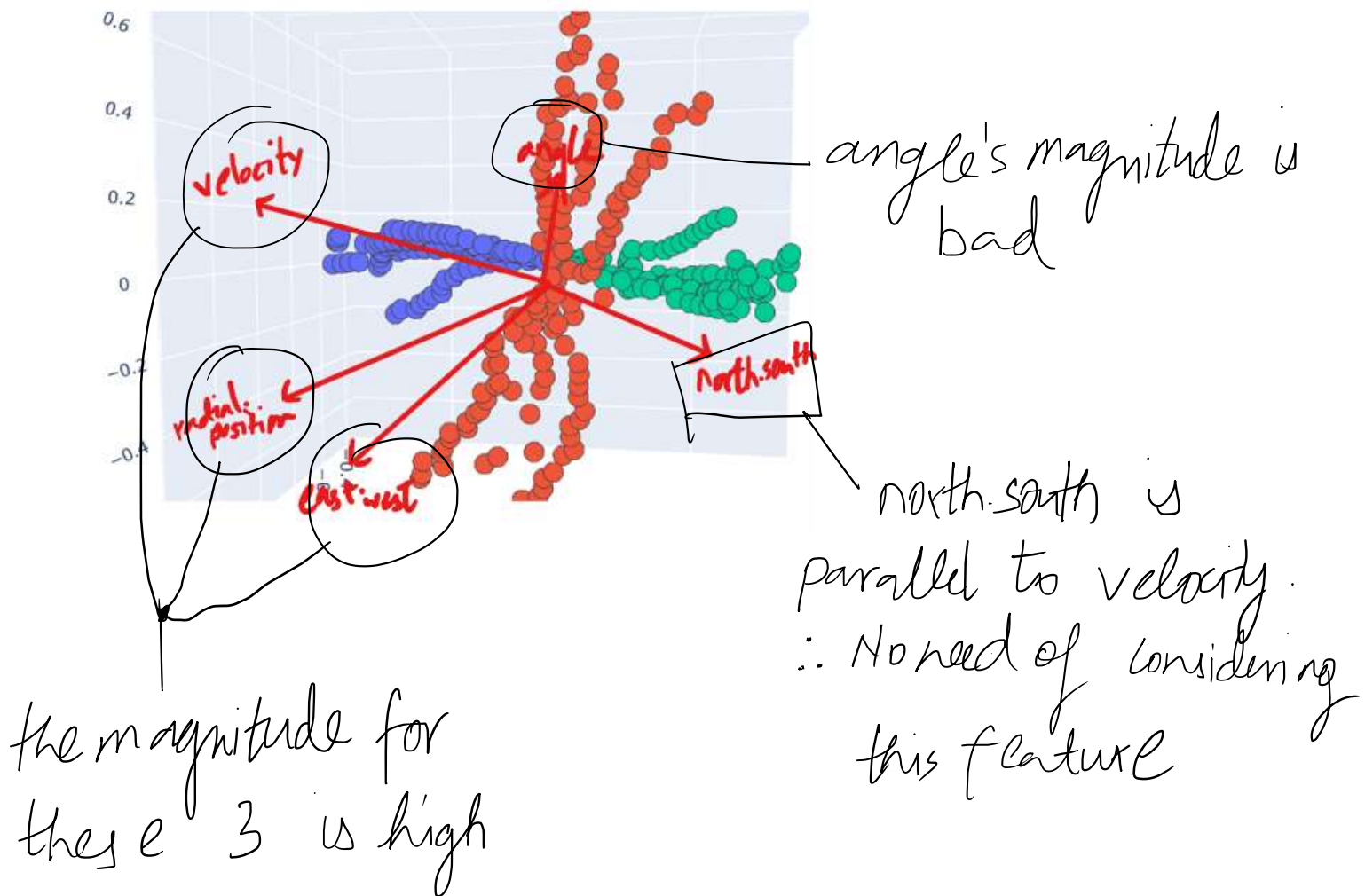
I Performed these dimensionality reductions too but they performed badly for this data set.

Even in Kernel PCA, I used different kernels like poly, sigmoid, rbf they all gave results similar to linear kernel.

Cluster Data:

Now let's talk about the data that's been clustered using *spectral clustering* and with the help of *PCA's* axes too!





So let's consider these 3 features when talking about the cluster labelled data

Properties of Label 0 (violet color points):

	east.west	radial.position	velocity	Class_label
count	101.000000	101.000000	101.000000	101.0
mean	-7.519775	-24.756436	1489.683168	0.0
std	6.890740	14.029686	40.585448	0.0
min	-29.666929	-52.400002	1409.000000	0.0
25%	-10.540609	-36.300000	1456.000000	0.0
50%	-6.001585	-24.799999	1486.000000	0.0
75%	-1.744776	-13.200000	1513.000000	0.0
max	-0.021644	-0.100000	1581.000000	0.0

Properties of Label 1 (red color points):

	east.west	radial.position	velocity	Class_label
count	125.000000	125.000000	125.000000	125.0
mean	0.492892	1.073600	1591.680000	1.0
std	15.373820	23.906974	37.986755	0.0
min	-29.093620	-50.400003	1519.000000	1.0
25%	-12.225820	-17.600000	1568.000000	1.0
50%	0.511948	0.700000	1587.000000	1.0
75%	13.475174	18.700001	1613.000000	1.0
max	29.484140	55.699999	1687.000000	1.0

Properties of Label 2 (green color points):

	east.west	radial.position	velocity	Class_label
count	97.000000	97.000000	97.000000	97.0
mean	6.087946	21.587629	1704.360825	2.0
std	5.794736	13.783771	49.652372	0.0
min	-0.519455	-2.400000	1583.000000	2.0
25%	1.428501	9.900000	1675.000000	2.0
50%	4.480300	21.100000	1713.000000	2.0
75%	8.493320	32.200000	1748.000000	2.0
max	23.869942	51.400000	1775.000000	2.0

