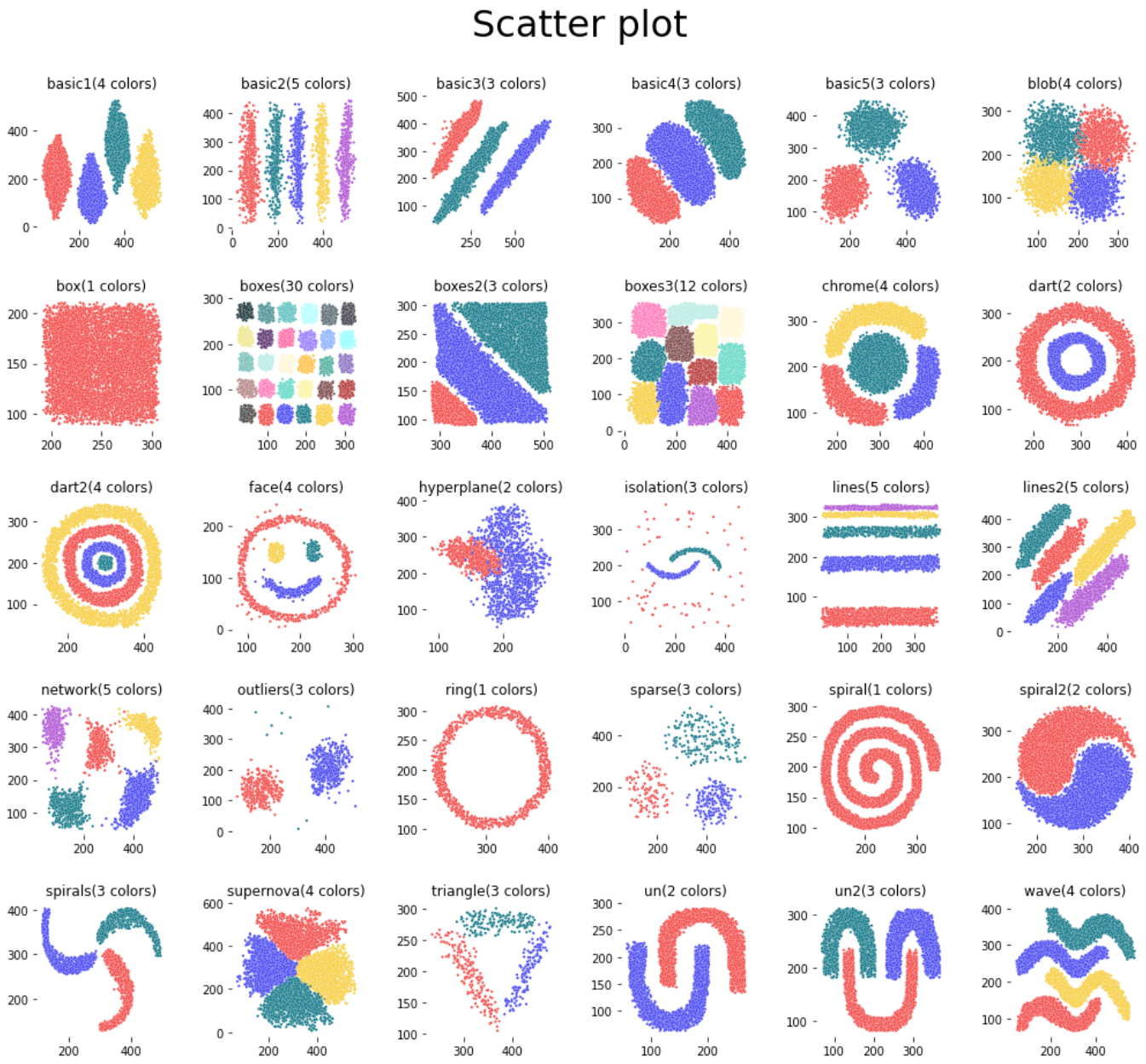
Visualize Cluster Distributions for 30 Datasets



In [ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Function to Visualize Cluster Distribution

In [126]:
```
# Function to Visualize the Clusters

def visualize_clusters(dataset, c_num, colors, name):
    #visulaizing the clusters
    for i in range(0, c_num):
        plt.scatter(dataset['x'][dataset['color'] == i],
                dataset['y'][dataset['color'] == i], s = 100,
                c = colors[i], label = f'Cluster {i+1}')

    plt.title(f'Clusters of {name} Dataset ({c_num} colors)')
    plt.xlabel('X')
    plt.ylabel('Y')
    plt.legend()
    plt.show()
```

Dataset 1 : basic1

In [28]:
```
basic1 = pd.read_csv('/kaggle/input/clustering-exercises/basic1.csv')
```

In [29]:
```
basic1.head()
```

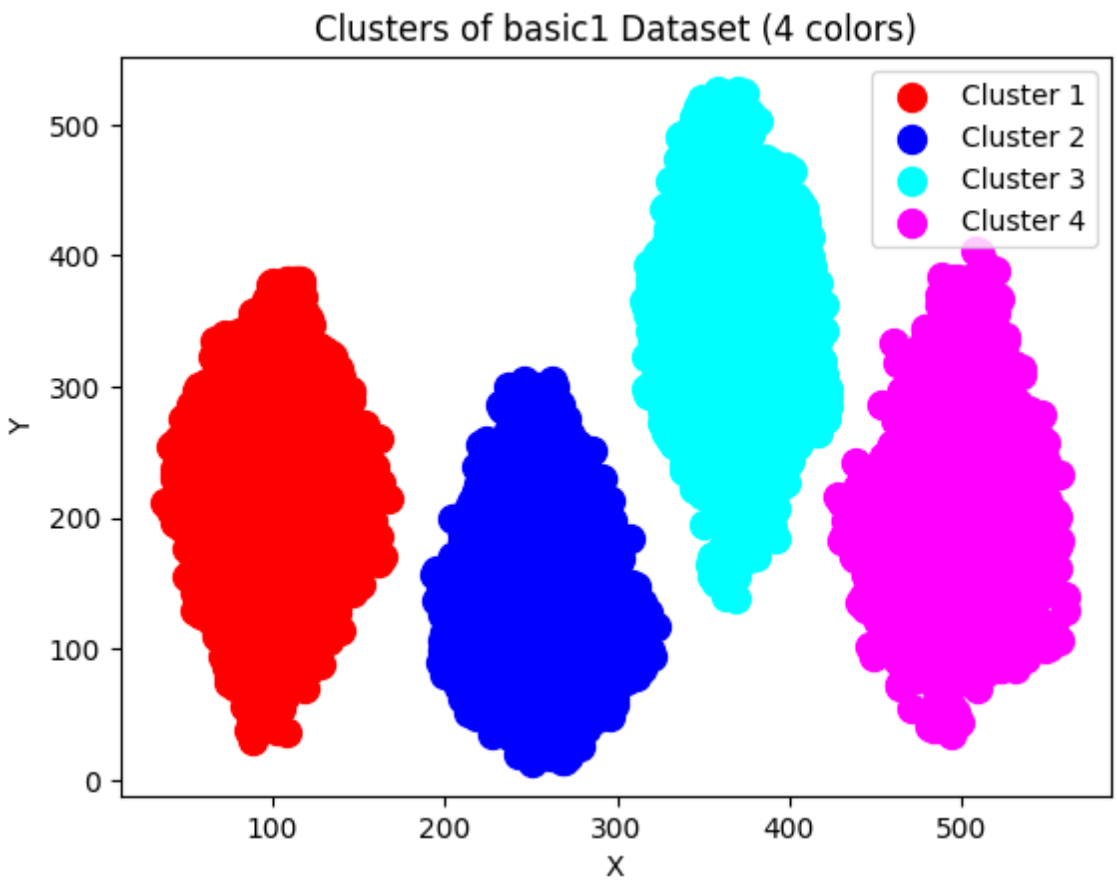
Out[29]:

	x	y	color
0	79.408289	152.834424	0
1	98.046263	186.910700	0
2	240.578979	48.473684	1
3	109.687183	277.945769	0
4	249.626082	229.753352	1

In [30]:
```
basic1['color'].unique()
```

Out[30]:
```
array([0, 1, 3, 2])
```

In [127]:
```
visualize_clusters(basic1, 4, ['red','blue','cyan','magenta'], 'basic 1')
```



Dataset 2 : basic2 Dataset

In [32]:
```
basic2 = pd.read_csv('/kaggle/input/clustering-exercises/basic2.csv')
```

In [33]:
```
basic2.head()
```

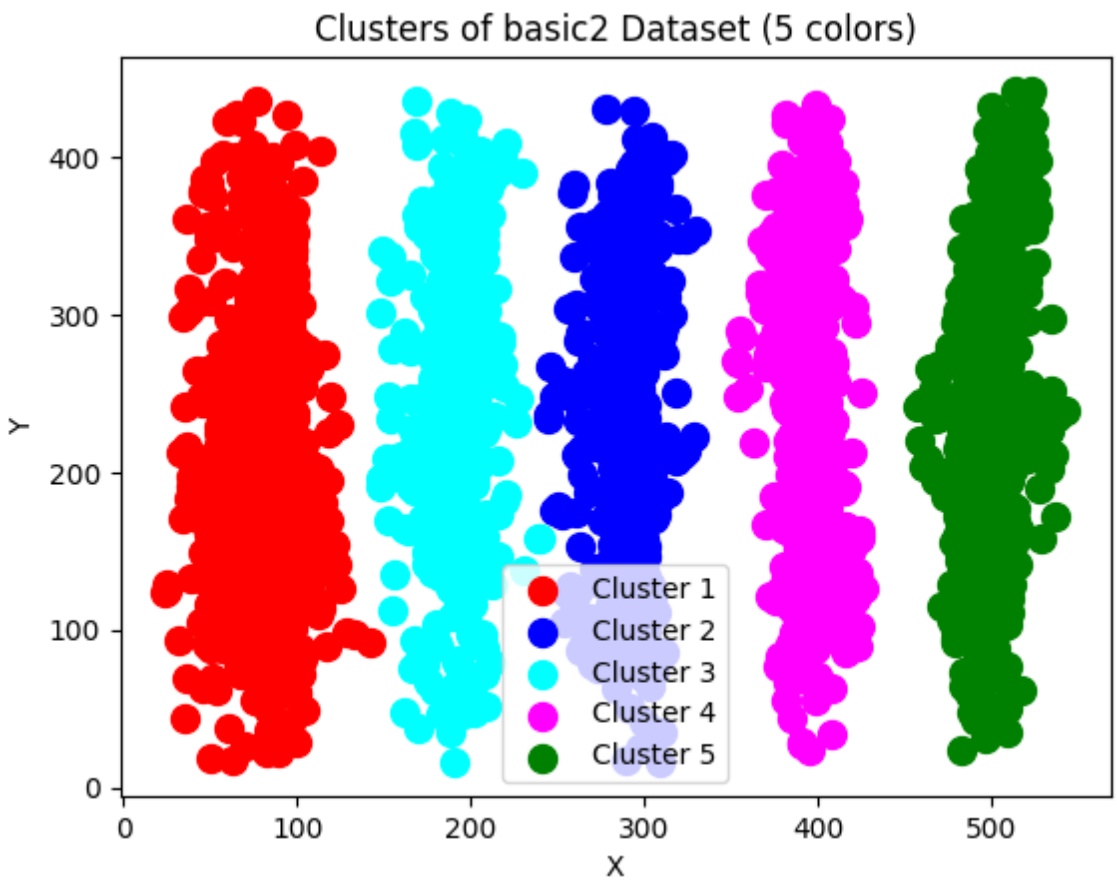
Out[33]:

	x	y	color
0	394.778257	162.929596	3
1	113.187672	136.498196	0
2	400.937252	194.179802	3
3	194.372621	208.106101	2
4	290.609779	135.674950	1

In [34]:
```
basic2['color'].unique()
```

Out[34]:
```
array([3, 0, 2, 1, 4])
```

In [128]:
```
visualize_clusters(basic2, 5, ['red','blue','cyan','magenta','green'],
'basic2')
```



Dataset : basic3

In [35]:
```
basic3 = pd.read_csv('/kaggle/input/clustering-exercises/basic3.csv')
```

In [36]:
```
basic3.head()
```

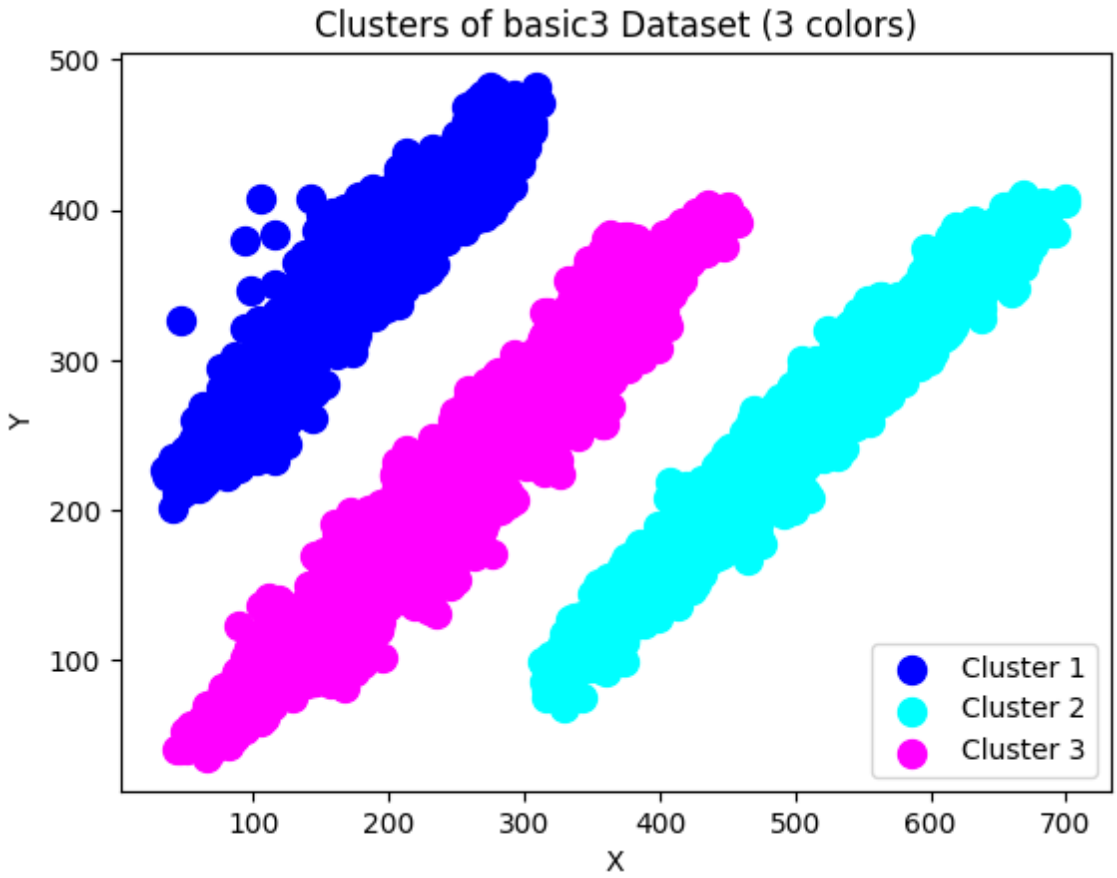
Out[36]:

	x	y	color
0	591.141815	339.509605	1
1	152.000000	337.000000	0
2	306.383225	304.362656	2
3	532.384613	279.305239	1
4	231.786211	216.340788	2

In [37]:
```
basic3['color'].unique()
```

Out[37]:
```
array([1, 0, 2])
```

In [129]:
```
visualize_clusters(basic3, 3, ['blue','cyan','magenta'], 'basic3')
```



Dataset : basic4

In [38]:
```
basic4 = pd.read_csv('/kaggle/input/clustering-exercises/basic4.csv')
```

In [39]:
```
basic4.head()
```

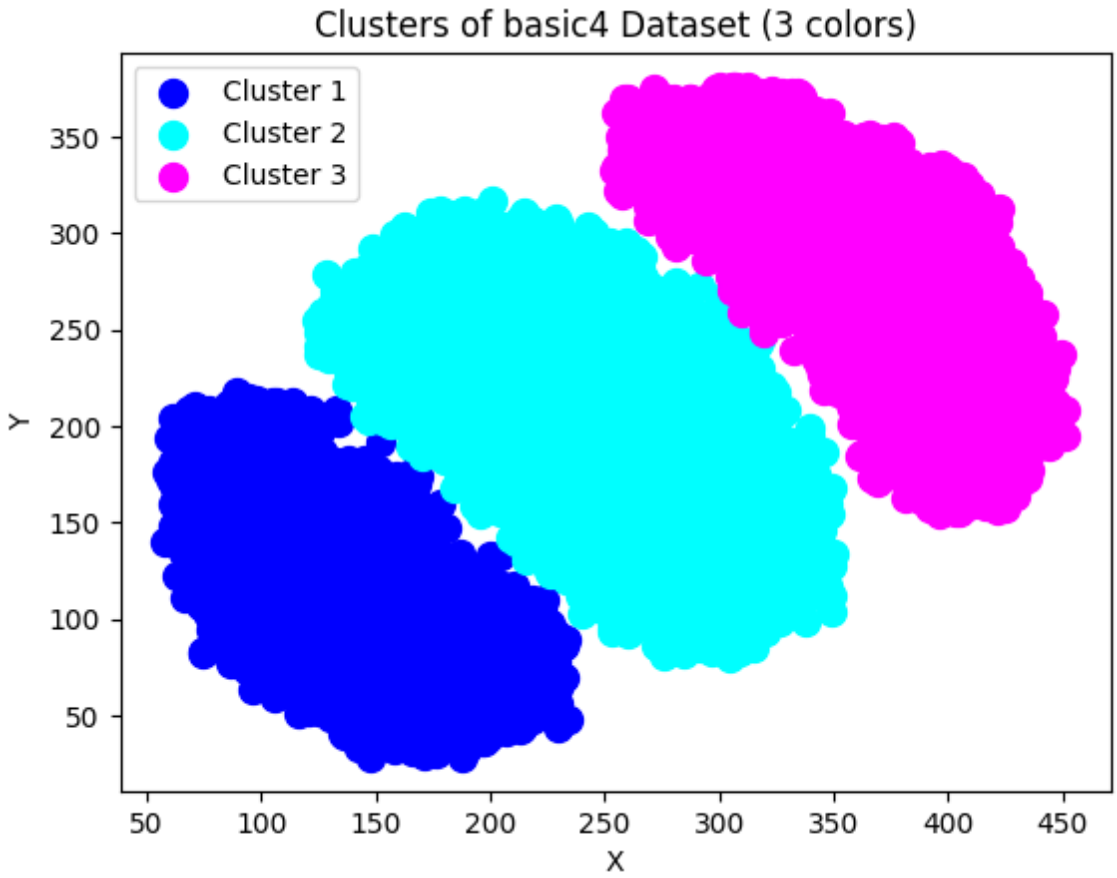
Out[39]:

	x	y	color
0	144.475109	101.450265	0
1	370.683312	298.003415	2
2	253.714046	234.079402	1
3	148.699165	78.654006	0
4	192.157973	252.837304	1

In [40]:
```
basic4['color'].unique()
```

Out[40]:
```
array([0, 2, 1])
```

In [130]:
```
visualize_clusters(basic4, 3, ['blue','cyan','magenta'], 'basic4')
```



Dataset : basic5

In [41]:
```
basic5 = pd.read_csv('/kaggle/input/clustering-exercises/basic5.csv')
```

In [42]:
```
basic5.head()
```

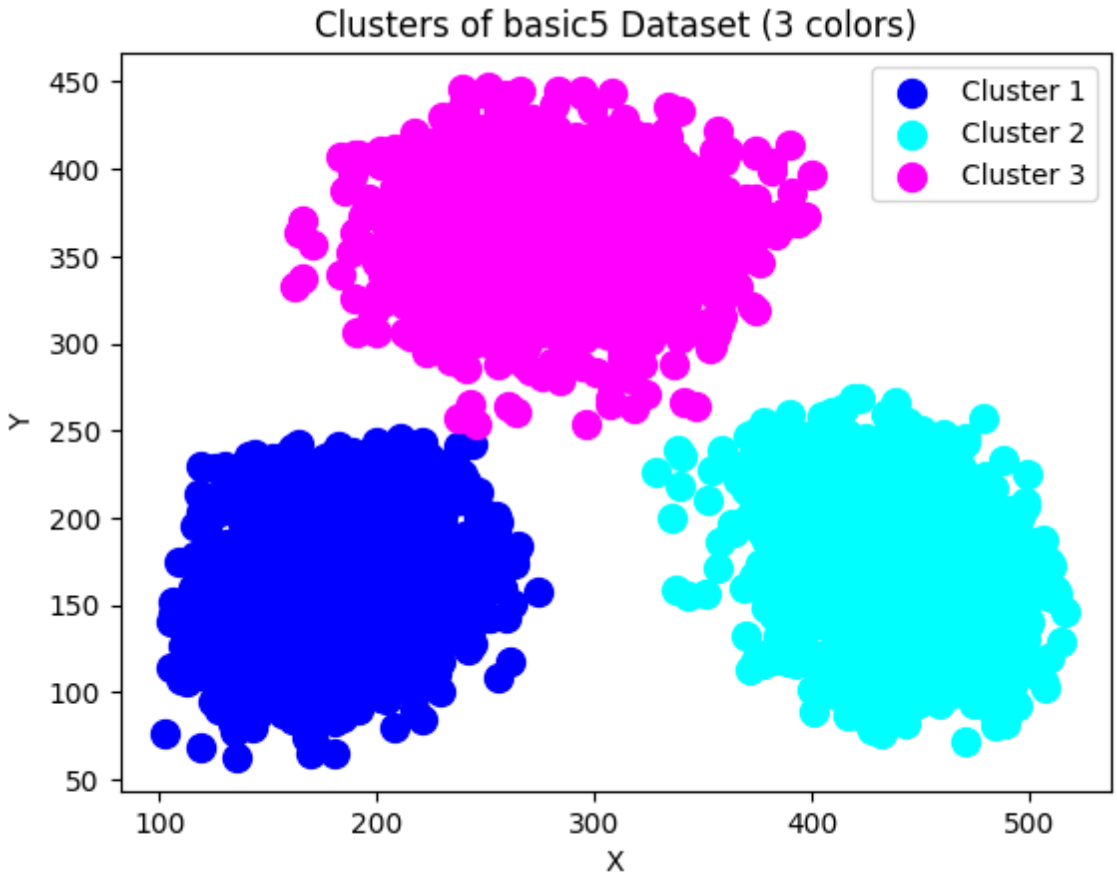
Out[42]:

	x	y	color
0	472.431845	133.637138	1
1	392.213650	248.151058	1
2	185.231588	128.911761	0
3	175.553180	234.373421	0
4	199.902134	373.492265	2

In [43]:
```
basic5['color'].unique()
```

Out[43]:
```
array([1, 0, 2])
```

In [131]:
```
visualize_clusters(basic5, 3, ['blue','cyan','magenta'], 'basic5')
```



Dataset : blob

In [44]:
```
blob = pd.read_csv('/kaggle/input/clustering-exercises/blob.csv')
```

In [45]:
```
blob.head()
```

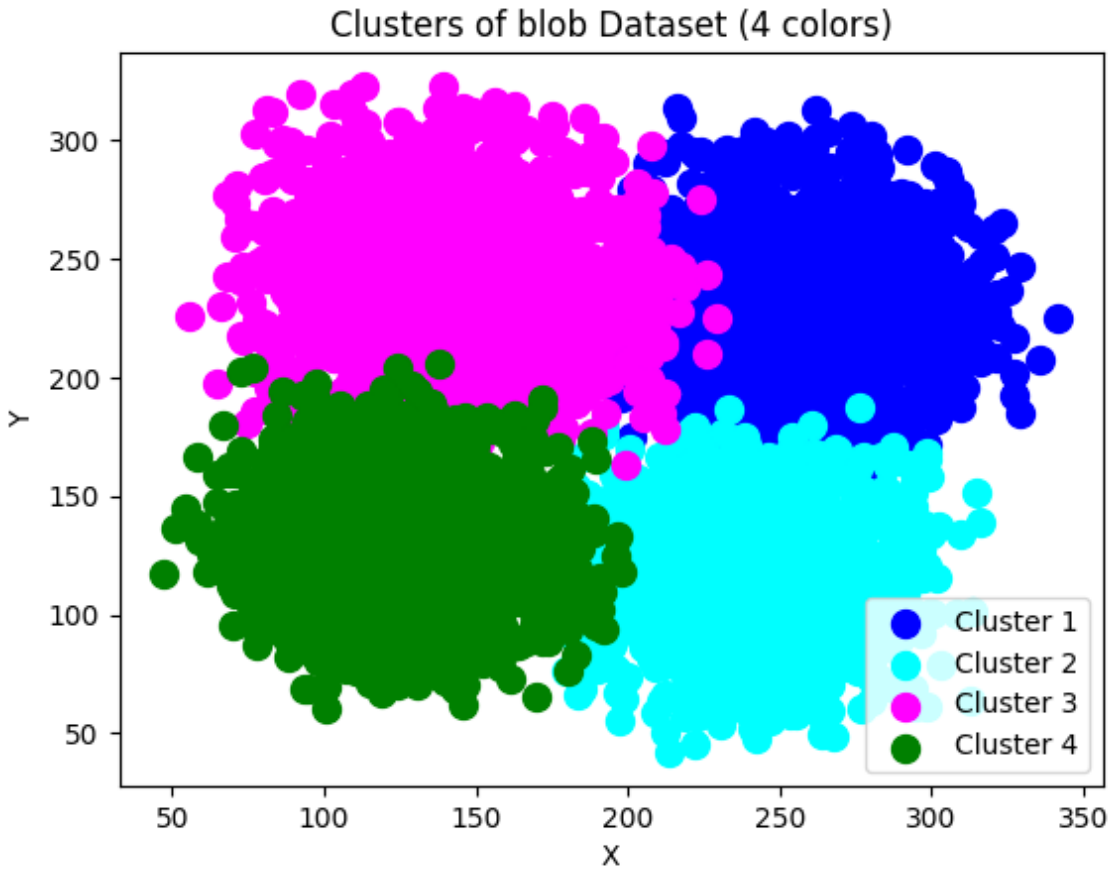
Out[45]:

	x	y	color
0	199.115034	72.641121	1
1	196.880523	247.744708	2
2	88.426271	233.962425	2
3	118.578789	264.044188	2
4	249.572490	260.896521	0

In [46]:
```
blob['color'].unique()
```

Out[46]:
```
array([1, 2, 0, 3])
```

In [133]:
```
visualize_clusters(blob, 4, ['blue','cyan','magenta','green'], 'blob')
```



Dataset : box

In [47]:
```
box = pd.read_csv('/kaggle/input/clustering-exercises/box.csv')
```

In [48]:
```
box.head()
```

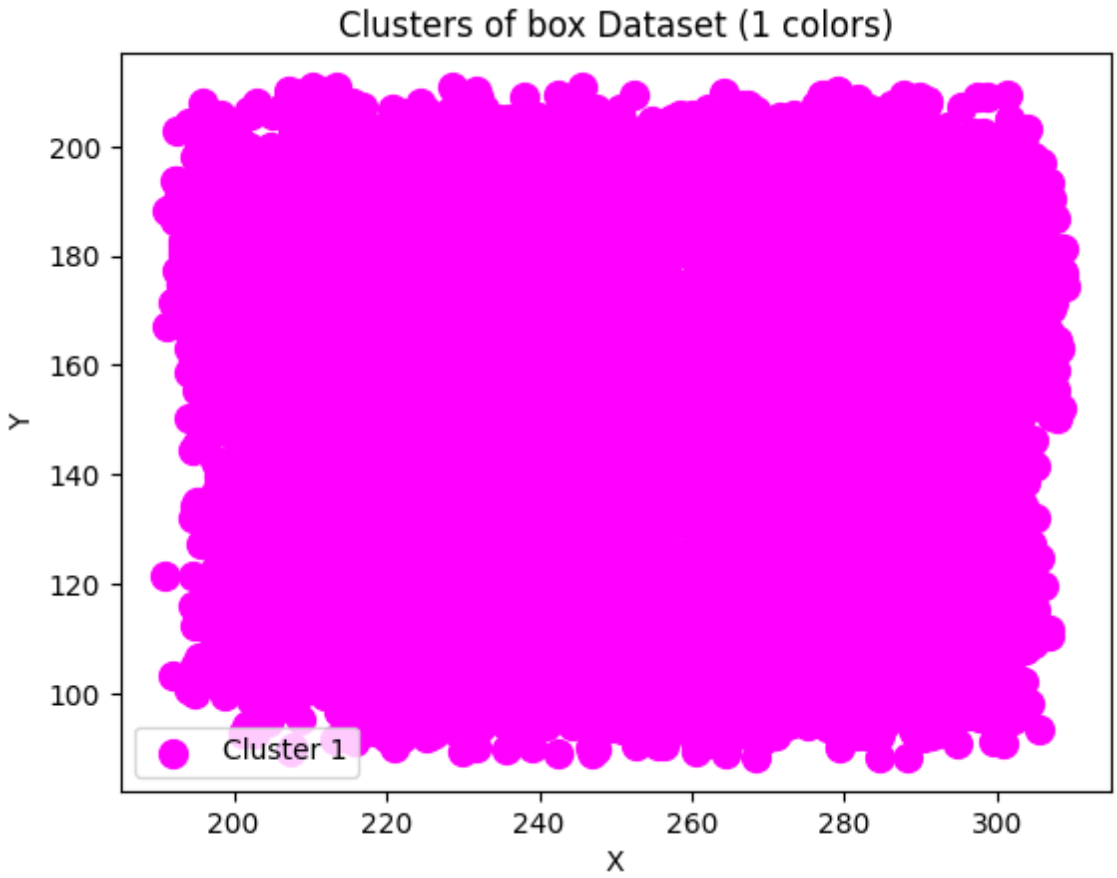
Out[48]:

	x	y	color
0	266.118821	175.407211	0
1	263.650932	109.278043	0
2	196.042446	155.582408	0
3	261.007069	121.454725	0
4	280.610358	160.586550	0

In [49]:
```
box['color'].unique()
```

Out[49]:
```
array([0])
```

In [134]:
```
visualize_clusters(box, 1, ['magenta'], 'box')
```



Dataset : boxes

In [50]:
```
boxes = pd.read_csv('/kaggle/input/clustering-exercises/boxes.csv')
```

In [51]:
```
boxes.head()
```

Out[51]:

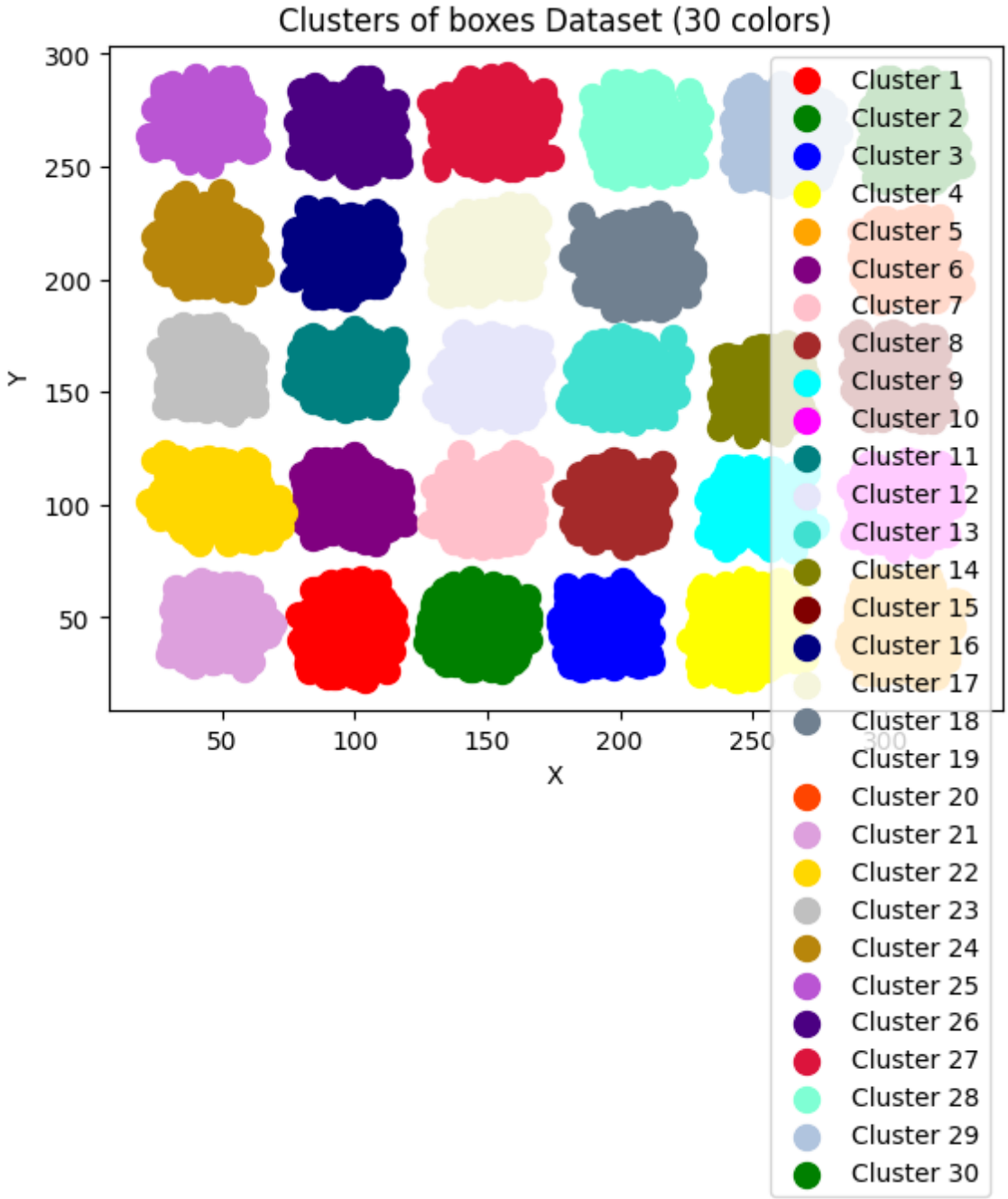
	x	y	color
0	181.635492	210.485307	17
1	198.440443	223.047363	17
2	43.552279	110.473015	21
3	313.263474	106.631427	9
4	194.493919	144.095507	12

In [135]:
```
len(boxes['color'].unique()),boxes['color'].unique()
```

Out[135]:
```
(30,
 array([17, 21,  9, 12, 24,  4, 10, 14, 26, 29,  0, 18,  1, 25, 15, 2
0,  8,
         3,  2, 19, 28,  6,  7, 22, 11, 27, 13, 16, 23,  5]))
```

In [142]:
```
colors = ["red", "green", "blue", "yellow", "orange", "purple", "pin
k", "brown", "cyan",
                "magenta", "teal", "lavender", "turquoise", "olive",
"maroon", "navy", "beige",
                "slategray", "white", "orangered", "plum", "gold", "s
ilver", "darkgoldenrod",
                "mediumorchid", "indigo", "crimson", "aquamarine", "l
ightsteelblue", "green"]


visualize_clusters(boxes, 30, colors, 'boxes')
```



Dataset : boxes2

In [53]:
```
boxes2 = pd.read_csv('/kaggle/input/clustering-exercises/boxes2.csv')
```

In [54]:
```
boxes2.head()
```

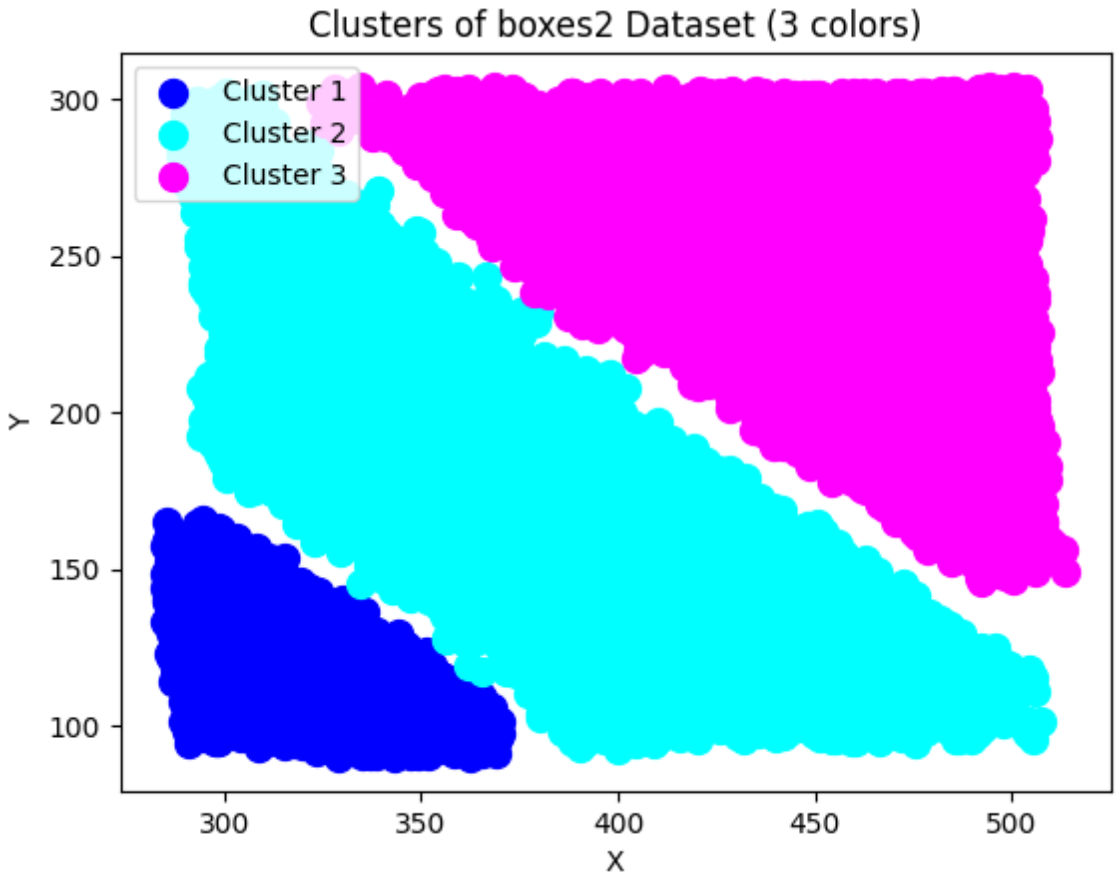
Out[54]:

	x	y	color
0	490.127172	115.109548	1
1	294.267870	158.581321	0
2	440.425055	203.140844	2
3	494.372290	112.077622	1
4	316.361087	238.759800	1

In [55]:
```
boxes2['color'].unique()
```

Out[55]:
```
array([1, 0, 2])
```

In [143]:
```
visualize_clusters(boxes2, 3, ['blue','cyan','magenta'], 'boxes2')
```



Dataset : boxes3

In [56]:
```
boxes3 = pd.read_csv('/kaggle/input/clustering-exercises/boxes3.csv')
```

In [57]:
```
boxes3.head()
```

Out[57]:

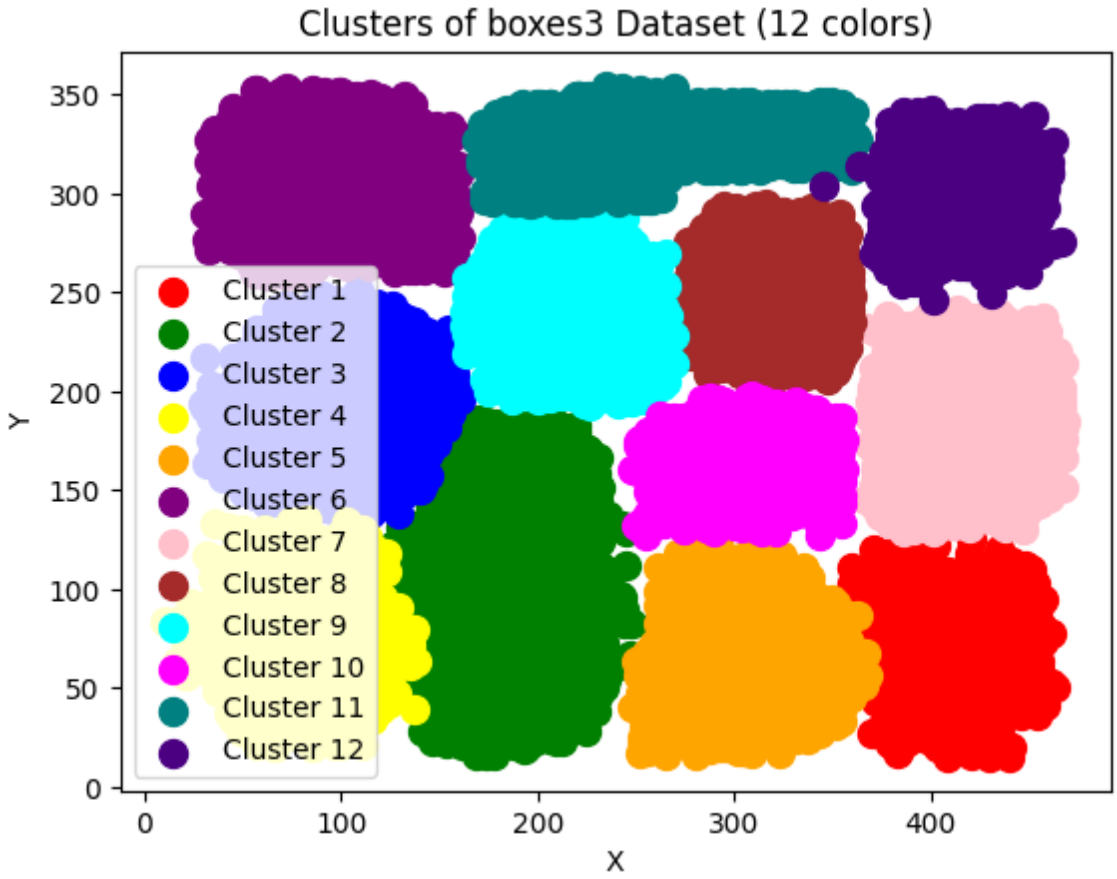
	x	y	color
0	400.806675	174.702750	6
1	401.154038	70.450386	0
2	294.722093	30.296820	4
3	415.366580	86.633522	0
4	181.464244	83.254630	1

In [58]:
```
boxes3['color'].unique()
```

Out[58]:
```
array([ 6,  0,  4,  1,  7,  3,  8,  5, 10,  2,  9, 11])
```

In [145]:
```
colors = ["red", "green", "blue", "yellow", "orange", "purple", "pin
k", "brown", "cyan",
                "magenta", "teal", "indigo"]


visualize_clusters(boxes3, 12, colors, 'boxes3')
```



Dataset : chrome

In [59]:
```
chrome = pd.read_csv('/kaggle/input/clustering-exercises/chrome.csv')
```

In [60]:
```
chrome.head()
```

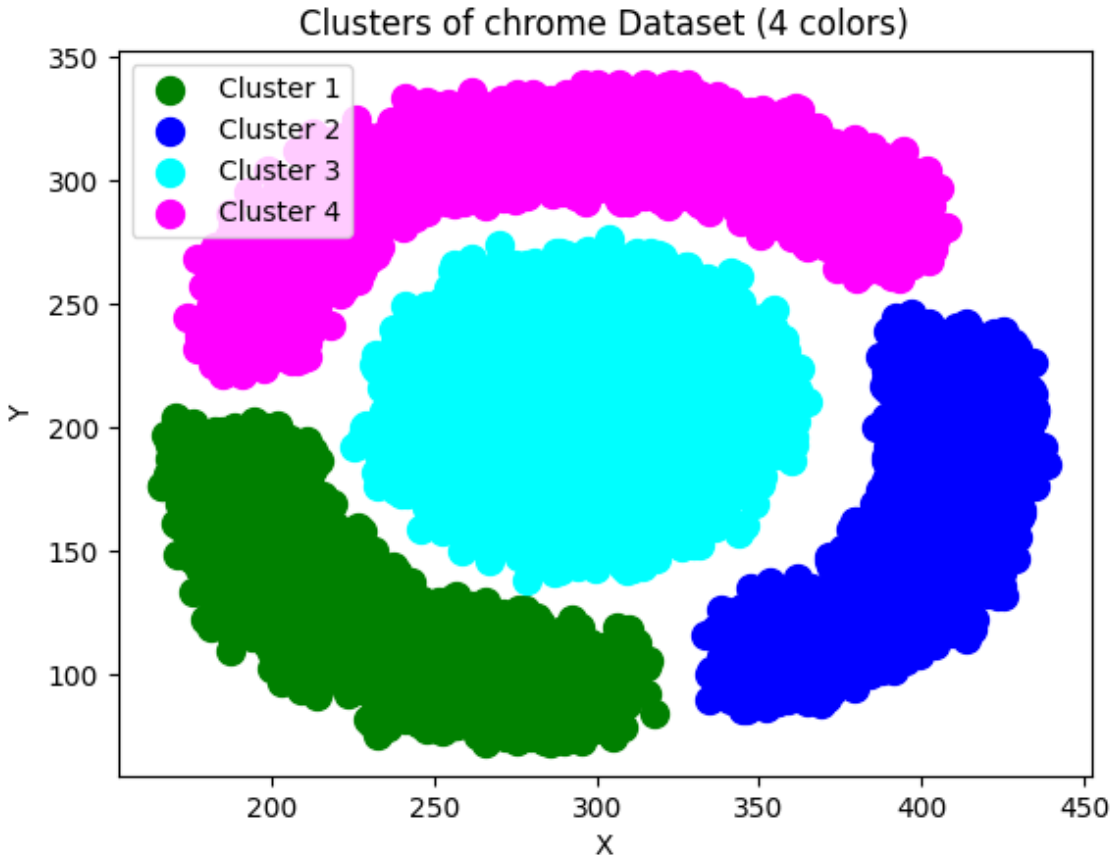
Out[60]:

	x	y	color
0	317.614913	197.197881	2
1	187.085153	164.653509	0
2	245.989615	233.783184	2
3	356.170303	206.712568	2
4	424.640194	141.090956	1

In [61]:
```
chrome['color'].unique()
```

Out[61]:
```
array([2, 0, 1, 3])
```

In [146]:
```
visualize_clusters(chrome, 4, ['green','blue','cyan','magenta'], 'chro
me')
```



Dataset : dart

In [62]:
```
dart = pd.read_csv('/kaggle/input/clustering-exercises/dart.csv')
```

In [63]:
```
dart.head()
```

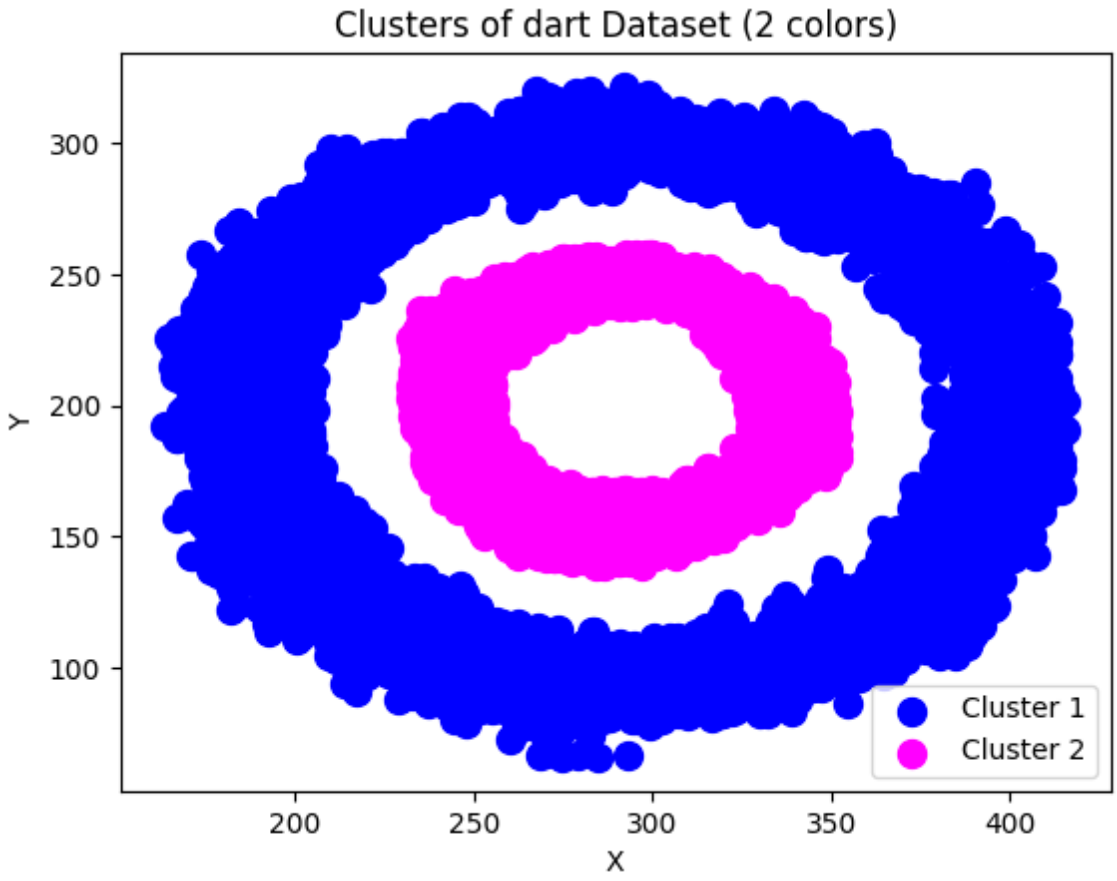
Out[63]:

	x	y	color
0	239.650438	300.176323	0
1	288.373881	102.813599	0
2	334.129409	165.764844	1
3	193.401962	202.979989	0
4	294.588391	98.569759	0

In [64]:
```
dart['color'].unique()
```

Out[64]:
```
array([0, 1])
```

In [148]:
```
visualize_clusters(dart, 2, ['blue','magenta'], 'dart')
```



Dataset : dart2

In [65]:
```
dart2 = pd.read_csv('/kaggle/input/clustering-exercises/dart2.csv')
```

In [66]:
```
dart2.head()
```

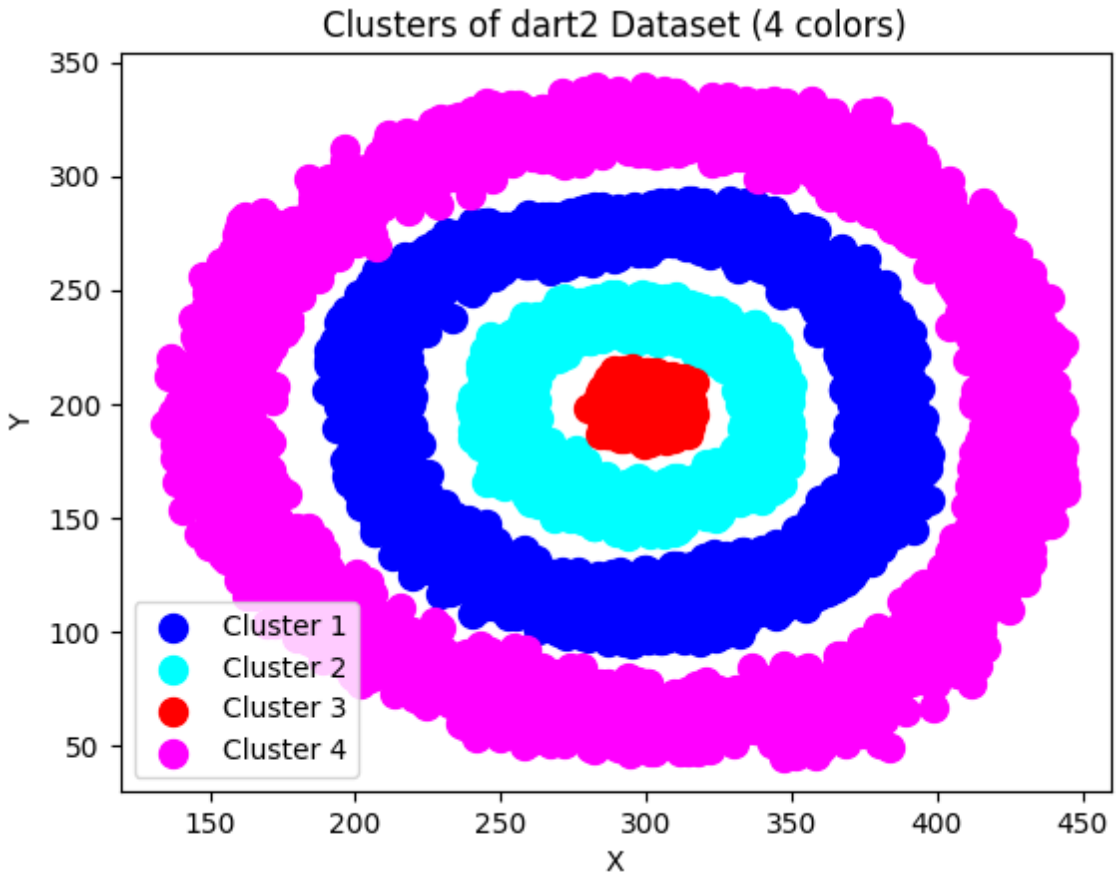
Out[66]:

	x	y	color
0	279.925028	310.327844	3
1	402.967849	84.953838	3
2	432.899756	241.304197	3
3	355.766547	72.821111	3
4	320.429621	240.766330	1

In [67]:
```
dart2['color'].unique()
```

Out[67]:
```
array([3, 1, 0, 2])
```

In [149]:
```
visualize_clusters(dart2, 4, ['blue','cyan','red','magenta'], 'dart2')
```



Dataset : face

In [68]:
```
face = pd.read_csv('/kaggle/input/clustering-exercises/face.csv')
```

In [69]:
```
face.head()
```

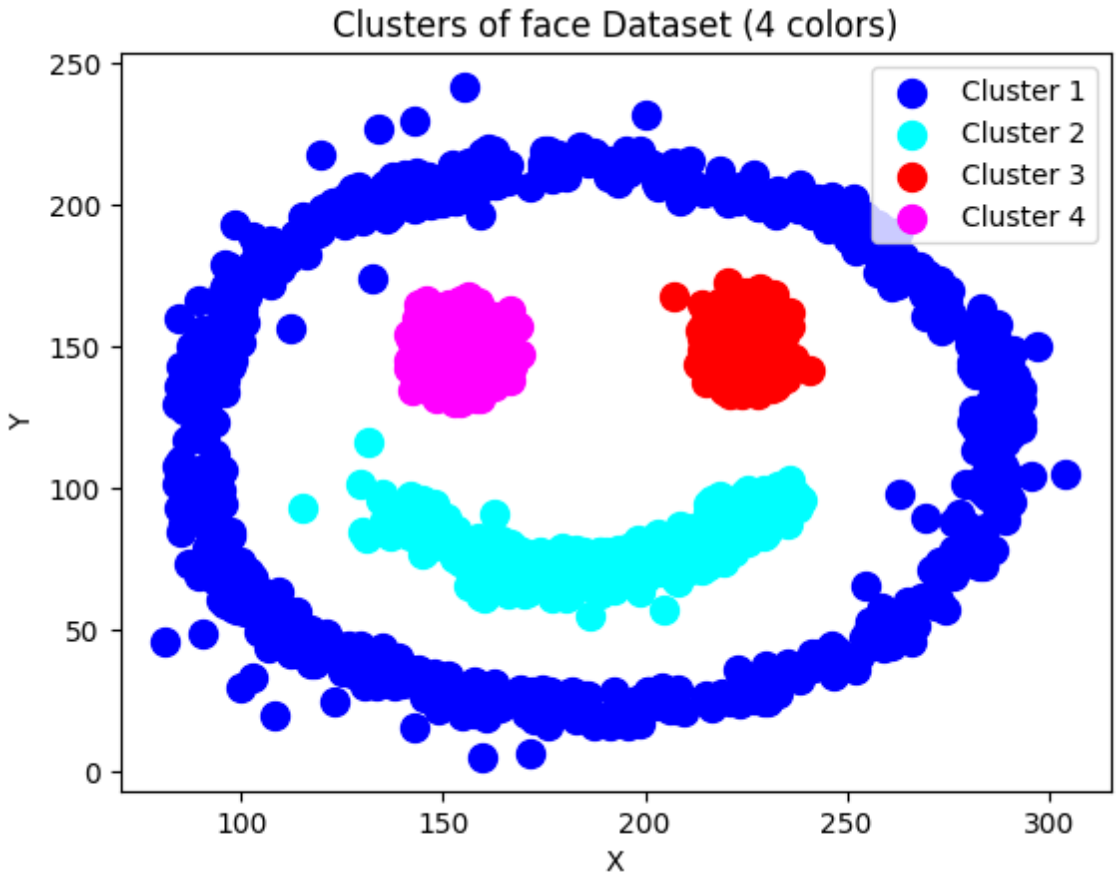
Out[69]:

	Unnamed: 0.1	x	y	color
0	1087	230.378162	141.270406	2
1	464	281.264318	127.419565	0
2	1237	157.530566	158.890252	3
3	68	107.354057	175.958260	0
4	1151	148.118641	161.809148	3

In [70]:
```
face['color'].unique()
```

Out[70]:
```
array([2, 0, 3, 1])
```

In [150]:
```
visualize_clusters(face, 4, ['blue','cyan','red','magenta'], 'face')
```



Dataset : hyperplane

In [71]:
```
hyperplane = pd.read_csv('/kaggle/input/clustering-exercises/hyperplane.csv')
```

In [72]:
```
hyperplane.head()
```

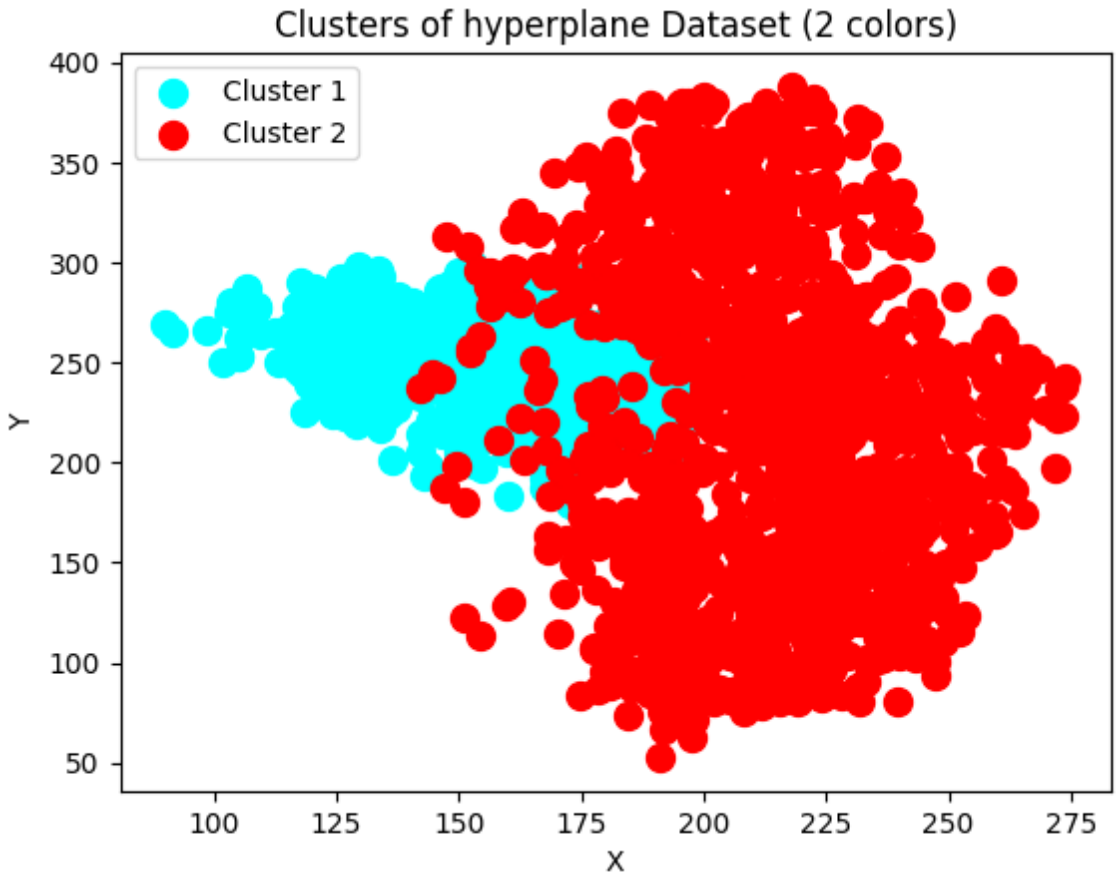
Out[72]:

	x	y	color
0	237.491363	136.898504	1
1	180.470910	202.233104	0
2	213.905287	310.329493	1
3	183.713515	172.680302	1
4	244.125483	214.833692	1

In [73]:
```
hyperplane['color'].unique()
```

Out[73]:
```
array([1, 0])
```

In [151]:
```
visualize_clusters(hyperplane, 2, ['cyan','red'], 'hyperplane')
```



Dataset : isolation

In [74]:
```
isolation = pd.read_csv('/kaggle/input/clustering-exercises/isolation.csv')
```

In [75]:
```
isolation.head()
```

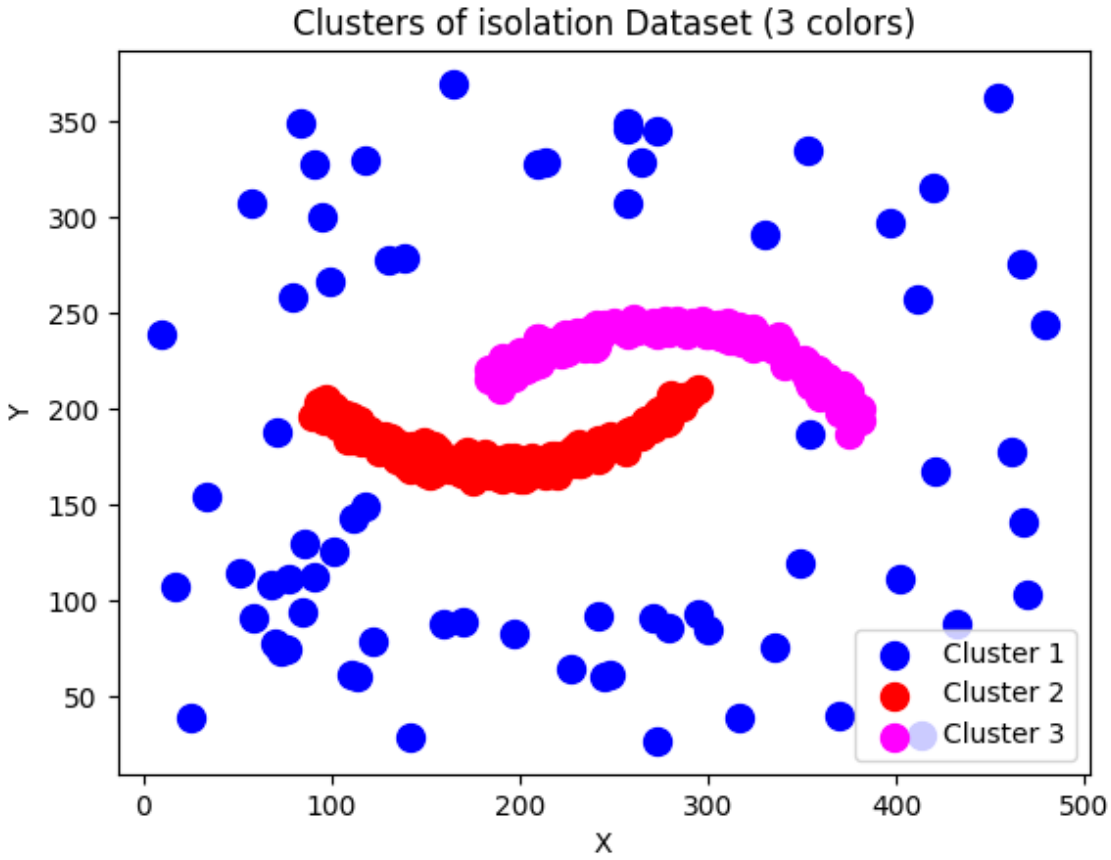
Out[75]:

	x	y	color
0	257.431260	345.345735	0
1	317.520132	234.087272	2
2	227.093467	64.139227	0
3	148.855905	174.524528	1
4	419.977732	315.164370	0

In [76]:
```
isolation['color'].unique()
```

Out[76]:
```
array([0, 2, 1])
```

In [152]:
```
visualize_clusters(isolation, 3, ['blue','red','magenta'], 'isolatio
n')
```



Dataset : lines

In [77]:
```
lines = pd.read_csv('/kaggle/input/clustering-exercises/lines.csv')
```

In [78]:
```
lines.head()
```

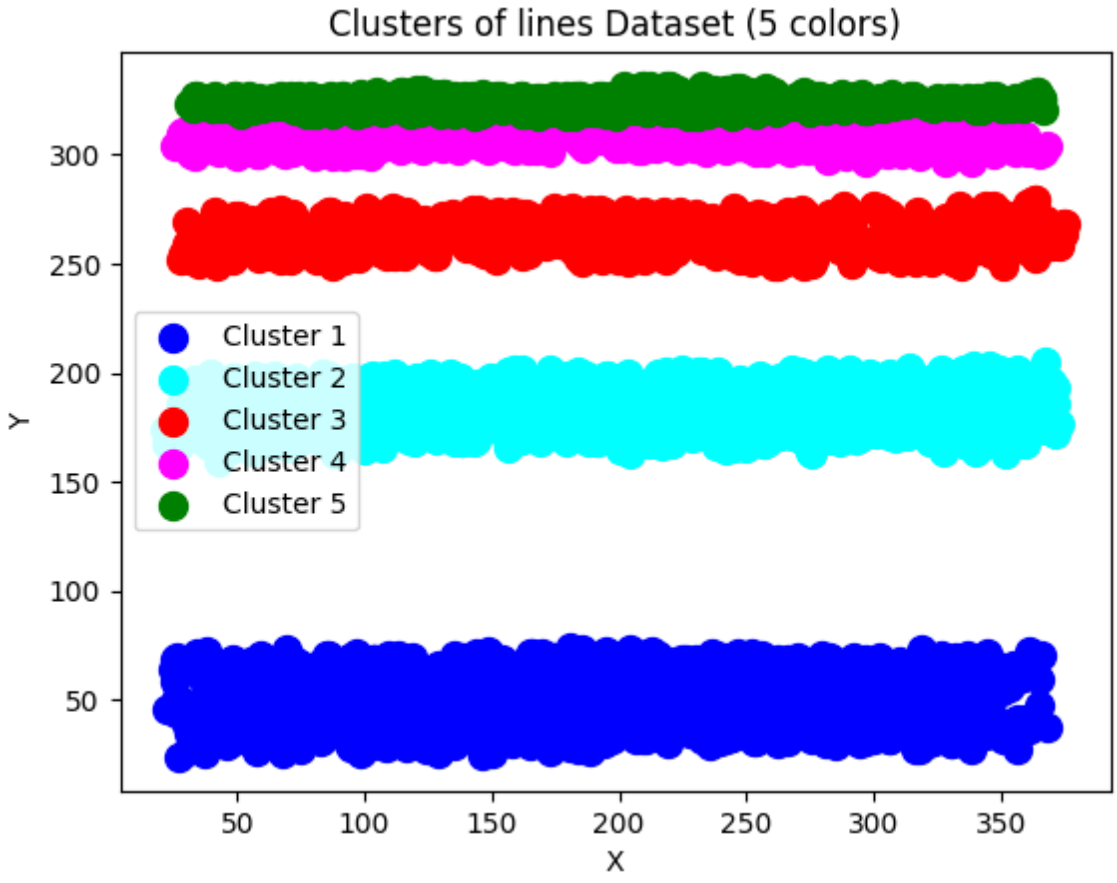
Out[78]:

	x	y	color
0	52.064627	324.023156	4
1	147.144185	31.042611	0
2	199.120406	48.122634	0
3	54.144343	56.319339	0
4	290.792804	55.428726	0

In [79]:
```
lines['color'].unique()
```

Out[79]:
```
array([4, 0, 1, 2, 3])
```

In [153]:
```
visualize_clusters(lines, 5, ['blue','cyan','red','magenta','green'], 'lines')
```



Dataset : lines2

In [80]:
```
lines2 = pd.read_csv('/kaggle/input/clustering-exercises/lines2.csv')
```

In [81]:
```
lines2.head()
```

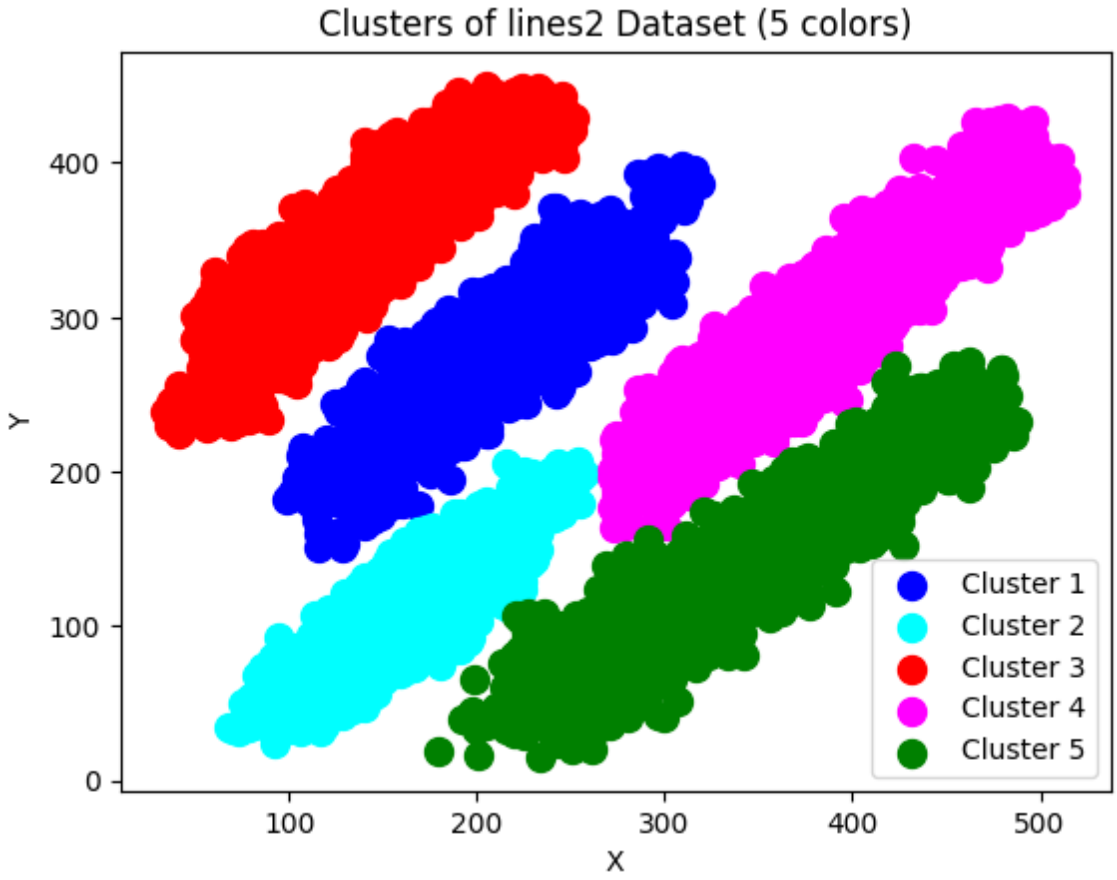
Out[81]:

	x	y	color
0	153.699456	94.381506	1
1	168.964194	103.066688	1
2	205.432316	303.114200	0
3	244.592738	69.336451	4
4	200.651645	248.183610	0

In [82]:
```
lines2['color'].unique()
```

Out[82]:
```
array([1, 0, 4, 3, 2])
```

In [154]:
```
visualize_clusters(lines2, 5, ['blue','cyan','red','magenta','green'],
'lines2')
```



Dataset : network

In [83]:
```
network = pd.read_csv('/kaggle/input/clustering-exercises/network.cs
v')
```

In [84]:
```
network.head()
```

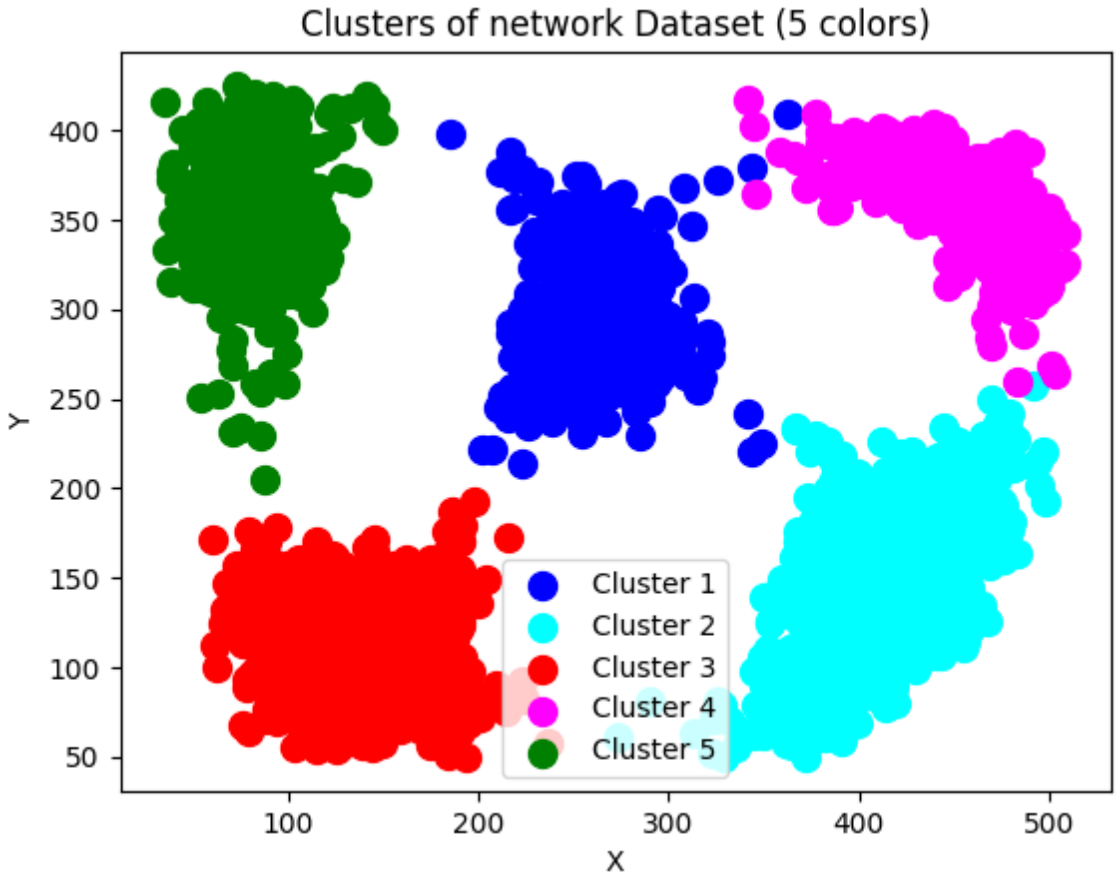
Out[84]:

	x	y	color
0	67.032380	131.777258	2
1	412.517529	156.706716	1
2	109.730290	364.408237	4
3	448.511829	343.291836	3
4	405.997147	128.725708	1

In [85]:
```
network['color'].unique()
```

Out[85]:
```
array([2, 1, 4, 3, 0])
```

In [155]:
```
visualize_clusters(network, 5, ['blue','cyan','red','magenta','gree
n'], 'network')
```



Dataset : outliers

In [86]:
```
outliers = pd.read_csv('/kaggle/input/clustering-exercises/outliers.cs
v')
```

In [87]:
```
outliers.head()
```

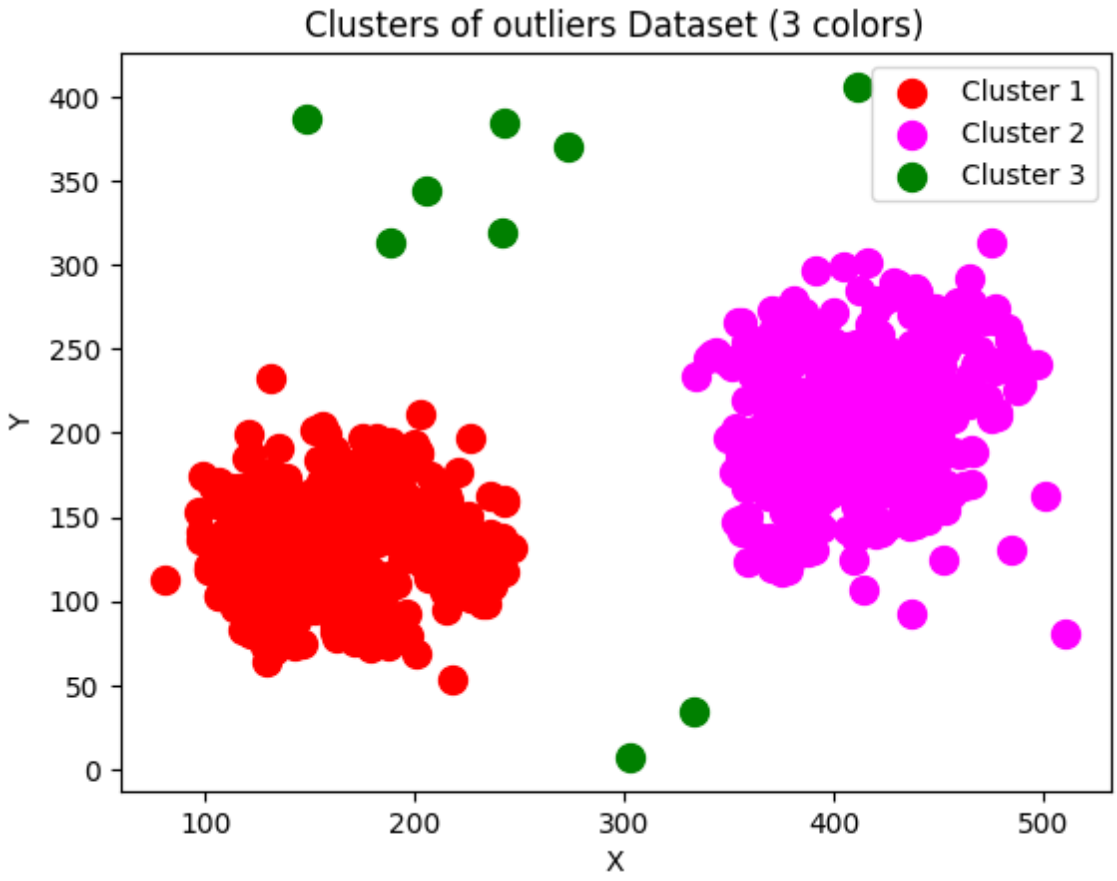
Out[87]:

	x	y	color
0	355.297197	148.880238	1
1	402.111403	244.566989	1
2	383.443863	224.191167	1
3	424.774071	207.348790	1
4	398.722062	214.262629	1

In [88]:
```
outliers['color'].unique()
```

Out[88]:
```
array([1, 0, 2])
```

In [156]:
```
visualize_clusters(outliers, 3, ['red','magenta','green'], 'outliers')
```



Dataset : ring

In [89]:
```
ring = pd.read_csv('/kaggle/input/clustering-exercises/ring.csv')
```

In [90]:
```
ring.head()
```

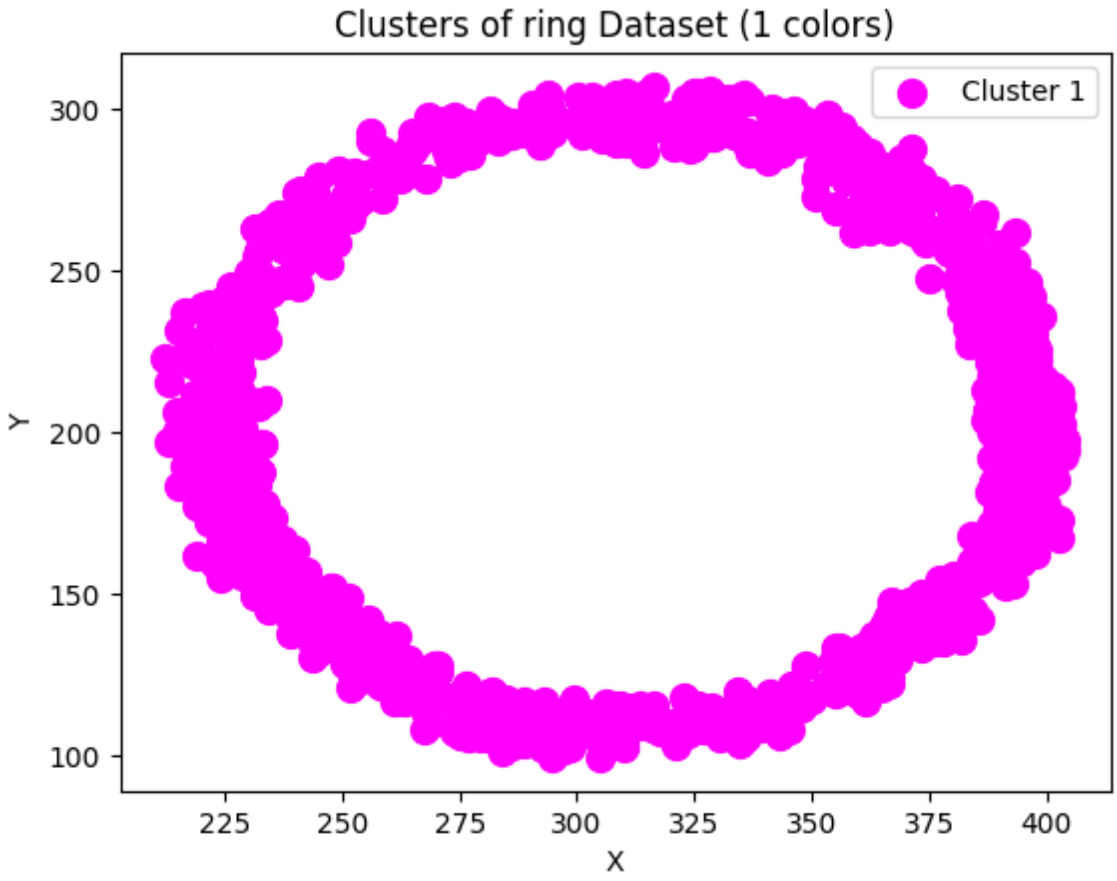
Out[90]:

	x	y	color
0	226.876818	221.537355	0
1	217.537777	193.984163	0
2	356.350749	294.262416	0
3	344.862025	118.975404	0
4	382.630427	254.557176	0

In [91]:
```
ring['color'].unique()
```

Out[91]:
```
array([0])
```

In [157]:
```
visualize_clusters(ring, 1, ['magenta'], 'ring')
```



Dataset : sparse

In [92]:
```
sparse = pd.read_csv('/kaggle/input/clustering-exercises/sparse.csv')
```

In [93]:
```
sparse.head()
```

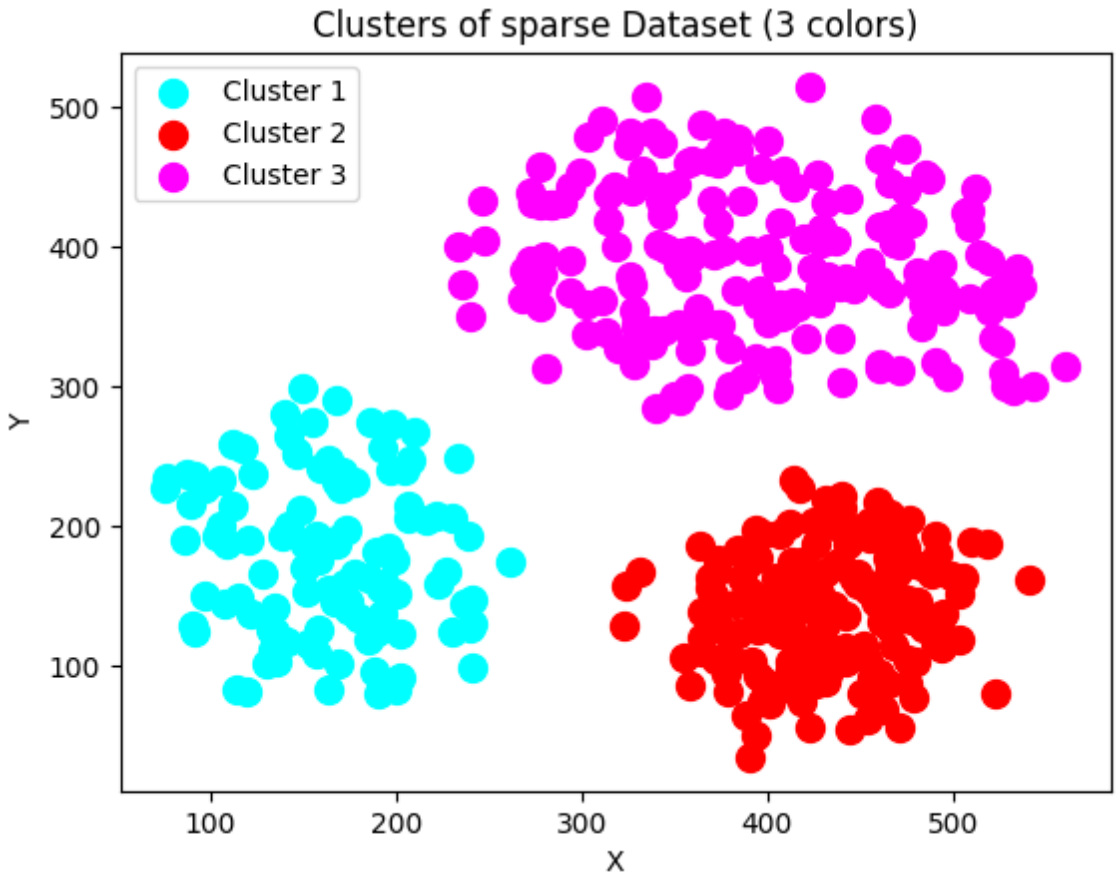
Out[93]:

	x	y	color
0	414.042384	360.539136	2
1	210.056031	267.597809	0
2	416.368288	357.085060	2
3	394.072632	51.351053	1
4	363.803924	186.693155	1

In [94]:
```
sparse['color'].unique()
```

Out[94]:
```
array([2, 0, 1])
```

In [159]:
```
visualize_clusters(sparse, 3, ['cyan','red','magenta'], 'sparse')
```



Dataset : spiral

In [95]:
```
spiral = pd.read_csv('/kaggle/input/clustering-exercises/spiral.csv')
```

In [96]:
```
spiral.head()
```

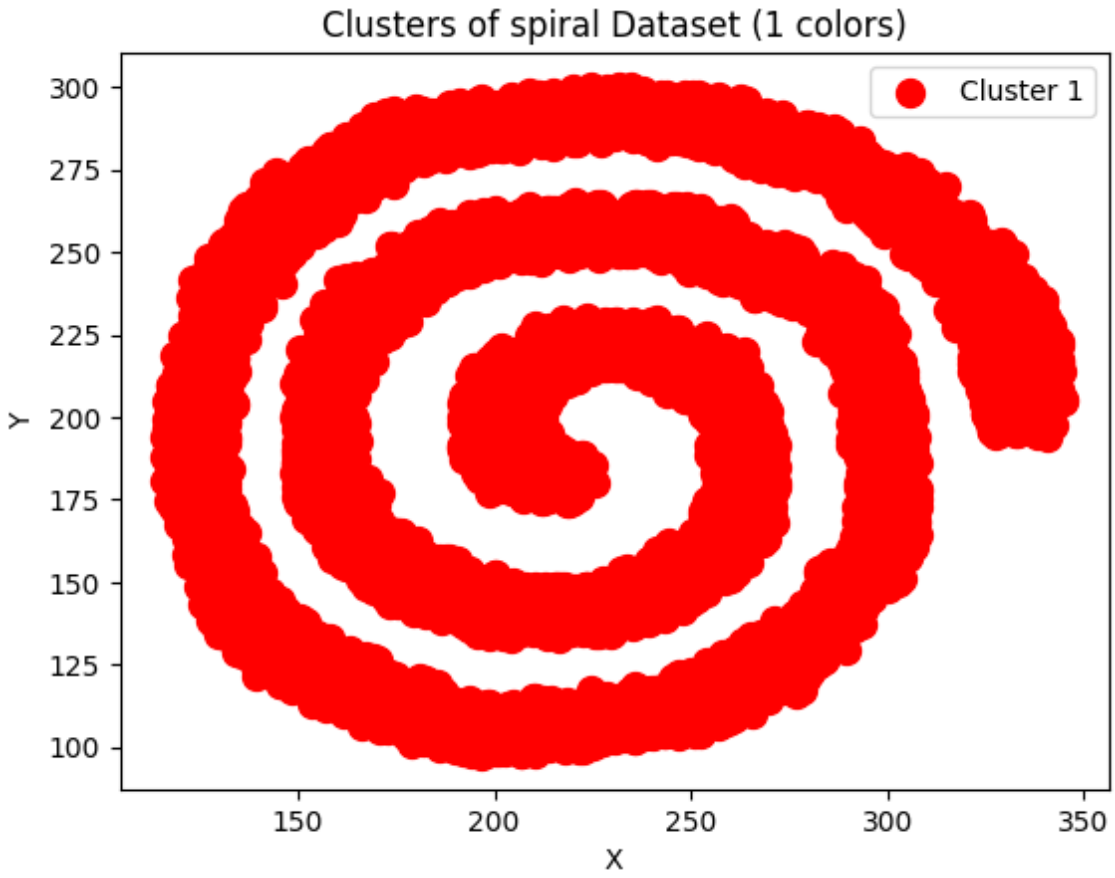
Out[96]:

	x	y	color
0	323.857377	225.428102	0
1	146.978425	257.287542	0
2	261.374770	253.624468	0
3	157.372865	162.127028	0
4	118.678976	188.888146	0

In [97]:
```
spiral['color'].unique()
```

Out[97]:
```
array([0])
```

In [160]:
```
visualize_clusters(spiral, 1, ['red'], 'spiral')
```



Dataset : spiral2

In [98]:
```
spiral2 = pd.read_csv('/kaggle/input/clustering-exercises/spiral2.csv')
```

In [99]:
```
spiral2.head()
```

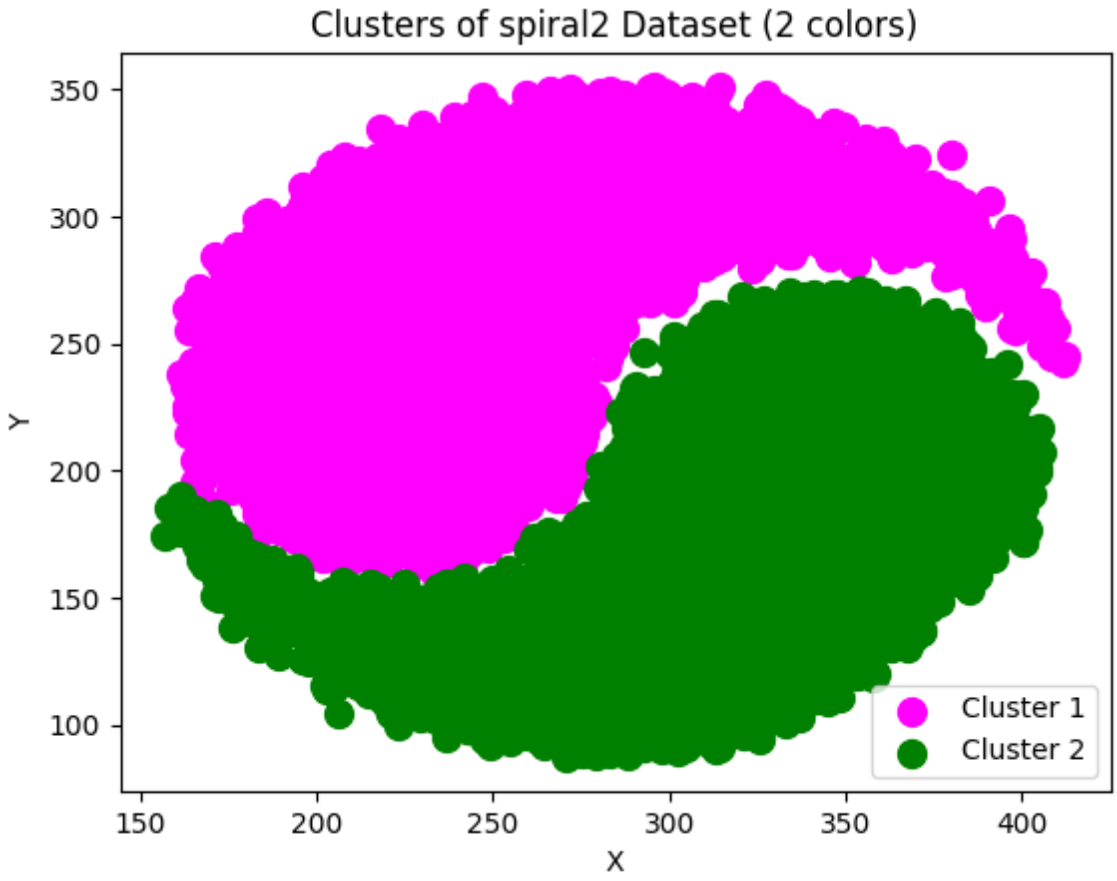
Out[99]:

	x	y	color
0	281.762010	186.024837	1
1	212.503675	199.031585	0
2	180.014895	259.379060	0
3	302.372404	284.465182	0
4	328.759727	176.945518	1

In [100]:
```
spiral2['color'].unique()
```

Out[100]:
```
array([1, 0])
```

In [161]:
```
visualize_clusters(spiral2, 2, ['magenta','green'], 'spiral2')
```



Dataset : spirals

In [101]:
```
spirals = pd.read_csv('/kaggle/input/clustering-exercises/spirals.csv')
```

In [102]:
```
spirals.head()
```

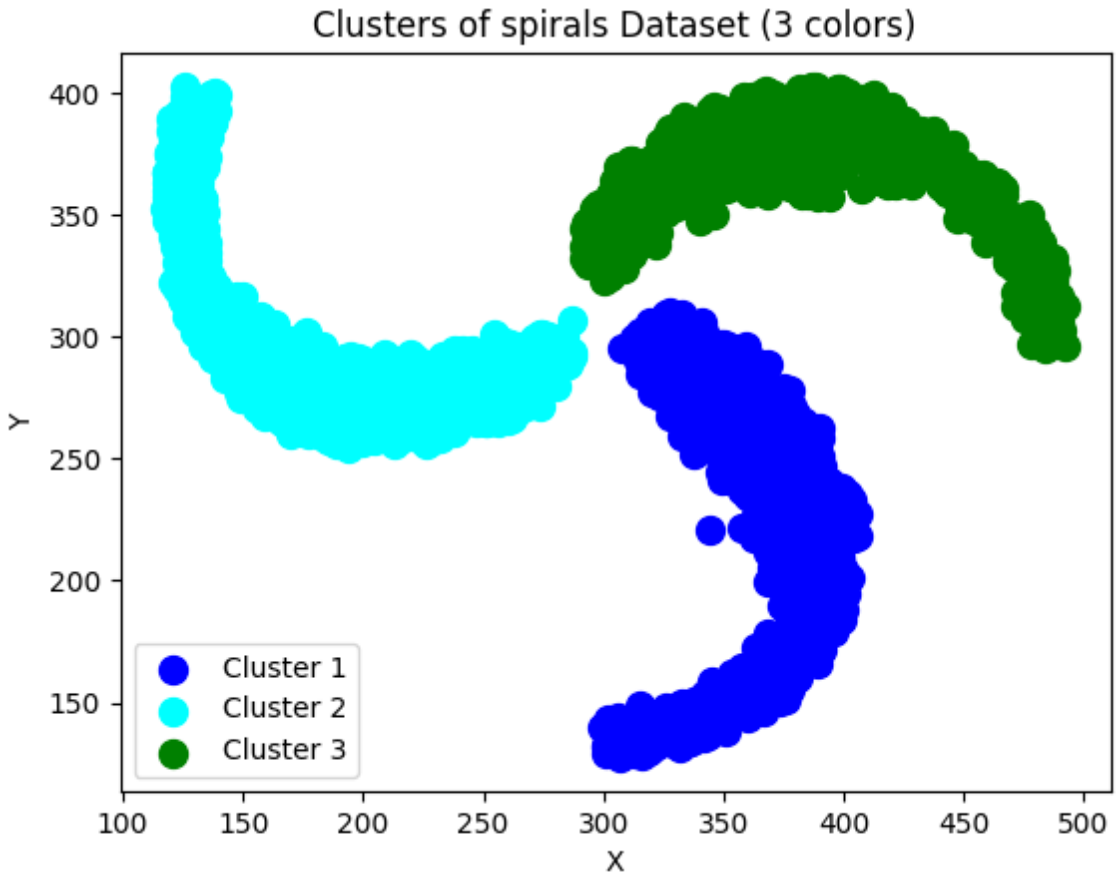
Out[102]:

	x	y	color
0	125.322566	386.764772	1
1	307.117266	337.307489	2
2	246.744149	275.732338	1
3	334.055047	291.824961	0
4	381.657454	251.359150	0

In [103]:
```
spirals['color'].unique()
```

Out[103]:
```
array([1, 2, 0])
```

In [162]:
```
visualize_clusters(spirals, 3, ['blue','cyan','green'], 'spirals')
```



Dataset : supernova

In [104]:
```
supernova = pd.read_csv('/kaggle/input/clustering-exercises/supernova.csv')
```

In [105]:
```
supernova.head()
```

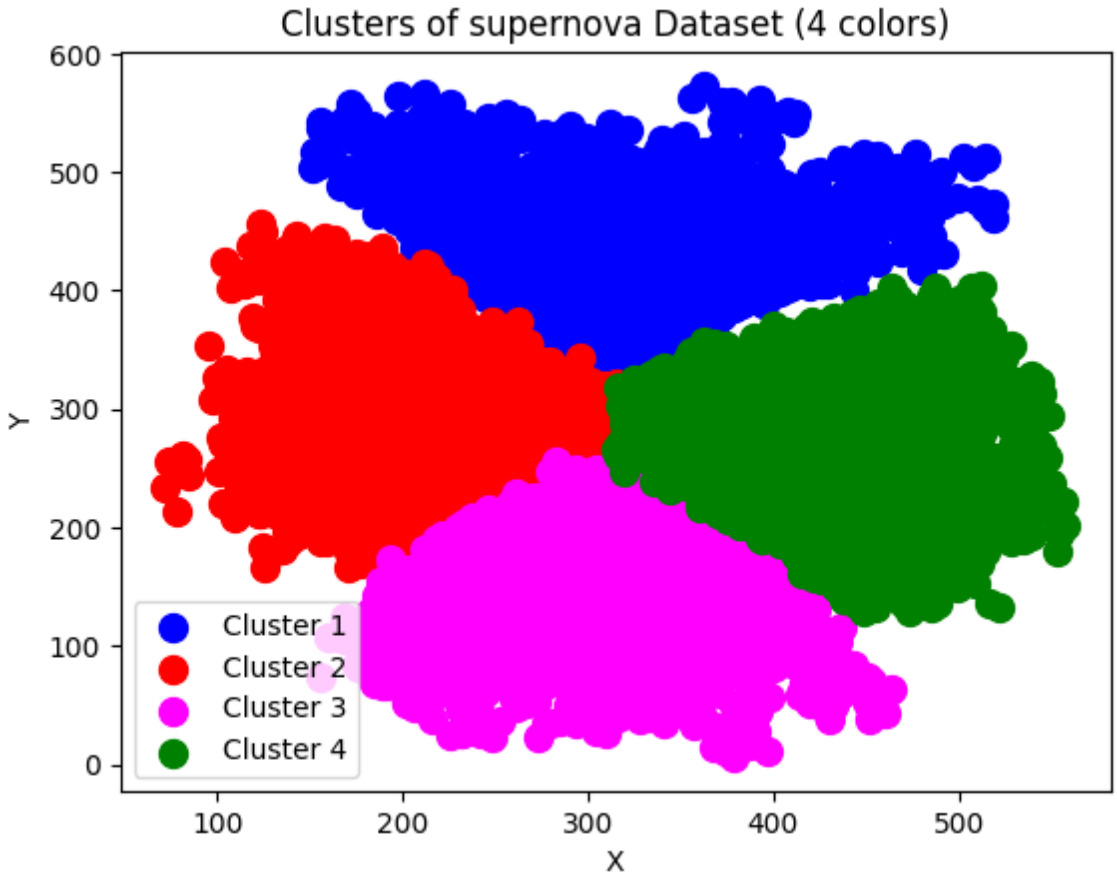
Out[105]:

	x	y	color
0	202.207191	333.390000	1
1	214.424501	374.531029	1
2	476.317977	255.623373	3
3	237.659471	148.173931	2
4	201.959301	355.790748	1

In [106]:
```
supernova['color'].unique()
```

Out[106]:
```
array([1, 3, 2, 0])
```

In [163]:
```
visualize_clusters(supernova, 4, ['blue','red','magenta','green'], 'supernova')
```



Dataset : triangle

In [107]:
```
triangle = pd.read_csv('/kaggle/input/clustering-exercises/triangle.csv')
```

In [108]:
```
triangle.head()
```

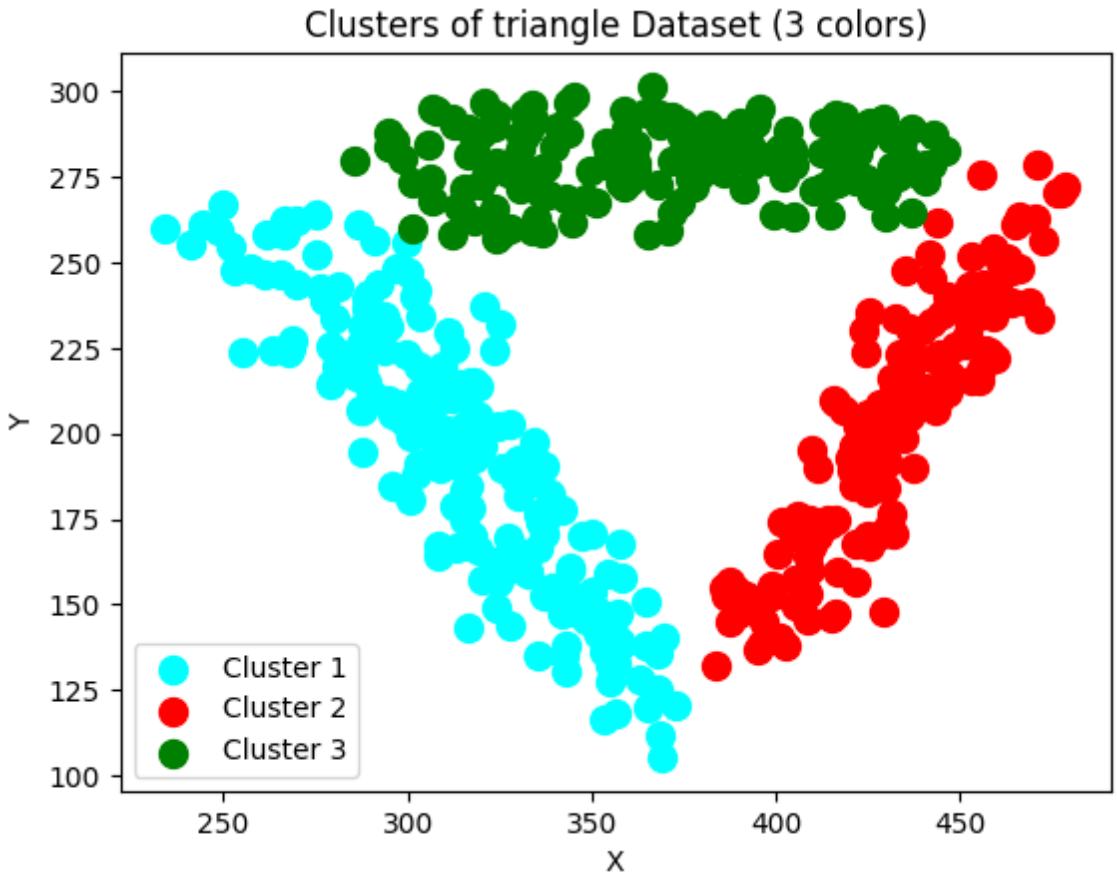
Out[108]:

	x	y	color
0	431.064943	210.021279	1
1	330.548001	287.098099	2
2	402.580348	137.726709	1
3	234.163225	259.808568	0
4	369.226618	105.374233	0

In [109]:
```
triangle['color'].unique()
```

Out[109]:
```
array([1, 2, 0])
```

In [164]:
```
visualize_clusters(triangle, 3, ['cyan','red','green'], 'triangle')
```



Dataset : un

In [110]:
```
un = pd.read_csv('/kaggle/input/clustering-exercises/un.csv')
```

In [111]:
```
un.head()
```

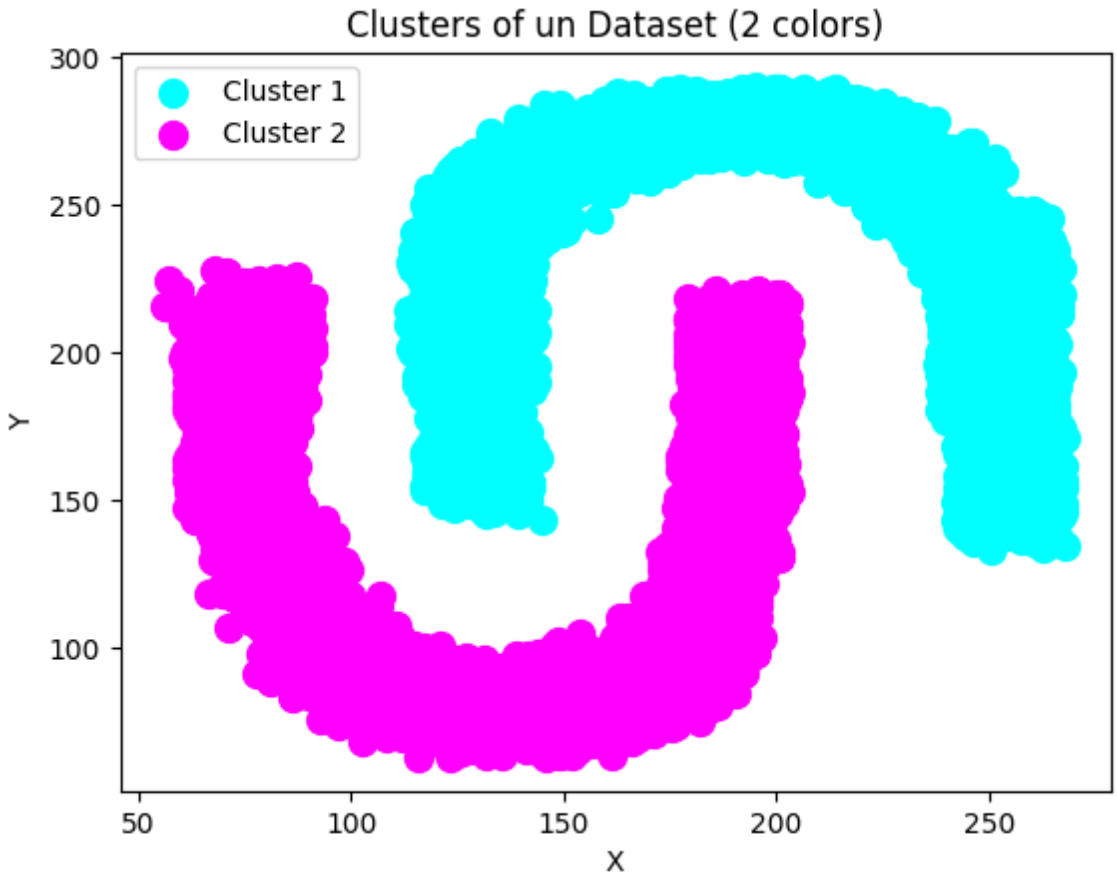
Out[111]:

	x	y	color
0	228.857880	272.467571	0
1	263.739784	144.066234	0
2	189.571408	137.603932	1
3	189.529792	276.010291	0
4	123.538840	208.087264	0

In [112]:
```
un['color'].unique()
```

Out[112]:
```
array([0, 1])
```

In [165]:
```
visualize_clusters(un, 2, ['cyan','magenta'], 'un')
```



Dataset : un2

In [113]:
```
un2 = pd.read_csv('/kaggle/input/clustering-exercises/un2.csv')
```

In [114]:
```
un2.head()
```

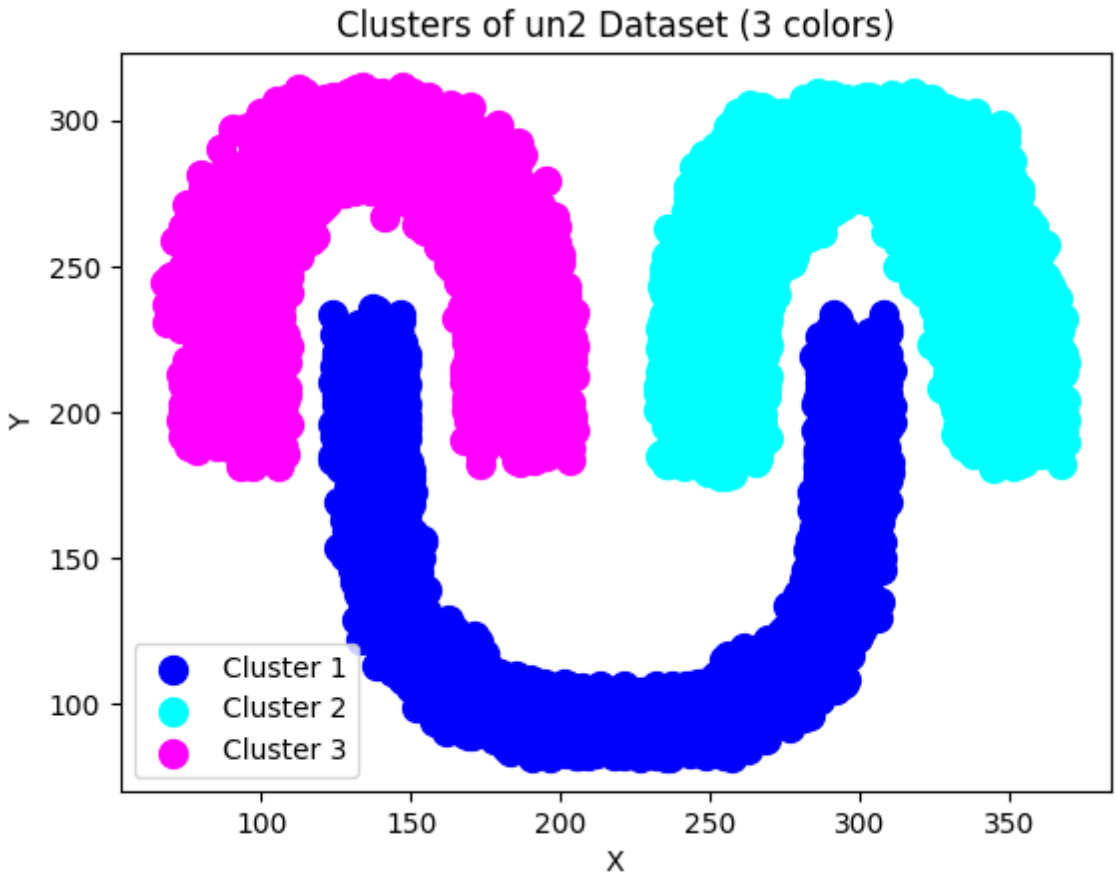
Out[114]:

	x	y	color
0	181.077591	212.126987	2
1	273.082908	100.085357	0
2	309.646417	265.791051	1
3	101.494244	268.526639	2
4	199.982098	225.778452	2

In [115]:
```
un2['color'].unique()
```

Out[115]:
```
array([2, 0, 1])
```

In [166]:
```
visualize_clusters(un2, 3, ['blue','cyan','magenta'], 'un2')
```



Dataset : wave

In [116]:
```
wave = pd.read_csv('/kaggle/input/clustering-exercises/wave.csv')
```

In [117]:
```
wave.head()
```

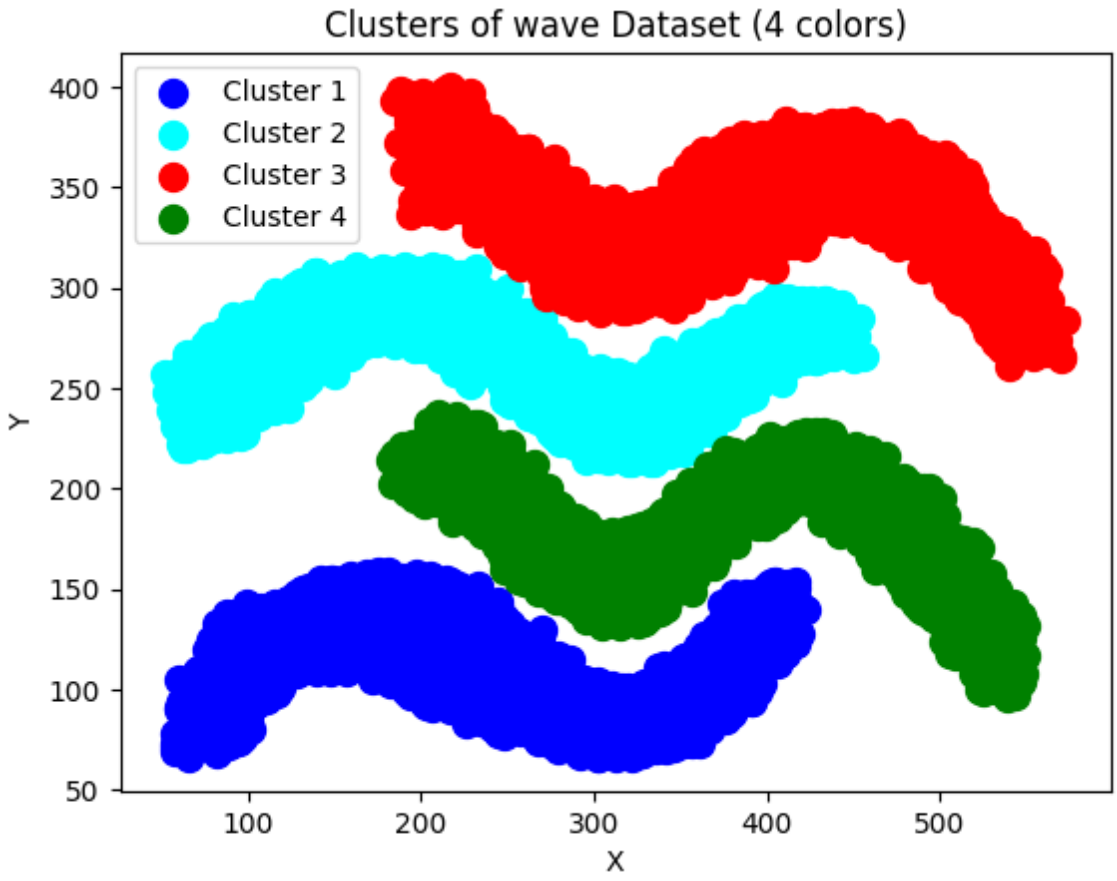
Out[117]:

	x	y	color
0	228.396722	90.875072	0
1	319.395295	81.619745	0
2	162.945088	125.089368	0
3	233.210908	122.902300	0
4	419.122186	334.096431	2

In [118]:
```
wave['color'].unique()
```

Out[118]:
```
array([0, 2, 3, 1])
```

In [167]:
```
visualize_clusters(wave, 4, ['blue','cyan','red','green'], 'wave')
```



In [ ]:
```

```