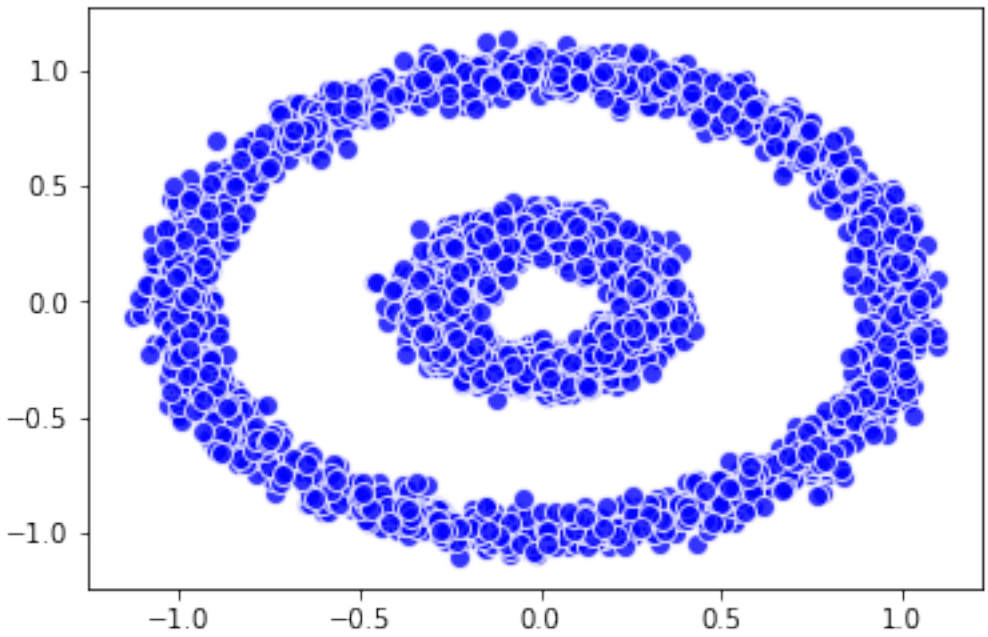
K_means_and_dbscan

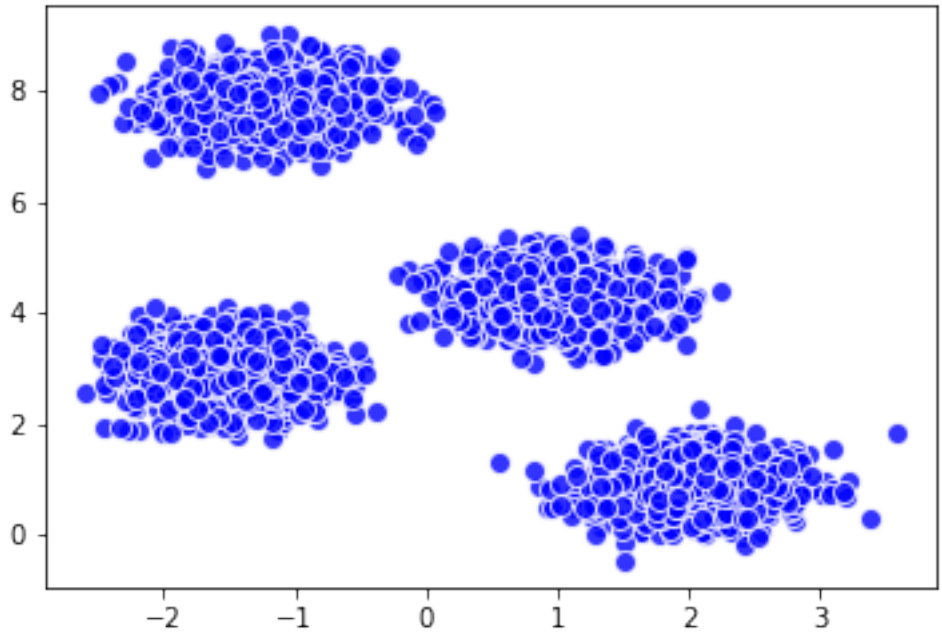
March 31, 2019

In [8]:
```
%matplotlib inline
import numpy as np
import matplotlib.pyplot as plt
from sklearn import datasets
N_samples = 4000
dataset_1 = np.array(datasets.make_circles(n_samples=N_samples, noise=0.05,factor=0.3)
dataset_2 = np.array(datasets.make_blobs(n_samples=N_samples,centers=4, cluster_std=0.
```

In [9]:
```
plt.scatter(dataset_1[:,0],dataset_1[:,1],alpha=0.8,s=64,edgecolors='white',c='blue')
plt.show()
```

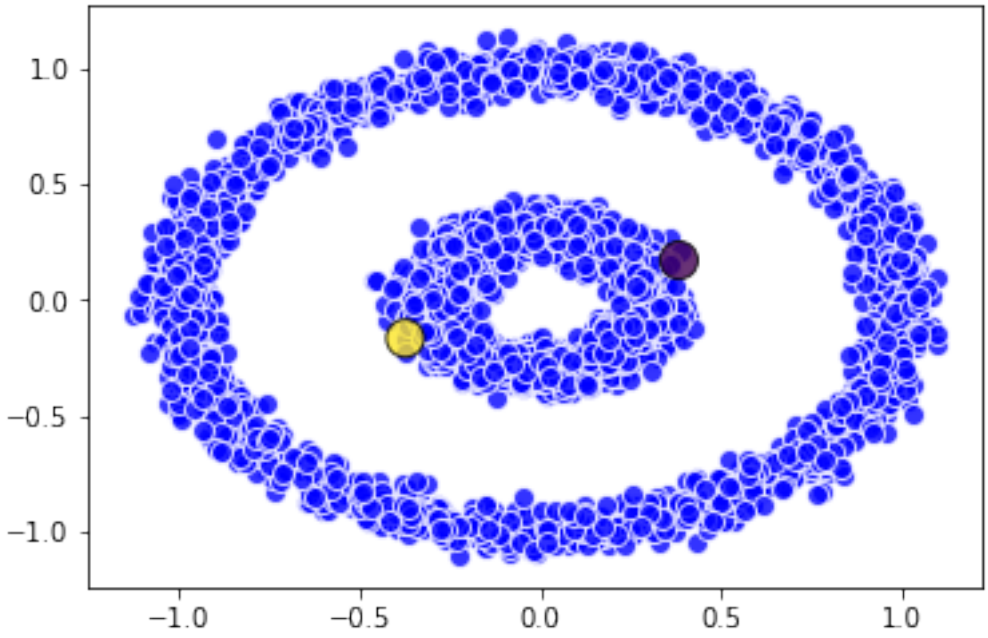


In [10]:
```
plt.scatter(dataset_2[:,0],dataset_2[:,1],alpha=0.8,s=64,edgecolors='white',c='blue')
plt.show()
```



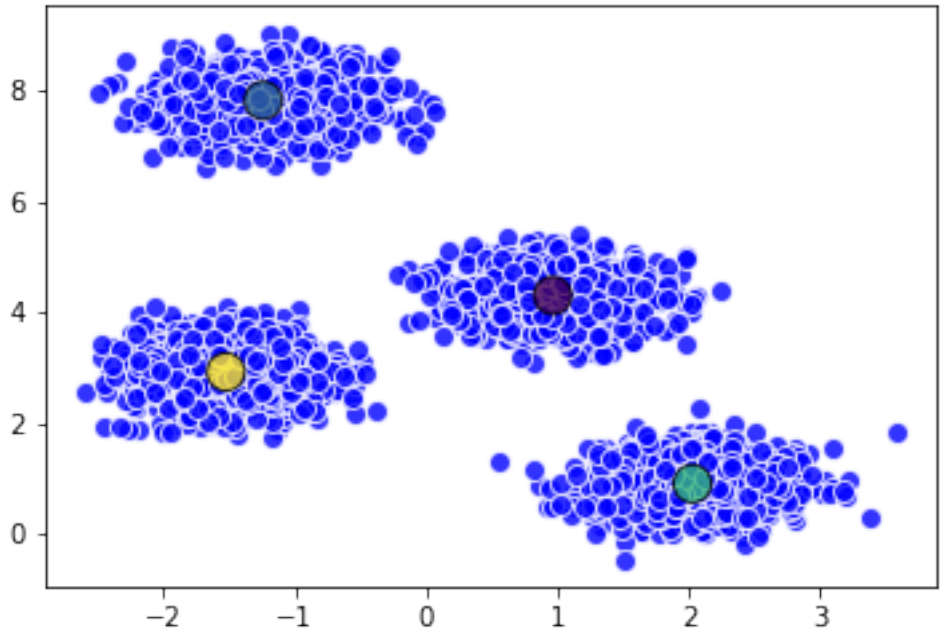

```
In [12]: from sklearn.cluster import KMeans
         k_dataset_1 = 2
         km_1 = KMeans(n_clusters=k_dataset_1)
         labels_1 = km_1.fit(dataset_1).labels_
```

```
In [15]: plt.scatter(dataset_1[:,0],dataset_1[:,1],alpha=0.8,s=64,edgecolors='white',c='blue')
         plt.scatter(km_1.cluster_centers_[:,0],km_1.cluster_centers_[:,1],alpha=0.8,s=200,edg
         plt.show()# Incorrect clusterization
```



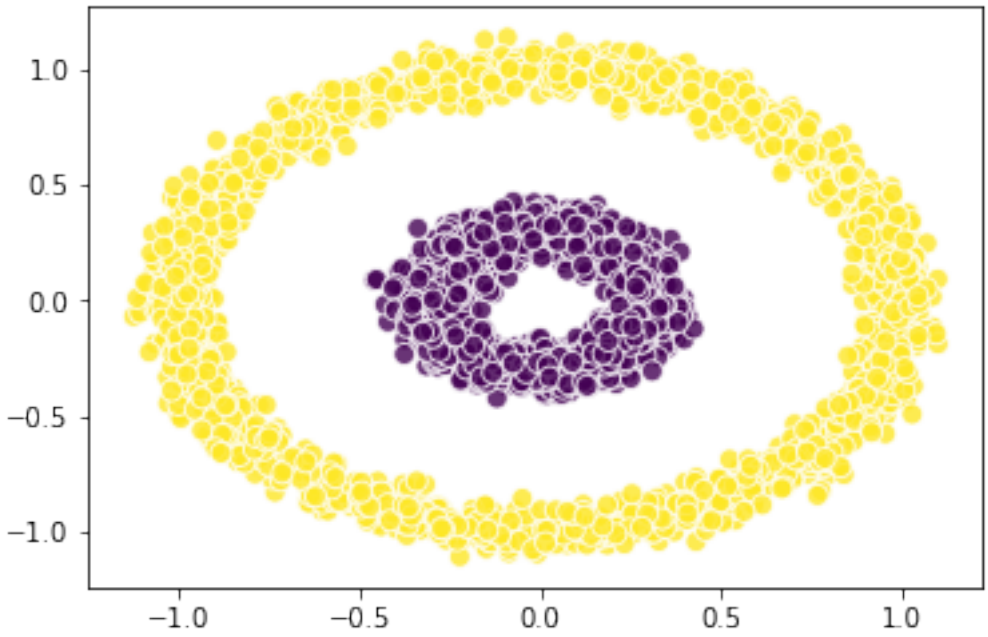
In [22]:
```
k_dataset_1 = 4
km_2 = KMeans(n_clusters=k_dataset_1)
labels_2 = km_2.fit(dataset_2).labels_
```

In [23]:
```
plt.scatter(dataset_2[:,0],dataset_2[:,1],alpha=0.8,s=64,edgecolors='white',c='blue')
plt.scatter(km_2.cluster_centers_[:,0],km_2.cluster_centers_[:,1],alpha=0.8,s=200,edg
plt.show()# Correct clusterization
```



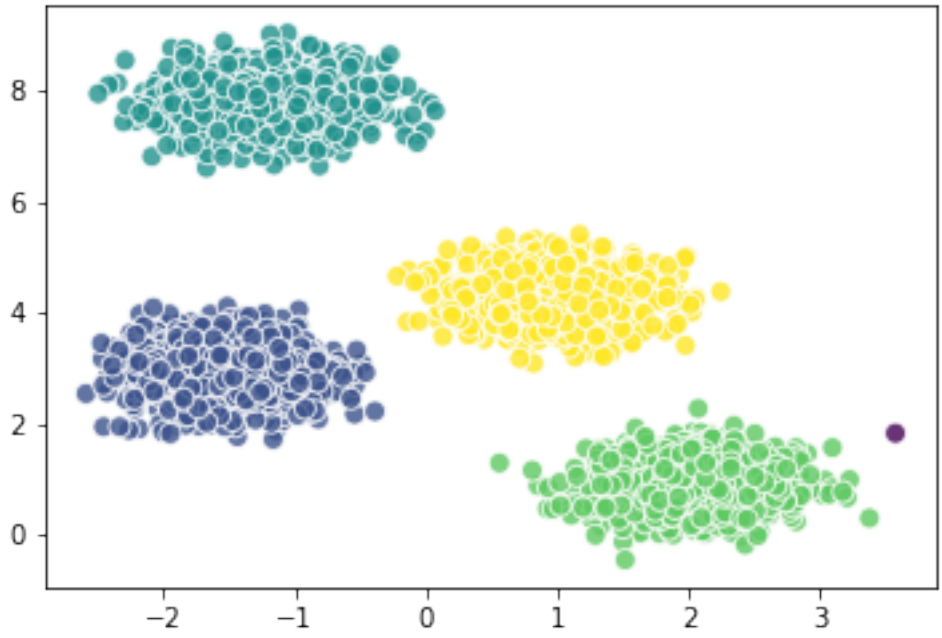
In [24]:
```
from sklearn.cluster import DBSCAN
dbs_1 = DBSCAN(eps=0.25)
labels_1 = dbs_1.fit(dataset_1).labels_
```

In [27]:
```
plt.scatter(dataset_1[:,0],dataset_1[:,1],alpha=0.8,s=64,edgecolors='white',c=labels_
plt.show()# Correct clusterization
```



In [30]:
```
dbs_2 = DBSCAN(eps=0.45)
labels_2 = dbs_2.fit(dataset_2).labels_
```

In [31]:
```
plt.scatter(dataset_2[:,0],dataset_2[:,1],alpha=0.8,s=64,edgecolors='white',c=labels_
plt.show()
```



In [ ]: