ML Assignment 2

Task 1 Link : M23MAC004_task1.ipynb

Subtask B

Sample image visuals from invivisual clusters from a total of 10 clusters

Total 10 Clusters

samples from cluster 1 out of total 7059 datapoint

samples from cluster 2 out of total 5960 datapoint

samples from cluster 3 out of total 5266 datapoint

samples from cluster 4 out of total 8021 datapoint

samples from cluster 5 out of total 4286 datapoint

samples from cluster 6 out of total 5499 datapoint

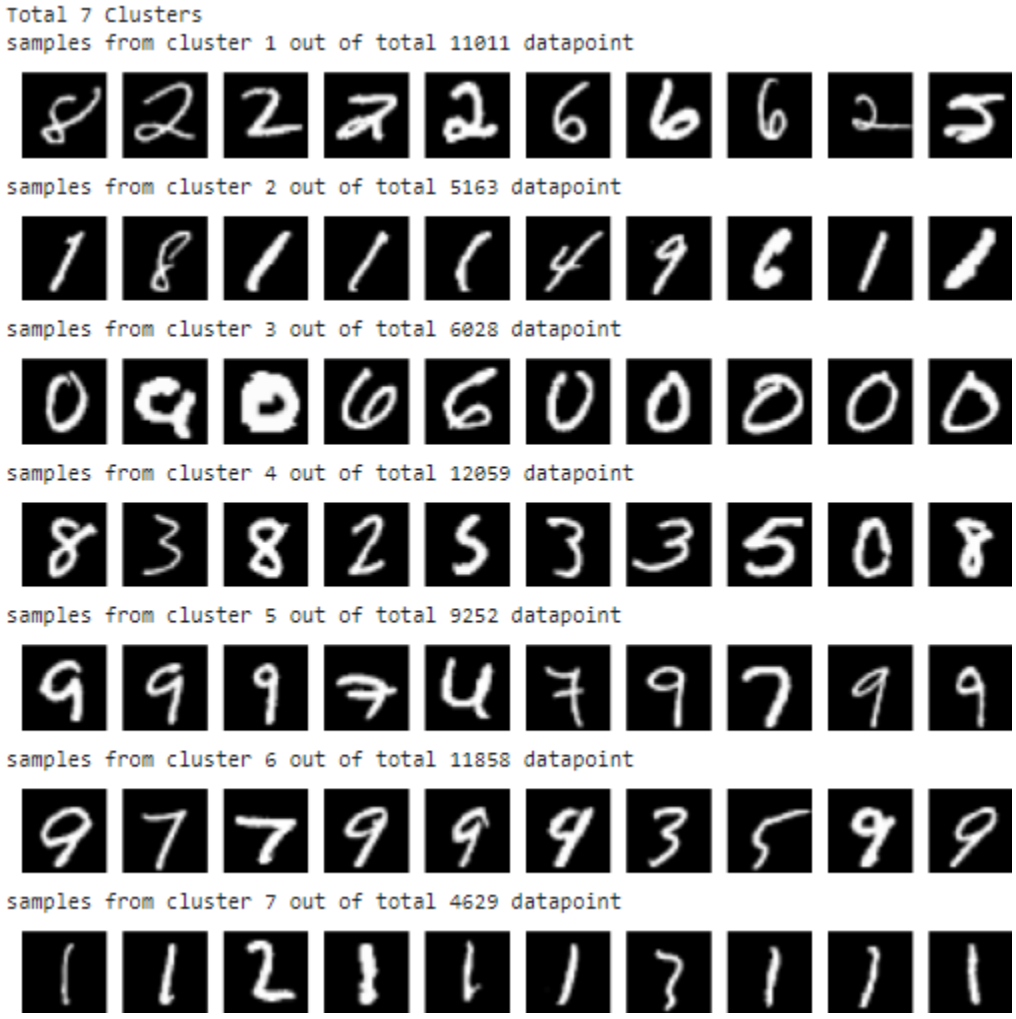
samples from cluster 7 out of total 6509 datapoint

samples from cluster 8 out of total 5151 datapoint

samples from cluster 9 out of total 8275 datapoint

samples from cluster 10 out of total 3974 datapoint

Sample image visuals from invivisual clusters from a total of 7 clusters

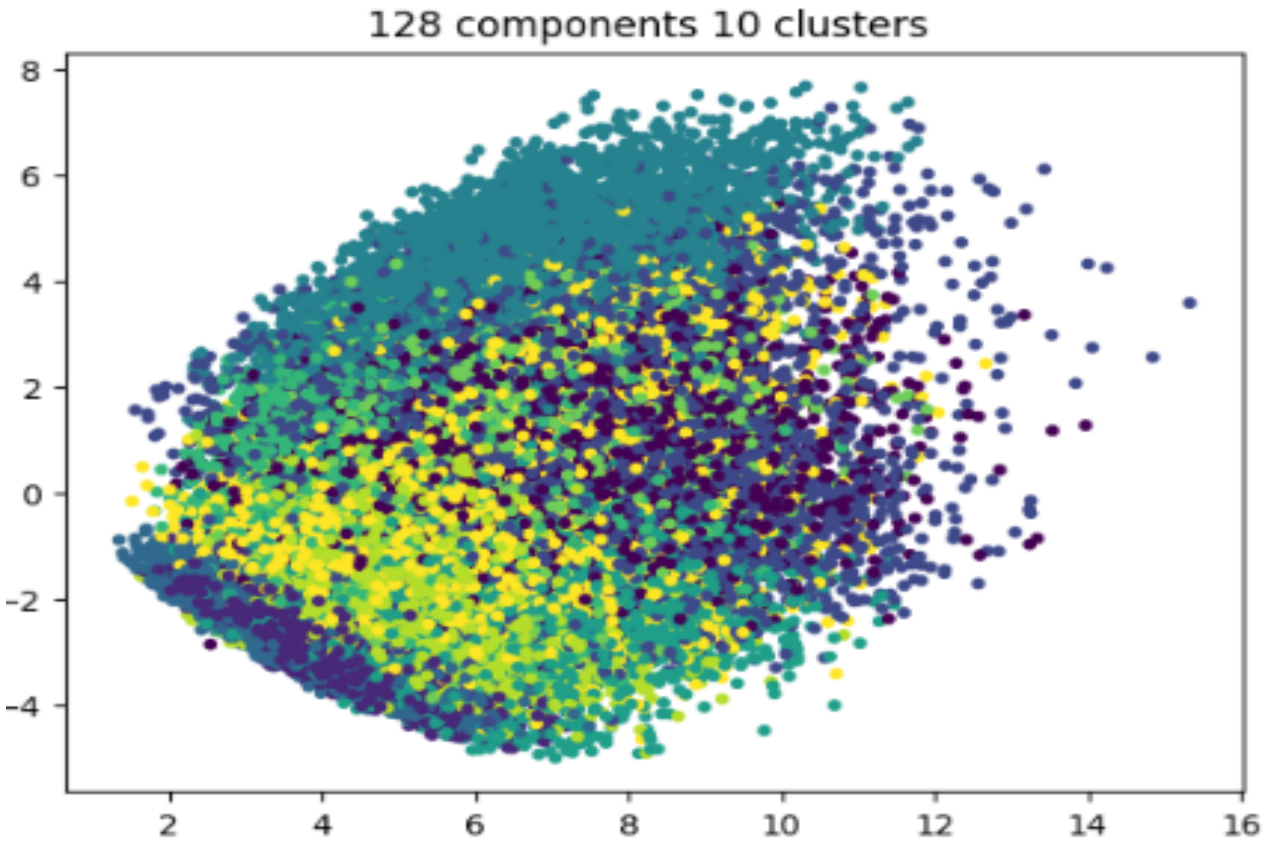


Sample image visuals from invivisual clusters from a total of 4 clusters

Subtask C
For 10 clusters, image visuals are displayed in the correct manner as compared to 7 and 4 clusters. In 10 clusters, it is more sensitive to data for some significant digits like 0, 1, 2, and 6. And it created an almost separate cluster for there digits.
While proceeding for 7 clusters, it has clustered digit 0,1 only correctly and produced a mixture for the rest.
In 4 clusters, it hasn't done clustering for any digit correctly.
From here, it can be concluded that moving toward less number of clusters than optimal won't produce a better result.

Task 2 Link : M23MAC004_task2.ipynb

Subtask B



Total 10 Clusters
samples from cluster 1 out of total 9214 datapoint

samples from cluster 2 out of total 2517 datapoint

samples from cluster 3 out of total 10253 datapoint

samples from cluster 4 out of total 3160 datapoint

samples from cluster 5 out of total 3798 datapoint

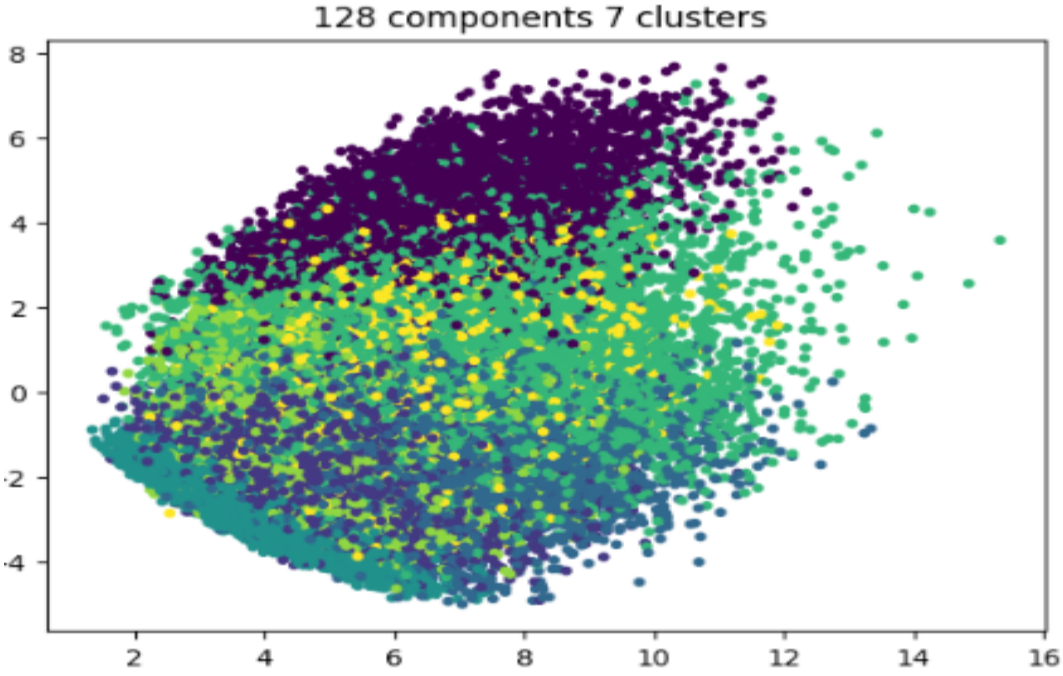
samples from cluster 6 out of total 6229 datapoint

samples from cluster 7 out of total 6207 datapoint

samples from cluster 8 out of total 4770 datapoint

samples from cluster 9 out of total 6548 datapoint

samples from cluster 10 out of total 7304 datapoint




```
Total 7 Clusters
samples from cluster 1 out of total 4638 datapoint
```

```
samples from cluster 2 out of total 7264 datapoint
```

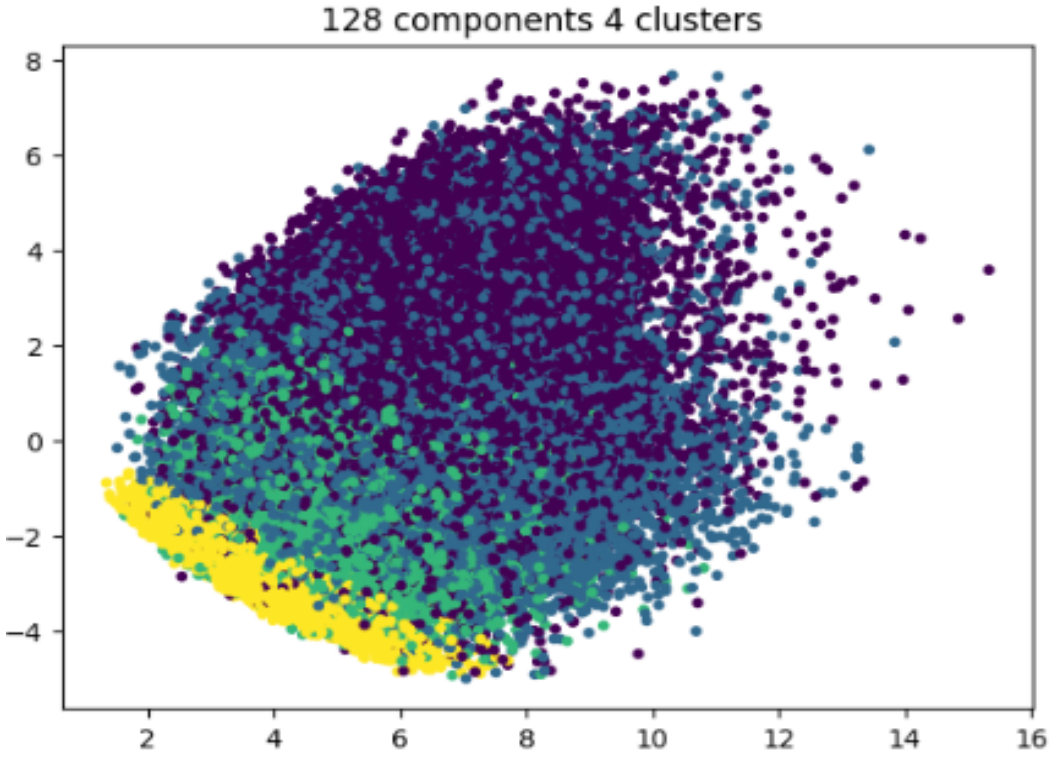
```
samples from cluster 3 out of total 10914 datapoint
```

```
samples from cluster 4 out of total 5223 datapoint
```

```
samples from cluster 5 out of total 18972 datapoint
```

```
samples from cluster 6 out of total 8173 datapoint
```

```
samples from cluster 7 out of total 4816 datapoint
```

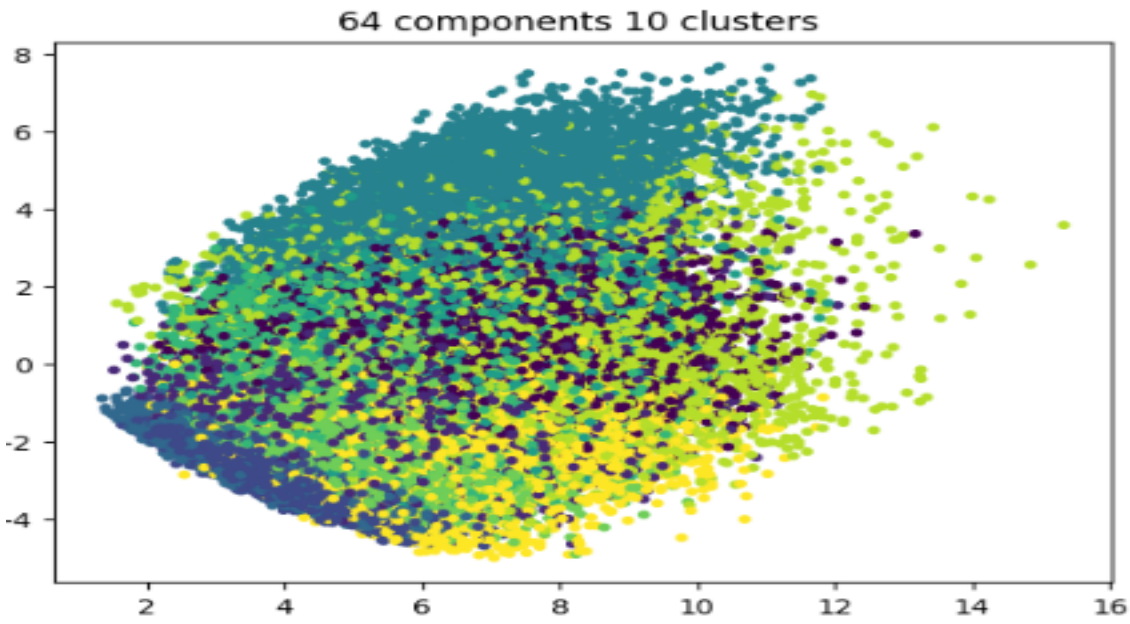



```
Total 4 Clusters
samples from cluster 1 out of total 23536 datapoint
```

```
samples from cluster 2 out of total 19778 datapoint
```

```
samples from cluster 3 out of total 11069 datapoint
```

```
samples from cluster 4 out of total 5617 datapoint
```



Total 10 Clusters

samples from cluster 1 out of total 8152 datapoint

samples from cluster 2 out of total 6140 datapoint

samples from cluster 3 out of total 2693 datapoint

samples from cluster 4 out of total 3009 datapoint

samples from cluster 5 out of total 4624 datapoint

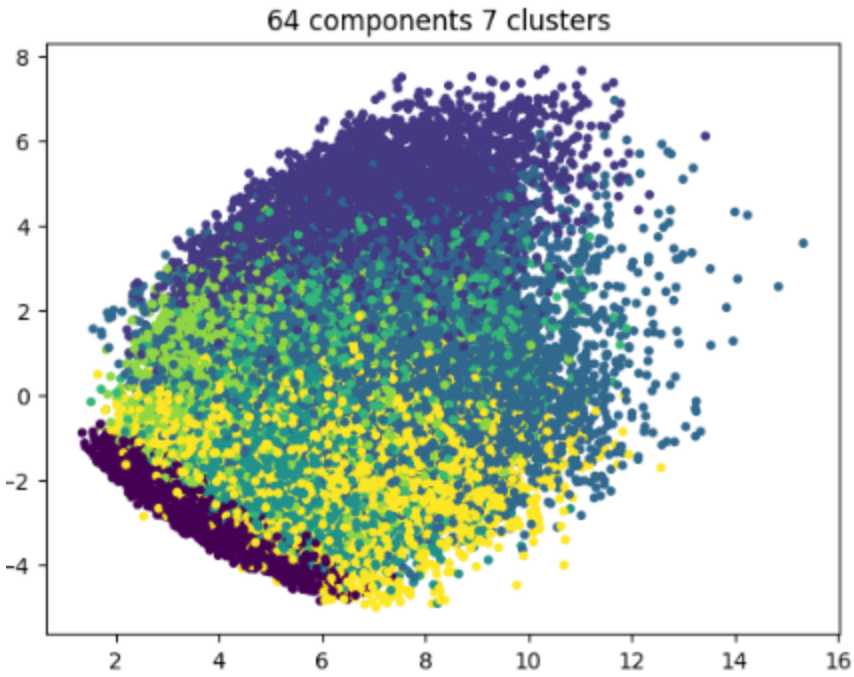
samples from cluster 6 out of total 4989 datapoint

samples from cluster 7 out of total 7139 datapoint

samples from cluster 8 out of total 6119 datapoint

samples from cluster 9 out of total 11536 datapoint

samples from cluster 10 out of total 5599 datapoint




```
Total 7 Clusters
samples from cluster 1 out of total 5256 datapoint
```

```
samples from cluster 2 out of total 5112 datapoint
```

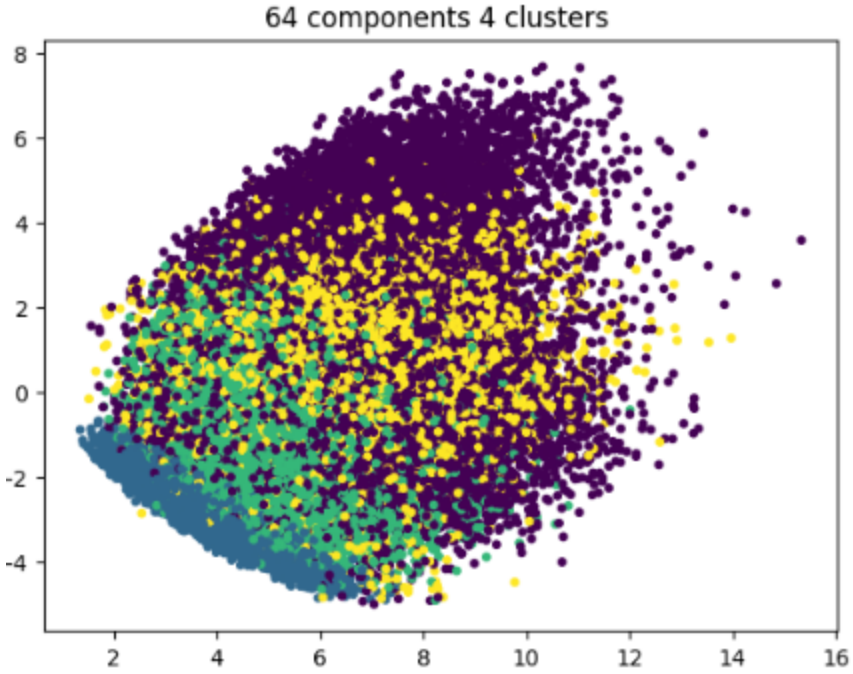
```
samples from cluster 3 out of total 18742 datapoint
```

```
samples from cluster 4 out of total 7595 datapoint
```

```
samples from cluster 5 out of total 5210 datapoint
```

```
samples from cluster 6 out of total 8226 datapoint
```

```
samples from cluster 7 out of total 9859 datapoint
```

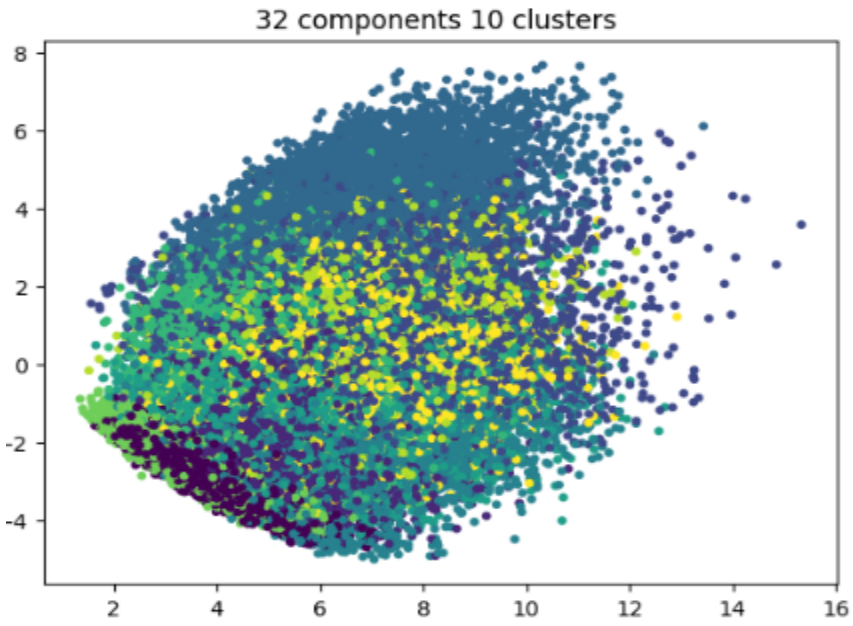



```
Total 4 Clusters
samples from cluster 1 out of total 23814 datapoint
```

```
samples from cluster 2 out of total 5876 datapoint
```

```
samples from cluster 3 out of total 15308 datapoint
```

```
samples from cluster 4 out of total 15002 datapoint
```



Total 10 Clusters

samples from cluster 1 out of total 2909 datapoint

samples from cluster 2 out of total 6218 datapoint

samples from cluster 3 out of total 10066 datapoint

samples from cluster 4 out of total 5091 datapoint

samples from cluster 5 out of total 6413 datapoint

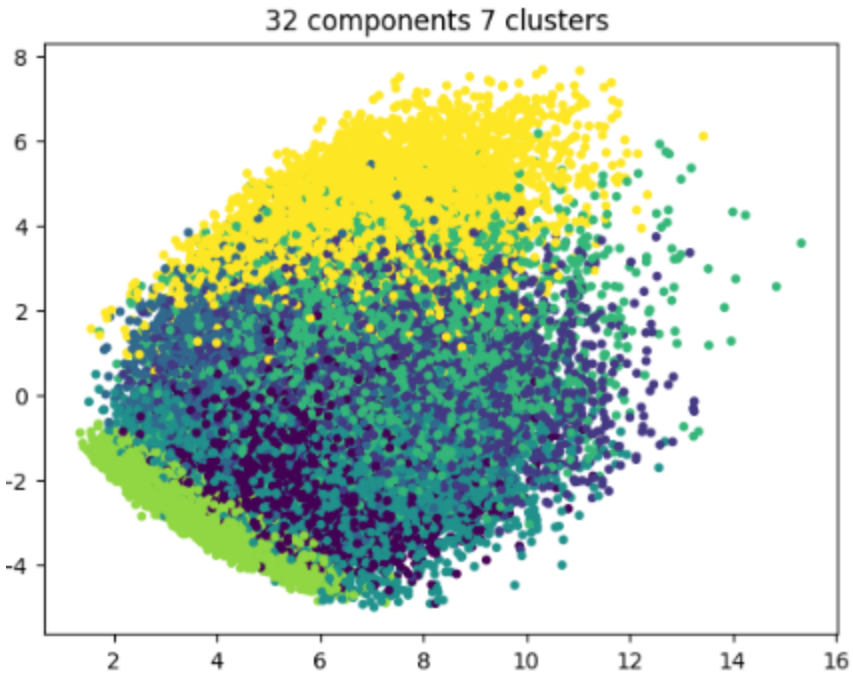
samples from cluster 6 out of total 6757 datapoint

samples from cluster 7 out of total 8261 datapoint

samples from cluster 8 out of total 3017 datapoint

samples from cluster 9 out of total 5376 datapoint

samples from cluster 10 out of total 5892 datapoint



Total 7 Clusters

samples from cluster 1 out of total 8513 datapoint

samples from cluster 2 out of total 10523 datapoint

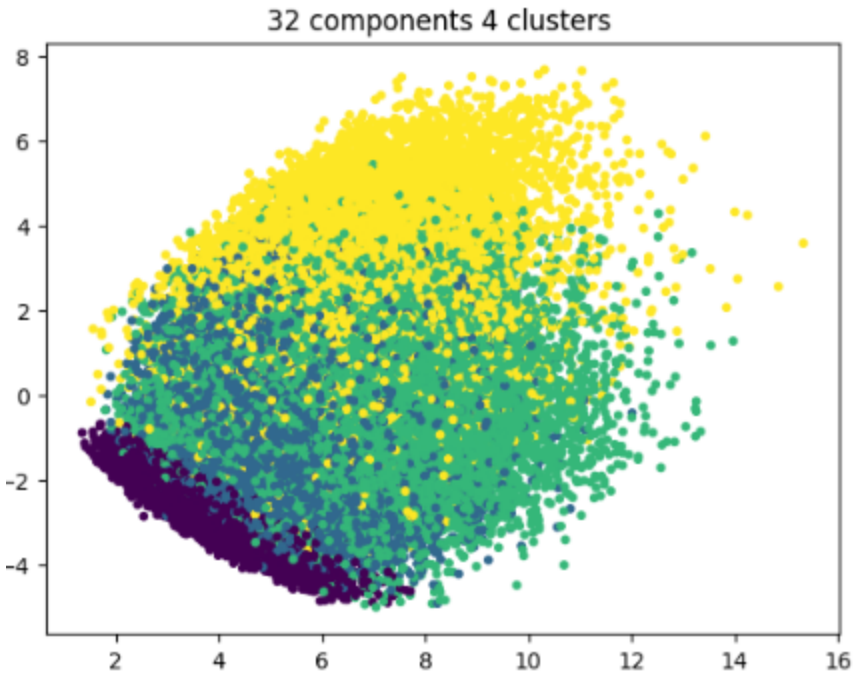
samples from cluster 3 out of total 8522 datapoint

samples from cluster 4 out of total 8840 datapoint

samples from cluster 5 out of total 12689 datapoint

samples from cluster 6 out of total 5630 datapoint

samples from cluster 7 out of total 5283 datapoint




```
Total 4 Clusters
samples from cluster 1 out of total 6037 datapoint
```

```
samples from cluster 2 out of total 16346 datapoint
```

```
samples from cluster 3 out of total 26105 datapoint
```

```
samples from cluster 4 out of total 11512 datapoint
```

```
CPU times: user 26min 50s, sys: 10min 19s, total: 37min 10s
Wall time: 21min 47s
```

Subtask c

For 10 clusters

On comparing with the previous task without clustering, the cluster created with pca is more perfect because it successfully clustered for each digit in groups than the task 1. Cluster with 64 components, which created a better result.

For 7 cluster
128 and 64 components clusters created a better result as compared to the previous task. But with 32 components, it’s not better than the previous one.

For 4 cluster
Clusters with 128, 64, and 32 components almost produce similar results. But better than task 1.

According to the generated sample
The result with 64 components and 10 cluster values is more appropriate than others.