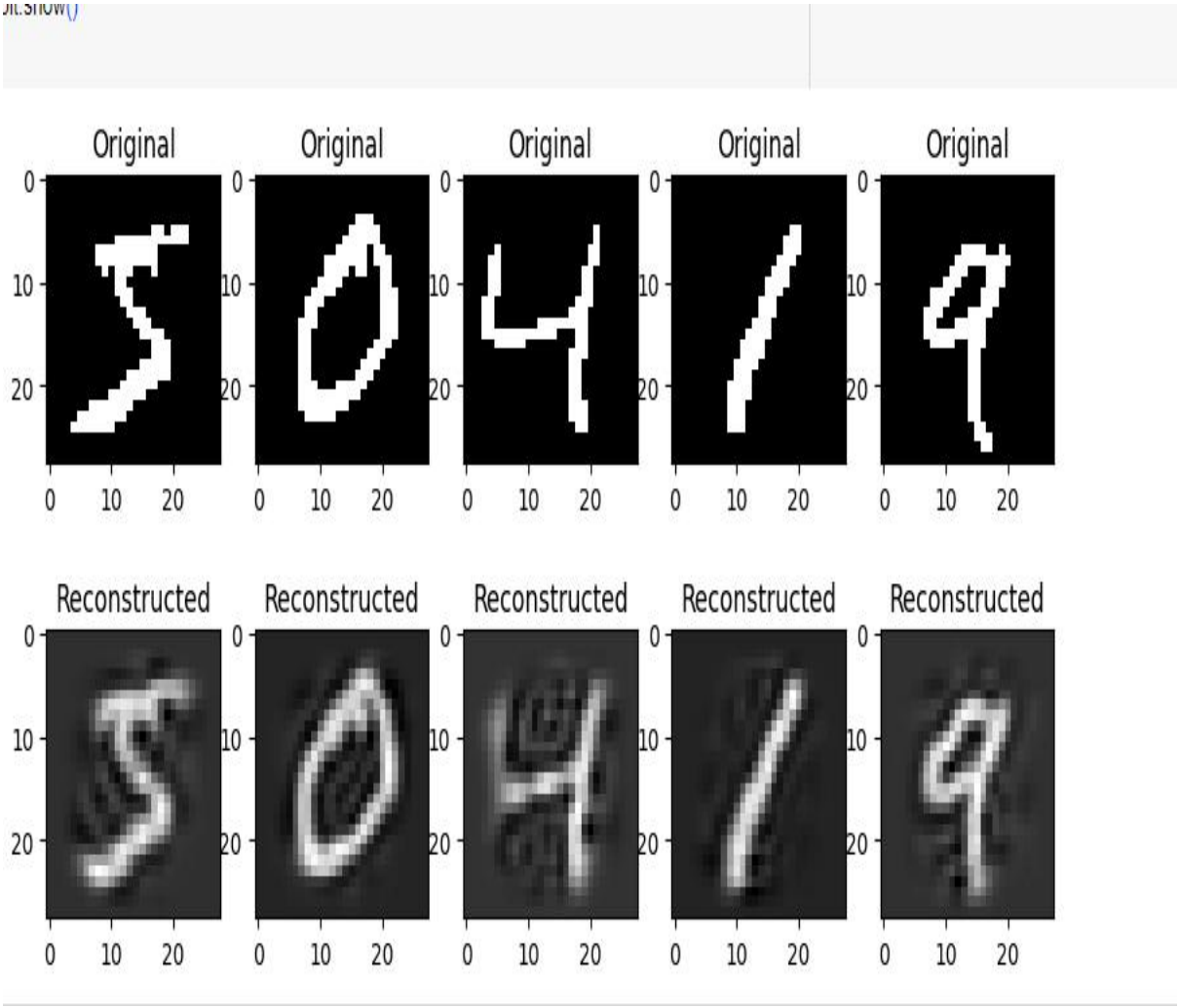
Group 11

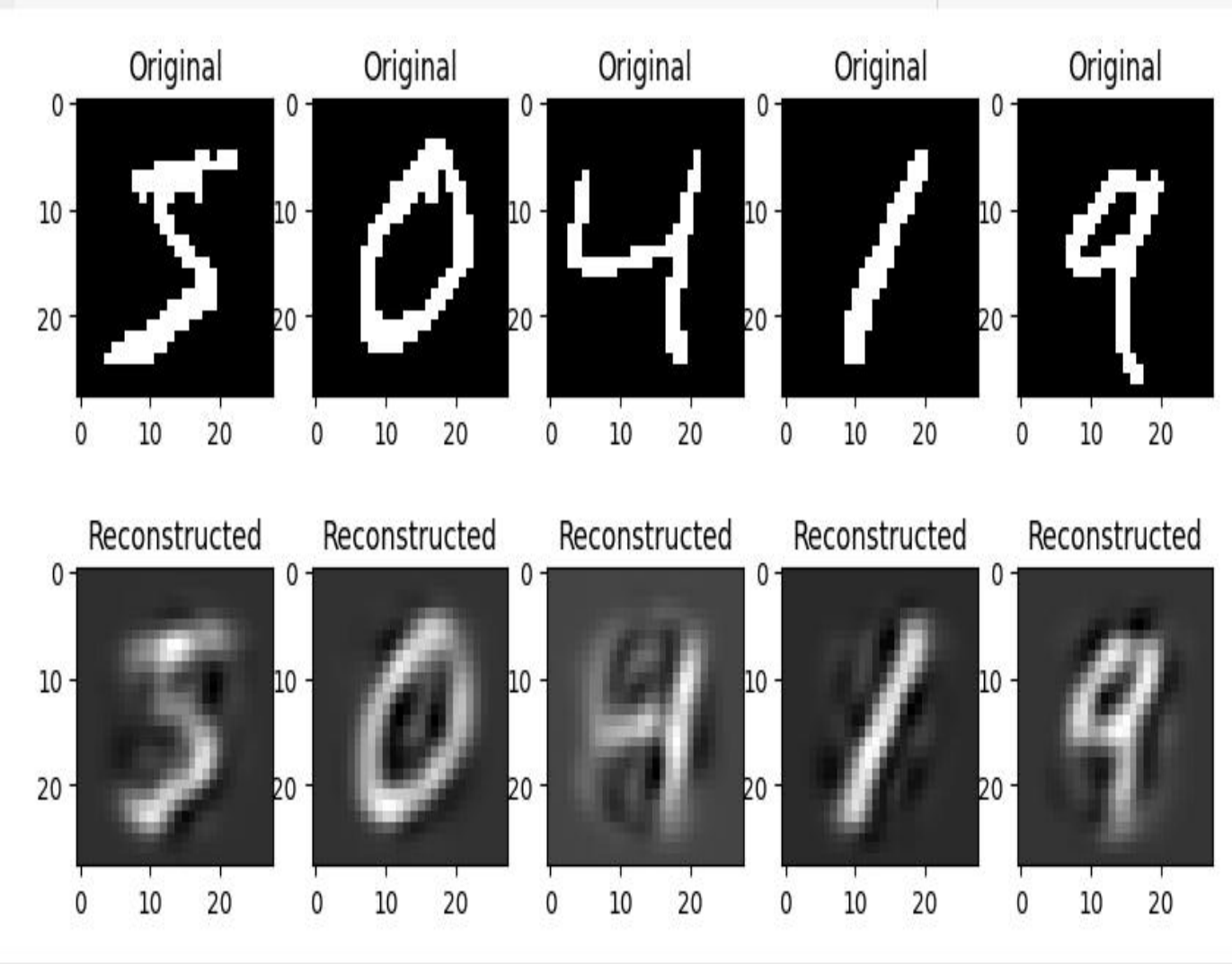
Name	ID
Alaa Amer Mohammed	20200090
Kareem Waleed	20200397
Marwa Shaaban	20200516
Mohammed Nasser	20200480
Sama Hussien Abo-Elala	20200232

First trail -->applying PCA with Q-matrix first 100 components



plot1

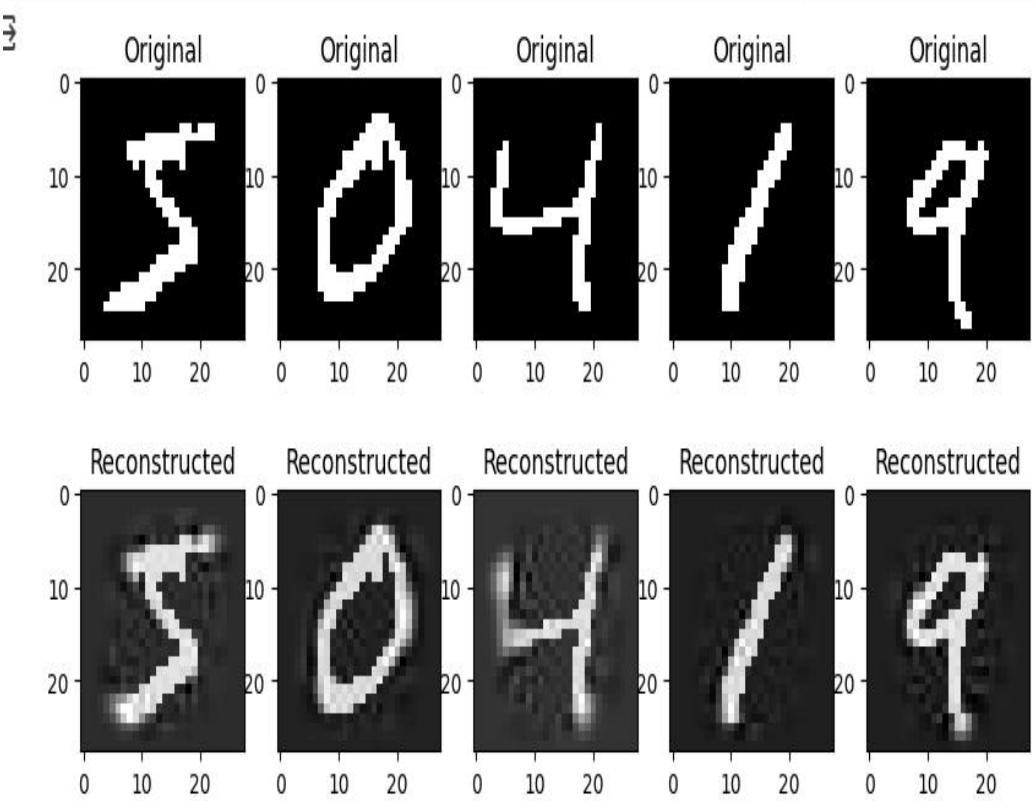
Second trail --> applying PCA with Q-matrix second 100 components



Plot2

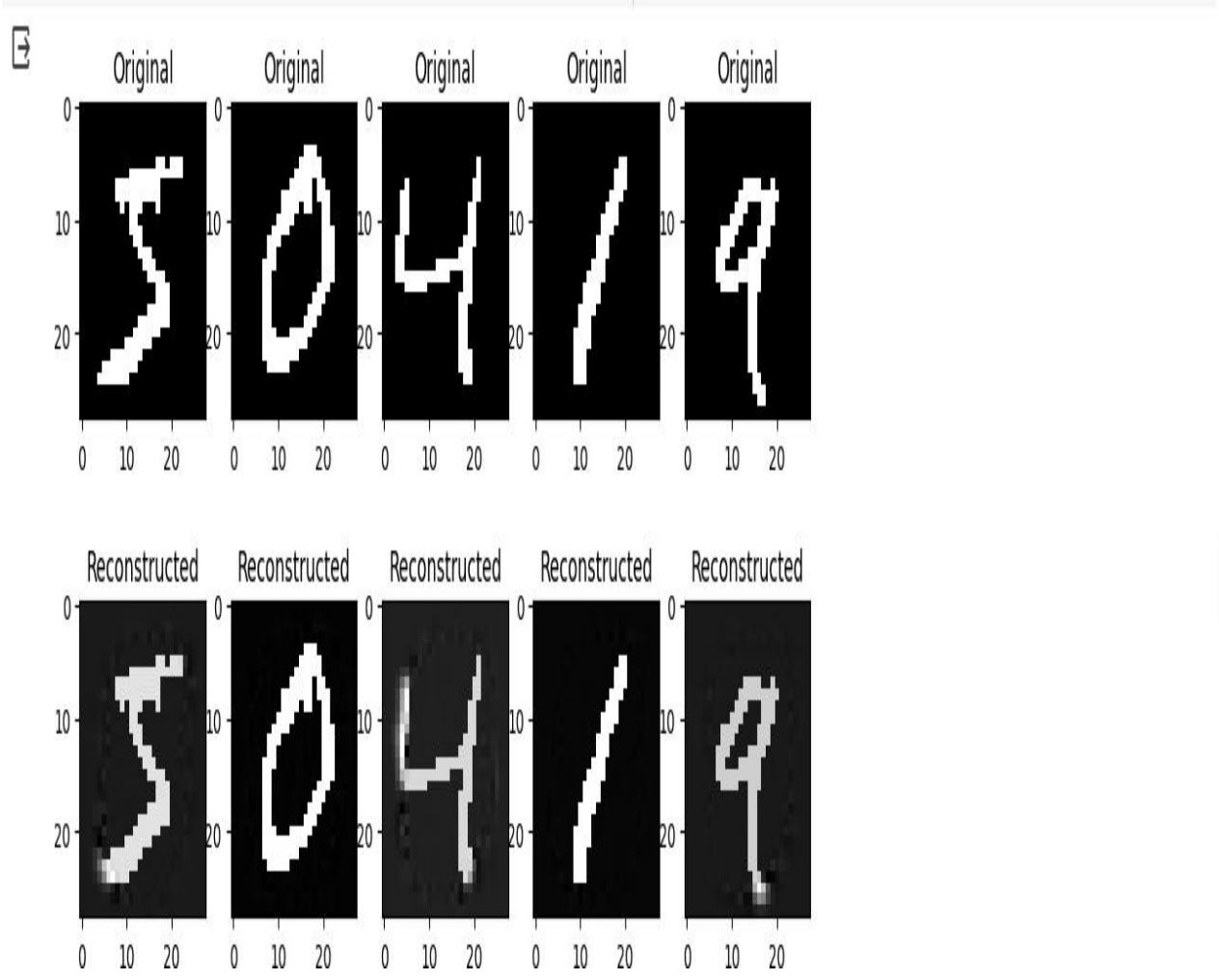
Third trail --> applying PCA with Q-matrix first 250 components

```
plt.show()
```



Plot3

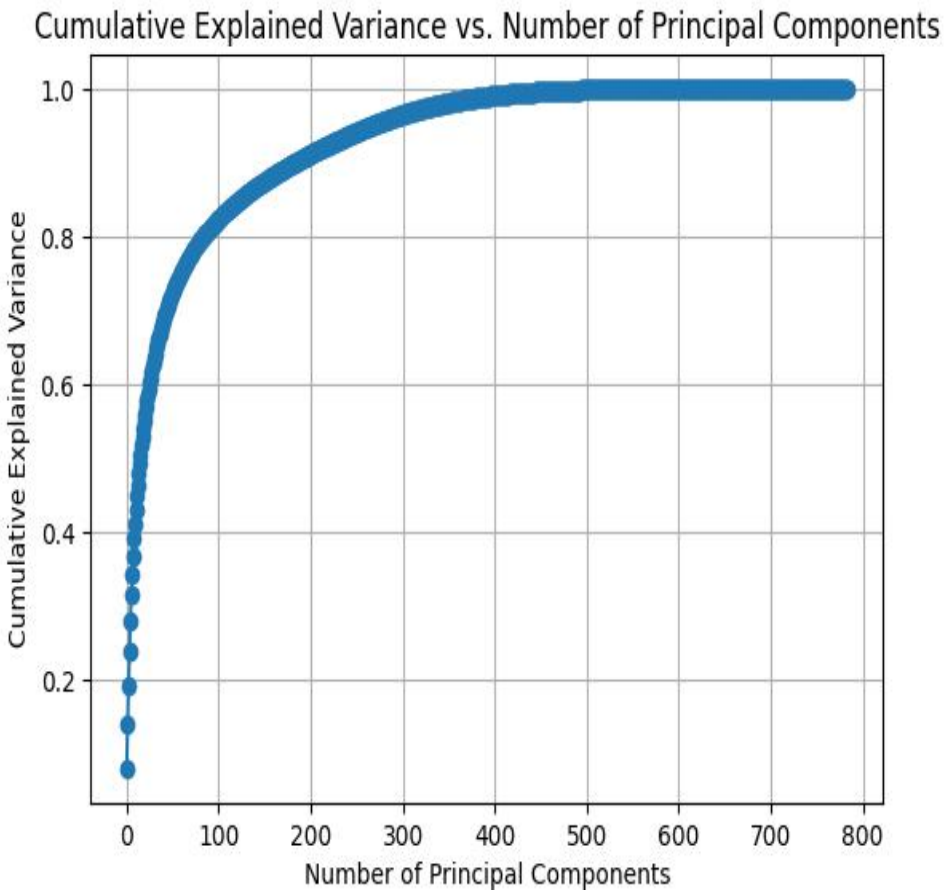
Fourth trail --> applying PCA with Q-matrix first 400 components



Plot4

This plot is the cumulative variance with number of principal components which shows that after 400 components there's no change in the cumulative variance. This means that the first 400 components preserve most of the information from the data features. So, if we used 400 as our Q-matrix we'll get the best result.

```
plt.show()
```



Plot

From previous plots (4,5) we can see that 400 Q-matrix is the best representation of the data.

The result we got from applying the PCA is great as it represents our data which has 784 features with only 400 features and still preserves most of our data information not causing much loss.

And from plot (1,2) we could see that the first 100 components in Q-matrix preserve more information than the second 100 as we sort them descending.

Phase 2:

Our choice in phase 0 was Hamming Network. First, we implemented the K-means clustering in order to get the centroids and labels. Then, we implemented the hamming network. Upon evaluating, we got a higher accuracy in PCA version than the original version.