Assignment 1 Part 1c)

Images per class in the sample of 12000 images with seed=5:

```
#Summary of number of images per class
from collections import Counter
Counter(y_test[:,0])
```

```
Counter({0: 1207,
         1: 1177,
         2: 1184,
         3: 1158,
         4: 1223,
         5: 1202,
         6: 1206,
         7: 1219,
         8: 1180,
         9: 1244})
```

Performance of the best model created on the new data set:78%

Accuracy of the model trained using 12000 images: 80%

The accuracy of the model is higher in the second case, because of the randomness in the sampling. This might result in overfitting as we are selecting random numbers that optimize for a small sample of data. The quality of such models are dependent on the quality of the dataset that has been held out. In the production, we need to select a model that has been trained on the entire training dataset, once we decide an algorithm and configuration.