SUMMARY POINTS:

1) The final test accuracy on using LeNet on the CIFAR10 dataset without any augmentations is 0.6889976038338658 ~ 0.69.

2) The final test accuracy on using LeNet on the CIFAR10 dataset with transforms (data augmentation) is 0.577276357827476 ~ 0.58.

3) The final test accuracy on using LeNet on the MNIST dataset without any augmentations is 0.9746405750798722 ~ 0.97.

4) The final test accuracy on using LeNet on the MNIST dataset with transforms (data augmentation) is 0.9568690095846646 ~ 0.96.

- Based on the above results, it can be seen that data augmentation was not useful for the CIFAR10 dataset and surprisingly, augmentation performed slightly worse than no augmentation on the MNIST dataset.

- LeNet worked fantastic on MNIST which was expected since the LeNet architecture was tailored to work on digit images (and it is also used in ATMs even now).

- CIFAR10 dataset, on the other hand, is more complicated than MNIST since the categories of images vary significantly. In MNIST, all images were of digits and were relatively simple.

- Results were surprising since augmentation enhances model's performance in most cases, but in this case neither dataset benefitted from it. This may most likely be due to the relatively simple LeNet model. Using a deep network like VGGNet, ResNet, or others would allow extraction of more features and hence better performance.

- The reason for the accuracy decrease from 1 → 2 may be because:
  a) CIFAR10 - There is huge variety of images and flipping, cropping, rotating or other augmentations might discard important information about images. Airplanes are relatively flat and maybe don't make much of a difference if such images are horizontally flipped or cropped, for example. On the other hand, augmenting dog or deer images, for example, might lessen the gap between them and confuse the model. This is what would have happened. Apart from this, it is difficult for the model to capture features even after augmenting due to the complexity of images.
  b) MNIST - A small decrease in accuracy results in going from 3 → 4. We may not blame augmentation here since the decrease in accuracy is small. Moreover, taking multiple test accuracies and averaging them may result in almost same accuracy without augmentation. The decrease in accuracy is smaller than the CIFAR10 case maybe because model was better generalized. But augmenting the dataset didn't prove massively beneficial.

Image Flipping – Horizontally flipping the images randomly reduced the test accuracy slightly on both the datasets. For CIFAR10, it went from 0.58 to 0.577 and for MNIST from 0.9738 to 0.9569. This decrease is very small for CIFAR and relatively larger for MNIST. Expectation was that on CIFAR,

flipping must reduce the accuracy more than on MNIST because of CIFAR having complex and varied images.

For situations where there was a slight difference in accuracies, it may be better to train and test multiple times and take the average to increase the confidence of our conclusions.