The problem statement for the Kaggle competition I chose was Histopathologic Cancer Detection to identify metastatic tissue in histopathologic scans of lymph node sections https://www.kaggle.com/c/histopathologic-cancer-detection

I started by analyzing the data and observed that there are 220025 pictures in train and 57458 pictures in test.

The train labels had a balanced label distribution
0 130908
1 89117

I then started preprocessing my data and building my model

1. Opened the image using openCV
2. Converted the image to rgb format as the default is in bgr
3. To optimize the data, I perform augmentation by rotation, flip, x-y coordinate flip, brightness and contrast change. I also perform cropping to get the data to the center
4. Normalize the data to a range between 0 to 1
5. I use three blocks of code to build my model - identity block, convolutional block and resnet block
6. I add 3 convolution layer components with batch optimization and a ‘relu’ activation filter for the first identity block and the second convolutional block each
7. I then add these two blocks as the first layers for my resnet block
8. I then average pool, perform a flatten and then use a dense layer for my output
9. I then compile my model and introduce early stopping and a checkpoint
10. With a batch size of 32 I run my model
11. I had an accuracy of 0.9671 and a loss of 0.3453
12. I then used the model to make predictions on the test data
13. I submitted my submission.csv file on kaggle successfully