I just begin with the first week’s report. In the first week, we tried LinearUCB on statlog dataset with different reward function, including linear, polynomial, relu, sigmoid, log and tanh. The outcome is that non-linear reward with Tanh function performs the best among all the rewards. Then, we further try to discover how should we choose the level of exploration versus exploitation by plotting the cumulative reward with different alpha and linear reward function in linearUCB. It turns out that linearUCB with alpha=1 or 10 should be the best set of parameters to maximize the cumulative reward compared to the optimal policy. And it also reaches the highest accuracy and F1-score. Then, Marco asked us to perform the different alpha with Tanh reward, and this will be in the report in the second week.

For the last part, we just performed the cumulative reward with different alpha. The outcome should be the same as in the linear case. We draw a conclusion that while low alpha emphasizes exploitation, using known information to maximize immediate rewards, high alpha emphasizes exploration, seeking new actions to potentially discover higher rewards. We still need to do exploration than just doing exploitation. That’s all about what we have done.