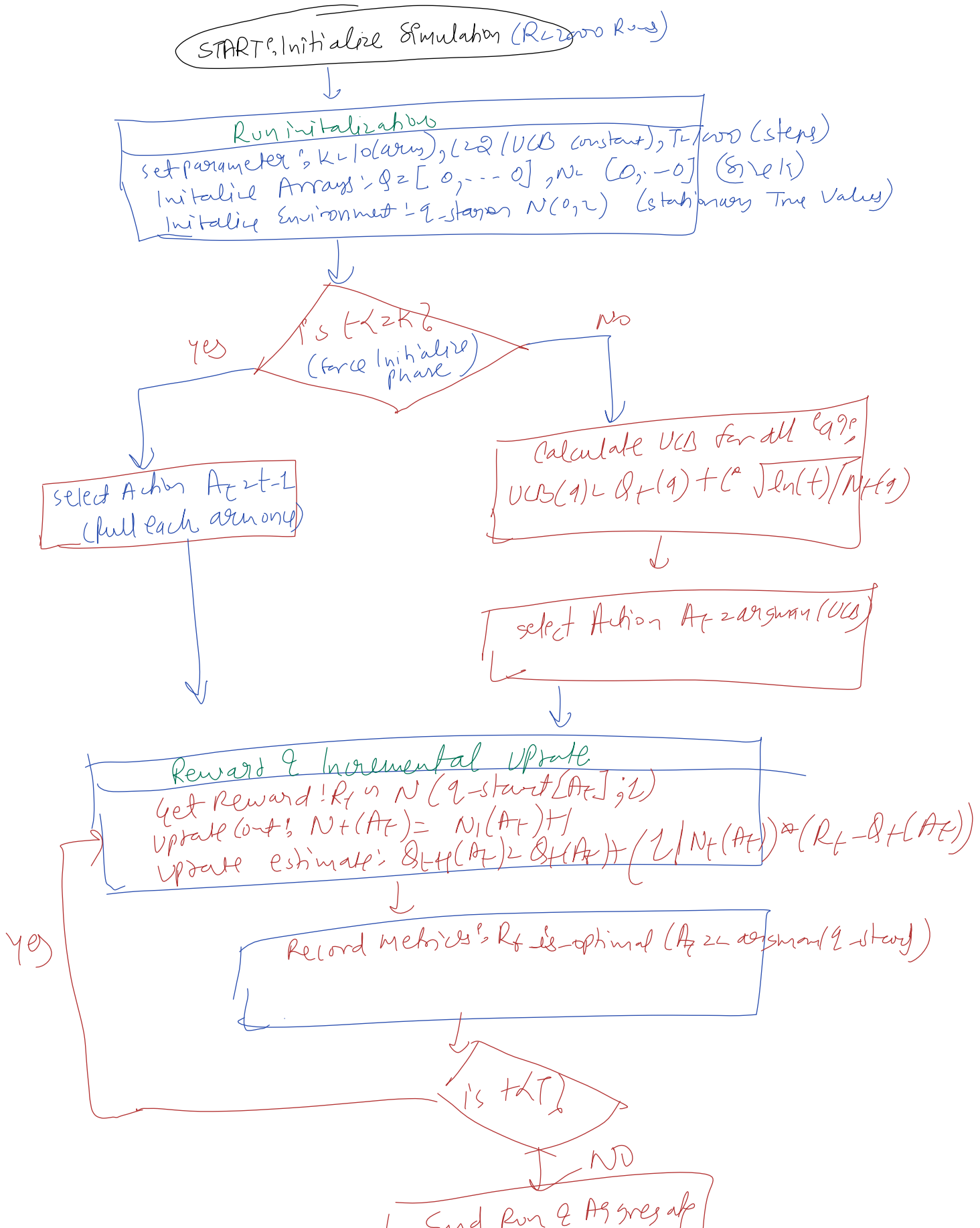
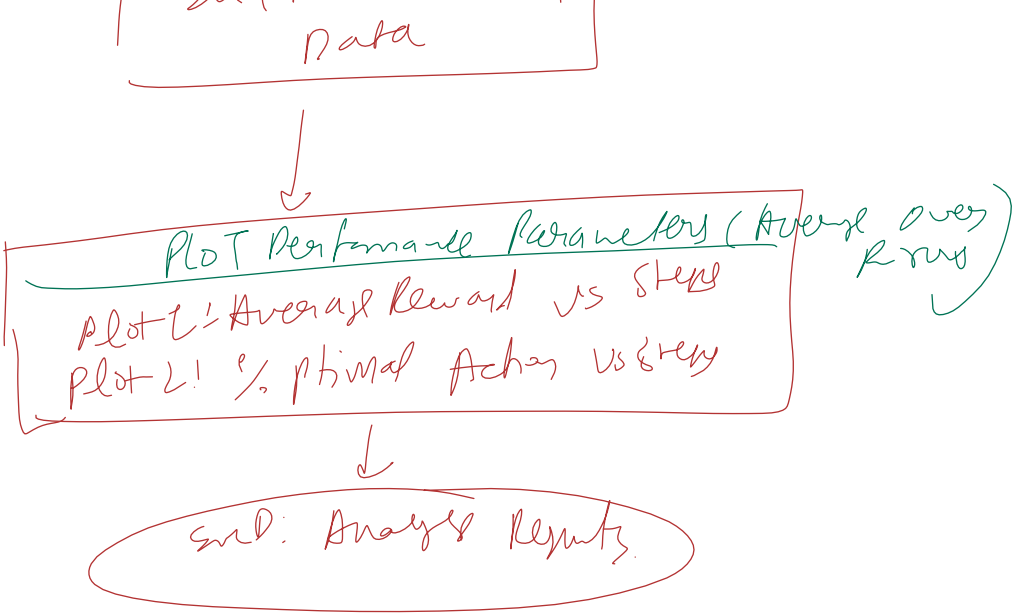

START: Initialize Simulation (R = 2000 Runs)
Run Initialization
set parameter: K = 10 (arms), c = 2 (UCB constant), T = 1000 (steps)
Initialize Arrays: Q = [0, ... 0], N = [0, ... 0] (Size K)
Initialize Environment: q_star = N(0, 1) (stationary True Values)
t = 1
Is t <= K ? (Force Initialize phase)
yes
No
select Action A_t = t - 1 (Pull each arm once)
Calculate UCB for all a:
UCB(a) = Q_t(a) + c * sqrt(ln(t) / N_t(a))
select Action A_t = argmax(UCB)
Reward & Incremental Update
Get Reward: R_t ~ N(q_star[A_t], 1)
Update Count: N_t(A_t) = N_t(A_t) + 1
Update estimate: Q_{t+1}(A_t) = Q_t(A_t) + (1 / N_t(A_t)) * (R_t - Q_t(A_t))
Record Metrics: R_t, is_optimal (A_t == argmax(q_star))
Is t < T ?
yes
No
End Run & Aggregate


Data
Plot Performance Parameters (Average over Runs)
Plot 1: Average Reward vs Steps
Plot 2: % optimal Action vs Steps
End: Analyse Results