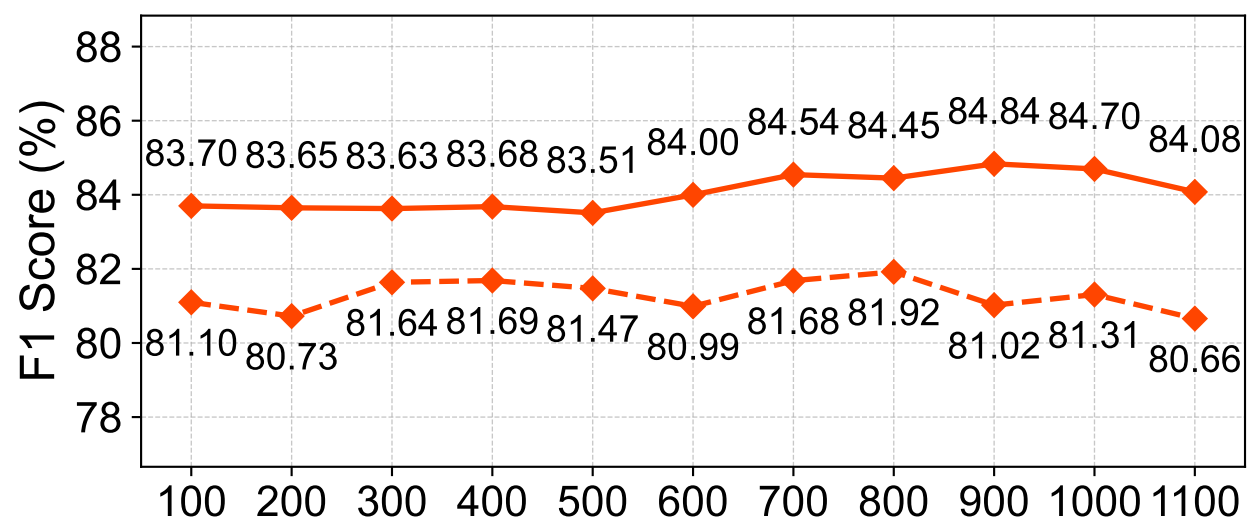


ACD

t-test, $p_{adj} < 0.001$ ***

Rest16

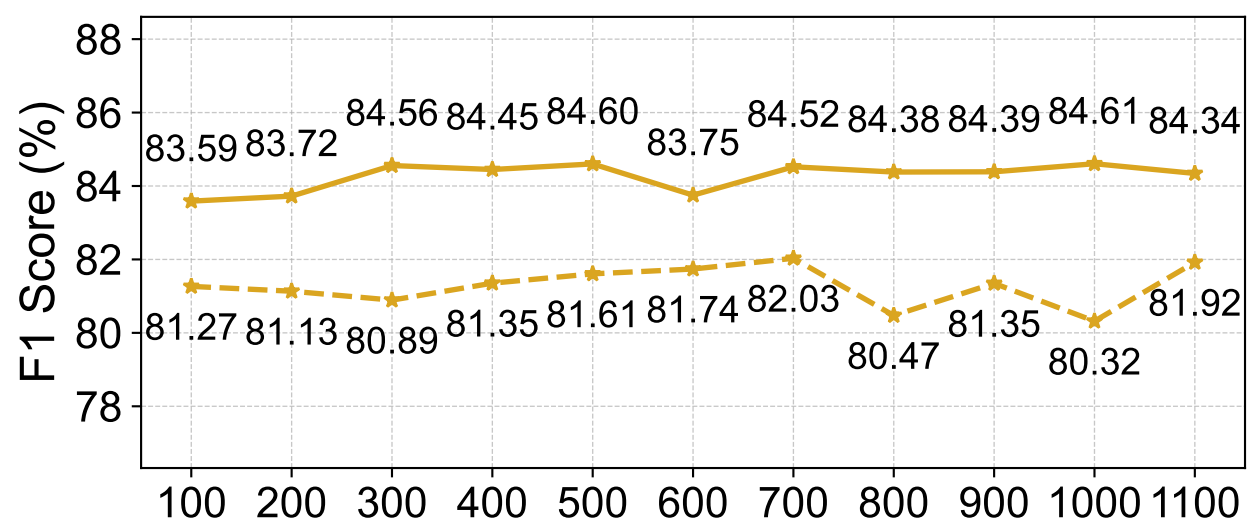
$M_{diff} = 2.78$



Wilcoxon, $p_{adj} = 0.005$ **

FlightABSA

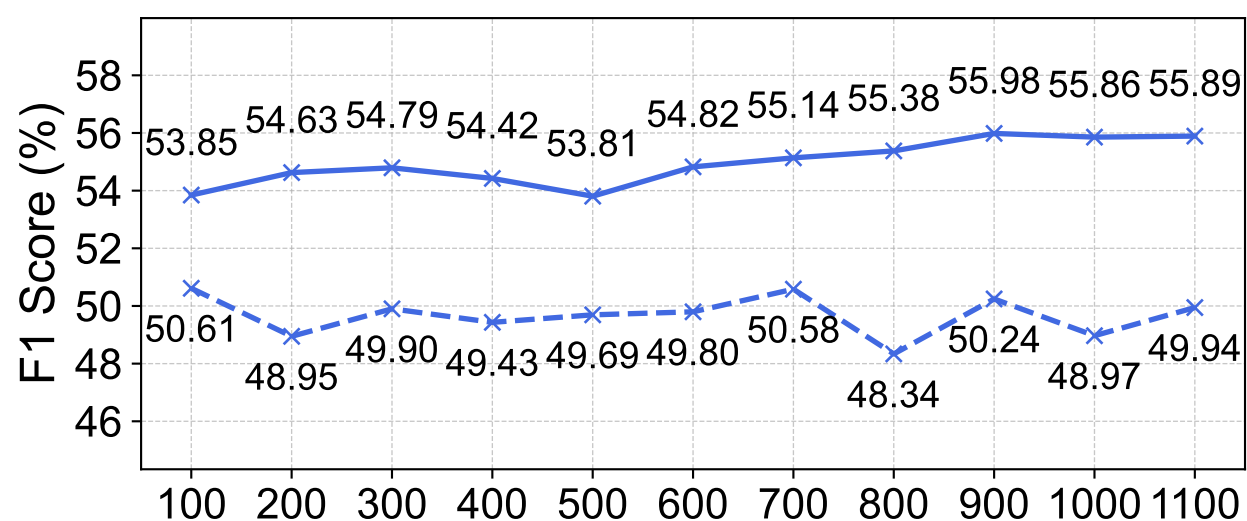
$M_{diff} = 2.98$



t-test, $p_{adj} < 0.001$ ***

Coursera

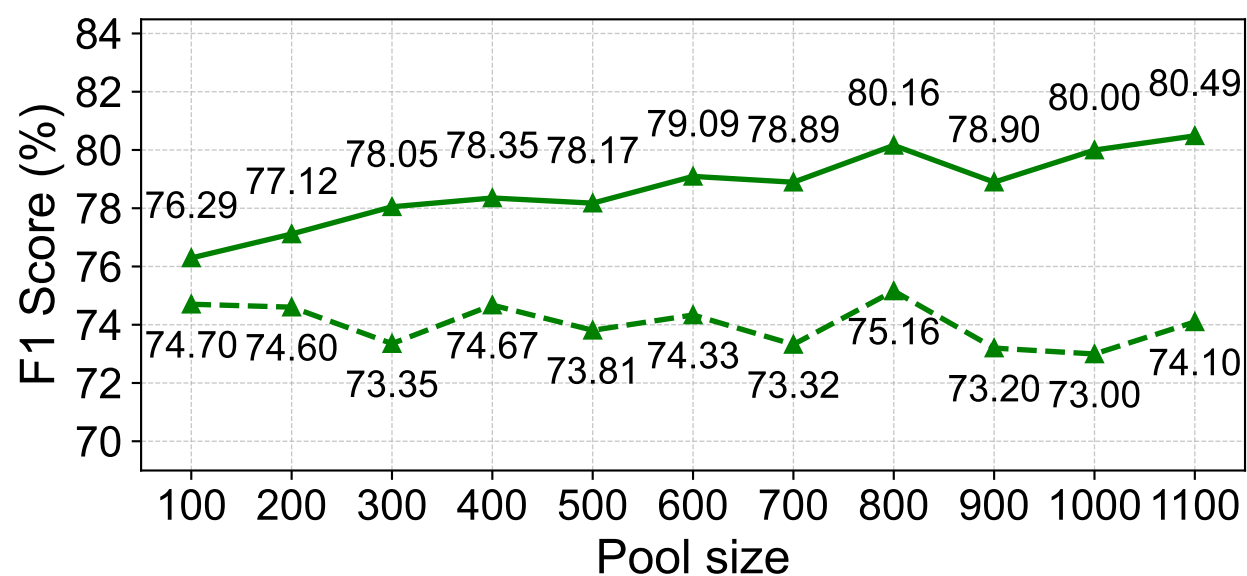
$M_{diff} = 5.28$



t-test, $p_{adj} < 0.001$ ***

Hotels

$M_{diff} = 4.66$

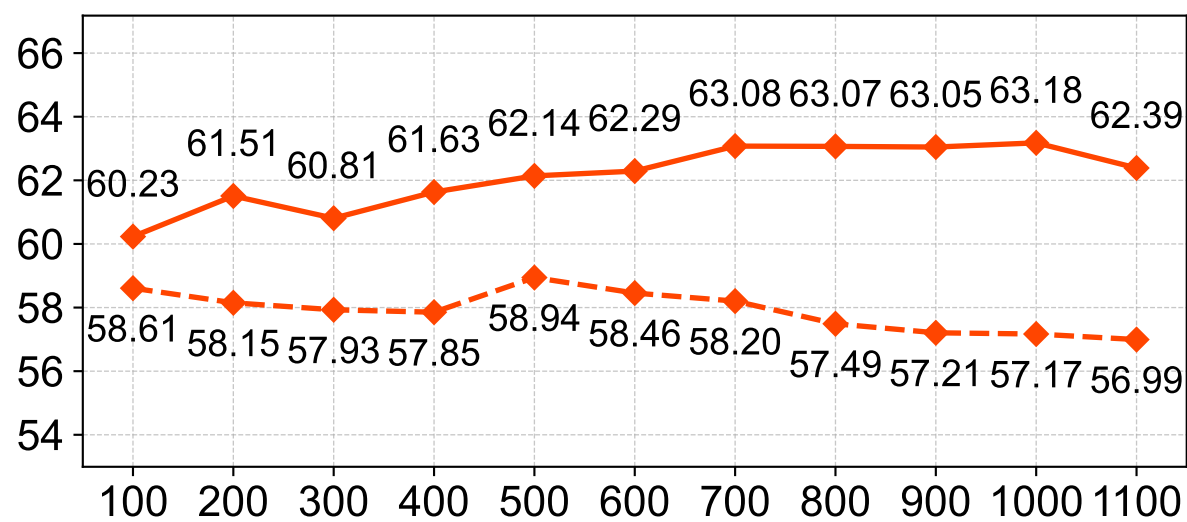


TASD

t-test, $p_{adj} < 0.001$ ***

Rest16

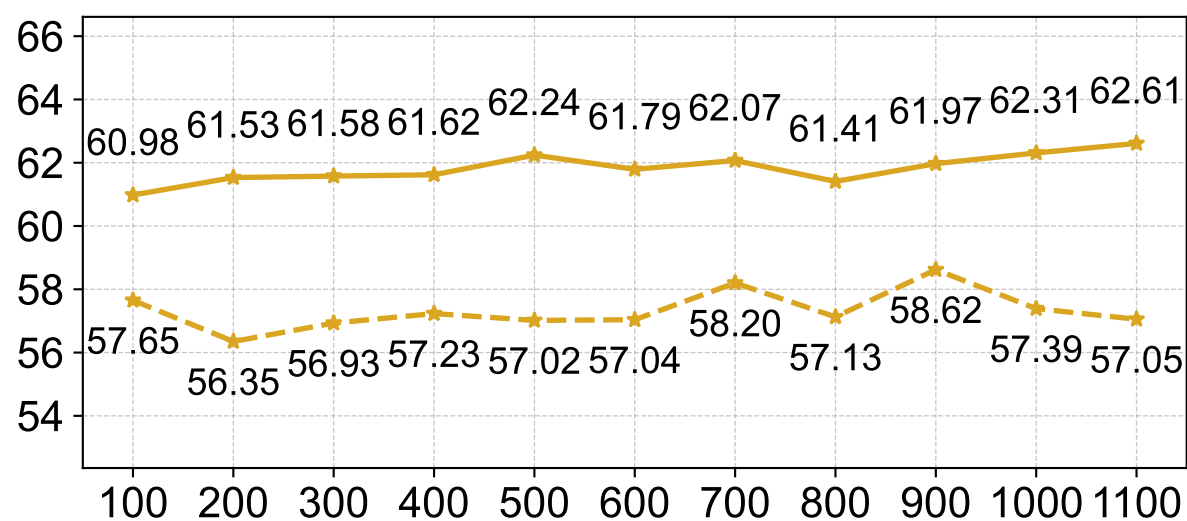
$M_{diff} = 4.21$



t-test, $p_{adj} < 0.001$ ***

FlightABSA

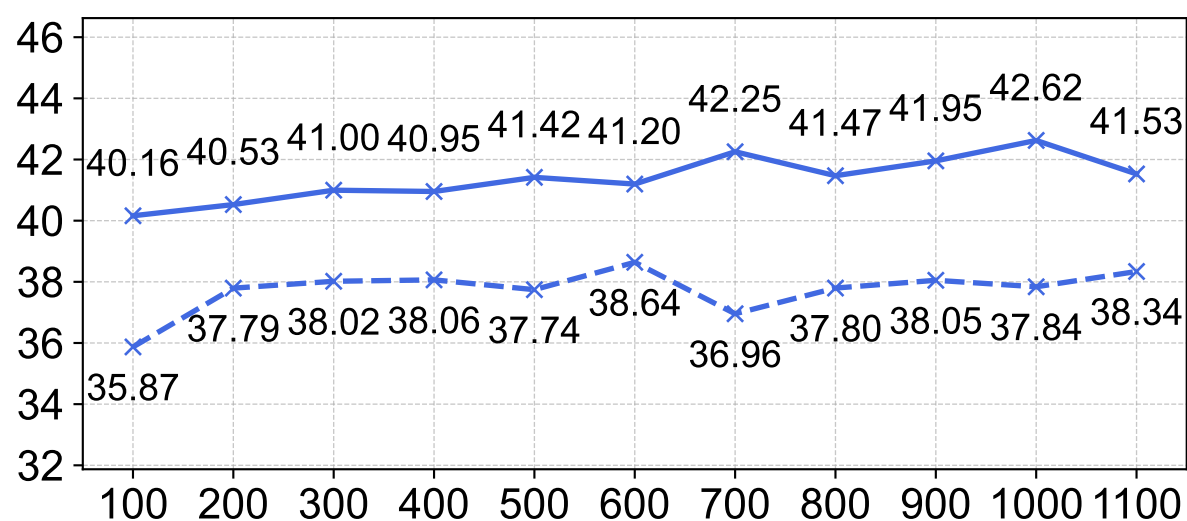
$M_{diff} = 4.50$



Wilcoxon, $p_{adj} = 0.005$ **

Coursera

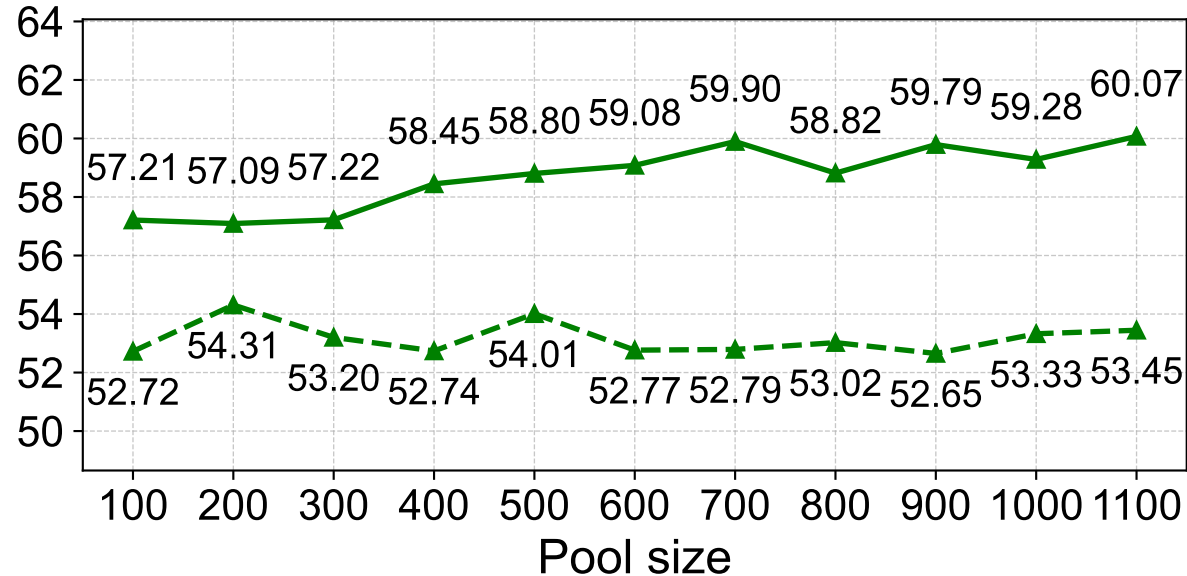
$M_{diff} = 3.63$



Wilcoxon, $p_{adj} = 0.005$ **

Hotels

$M_{diff} = 5.52$

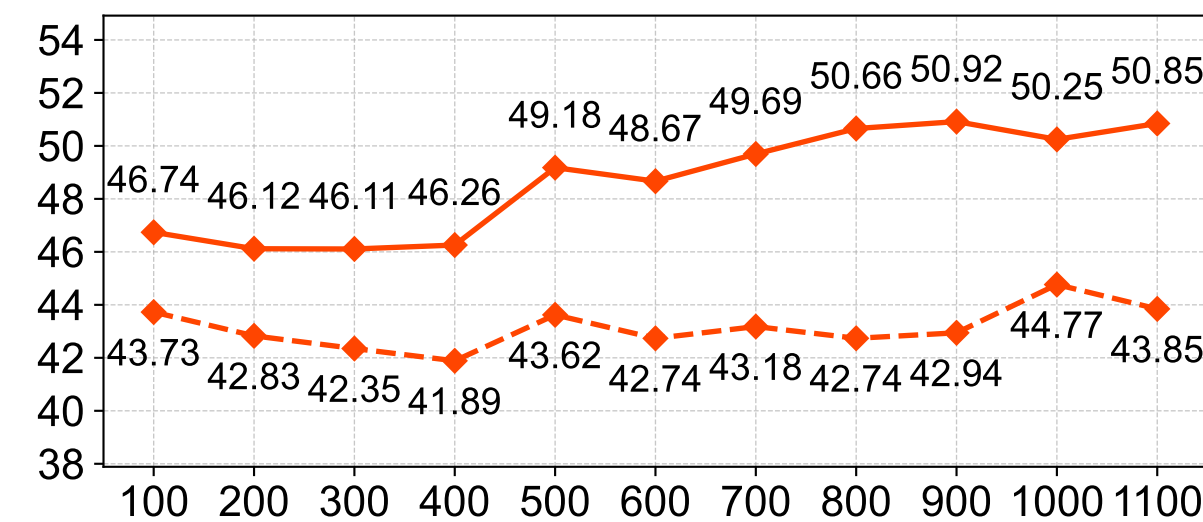


ASQP

Wilcoxon, $p_{adj} = 0.005$ **

Rest16

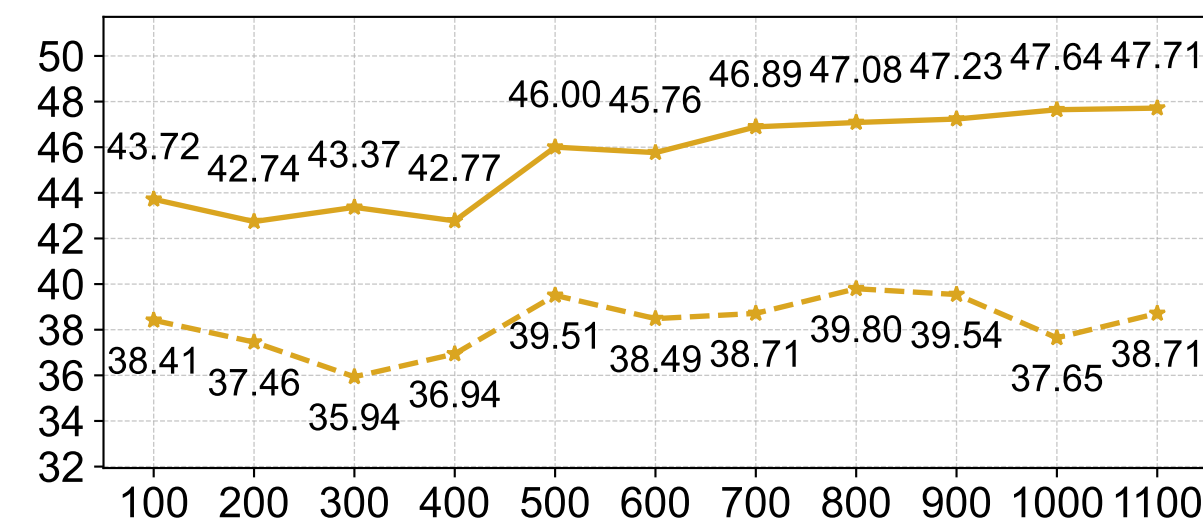
$M_{diff} = 5.53$



Wilcoxon, $p_{adj} = 0.005$ **

FlightABSA

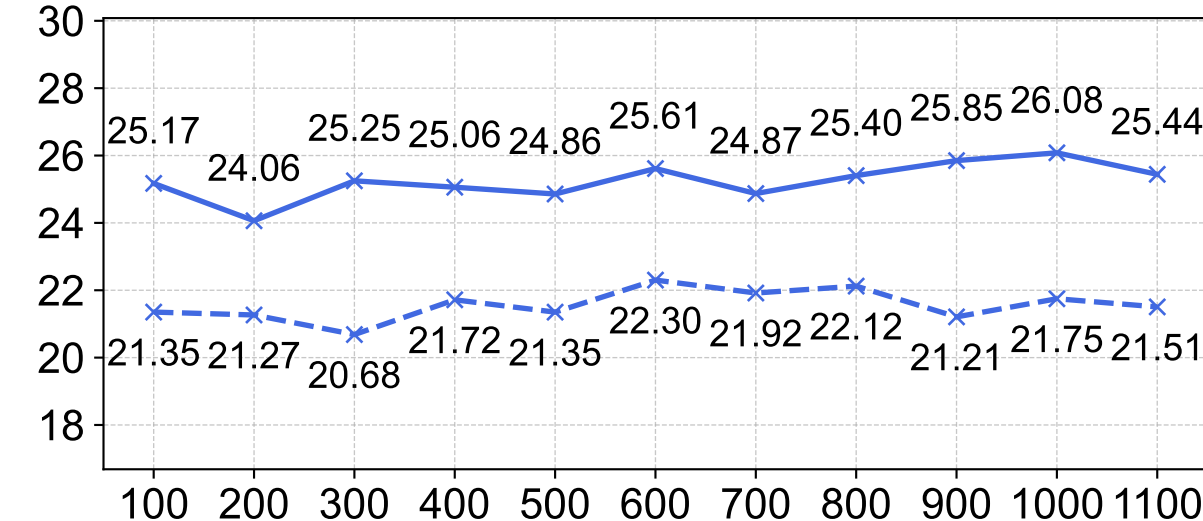
$M_{diff} = 7.25$



t-test, $p_{adj} < 0.001$ ***

Coursera

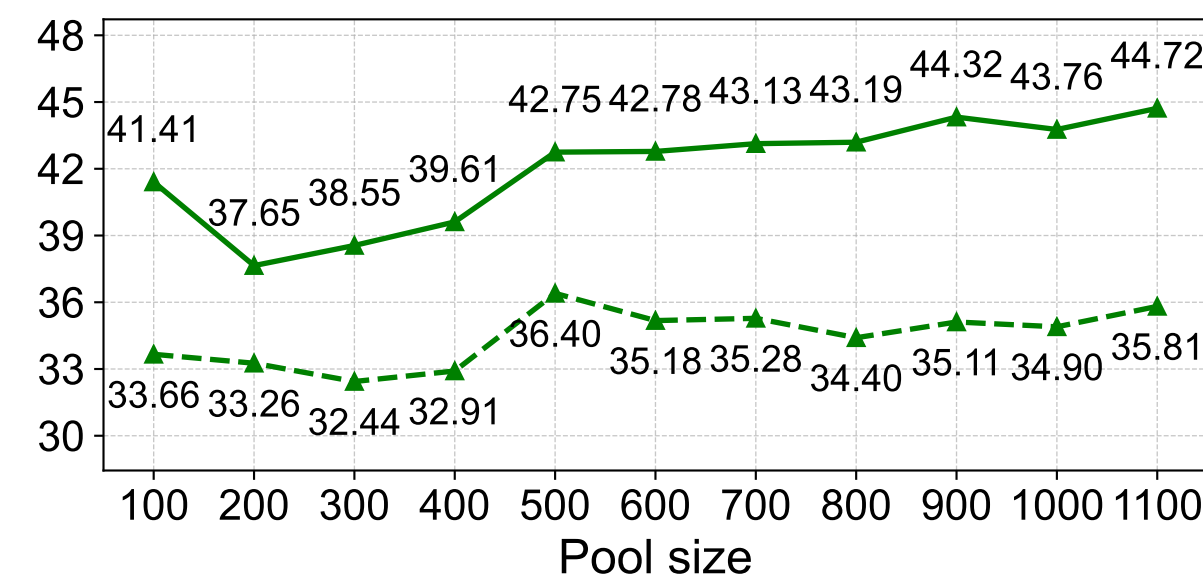
$M_{diff} = 3.68$



t-test, $p_{adj} < 0.001$ ***

Hotels

$M_{diff} = 7.50$



Rest16 FlightABSA Coursera Hotels

— RAG ---- Random