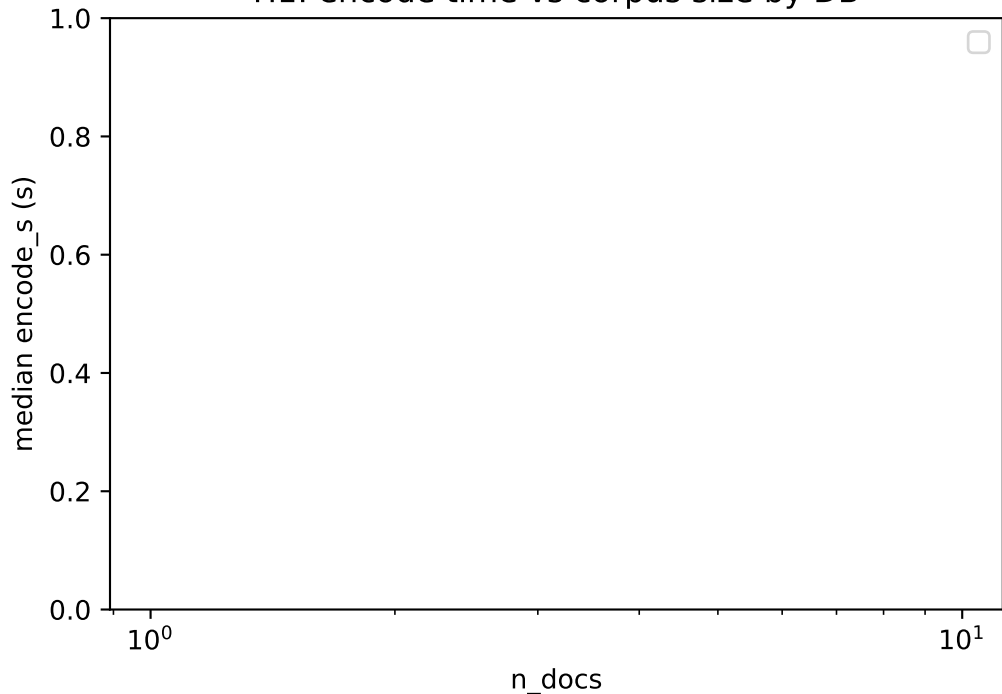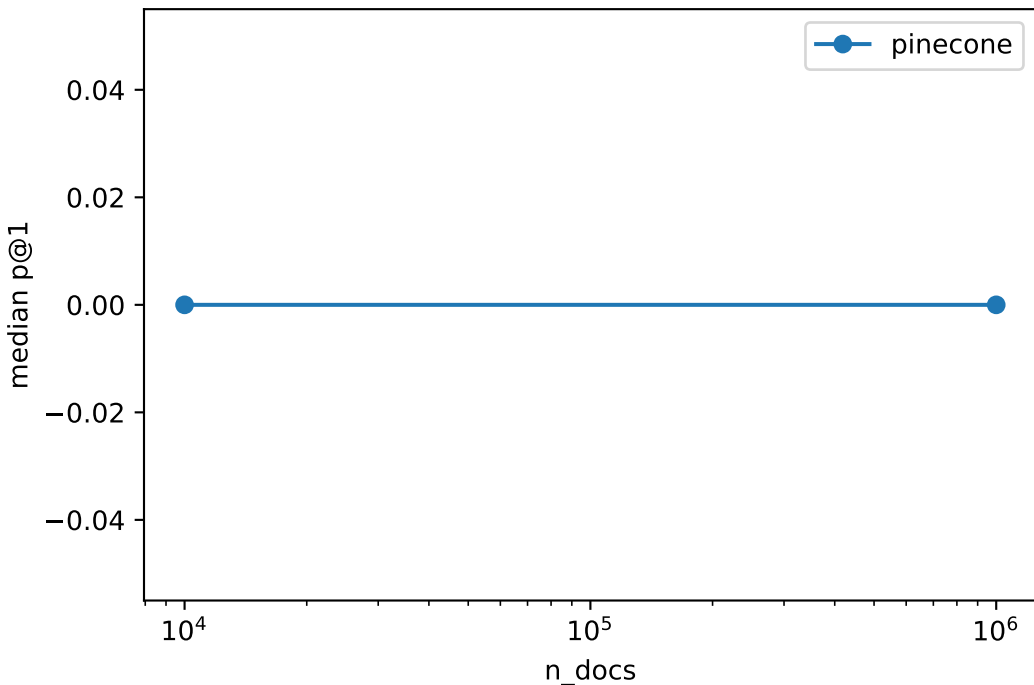H1: encode time vs corpus size by DB

H2: p@1 vs corpus size by DB

H3: model p@1 vs corpus size