Individual Task Completion Time Vs. $F_1^{\mathcal{D}*}$ Score Across Models