mean_f1_score for Draw Class Across Models and Datasets