

Comparison of bleu\_score across models  
Friedman  $\chi^2 = 38.9200$ ,  $p = 0.0000$  (significant)

