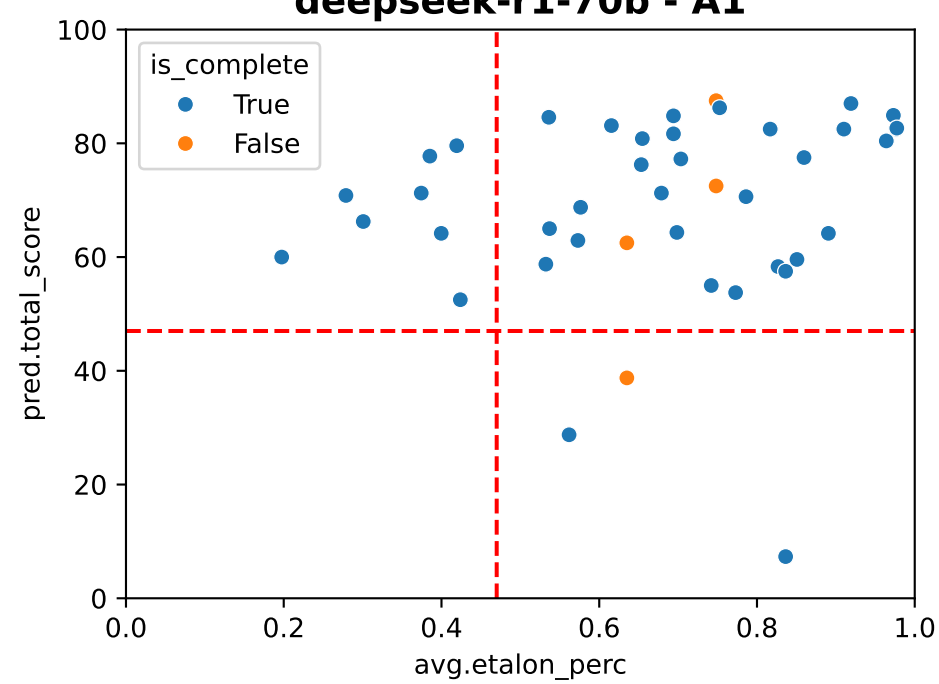
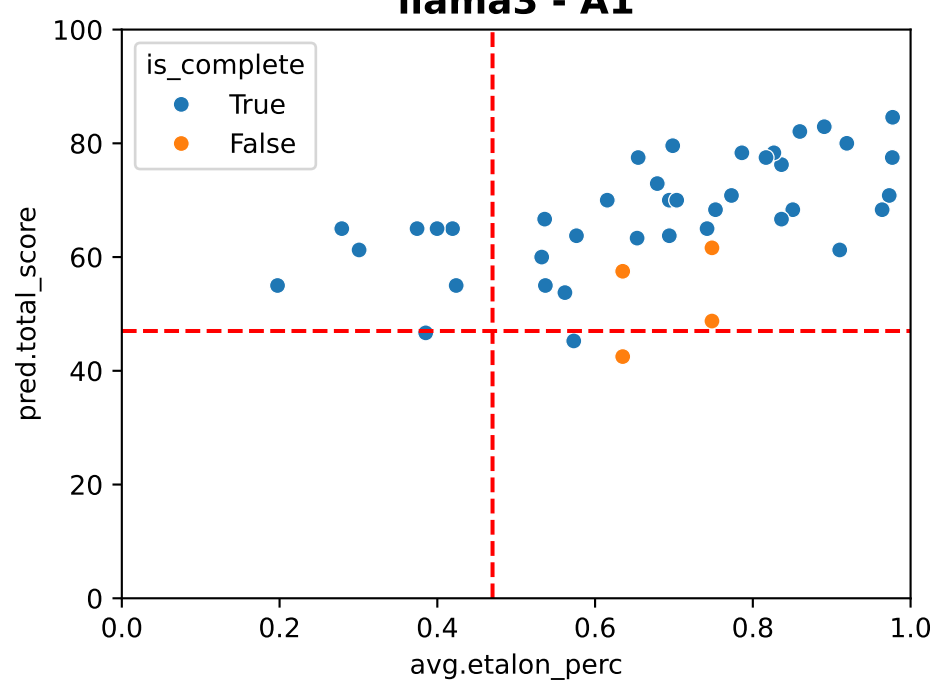


**deepseek-r1-70b - A1**



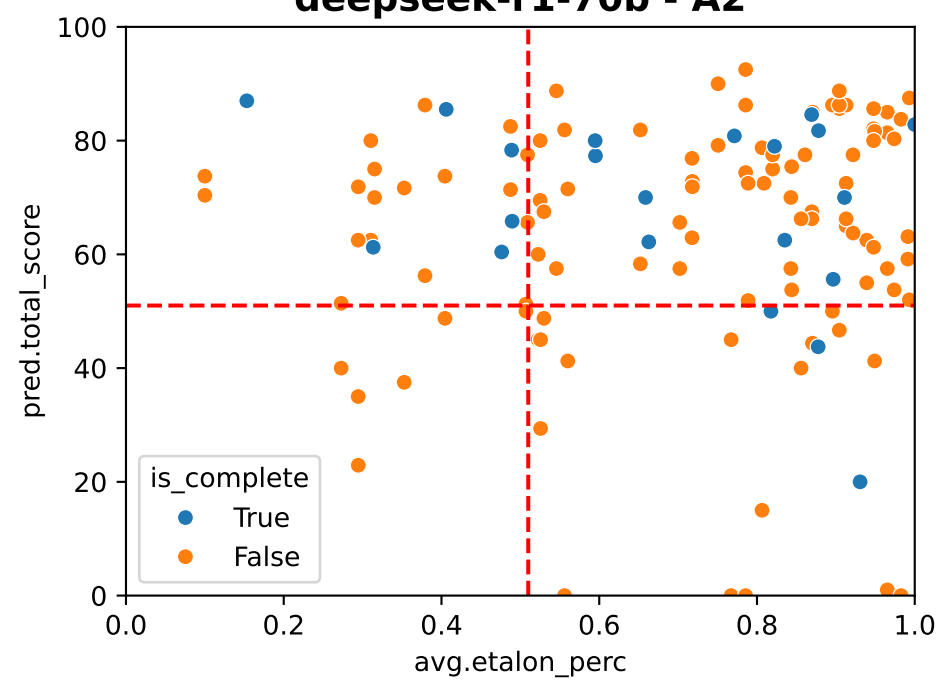
precision: 0.00 0.80  
recall: 0.00 0.91  
f1-score: 0.00 0.85  
support: 8 35  
QWK: -0.11

**llama3 - A1**



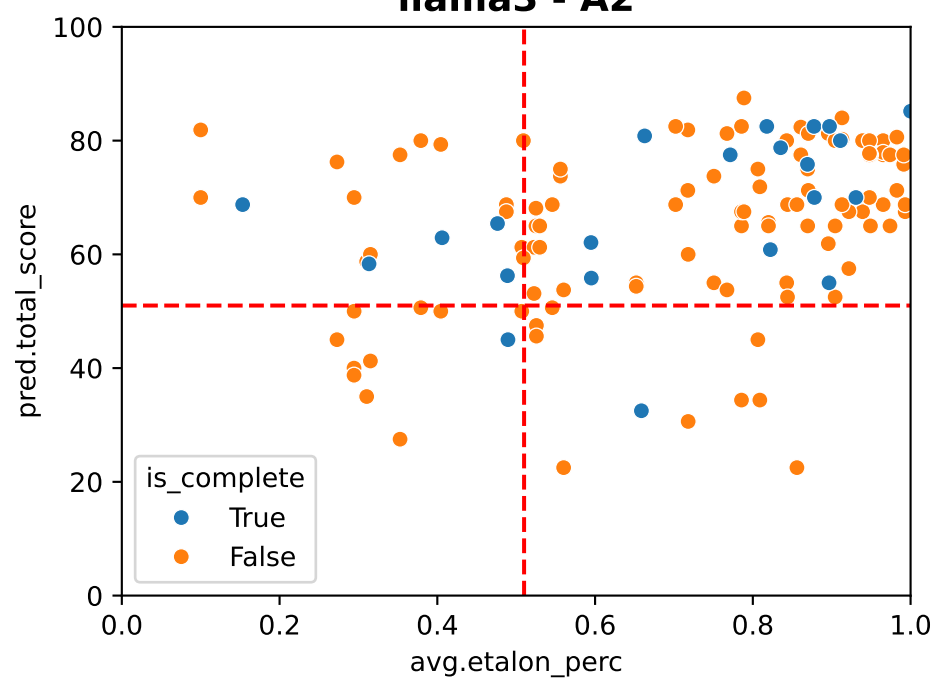
precision: 0.33 0.82  
recall: 0.12 0.94  
f1-score: 0.18 0.88  
support: 8 35  
QWK: 0.09

**deepseek-r1-70b - A2**



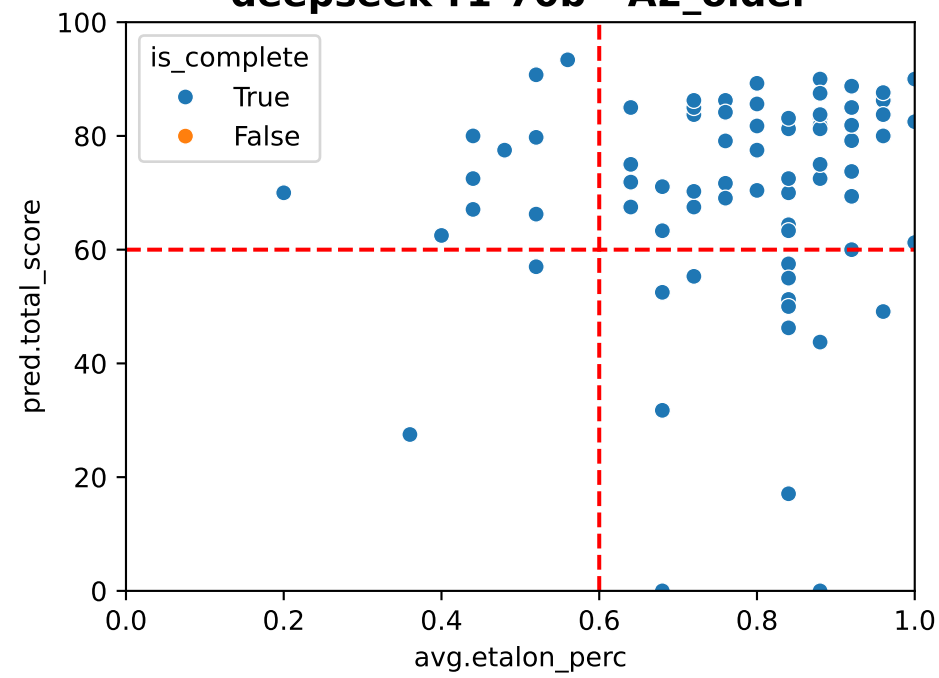
precision: 0.23 0.76  
recall: 0.20 0.79  
f1-score: 0.21 0.78  
support: 30 96  
QWK: -0.01

**llama3 - A2**



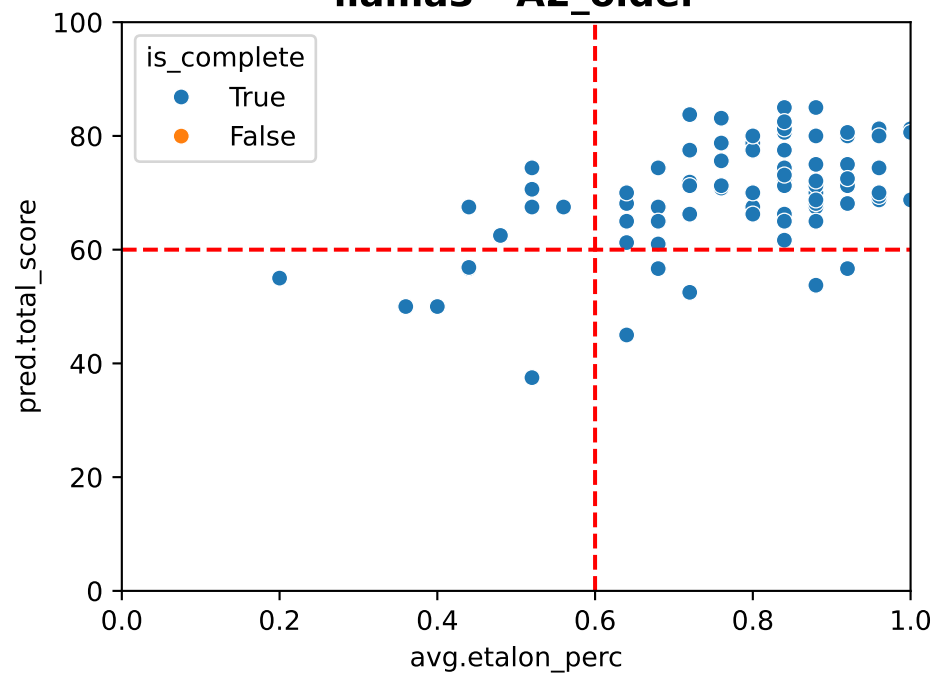
precision: 0.52 0.82  
recall: 0.37 0.90  
f1-score: 0.43 0.86  
support: 30 98  
QWK: 0.30

**deepseek-r1-70b - A2\_older**



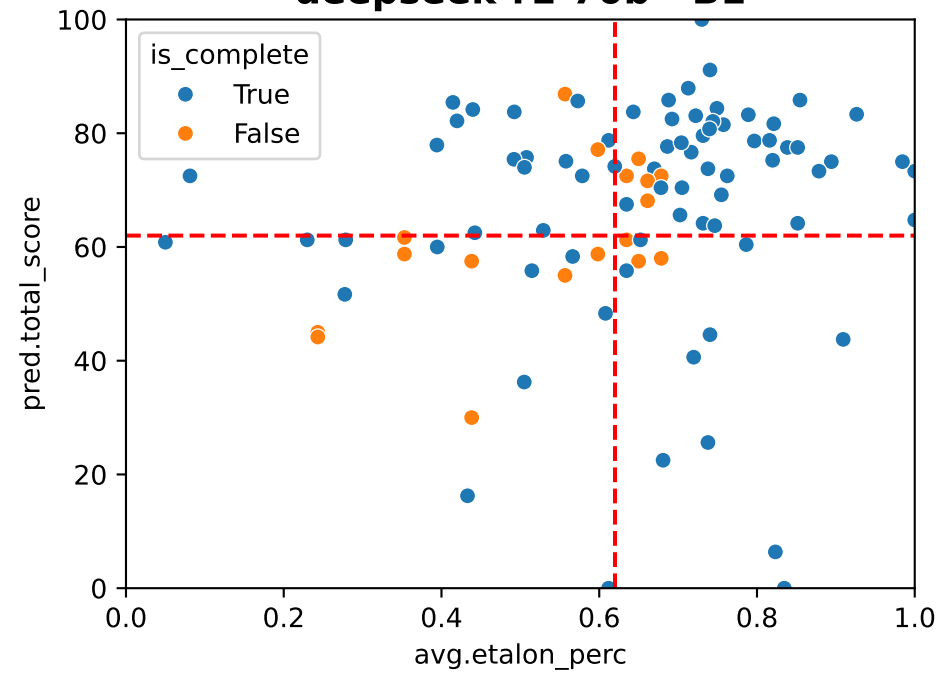
precision: 0.13 0.84  
recall: 0.17 0.81  
f1-score: 0.15 0.82  
support: 12 67  
QWK: -0.02

**llama3 - A2\_older**



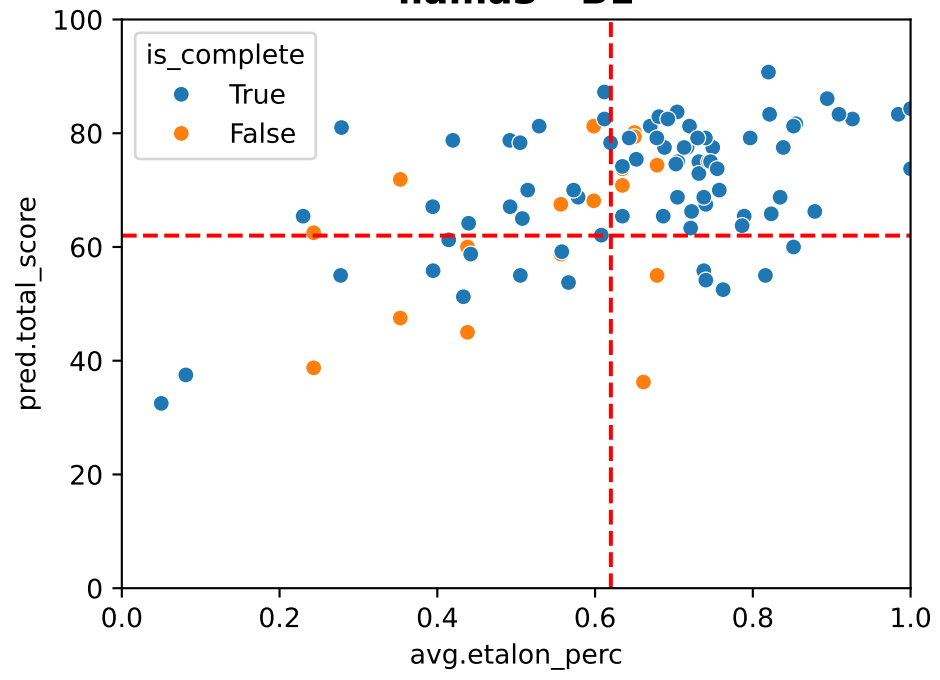
precision: 0.55 0.91  
recall: 0.50 0.93  
f1-score: 0.52 0.92  
support: 12 68  
QWK: 0.44

**deepseek-r1-70b - B1**



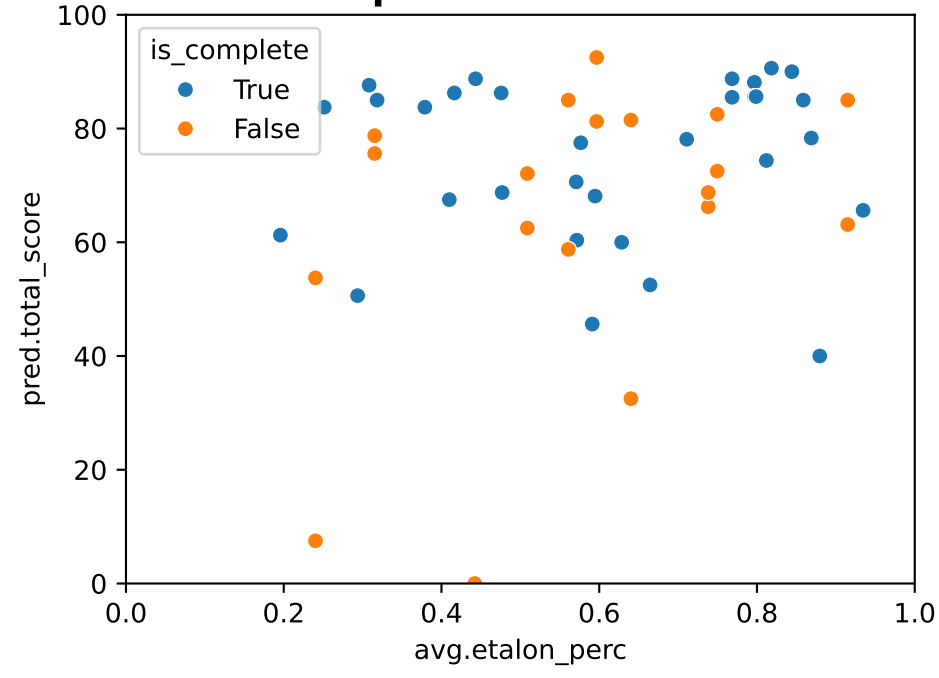
precision: 0.59 0.71  
recall: 0.51 0.78  
f1-score: 0.55 0.74  
support: 37 58  
QWK: 0.30

**llama3 - B1**



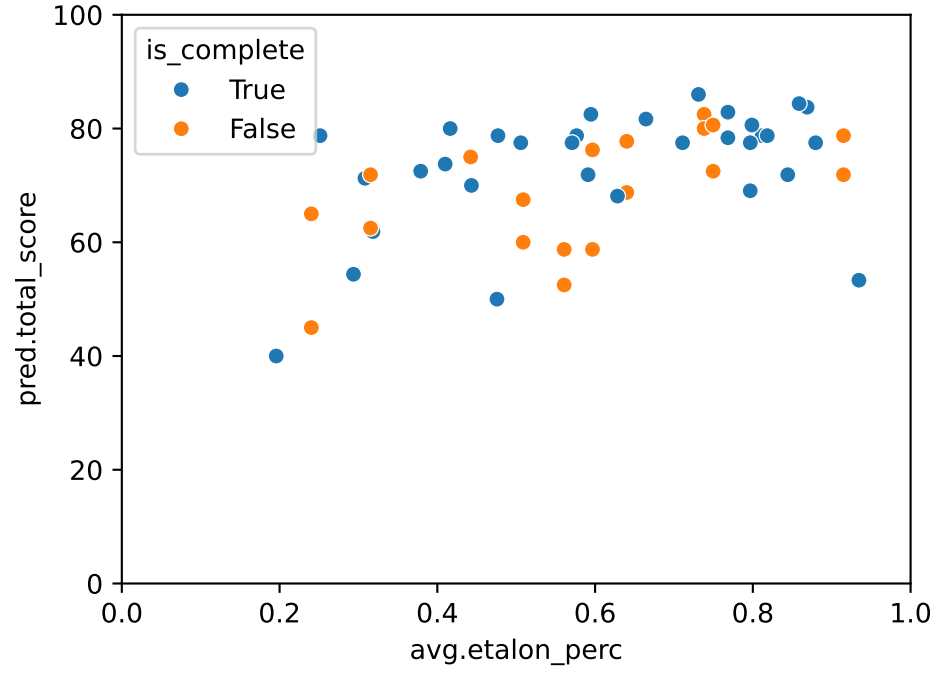
precision: 0.68 0.70  
recall: 0.41 0.88  
f1-score: 0.51 0.78  
support: 37 59  
QWK: 0.31

**deepseek-r1-70b - B2**



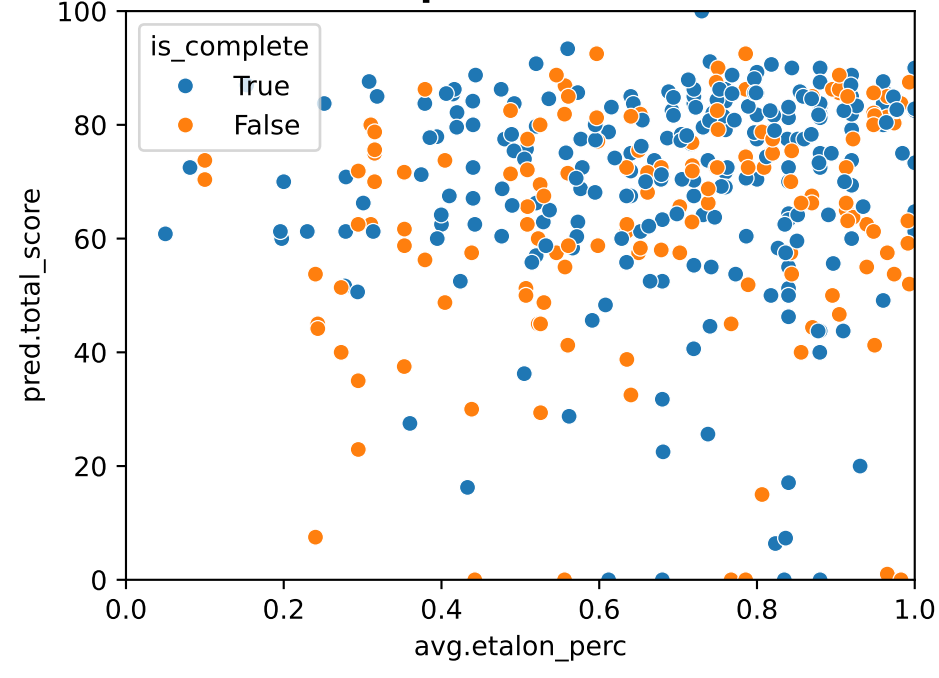
precision: 0.50 0.69  
recall: 0.18 0.91  
f1-score: 0.26 0.78  
support: 17 34  
QWK: 0.11

**llama3 - B2**



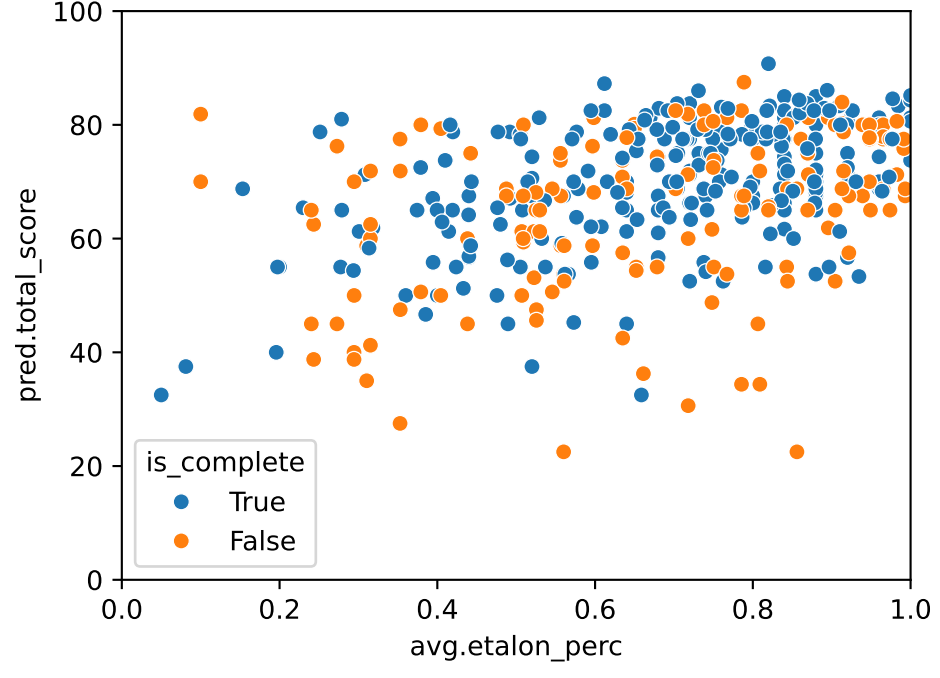
precision: 1.00 0.71  
recall: 0.12 1.00  
f1-score: 0.21 0.83  
support: 17 36  
QWK: 0.15

**deepseek-r1-70b**



precision: 0.24 0.81  
recall: 0.17 0.87  
f1-score: 0.20 0.84  
support: 78 316  
QWK: 0.04

**llama3**



precision: 0.50 0.83  
recall: 0.19 0.95  
f1-score: 0.28 0.89  
support: 78 322  
QWK: 0.19