Data Cleansing

Removal of all Carpet matches since these tournaments stoppen in 2009?

|  |  |
| --- | --- |
| Carpet | mean(loglossout of sample 1:40) for Grass en Hard |
| Only for rating | 0.5603055 |
| Not for rating | 0.5500794 |

So yeah, fuck this noise, let’s see if I also should remove Clay since it sucks probably

THE FOLLOWING DATA IS INVALID BUT USEFUL TO HAVE AN IDEA OF PLANS

|  |  |
| --- | --- |
| Clay | mean(loglossout of sample 1:40) for Grass en Hard |
| Only for rating | 0.5500794 |
| Not for rating | 0.5337872 |

This one is more difficult if you compare BR with Separate models predictions last graph, maybe ask the grand R. Janssen, but probably LogLoss wins, BR is very unreliable anyways

|  |  |
| --- | --- |
| Clay | mean(loglossout of sample 20:40) for Grass en Hard |
| Only for rating | 0.567 |
| Not for rating | 0.5401853 |

|  |  |
| --- | --- |
| Clay | Mean(Br of sample 5:40) for Grass en Hard |
| Only for rating | 7.263987 |
| Not for rating | 5.602273 |

|  |  |
| --- | --- |
| Clay | Mean(Br of sample 20:40) for Grass en Hard |
| Only for rating | 8.048981 |
| Not for rating | 8.443847 |

Hmm, never mind probably, even though I get beat very hard from 5:20 when removing them, this won’t happen for 20:40 and more data is probably better for similar Bankroll levels