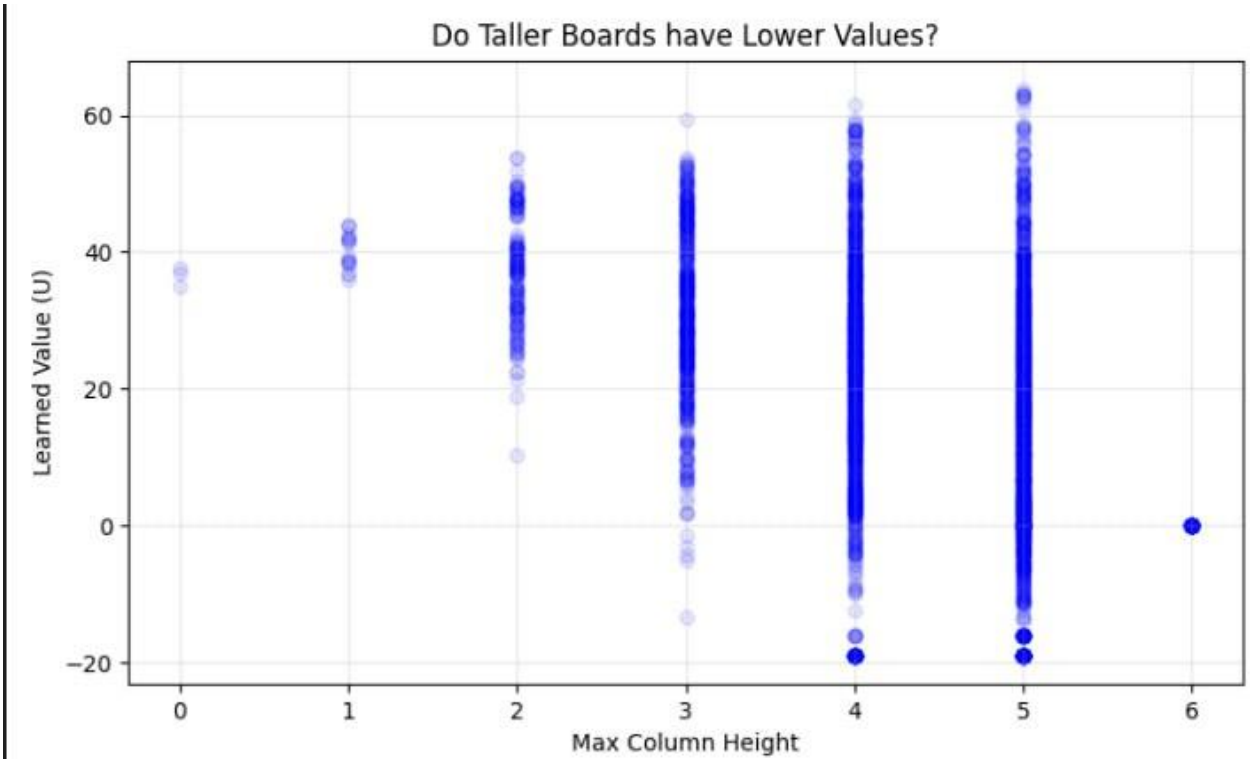
آرمان بیجاری – رضا فراستی

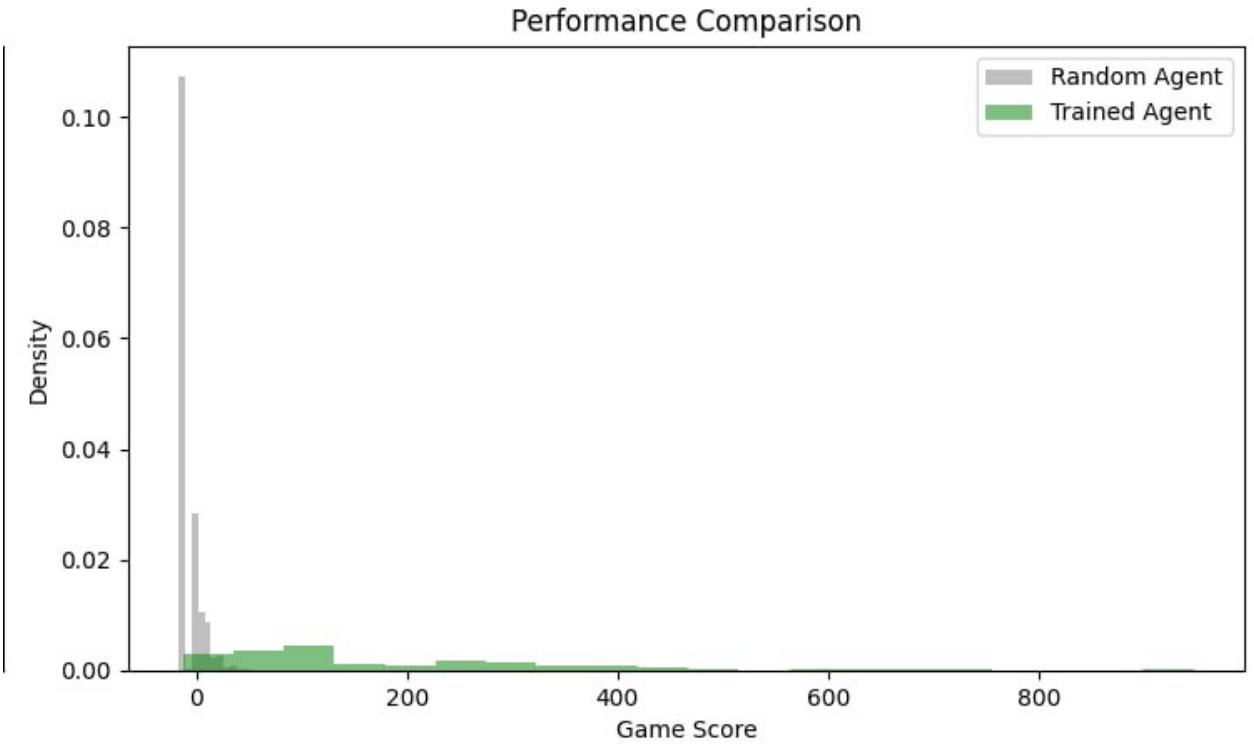
٤٠٢١٢٦٢١٣١ - ٤٠٢١٢٦٢٤٥٩

سوال ۱:



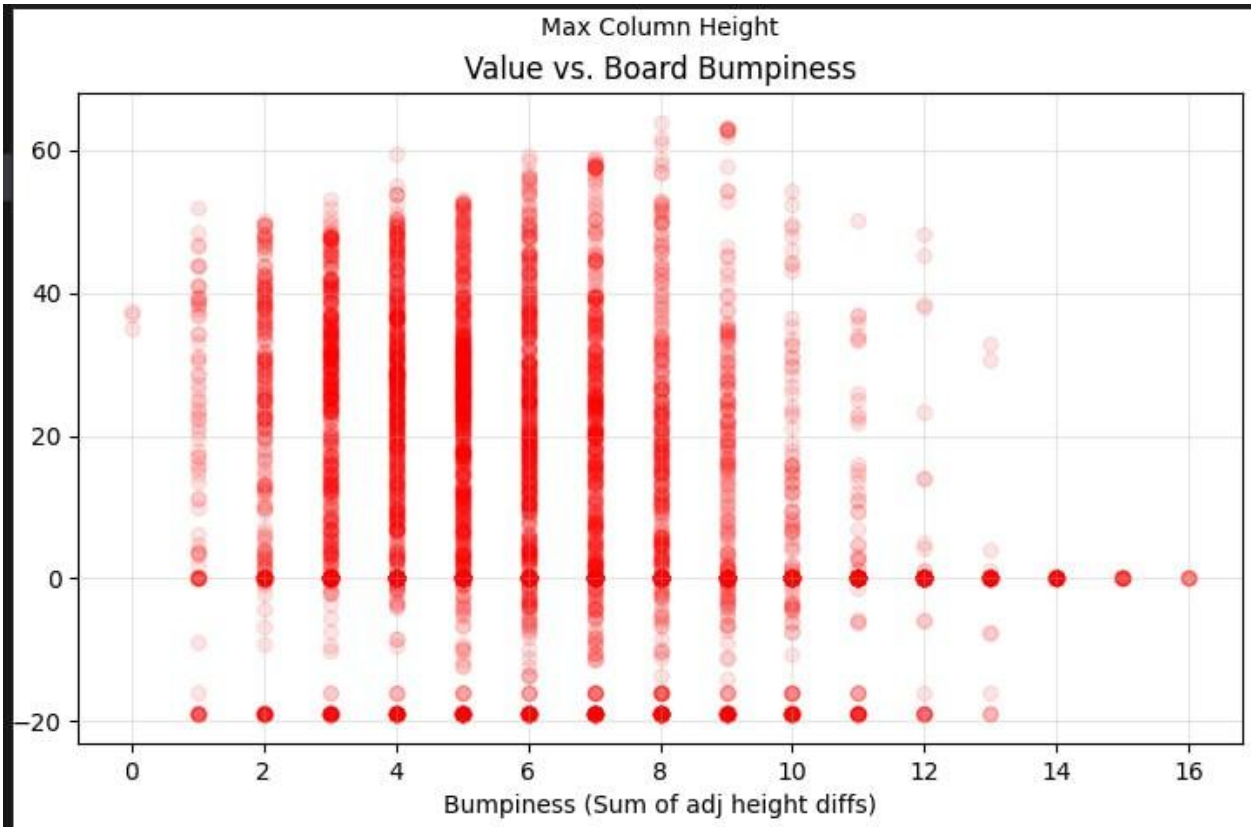
با افزایش ارتفاع ماکسیمم، Learned Value رنج بیشتری را شامل میشود؛ که این یعنی میانگین کاهش میابد و واریانس افزایش میابد. و این یعنی مدل یاد گرفته است که ماکسیمم ارتفاع بالا خوب نیست ولی همزمان همه آنها هم بد نیستند. این موضوع دلیلش این است که مدلسازی حالت ما خوب نیست که در سوال ٦ به آن پرداختیم.

سوال ۲:



همانطور که انتظار داشتیم مدل trained نتیجه بهتری دارد میگیرد و طبق نمودار مدل random اکثرا امتیاز نزدیک به صفر میگیرد و density آن در این نواحی بالاتر است اما مدل trained توزیع شده نتیجه میگیرد و امتیاز ٤٠٠ هم دارد.

سوال ۳:



طبق تعریف bumpiness اگر این مقدار صفر باشد، در حرکت بعدی نمیتوانیم چیزی را حذف کنیم؛ همینطور اگر خیلی زیاد باشد هم یعنی احتمالا در استیت ناهنجاری داشتیم. ولی وقتی متوسط است مثلا ٦، واریانس value بالاتری داریم یعنی احتمال این که حرکت خوبی انتخاب شود بیشتر است.

سوال ٤:

به طور کلی اگر گاما صفر باشد یعنی فقط reward لحظه ای همان حرکت را پس اکشنی را انتخاب میکند که بیشترین ریوارد را بدهد یا کمترین ریسک جریمه را داشته باشد. اما اگر گاما مثلا ۰/۹ باشد باعث میشود علاوه بر ریوارد فعلی، آینده برد را نیز میبیند. برای گامای صفر احتمالا اکشن صفر را انتخاب میکند اما برای گامای بالاتر احتمالا اکشن هایی که کمتر فضای خالی بذارن را انتخاب میکند؛ مثلا اینجا همان صفر را؛ در حالت اول ممکن است یک را هم انتخاب کند.

سوال ٥:

به دلیل منفی شدن reward، ایجنت تلاش میکند تا برای جلوگیری از منفی شدن بیشتر reward، زود بازی را تمام کند و ببازد. برای راستی آزمایی تحلیلمان، تغییر را اعمال کردیم و در مدل های تست شده، با تغییر این مورد، نتایج هم راستا با این تحلیل دیده میشود.

سوال ٦:

state بازی فقط شامل ارتفاع ها و شکل در حال آمدن است؛ بنا بر این مشخص نیست که هر ستون شامل چیست. مثلا ممکن است ارتفاع یک ستون ٤ باشد اما زیر آن خانه خالی وجود داشته باشد. کاری که میتوانیم بکنیم این است که علاوه بر ارتفاع ها و شکل در حال آمدن، میتوانیم تعداد خونه های خالی هر ستون رو نگه داریم و این موضوع میتواند باعث شود که دانش ما نسبت به حالت بیشتر شود.