“It’s a make-or-miss league.”

*—origin unknown*

"They tell us every game, every day, 'Trust the Process.'"

[*—Tony Wroten*](https://www.espn.com/nba/story/_/id/12318808/the-philadelphia-76ers-radical-guide-winning)

It’s been a pretty weird last three years for Jayson Tatum.

Well, to be fair, the last three years have been pretty bizarre just sort of universally… but particularly with regards to Jayson Tatum’s career trajectory, as dictated by how people feel about his shot-taking and his shot-making. In his lone collegiate season at Duke, Tatum demonstrated a proclivity for — and in fairness, something of a knack for — operating in isolation. Per Synergy Sports, [nearly a quarter of Tatum’s offensive usage](https://twitter.com/ChrisForsberg_/status/878049660786548736) came from him isoing, with his points-per-possession average on isos (0.896) actually surpassing his average on off-ball actions in the half-court (0.893). [In part due to his iso-heavy style](https://twitter.com/Tim_NBA/status/884851253834129408), Tatum also developed something of a reputation for launching shots from just inside the 3-point line, a blend that inspired comps in the mold of a Rudy Gay or Harrison Barnes, Danny Granger on the high end.

Needless to say, [I had](https://twitter.com/RetepAdam/status/882019541425041409) [some doubts](https://twitter.com/RetepAdam/status/864651885541810177) [at the time](https://twitter.com/RetepAdam/status/875852746662047745).

That’s not to say that the consensus on him at the time wasn’t justified by the data. Anecdotally, a word cloud of Tatum talk at the time would have featured “ISO” and “MID-RANGE” in large block letters, but a glance at FiveThirtyEight’s CARMELO [projections for his rookie season](https://projects.fivethirtyeight.com/carmelo-2017-18/jayson-tatum/) indicates that statistically, among his strongest comparables ranked players like Brandon Ingram, Andrew Wiggins, Carmelo Anthony and DeMar DeRozan.

A summer spent showing off his Drew Hanlen-crafted footwork in Vegas elevated expectations heading into his NBA debut, though questions about his shot selection still lingered. As it turned out, that would quite literally be the least of anyone’s worries. A player tabbed as the next mid-range volume scorer turned in an [uncanny Otto Porter Jr. impression](https://twitter.com/DuncanSmithNBA/status/942257486379307008) in the season’s first few months, racking up the kind of scoring efficiency that had him nipping at Ben Simmons’s heels for Rookie of the Year. People may have expected him to be good, but nobody saw it coming in this fashion. His Synergy numbers would eventually settle into [merely very, very good territory](https://twitter.com/DuncanSmithNBA/status/1017982624487411713), but the narrative had been ravaged. Whether everyone had underestimated his unselfishness or he just needed a little ‘Brad Stevens magic’ or whatever the reason we could find to invent, he was just a thoroughly different player than the one he’d shown at Duke. Gone were the mid-range marauders [from his CARMELO comps](https://projects.fivethirtyeight.com/carmelo-2018-19/jayson-tatum/). If he was going to be a star, it wasn’t going to be as a guy who dribbled the air out of the ball and jacked up contested fadeaways; it would be in the mold of fellow St. Louis native OPJ, who had begun to assuage some of the doubts about his max contract by just being really, really efficient at putting the ball in the hoop when passed to.

Then he met with Kobe.

Or so the narrative goes. Tatum spent time [working out with his childhood idol](https://www.gq.com/story/jayson-tatum-next-mamba), Kobe Bryant, and returned with a different… let’s say ‘mentality.’ In the season that followed, Tatum’s offensive profile underwent a marked shift. His spot-up frequency fell from 27.3% to 18.2%, while his iso frequency increased and his post-ups nearly doubled. His mid-range attempts increased, and with it, his eFG% declined sharply. In essence, he started to look more like the player many expected him to be coming out of Duke. His top CARMELO comps [for this upcoming season](https://projects.fivethirtyeight.com/carmelo/jayson-tatum/) include Rudy Gay, Harrison Barnes and Carmelo Anthony.

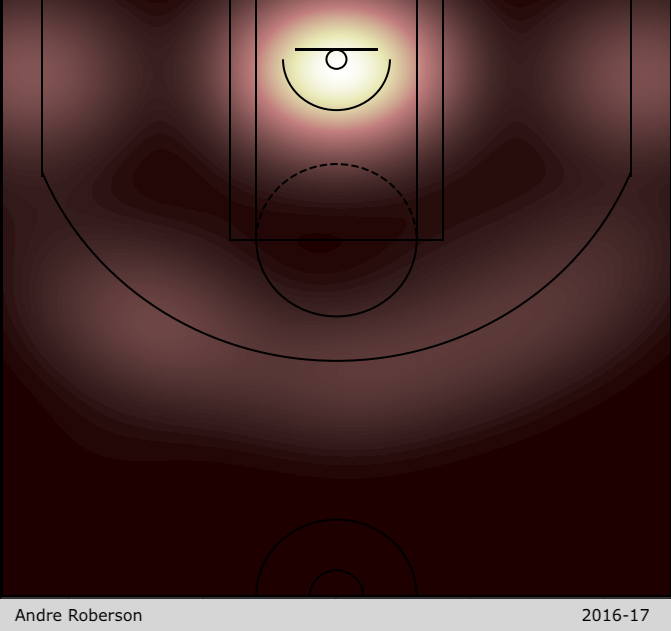
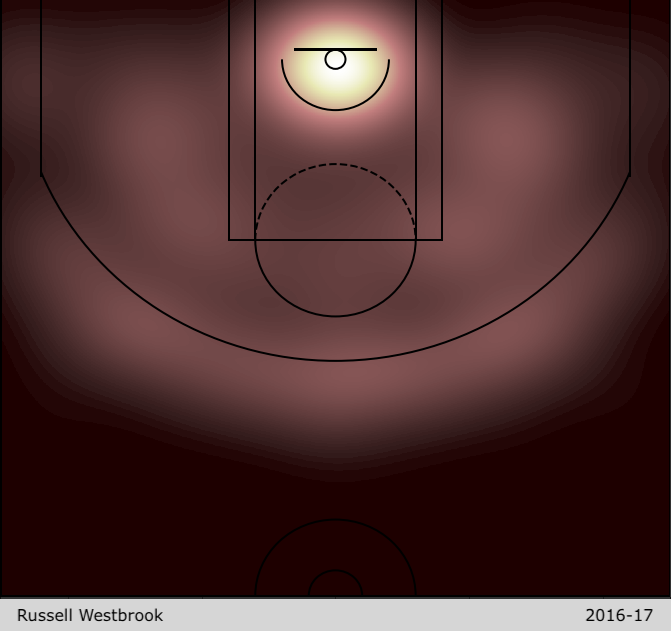
Throughout this transformation, though, from prospective volume scorer to paragon of efficiency back to Massachusetts Mamba, there’s been an underlying question of what is actually best for Jayson Tatum’s development. Yes, it would be super helpful for the Celtics, in the midst of a title push, to have the version of Tatum that stands where they need him and drains everything. But what about the Celtics of 2025? Tatum has a long career ahead of him, and though stars come in many forms, the type of player we think of as a franchise pillar usually bears several hallmarks. Additionally, lower-usage players tend to be more sensitive to changes in the context around them, which is likely one of the reasons Otto Porter Jr. doesn’t play for the Wizards anymore.

Many who worship at the altar of efficiency were quick to rag on Kobe’s influence, namely the advice he gave Tatum in an episode of ‘Detail’ where he [encouraged the young Celtic](https://www.youtube.com/watch?v=NZmEY3n97P4) to eschew popping out to the 3-point line off screens in favor of curling into pockets just inside the arc for catch-and-shoot opportunities that just so happened to be long, turnaround 2s.

And looking at the advice on its face… I mean, yes, prioritizing a difficult shot over a more efficient one is antithetical to how we think about basketball in 2019. But it’s important to note that the line of thinking in play here didn’t come from nowhere. I’ve long held the position that as much as volume-for-volume’s-sake scoring is not *good*… there’s likely some merit to that kind of shot diversity. It’s an idea we toy with when we talk about the usage-efficiency curve. It’s something we saw directly addressed by Houston’s acquisition of Chris Paul, an all-timer of a mid-range assassin, several years ago. There’s *something* there, even if we’re not necessarily sure of what it is. Put simply: if we’re able to explain a decent chunk of what’s happening on the court using the “what” as our inputs, it seems plausible that the remaining unexplained components have more to do with the “how” of what’s happening.

In an effort to get at the heart of this (and with some help from Nylon’s own Krishna Narsu), I compiled shot location data for every player from each season dating back to 1996-97 and visualized the data in the form of a heat map of each player’s shot distribution. Unlike most shot charts, the goal here was not to show where players were most proficient but simply to identify *where* each player’s shots were coming from. The hypothesis in mind was that players destined to carry the load on offense would make themselves known to some extent by their shot selection, specifically that they’d be more strongly represented in the areas that were typically darker for players with more regimented roles. The thinking was that this would either be a reflection of the leeway entrusted and responsibility tasked to players capable of carrying such a load — or that some of these would be areas that these players were the only ones consistently getting up shots from because they were the only ones who could routinely get to those spots ahead of their defenders.

And at a glance, there seemed to be some merit to the idea. To demonstrate, here’s perhaps a more extreme example from the 2016-17 season. The brighter an area is, the greater proportion that player’s attempts came from there.



On the left is Russell Westbrook, given a green light from just about anywhere in the county. To the right is his teammate Andre Roberson, who shot so infrequently from the foul line extended that there’s a blip right at the center that’s the same shade of squalor as out at midcourt.

It also gave me a small window into how player tendencies and roles developed over the course of their career. Not just with the Brook Lopezes of the world, but even with a player like Ray Allen, his shot distribution charted a path from slasher to shooter to specialist — with the late 90s in particular almost outright tracing lines along the contours of set plays one imagines Milwaukee was diagramming for him.

<img> https://i.imgur.com/yXjuM7k.gif</img>

From there, the goal was simply to translate what seemed to be bearing itself out anecdotally into a quantifiable input variable for projecting a young player’s development track.

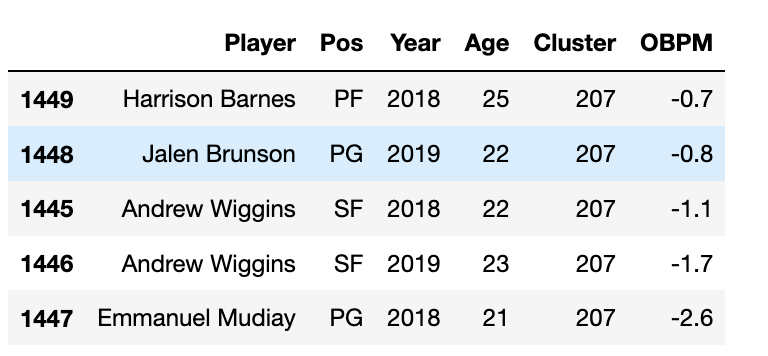
After tinkering around with trying to regress on the individual shot charts themselves, I decided a better approach might be to cluster them first, then use those clusters as inputs for the actual model. Not only would this simplify the process, but it would both allow me to introduce other variables more easily without having to worry about weighting them appropriately against so many shot chart inputs as well as prove more forgiving than the direct regression, which was trying too hard to FIT-OUT, as LeBron might say.

After initially including all players from within the span, I eventually filtered the sample down to only those who had played at least 500 minutes in the season in question after hierarchical clustering based on cosine similarity was yielding the strongest results using that as a cut-point.

For some young players, the clustering introduced some welcome context regarding which of their peers, past and present, shared a similar shot profile.



For others, perhaps a little less so.



But across the board, the clusters passed the smell test, and I felt comfortable moving forward with them.

From there, it was a matter of deciding how to approach the problem and which years made most sense to focus on for player development. I felt pretty strongly that I wanted to have the starting point be the player’s rookie year, but in order to quiet some of the bias that would inevitably introduce, I decided to include every player’s first year of hitting the 500-minute threshold in the NBA, provided they were 25 or younger (based on basketball-reference’s February 1st cut-off) that season. Determining to place the target turned out to be a bigger challenge, as initial passes based on single seasons were understandably flummoxed by Trey Burke posting a freakin’ 4.2 Offensive Box Plus/Minus (OBPM) rating in his fifth NBA season. Ultimately, a cumulative average of the player’s third, fourth and fifth seasons *following their first qualifying season* made the most sense, as it bolstered the model’s ability to remain stable regardless of contract-year surprises and injuries. Even then, there were still some surprising data points (Did you know Ryan Anderson had four straight seasons with a 4+ OBPM, including one where he nearly broke into the 6-point range?), but the results being generated looked increasingly sane.

In order to qualify on the back end, the player simply had to play at least two of the three seasons in their 3-4-5 years and amass at least 500+ combined minutes played in that span. While having players *qualify* by still being around to generate the relevant outputs inherently creates some survivorship bias, I wasn’t able to come up with a dummy value for those who failed to qualify on the back end that meshed well with the attempt to regress on them as Y-values, so I decided to just bite the bullet and exclude them for the time being.

The model itself was an XGBoostRegressor that took the shot clusters from the entire group of players in the 1996-and-on database with at least 500 minutes played but then *filtered out* everyone but the first-year players in the database as it gathered all the other inputs. What this means is that the clusters would be grouped based on the entire breadth of the data, but then they would be pared down to only the players who the model would actually be incorporating. This would make the clusters as true a representation as possible and also create the foundation to expand this research should I wish to examine the totality of players’ careers.

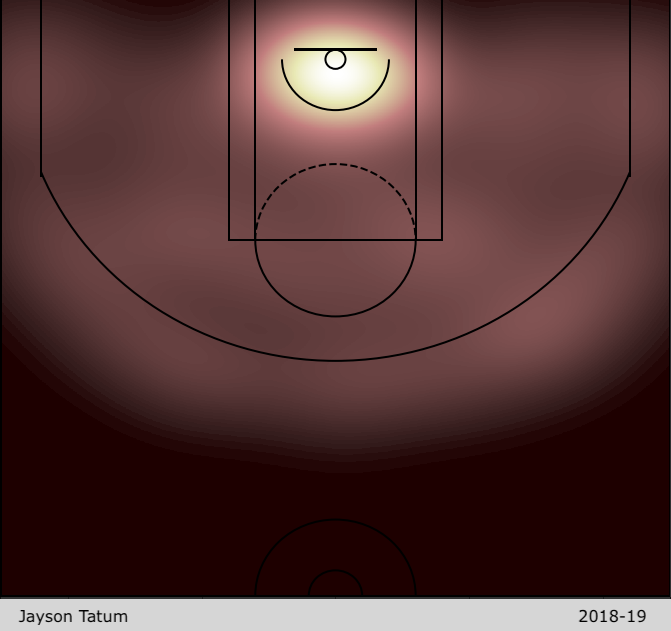
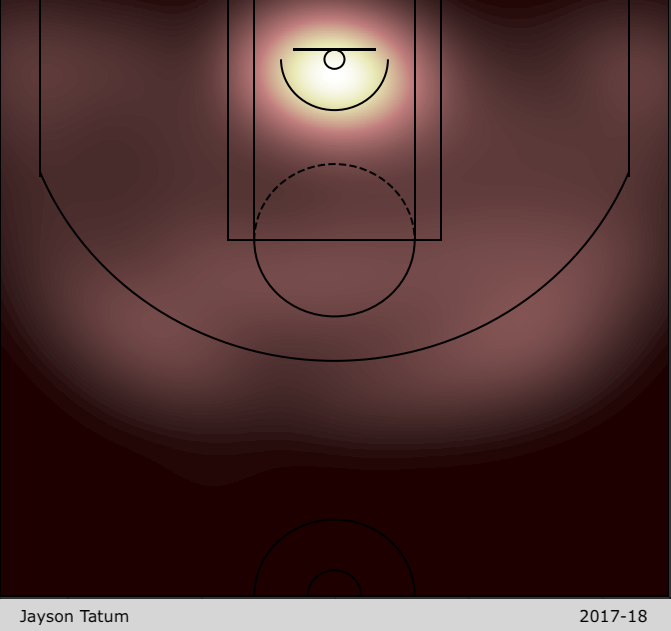
Returning to the hypothesis, the two tacks I was looking to get at were whether this shot distribution data could tell us anything about how *potent* a player was going to be in the future as well how much the team would be able to *play* *through them*. With these two endpoints in mind, I set to regressing with two different versions of the model; one would gauge its ability to predict OBPM for years 3-4-5 and the other would try to forecast assist rate (AST%) in years 3-4-5.

Clusteralone proved a fairly strong classifier of a player’s position, so where its predictive power stood out here was in recasting positions based on where each player had actually spent their time on the court on offense. By itself, **Shot Distribution Cluster** against **AST% 3-4-5** produced an **R2 of 0.3553** with a **mean average error of 5.4334**, both numbers I could feel reasonably good about. However, it was as an interaction variable where the clusters really shone. Using **Shot Distribution Cluster**, **Age** and **Position** as inputs, the model produced an **R2 of 0.5470** with a **mean average error of 4.2342**. While the other two variables fared well on their own, the inclusion of the clusters improved the R2 and mean average error from 0.4959 and 4.5898, respectively. On the flip side, incorporating the other variables helped round out some of the rougher edges of using shot clustering alone; a notable example being the clusters-only model awarding its lowest AST% projection to Ben Simmons, having looked at his data and said “I’d recognize Joel Anthony anywhere!”

As with assist rate, the OBPM model performed better with a little help from friends, topping out its solo effort of **Shot Distribution Cluster** against **OBPM 3-4-5** at an **R2 of 0.1759** with a **mean average error of 1.5959**. But with the addition of **Age** and **Current OBPM**, the model improved to producing an **R2 of 0.5039** with a **mean average error of 1.2580**. Now, yes, it may seem like cheating a little bit to give the model a peak at each player’s first-year OBPM as part of the process of calculating future OBPM. However, if shot clustering is removed, the same model, using **Age** and **Current OBPM** alone, saw its performance fall to an R2 of 0.3921 with a mean average error of 1.3475. So, while the clusters themselves are imperfect prognosticators of long-term success, they do appear to provide valuable context not captured by the other data.

Which brings us back to Jayson Tatum.

Below is a side-by-side comparison of Tatum’s shot distribution for each of his first two seasons.



The two are honestly pretty similar. Of the 2,119 unique clusters in the entire dataset, they stand only one cluster apart. The two are indistinguishable enough that they produce the exact same AST% projection after controlling for age. They do not, however, produce the same results for long-term OBPM. Whereas the model forecasts that based on his rookie season, his OBPM in years 3-4-5 will round out to 1.07, if you input his sophomore year data instead, it projects that he will have an OBPM of… 0.70. So, as it turns out, a year older and a little less efficient isn’t actually the best combination.

Somewhere along the line, though, there appears to be a seam. In a sea of players whose projections rise and fall based on age and production, there are examples to be found of players whose projections outpace the trendline of their age and production. Players like Kyle Kuzma and Spencer Dinwiddie, who saw their OBPM drop this season but had their projections improve — or Davis Bertans, who radically changed his cluster between 2016-17 and 2018-19, precipitating a spike in OBPM 3-4-5 that, while partially explained by his actual OBPM having risen 0.4, dramatically outpaces similar growth spurts that occurred while a player aged two full years.

Those three will be hoping to follow in the footsteps of Jusuf Nurkic, whose similarly unexplained rise in OBPM 3-4-5 a season ago presaged a career-best year on offense in 2018-19.

While there haven’t yet been as many answers here as I’d perhaps hoped yet, it does seem that there is, in fact, *something* there. At minimum, it requires further exploration. In addition to expanding the study to include players at other junctures of their careers, it may be possible to hone in more on what exactly it is about some of the specific clusters that are creating the signal being found. Whether there are certain areas on the court in specific that provide outsize predictive power or if it’s something that has to be taken in its entirety. Whether the clustering can interact with variables like free throw percentage or pre-NBA statistics to forecast how players will grow into their roles over time. And whether it’s a sign of growth for Jayson Tatum to find that pocket just inside the arc or a sign of growth for him to avoid it.