Enhancing Team Performance Through Baseball Analytics

A story about how I played R.B.I. Baseball 2 for the first time in over 30 years, and inspiration struck.

David Bryson
In 1989, two things happened; The Oakland Athletics won the World Series, and Nintendo released R.B.I. Baseball 2. That Christmas I got a brand-new Nintendo Entertainment System, and my parents gave me the game. Playing T-ball at the time, I was obsessed with all things baseball.

I remember R.B.I. Baseball 2 fondly, and I also remember how ridiculously good the Oakland A’s were in the game. They were unbeatable. Mark McGwire, Jose Canseco, Rickey Henderson and Dennis Eckersley were all superstars in the game and went on to have hall of fame careers in Major League Baseball. But there was another member of that 1989 team that would later become famous not for World Series titles, but how he embraced analytics and changed how the game was played. That player was outfielder Billy Beane.

The Idea

The story of Billy Beane and Paul DePodesta with the 2002 Oakland A’s was immortalized in the book Moneyball by Michael Lewis, and in the Hollywood movie adaptation starring Brad Pitt and Jonah Hill. But this story isn’t about that. This story is about how one weekend while on a trip with friends, I played R.B.I. Baseball 2 for the first time in over 30 years, and inspiration struck.

Early Nintendo sports games were notorious for having ridiculously powerful and seemingly invincible players. For anyone that played Techmo Super Bowl, and decided to play with Bo Jackson, you know exactly what I’m talking about. Bo Jackson was unstoppable. The perceived imbalance between star players and ‘regular’ players in those early games made me wonder. Why were the A’s in R.B.I. Baseball 2 so good, or Bo Jackson so unstoppable?

Going down the rabbit hole of sports video game talent weighting and physics was a fun few hours, but ultimately what I learned was that the programmers back then simply based the strength and speed characteristics of players on real world player stats. Not too exciting, but that led to another stroke of inspiration.

The Oakland A's went from winning the World Series in 1989, and having an amazing video game made where they were unbeatable, to being terrible 13 years later. This got me thinking, what if I was able to build a team using Alteryx with data on players from 1989 using the methods described to qualify players in the book Moneyball? What would that team look like, and more importantly, could I figure out how much money the A's might have saved by using the method they applied years later?

The Analysis

In 1989, the Oakland A's had the third highest payroll in Baseball. In 2002, Billy Beane and his team ranked 6th from the bottom in total team payroll. I was able to find this out by leveraging a well-known database in baseball statistics known as the Lahman database. It contains baseball statistics on hitting, fielding, pitching, salaries and more all the way back to the 1800's. And it's free to use... incredible.

This "database" is made up of dozens of .csv files that are interconnected by key variables. Rather than combing through the tables one by one, I wanted to be able to easily navigate these tables, and more importantly, be able to quickly query and scale this massive dataset securely in the cloud. After all, we are about to do some very valuable analytics here, so I want to make sure I'm the only one that can see the results. So, I turned to our trusted partner, Snowflake. They hooked us up with the platform I needed to manage my data and run my analysis with lightning-fast speed and rock-solid security. Thanks to Snowflake, I hit the ground running!

With the data now loaded in Snowflake, I set out to calculate what the A's team salary total would have been in 1989 if they were the 20th richest team (of 26 teams). That was easy to find with a little joining and pivoting. Here are the bottom team salaries from 1989:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Year</th>
<th>Team</th>
<th>Sum_salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>1989</td>
<td>MIL</td>
<td>11596000</td>
</tr>
</tbody>
</table>

The 20th richest team in 1989 spent $11,596,000 in payroll. For reference, the Oakland A's payroll that year was $18,688,460. The team with the highest payroll in 1989 was the New York Mets at $25,063,737, and they didn’t even make the playoffs. Sorry Mets fans.

We now know that our fictitious team, let’s call them the ‘Alteryx A's’, has a little over $11 Million dollars to spend. Now it’s time to calculate the statistics used in Moneyball to find out the offensive efficiency of all players in 1989. To do that, we need to know the formula that was used to calculate this statistic.

---

ii The Lahman Database. Sean Lahman. seanlahman.com/download-baseball-database/
The statistic that Paul DePodesta (the analyst for the A’s) was primarily looking at was something called On-base plus slugging, or OPS. OPS is a Baseball statistic which is calculated as the sum of a player’s on-base percentage and slugging percentage. Both the ability of a player to get on base and to hit for power, two important hitting skills, are represented, making it an effective way of measuring the offensive worth of a player. An OPS of .900 or higher in Major League Baseball puts the player in the upper echelon of offensive ability. Typically, the league leader in OPS will hover near the 1.000 mark.

All the raw ingredients to calculate OPS are in the Lahman database, and the formula tool makes calculating OPS a breeze. Below you can see a snapshot of the formula tool configuration to calculate OPS.

It is amazing that it takes just one tool in Alteryx to calculate such an impactful Baseball statistic. I bet Paul DePodesta would have loved to have had Alteryx back then instead of Excel.
Using the Alteryx Optimization Tool

We now know our salary cap ($11 Million) and the OPS and salary of every player in 1989. So next we need:

9 total players optimized for the absolute best OPS for each position whose total salary cannot exceed $11 Million dollars.

This sounds like a job for Alteryx, and the Optimization tool.

If you've never heard of the Optimization tool, then you are in for a treat. The Optimization Tool is a member of the Prescriptive Tools (included with the Predictive Tools installation) and allows you to solve optimization problems. Mathematical Optimization is the selection of the best possible option(s), given a set of alternatives and a selection criterion. Optimization is used for a wide variety of applications across many different industries.

One common example of a problem that can be solved with optimization is the Knapsack Problem, where given a collection of items (each with a weight and value) optimization determines which combination of items you can take in your knapsack, maximizing value without going over the maximum weight you are able to carry.

Configuring the tool can be a bit of a chore, but there are a ton of examples and help docs out there.¹

Using the Optimization tool, I was able to feed in the criteria I wanted, and the tool produces a brand-new roster for the 1989 Alteryx A's. This is the roster:

1 You can find a really great writeup here: community.alteryx.com/t5/Alteryx-Designer-Desktop-Knowledge-Base/Tool-Mastery-Optimization/ta-p/306117

An Alteryx created roster for the 1989 Alteryx A's
Notice how good the OPS scores are here, all over the .900 mark, and we get that for $7 Million dollars less than what the A’s originally paid for their 1989 team. And we got some pretty good players out of it too:

<table>
<thead>
<tr>
<th>Barry Larkin</th>
<th>Rhyne Sanburg</th>
<th>Will Clark</th>
</tr>
</thead>
<tbody>
<tr>
<td>was a 12x All Star winning a World Series, NL MVP Award, 9 Golden Gloves and over 2,300 career hits, most of this success happening in the 1990’s.</td>
<td>had a Hall of Fame career with the Chicago Cubs with 10 All Star appearances, 9 Golden Gloves and was the home run leader in 1990.</td>
<td>was a 6 time All Star and was the NLCS MVP in 1989. He also turned in over 2,000 career hits before retiring in 2000.</td>
</tr>
</tbody>
</table>

This shows you how powerful this analysis can be. All these players (and this is just three of nine players) were at the start of their careers in 1989 and went on to have stellar careers in the 1990’s. Finding diamonds in the rough is exactly what the Moneyball analysis is all about, and this proves that. Had the real Oakland A’s had this team through the 1990’s, they would have gotten in early on some serious talent without the price tag.

Now that we’ve run this amazing analysis, we must make sure that our super-secret player roster doesn’t get in the wrong hands. Data governance and security are critical in Snowflake for securely storing data and managing data access and privacy—without sacrificing collaboration. By pushing data processing and output to Snowflake, our player analytics data doesn’t leave the Data Cloud and our secret player recipe is safe from competition.

The Analysis

With this fun example now complete, start to think about how this relates to your business, and the analytics you do every day. Wouldn’t it be amazing if you could apply this type of optimization, integrated seamlessly with your scalable cloud data infrastructure, for your business operations? Knowing how to select a set of options that are optimized for quality and capped at a certain cost is something every business wants to know. Imagine being able to optimize your product offerings based on quality and customer satisfaction metrics while keeping total costs low.

In this project, Alteryx and Snowflake were an all-Star team that knocked it out of the park! With Alteryx’s Analytics Platform and the Snowflake Data Cloud, you’ll have an unbeatable combination of powerful optimization functions, governed and secure data management and scalable, lightning-fast performance. Together, Alteryx and Snowflake are a winning combination providing an unparalleled solution for organizations looking to optimize their data-driven workflows and achieve maximum efficiency.

Try this for yourself. Use this workflow and see how we did the analysis I performed above. You won’t be disappointed.
Fanalytics by Alteryx is our "highlight reel" featuring elite athletes and teams who use analytics to improve performance, perfect the fan experience, and drive the business strategy of the game.

To compete and win, athletes and teams must use analytics for a competitive edge, and businesses can learn from their approach.

Fanalytics. Analytics for the win.

Learn More