



New York Mets

Midseason Talent Acquisition Strategy

Prepared for Don Wedding, GM
July 15, 2018

New York Mets Analytics

Alexander Booth, Justin Benson, Noah Lieberman, Michael Pallante,
Thomas Popeck Spiller



Table of Contents

Executive Summary.....	2
Business Objectives Explained.....	3
Data Overview.....	4
Modeling Approach.....	6
Early Findings.....	8
Next Steps & Project Schedule.....	14
References.....	15



Executive Summary

Dear Prof. Wedding,

Congratulations on your appointment as New York Mets General Manager. On behalf of the Analytics department, the purpose of this prospectus is to outline recommendations for developing the best possible on-field product for the Mets, while maintaining a healthy farm system. As of the halfway mark in the 2018 season - which occurred last week upon completion of our 81st game - we have a record of 33-48 and lie 15 games behind the Atlanta Braves in the National League East Division. Despite a strong start to the year, the season has overall been a great disappointment and we have the fewest wins in the National League. In addition, division rivals Atlanta and Philadelphia have bested their own expectations, while we expect Washington to improve upon their first half results. To correct course for our organization, we believe it is time to sell off underperforming veteran assets at the Major League level and acquire promising young talent throughout the minor league ranks.

As you are aware, the Seattle Mariners have expressed interest in several players including infielders Asdrubal Cabrera and Todd Frazier, outfielders Yoenis Cespedes, Jose Bautista and Jay Bruce, and pitchers Noah Syndergaard, Jacob DeGrom, and Jeurys Familia. The Mariners are in a close race with the Houston Astros for the American League West to avoid a one-game wildcard playoff and would consider trading minor league prospects.

We plan to develop a robust predictive model to forecast the likelihood of minor league talent within the Mariners organization to reach the Major Leagues as well as the on-field value the players will deliver. We may utilize this model and compare results relative to Seattle's top prospects rankings to find opportunities in which we may exploit potential value. Upon acquiring the talent, future New York Mets on-field talent will be optimized, ultimately leading to increased ticket sales and thus revenue for the organization as fan engagement will increase based on future expectations.

In addition to our internal databases, there are vast amounts of supplemental sources publically available to be used. Using these data sources, we will build the models that can be utilized in a dashboard to explore trade possibilities, along with a mobile app to view insights and results on the go.

We look forward to working with you on this exciting journey. Sincerely,

New York Mets Analytics

Alexander Booth

Michael Pallante

Justin Benson

Thomas Popeck Spiller

Noah Lieberman



Business Objectives Explained

To enable our midseason talent acquisition strategy, there are three key business objectives the analytics department aims to achieve. These objectives are outlined in the following table along with their respective deliverables and what defines success. A further explanation of the planning and features for each objective is explained in greater detail in subsequent sections of this report.

NYM Analytics Midseason Talent Acquisition Objectives:

Objective	Deliverables	What Defines Success
Prospect Value Projections*	<ul style="list-style-type: none">· Robust data infrastructure of historical minor league performance· Predictive models for likelihood to reach majors (AKA 'Make it') and projected career value· Final report and trade recommendations	<ul style="list-style-type: none">· Successful acquisition and organization of data model· Accurate testing of model within agreed upon error bounds
Trade Scenario Dashboard*	<ul style="list-style-type: none">· Dashboard that incorporates current rosters, minor league projections, and can facilitate what-if trade scenarios	<ul style="list-style-type: none">· Usability and successful sign-off from the GM
Mobile Application	<ul style="list-style-type: none">· Mobile app that enables the analytics team and the GM to visualize results as well as the dashboard on the go	<ul style="list-style-type: none">· Usability and successful sign-off from the GM

**Note: to be tailored to facilitate trades with the Mariners' organization, with the ability to go broader for other clubs*

 Key Project Deliverable



Midseason Talent Acquisition Strategy

Data Overview

In order to deliver on the above items, we plan to develop models based upon the following data sources:

Source Name	Description	Location	Acquisition
Fangraphs	Minor League Player batting and pitching data from 2006-Current	https://www.fangraphs.com/minorleaders.aspx	Download
Baseball Reference	Minor League Player batting and pitching data from 1977 - 2017	https://www.baseball-reference.com/register/	Web scraping
The Baseball Cube	Major League Data Player batting, fielding and pitching data from 1865 - 2017 & Minor League batting and pitching data from 1977- 2017	http://www.thebaseballcube.com	Download
Lahmans' MLB Database	Major League Data Player batting, fielding and pitching data from 1865 - 2017	http://www.seanlahman.com/baseball-archive/statistics/	Download
The Baseball Prospectus	Scouting reports for recent prospects	https://legacy.baseballprospectus.com/prospects/eyewitness.php	Web scraping
Sentiment Analysis - word list	List of words implying positive/negative sentiment analysis, will be augmented to include "baseball terms"	https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html	Download

Primary Data Source

Using the Baseball Cube dataset as our primary source, as it contains the primary playerID key, we will first construct a data set based on joining together the disparate data from above to create a single dataset from which to model. We will utilize the additional sources (Fangraphs, Baseball Reference, and Lahmans') as model validation and as a robustness check.

We will then define a player as having "made it" as any player that stayed in the majors for any significant time (currently defined at more than 3 seasons, although we may change this based on data exploration or executive directive). This will create a binary variable which will be defined as 1 for "made it" and 0



Midseason Talent Acquisition Strategy

for “Didn’t Make it”. Additionally, we will cluster on MLB Players stats to generate a profile for the type of player they are forecasted to be. We will use the following standard WAR grouping to assess player’s forecasted role within the MLB.

Projected Value	Associated WAR (Career)
Scrub	<0
Replacement Player	0-1
Role Player	1-4
Solid Starter	4-6
Good Player	6-8
All-Star	8-10
Superstar	10-12
MVP	12-16+

Additionally, we will perform sentiment analysis on scouting reports by augmenting traditional positive/negative word lists with baseball terminology. This data will be used to augment final recommendations, and depending on the volume of scouting reports available may not affect the predictive model. To do this, we will scrape the data from scouting report websites and parse out the free text to get a count for each word across a population of top prospects. We will then give a score of positive or negative words for each prospect as available.

As a note, while we are not focusing on college ball data, we will use some level of pre-minor league information to inform on this space. As we are close to the trade deadline and this data, while available, is sparse and inconsistent we will elect not to use it at this time.



Modeling Approach

The intent of this analysis is to identify minor league players that are valuable pieces to improve the future standing of the Mets. We will identify players likely to stick in the major leagues and model their expected WAR values, based on their past play in the minors, physical attributes, and scouting reports. We will generate several models to predict sticking ability and expected WAR, using more traditional models such as linear and logistic models, as well as more advanced algorithms such as support vector machines (SVM) and gradient boosting.

Among the suite of models built, we will use the Blinder-Oaxaca decomposition to address the Mets' development program as it compares to the league overall. Building models separately for the team's prospects and using the Oaxaca method we can identify players that will be more likely to succeed in the Mets' system. By trying to trade for players more likely to succeed under the Mets' development program we can gain extra value out of the deal.

Using a variety of tools to project prospect values and success will help the analytics team to build a robust measure of value for the prospects, tailored to different situations.

Given historical difficulties in predicting major league success based on minor league competition, we are making a strong effort to minimize the costs associated with this analysis. The lowered cost of model development will aid the team in better taking advantage of their prospect pool. To this end we will be primarily using open source and already-licensed tools such as R, Python, and Tableau (already in use within the Mets organization).

The table on the next page shows the software and packages we plan to use for this analysis. We will allow for other open-source packages to be included, but will not be requesting budget for any additional software.



Midseason Talent Acquisition Strategy

Activity	Tool	Description
Data Wrangling	Python	Common open source software with packages for data manipulation and extensive community support. - Pandas, numpy
	R/R Studio	Common open source software with packages for data scraping, manipulation, and extensive community support. - rvest, tidyverse, ggplot2
Model Development	Python	Common open source software with packages for data modeling and extensive community support. - SciKit-Learn, spAcy
	R/R Studio	Common open source software with packages for data modeling and extensive community support. - tidyverse, caret, e1071, rpart, tm
	SAS	Proprietary software already embedded in the organization.
Reporting	R Shiny	Online dashboard package from R Studio.
	Tableau	Proprietary software already embedded in the organization.



Early Findings

We began processing the data and across the several datasets there is consistency in that 13% of minor league players have at least one at bat or inning pitched in the major leagues. This is consistent with the findings of other baseball analysts, suggesting good quality data. There are minor discrepancies between the data, due in part to the length of history. As the data goes further back, different sources have different methods of collecting the historical data. The Baseball Cube data identified a total of 64,007 minor league players from 1977-2017, but Baseball Reference identified 63,601.

It is imperative that we pay attention to facts and trends regarding minor leaguer and draftee success rates in the major leagues. This will assist us in narrowing down prospects pools to find a trade that will help shape the long-term future of our franchise.

First, we must note the probability of baseball players playing college and professional baseball. Approximately 5.6% of high school baseball players will go on to play baseball at an NCAA institution and approximately 0.5% of them will be drafted by an MLB team (High School Baseball Web). If they go the college route, approximately 10.5% of NCAA baseball players will get drafted by an MLB team (High School Baseball Web). Recent draft history shows that majority of the players drafted from NCAA schools are Division I student athletes. There were 1,215 draft picks in the 2017 draft and 735 of them were Division I players, as opposed to only 73 Division II players and only 12 Division III players (NCAA Research, 2018). We can derive from all of this information the incredibly long odds of players that are good enough to even be drafted into the league.

Taking it a step further, we can study the rates of players reaching the major league level by their respective draft rounds. The estimated percentage of MLB draft picks who reach the major leagues are as follows: 66% of first round picks, 49% of second round picks, 32% of players from rounds 3-5, 20% of players from rounds 6-10, 11% of players from rounds 11-20, and 7% of players from rounds 21-40 (Rosenbaum, 2012). Below are some additional statistics of relevance from an average MLB draft (Source: OKennedy, 2013):

- 34.2% of 1,560 players chosen in the first ten rounds make it to major leagues
- Over 80% of players selected in the first round make it to the major leagues
- 44% of players chosen from the supplemental first round through the fourth round make it to the majors
- 21% of 900 players chosen from 5th through 10th rounds makes it to majors
- 41% of all players & 64% of first rounders who have made it have WAR of 1.0+
- Over half of all first round selections have a WAR of 1.0 or better
- 16.5% of 1s to 4th rounders and 6.4% of 5th- 10th rounders have a WAR of 1.0+
- 107 players, or 3.5 per club, selected in the first ten rounds make it to the majors from each draft



Midseason Talent Acquisition Strategy

- 43.8 players, or just 1.5 per club, selected in the first ten rounds has a WAR of 1.0 or more
- 10 players per draft have produced 10+WAR and 20 players per draft produced 5+ WAR thus far
- 50 of the 54 players (57 counting 11-20) with 10+ WAR were selected in the first four rounds
- Over 40% of players chosen after the first round have a negative WAR

None of the above information means we cannot find talent that has been previously drafted in deeper rounds, but rather that our odds in finding major league talent decreases as we look at prospects that were once drafted later in rounds.

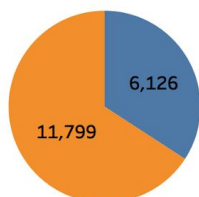
It is believed that the gap in talent from the high minor leagues to the MLB is increasing, meaning it has taken prospects longer on average to make it to the MLB level. Of course, this varies with some prospects, as we see a bunch that are ready in shorter times. However, on average it takes most players 4-6 years in the minor leagues to make it to the MLB (Gaines, 2013). It is normal to see approximately only 15% of a draft class in the major leagues 3 years after their respective draft (Gaines, 2013). Prospects generally spend 1-2 years at each level of the minor leagues and if they have not reached the major leagues after seven years, it is highly unlikely that they will ever reach the MLB. We need to pay attention to the number of years prospects have spent in the minors and at what specific levels. It is our goal to net major league talent in whatever trade we make. The closer prospects are to reaching the major league level, the harder they will be to acquire in a deal and the higher the cost we will have to pay to get them.

The chart below displays the percentage of minor league players that will ever reach the major leagues as well as the level they finished at if they did not make the major leagues:



Midseason Talent Acquisition Strategy

Drafted Players



Signed Players



(Source of information used in this chart: Owens, 2017)

Only 11.2% of drafted players have reached the MLB level (Owens, 2017). This is an incredibly low rate of players that have reached this high of a level. While this differs from our above 13%, it reflects minor differences in data sources and definitions of MLB status. It is easy to see how difficult it is to advance through the levels of the minor leagues before even reaching the majors. It is our job to find major league talent and value in each level of professional baseball.

Next, we examined the Baseball Cube data to see how the number of unique players changed across the various levels of the minor leagues, and across years. The below table depicts the number of unique players who played in each professional league and level between 1977-2017:

Level	Unique Players
MLB	19,019
All Minors	64,007
AAA	16,247
AA	20,679
A+	29,745
A	30,829
A-	26,344
RK	43,764

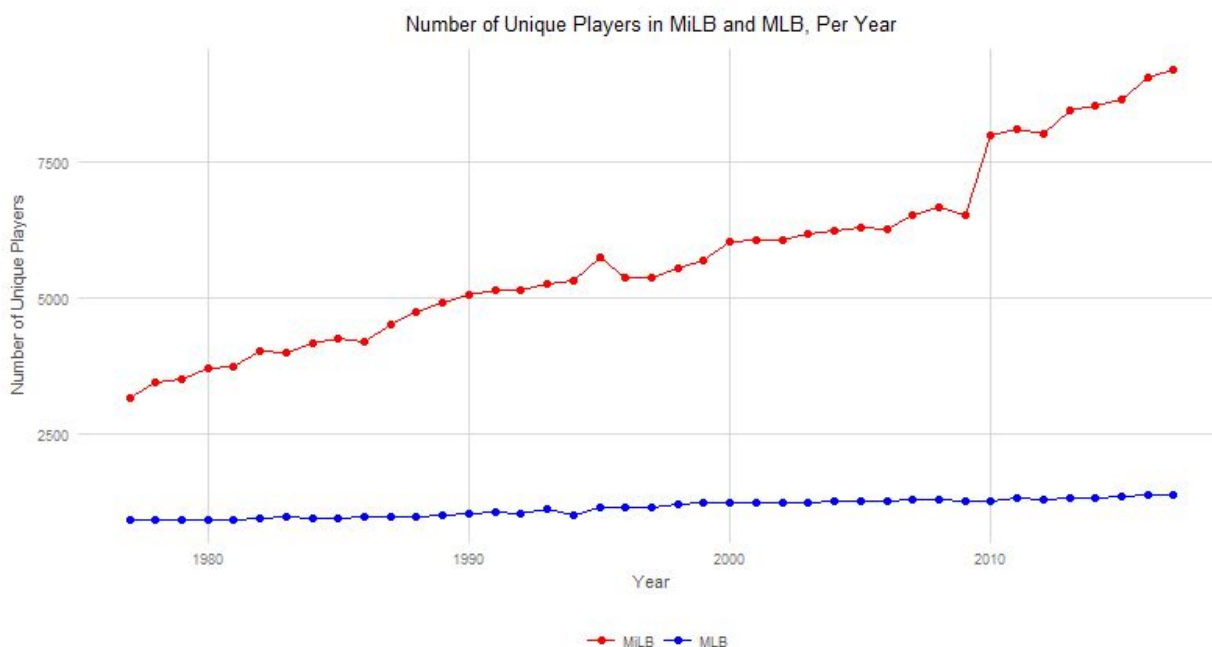


Midseason Talent Acquisition Strategy

As expected, a much larger population passed through some level of the minor leagues than the population of unique players that played in the major leagues. Surprising to us at this early stage, there were actually more unique players in the MLB across the time period examined than unique players who appeared in AAA, the highest level of the minor leagues. This implies that perhaps there is a large proportion of minor league players that skip AAA in their development.

Additionally, the largest number of unique players played in the single A levels as well as Rookie ball. Because of the large number of players in these lower levels, determining who has the highest potential to reach the majors can provide the Mets with a competitive advantage.

We continued breaking down this data by viewing how the number of unique players per league varied across years. That graphic is depicted below:

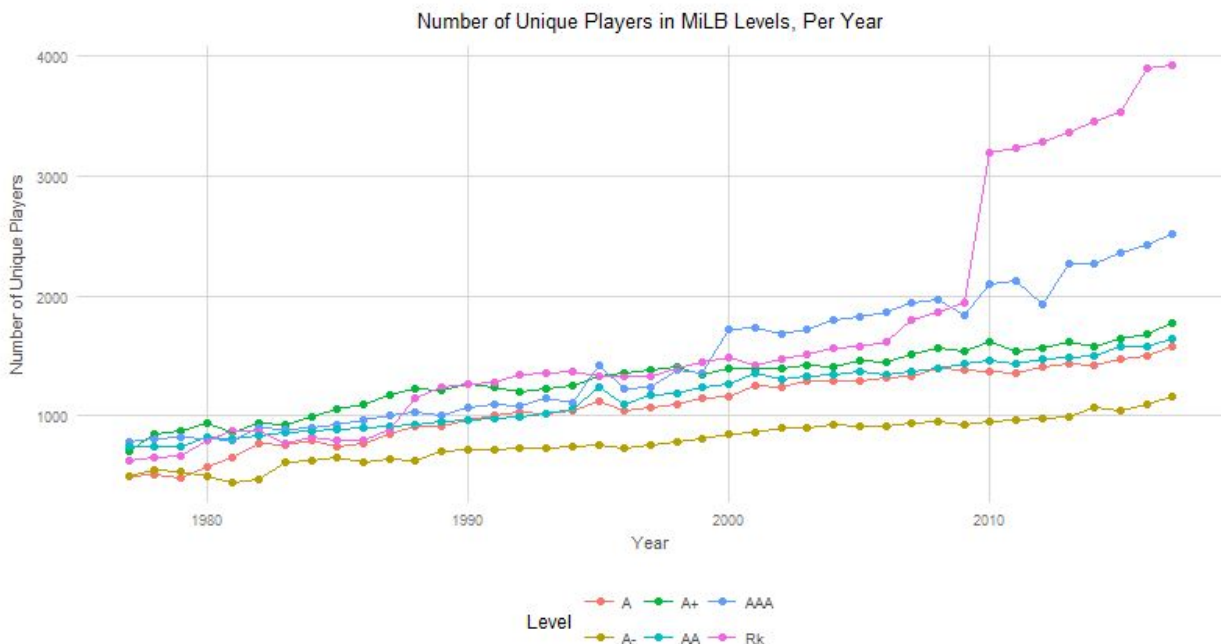


The rate that the MiLB has added unique players per year far exceed that of the MLB. This makes sense, as each MLB team has a set roster size, that only expands in September, or when players are injured. The biggest jump in the MLB actually happened when expansion teams were added to the league in the '90s. On the other hand, a large spike from 2009-2010 increased the number of unique players in the MiLB by approximately 1,000 players. This has made the difference in unique players between the two leagues more stark in recent years.



Midseason Talent Acquisition Strategy

We then broke down the number of unique players in the MiLB by the levels enumerated in the table above. This graph is below, and seems to provide an explanation for the spike in 2009-2010.



All of the levels increased at the same rate from the start of the timeframe examined, 1977, until 1999, when AAA spiked. However, the levels continued at the same rate again until 2009, when the number of players in rookie ball skyrocketed. In 2009 there were 17 different rookie ball leagues, with a total of 200 teams, in 2010 the data included 18 leagues and 234 teams by adding the Dominican Summer League (DSL) data. While the DSL league has existed since the 1980's, it has been problematic to get accurate statistics until recently.

Finally, as we have begun to process through the scouting report data, it was appropriate to develop an initial word cloud to help visualize the type of terminology the scouts are using to describe prospects. This will help not only familiarize the data set, but also give the analysts an opportunity to see the most used terms that we should look out for when coding for sentiment. Non-applicable words such as “the, it, a” etc have been excluded and only words which occur 100 times or more have been included on the next page.



Midseason Talent Acquisition Strategy



Words such as “Speed, strength, arm, hands, quick, frame, plus, and build” jump out immediately as characteristics that scouts are looking for. This indicates that it may make sense to also add body type metrics such as BMI into our modelling to help include some of the scouting knowledge that normal stats will miss. Additionally, the prevalence of words like “average” and “aboveaverage” indicate that more logical processing may be needed to determine what feature is being described this way.



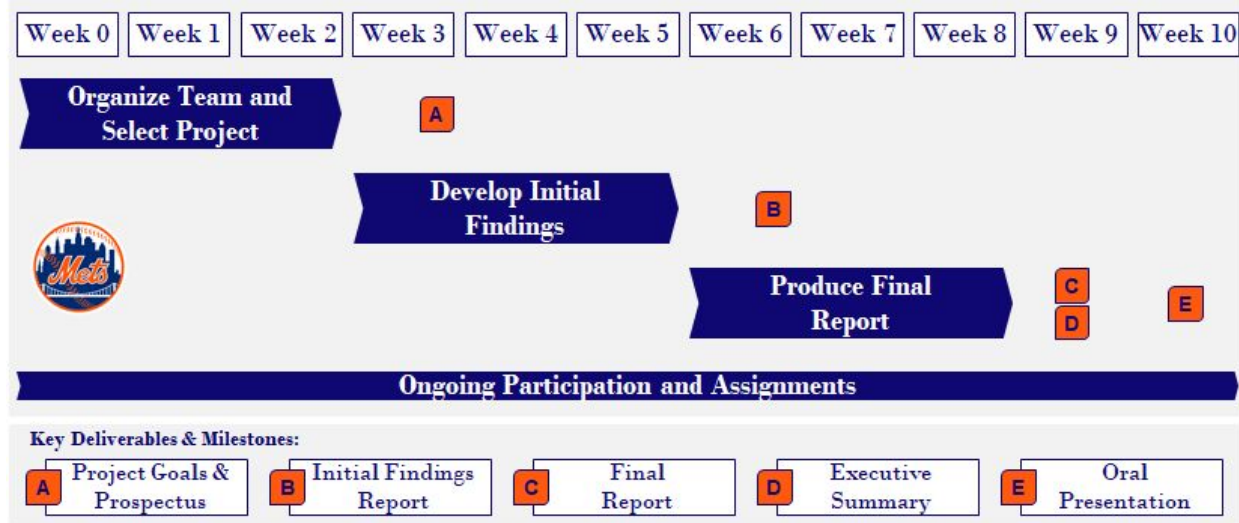
Midseason Talent Acquisition Strategy

Next Steps & Project Schedule

The Midseason Talent Acquisition Strategy has a ten week timeline as shown in the chart below. There are three key phases; team organization, initial findings, and final report production. The report you are currently reviewing - as it is now the end of week 3 - is the culmination of our first phase to organize and select a project topic. Upon approval of our strategy, we will jump into modeling analysis in weeks 4-6 to produce a report of initial findings. In weeks 6-9 we will focus on finalizing the report output and generating key suggestions and prospect trade scenarios from the model. The key deliverables from this final phase will be aligned to the business objectives. The project will conclude with an oral report in week 10.

In addition to the 3 main phases of work, the analytics team will produce ongoing status reports as requested along with ad-hoc project requests.

NYM Analytics Midseason Talent Acquisition Timeline:





References

Baseball Reference. (n.d.). MLB Stats, Scores, History, & Records. Retrieved July 9, 2018, from <https://www.baseball-reference.com/>

Fangraphs Baseball. (n.d.). Baseball Statistics and Analysis | FanGraphs Baseball. Retrieved July 9, 2018, from <https://www.fangraphs.com/>

Gaines, C. (2013, June 7). Most Baseball Draft Picks Will Still Be In The Minors Four Years From Now. Retrieved July 7, 2018, from <http://www.businessinsider.com/chart-how-long-it-takes-a-draft-pick-to-reach-major-league-baseball-2013-6>

High School Baseball Web. (n.d.). Probability Of Playing College and Professional Baseball. Retrieved July 9, 2018, from <http://www.hsbaseballweb.com/probability.htm>

Major League Baseball. (n.d.). The Official Site of Major League Baseball. Retrieved July 9, 2018, from <https://www.mlb.com/>

Mitchell, C. (2014, December 30). KATOH: Forecasting Major League Hitting with Minor League Stats. Retrieved July 9, 2018, from <https://www.fangraphs.com/tht/katoh-forecasting-a-hitters-major-league-performance-with-minor-league-stats/>

NCAA Research. (2018, April 20). Estimated probability of competing in professional athletics. Retrieved July 9, 2018, from <http://www.ncaa.org/about/resources/research/estimated-probability-competing-professional-athletics>

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693-709. Retrieved July 9, 2018, from https://www.jstor.org/stable/2525981?seq=1#page_scan_tab_contents.

OKennedy, P. (2013, March 5). What are the odds of making it to the major leagues? Retrieved July 9, 2018, from <https://www.blessyouboys.com/2013/3/5/3977782/what-are-the-odds-of-making-it-to-the-major-leagues>



Midseason Talent Acquisition Strategy

Owens, F. (2017, June 26). What percentage of minor league baseball players will ever reach the majors? Retrieved July 9, 2018, from <https://www.quora.com/What-percentage-of-minor-league-baseball-players-will-ever-reach-the-majors>

Rosenbaum, M. (2012, June 12). Examining the Percentage of MLB Draft Picks Who Reach the Major Leagues. Retrieved July 9, 2018, from <https://bleacherreport.com/articles/1219356-examining-the-percentage-of-mlb-draft-picks-that-reach-the-major-leagues#slide0>

The Baseball Cube. (n.d.). The Baseball Cube - MLB, Minor League, College Statistics, Data and the Draft. Retrieved July 9, 2018, from <http://www.thebaseballcube.com/>