ISM 6419 Summer 2021 Final Project Report Draft

Mark Morey U36467084

**Introduction:**

In the world of professional baseball, the importance of analytics has become more and more evident with each passing year. One could say the popularity of analytics in baseball started after 2003 when Michael Lewis published the book Moneyball: The art of winning an unfair game. The book follows the Oakland athletics on how they used statistics, and analytics to keep up with teams with much higher payrolls than themselves. This is the background behind why I chose the to cover the topic of baseball analytics. My favorite baseball team is the Tampa Bay Rays play in a division in baseball where some teams have a much higher payroll to pay their players. The Rays are widely regarded as being great at applying analytics to help close the gap in wins and talent against their opponents with a larger payroll. In this project I played the role of an analyst in the professional baseball field, exploring player performance and salary data from the 2020 baseball season where the Tampa Bay Rays placed 2nd overall in the MLB to report back to my manager to recap the season / strategically plan for the next one. (Lost in the 2020 world series to the LA Dodgers sadly.) Sources I utilized were; <https://www.baseball-reference.com> for player performance statistics in the 2020 year, and <https://www.spotrac.com/mlb/> for the salaries of the players in the American League East division of the MLB. These sources allowed me to explore, and research the answer to a few key research questions. Questions I looked to visualize and be able to answer for my boss is how does our payroll stack up against the teams in our division? How are our offensive players performing and how does that compare by position to the other teams in our division? How valuable are our players and are they performing well enough to warrant the money we are giving them with our limited payroll? (Here I am looking to say if we should move on from expensive players or not).

**Methodology:**

To collect data for this project I utilized the library RVEST within the programming language R, to web scrape html tables from 2 sources: <https://www.spotrac.com/mlb/>, and <https://www.baseball-reference.com>. Baseball reference provides numerous performance indicators for players, whether it be hits/walks for batters, or strikeouts/Innings pitched for pitchers. Batting performance indicators represent the offensive side of the team, where you walk or get hits to get on base to score. Pitching performance indicators represent a portion of the defensive side of the team, where pitchers perform in innings at a time to strike out opposing batters and prevent runs from being scored by the other team. This website had an extensive list of player performance data I was able to pair with the salary data from spotrac. Professional athletes are only worth the salary you pay them if they perform well and win your organization games. Some players are paid far more than other players because they are perceived to be that valuable. The New York Yankees have a $308 million payroll to pay their players vs the Tampa Bay Rays who had $90 million in 2020. When there is that much money being paid to these players, their performance needs to be watched closely. To wrangle the 2 data sources together for the 5 teams in this division (10 links in total, for 5 tables) I utilized R’s Tidyverse package for data cleaning and used a union for the player data (Batters & Pitchers csv documents, see photo below) and joined the player data on player name to the salary data which was a union of the 5 teams’ salaries for 2020.

Player Data Tables:

Table

Description automatically generated

Salary Data:

Graphical user interface, table

Description automatically generated

Data Source Joins on Player’s Name:

Shape, rectangle

Description automatically generated with medium confidence

Graphical user interface, application

Description automatically generated

**Analysis:**

*Question:* How does our payroll stack up against the teams in our division?

Chart, pie chart

Description automatically generated

Above we are analyzing the Tampa Bay Rays payroll, allowing us to see which players are making up the bulk of our payroll before we start to compare across teams. Blake Snell, Charlie Morton, and Kevin Kiermaier are our 3 highest paid players, contributing **55.92%** to the payroll. If we are in a cost cutting, or potential trade scenario I would recommend the Rays look to move 1 of these 3 players to help free up some payroll space.

The Tampa Bay Rays payroll sits at the 2nd lowest out of the 5 teams in the AL east division (Per Team level of Salary Drill Down sheet, see screenshots that follow). The highest payroll in the division is the New York Yankees. Both teams were playoff teams in the 2020 MLB season. After drilling down from the team level to the position level we see the position earning the highest salary for every team is SP (Starting pitcher). We start by comparing the players who are earning the highest salary across every team to the Rays highest paid player Charlie Morton (SP) at $15 million. Gerrit Cole (SP) for the New York Yankees is the highest paid player in the division with over twice the amount of Charlie Morton’s salary at $36 Million. The Rays highest paid player has the 2nd lowest salary for a highest paid player on a team in the division, beating only Alex Cobb of the Baltimore Orioles by $1 Million. After reviewing the payrolls of the 5 teams in the division it is clear the Rays cannot compete on payroll alone (paying the best players the most to come to our team). The Rays must focus their business model of efficient, high performing and relatively cheap players. The Rays do this well by relying heavily on young players, who on their rookie deals do not require a high salary. The average age of Rays player in the 2020 MLB season was 27.79 years old.

Average Age Visual

Table

Description automatically generated

Salary Screenshots:

By Team:

Chart, bar chart

Description automatically generated

By Position: (Focused on New York and Rays teams due to width of visual)

Chart

Description automatically generated

By Player: (Depicting the highest paid players for the Yankees and Rays due to visual width)

Chart

Description automatically generated

*Question:* How are our offensive players performing and how does that compare by position to the other teams in our division?

On the offensive side of the ball, the Tampa Bay Rays were last place in total bases collectively as a team in the AL east division for the 2020 MLB year at 839 total bases. The total bases metric represents anytime an offensive players get a hit or walk, giving them the potential to score a run in that inning. The positions which contributed the most to the Rays total bases team total were 2B (second baseman), SS (shortstop), and players who are OF (outfielders which can play all positions in the outfield LF, RF, and CF). Our 2B position, played by Brandon Lowe, produced the second highest total bases count in division, at 107 total bases. This was second only to DJ Lemaheiu of the New York Yankees. The outfield position for the Tampa Bay Rays outpaced all other teams due to the Rays having 3 different players who play multiple outfield positions (100 total bases vs Toronto’s 36 next closest). These 3 players, appear to be very valuable and outpace the entire division: Randy Arrozarena, Manuel Margot, and Brett Phillipps. The Rays are known around the league for having numerous players who can player multiple positions. Finally, the shortstop Willy Adames comes in second place in the division for total bases by a shortstop. Second only to Xander Bogaretes of the Red Sox.

By Team Total Bases:

Chart, bar chart

Description automatically generated

By Position Total Bases: (Condensed due to width)

Chart, bar chart, histogram

Description automatically generated

*Question:* How valuable are our players and are they performing well enough to warrant the money we are giving them with our limited payroll?

When evaluating the value of an offensive player for the Rays I have chosen to focus on the metric Bases per 1000 salary dollars. As mentioned earlier, the emphasis for the Rays is high performance from less expensive players due to their lower payroll compared to the other teams in the division. Those who did this best for the Rays are Willy Adames (SS), Joey Wendle (IF, stands for infielder since he can play multiple positions), and Austin Meadows (LF). These are the offensive players the most valuable to the Rays. Meaning, these 3 players produce the most offense for the Rays while costing the team less in payroll salary. These 3 players need to be retained for the next season considering their value to the team.

Offensive Players:

Chart, bar chart

Description automatically generated

On the defensive side of the game, we focus on the WHIP metric, and its relationship with payroll. The WHIP metric is Walks + Hits / by Inning pitched, this measures how many opposing offensive players the pitchers allow on base per inning they are in the game pitching. We want the pitcher’s WHIP to be as low as possible. While also considering the salary of a player, our most valuable relief pitchers (RP) and Closing pitchers (CL) are Nick Andersen, Aaron Loup and Aaron Slegers. Of the starting pitchers (who are expected to pitch more innings than the relievers/closers), the most valuable was Josh Fleming when considering how much less he mad than the other starting pitchers with low WHIPs. These 4 pitchers should be retained next season after considering their value to the team.

Defensive Players (Pitchers):

Chart

Description automatically generated

**Conclusion:**

Here are the key takeaways for the Tampa Bay Rays organization after reviewing the business questions. Our payroll is the 4th lowest in the division, meaning we will need to always have a focus on player performance of young, and cheap players. Younger players on earlier rookie contracts will give us the edge against our higher paying divisional opponents such as the New York Yankees. Our top performing offensive players in terms of total bases are in the position groups; 2B, SS, and Outfield. We outperform most of the divisions in these positions, and this should be the focus for offensive performance in the organization going into the next season. The most valuable offensive players for us are those who are producing on offense in total bases and following our need for cheaper players. These offensive players are Joey Wendle, Willy Adames and Austin Meadows. The most valuable defensive players who put the least number of opponents on base while being cheaper players are Nick Andersen, Aaron Slegers, Aaron Loup and Josh Fleming. Going forward into the next season, with the strategy of focusing on these cheaper, young, and well performing players will help the Rays organization stay competitive within the American League East Division of the MLB.