

Week 10-11 Report

Data Cleaning

For our project, we utilized mathematical reasoning to determine the optimal salary for players that are 2022 free agents. We did this by using the 2022 Free Agent list from baseballprospectus.com. We populated hitters and (starting) pitchers who were already free agents into a Google Sheet and removed all players within the list who were not free agents. We separated hitters and pitchers to ensure players in each list were comparable and made sure each player had enough at bats/innings pitched to be starter level players. For pitchers, we chose to use starting pitchers only because they have more similar roles and therefore more comparable stats than relief pitchers since relievers can be used as lefty specialists, mop-up men, middle relievers, setup men, closers and other situational faceoffs. We also removed players with team/player option contracts since we do not know if they will be free agents in 2022 or not.

The metrics that we utilized to evaluate the hitters were wOBA, wRC+, BsR, and Def. The metrics that we utilized for the pitchers were FIP, SO, BB%, and Hard %.

We ranked the players in each list using the formulas below:

$$\text{Hitters Formula: } \frac{7}{20}(\text{wOBA}) + \frac{7}{20}(\text{wRC+}) + \frac{3}{20}(\text{BsR}) + \frac{3}{20}(\text{Def})$$

$$\text{Pitchers Formula: } \frac{2}{5}(\text{FIP}) + \frac{1}{5}(\text{SO}) + \frac{1}{5}(\text{BB}\%) + \frac{1}{5}(\text{Hard \%})$$

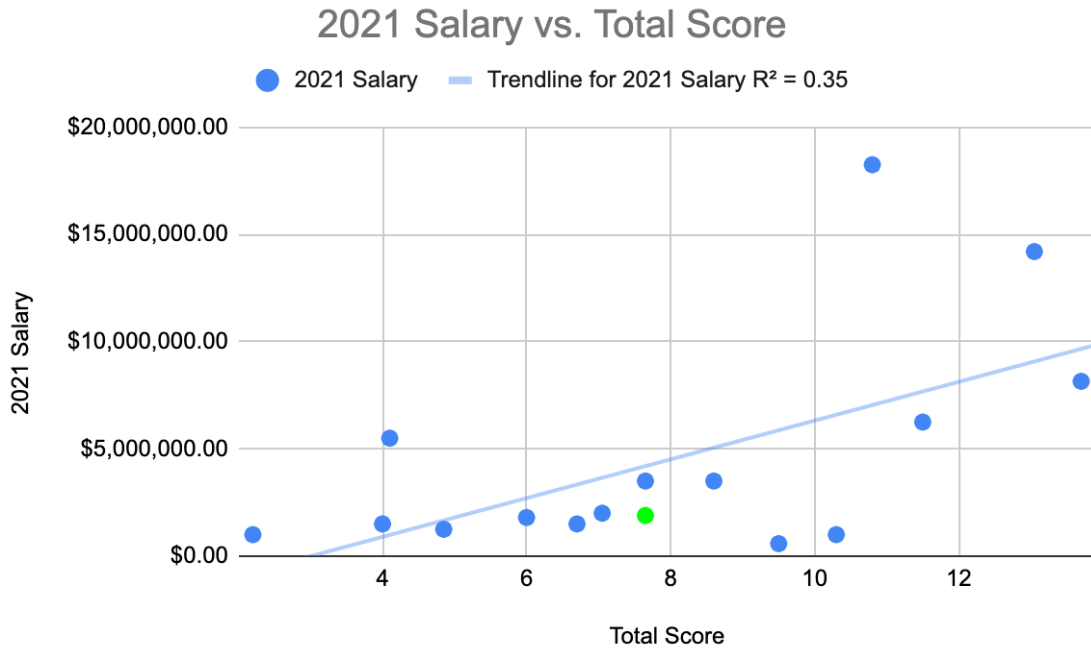
After we ranked the players within both the list of hitters and pitchers, we identified the midpoint within each list and selected all the players at the midpoint and below. Then, we identified 3 upcoming free agent players, who had signed free agent deals in the past, (2 pitchers and 1 hitter) from these at the midpoint or below-the-midpoint lists. We selected Martin Maldonado (American League Catcher), Eduardo Escobar (National League 3rd Baseman), and Danny Duffy (American League Starting Pitcher) as our free agent players.

Next, we compiled distinct lists of all the (starting level) American League catchers, National League 3rd basemen, and American League starting pitchers that were on free agent contracts in 2021, and applied the formulas stated above to rank the players within these lists according to the score assigned to them based on our formula.

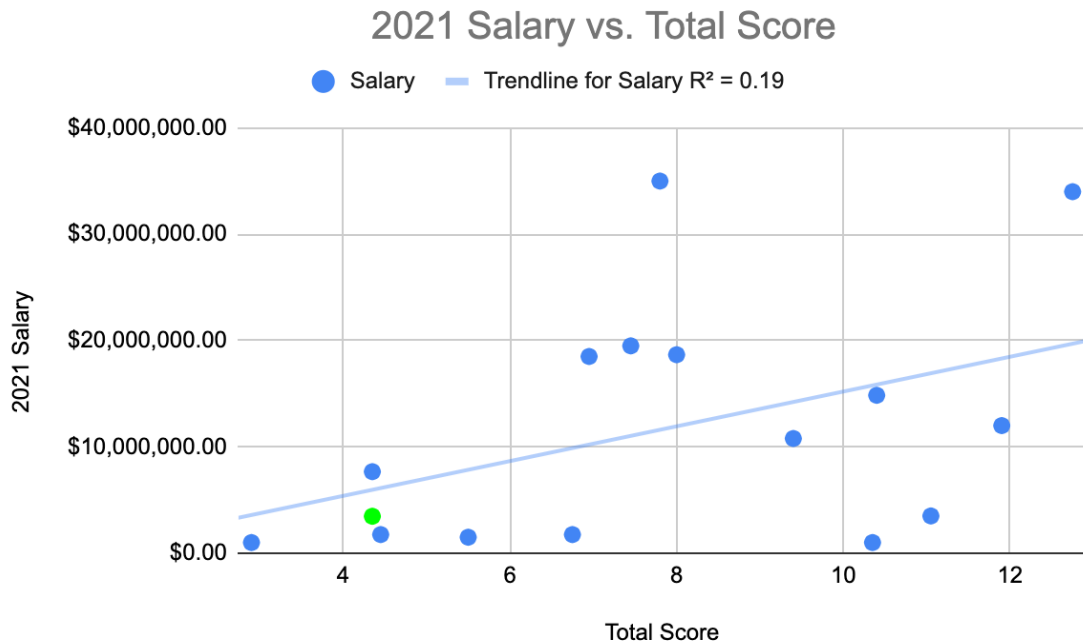
Alongside the metrics we used for both hitters and pitchers, we also captured the age of these free agent players in 2022 by simply adding 1 to their current age, as well as identifying their 2021 salaries. To calculate the optimal salary for our 3 free agent players, we decided to take the moving average by taking the average of the two salaries above plus the two salaries below each one of our free agent players and setting the result equal to the optimal salary of our free agent player.

We then created 3 scatter plots to visualize the data:

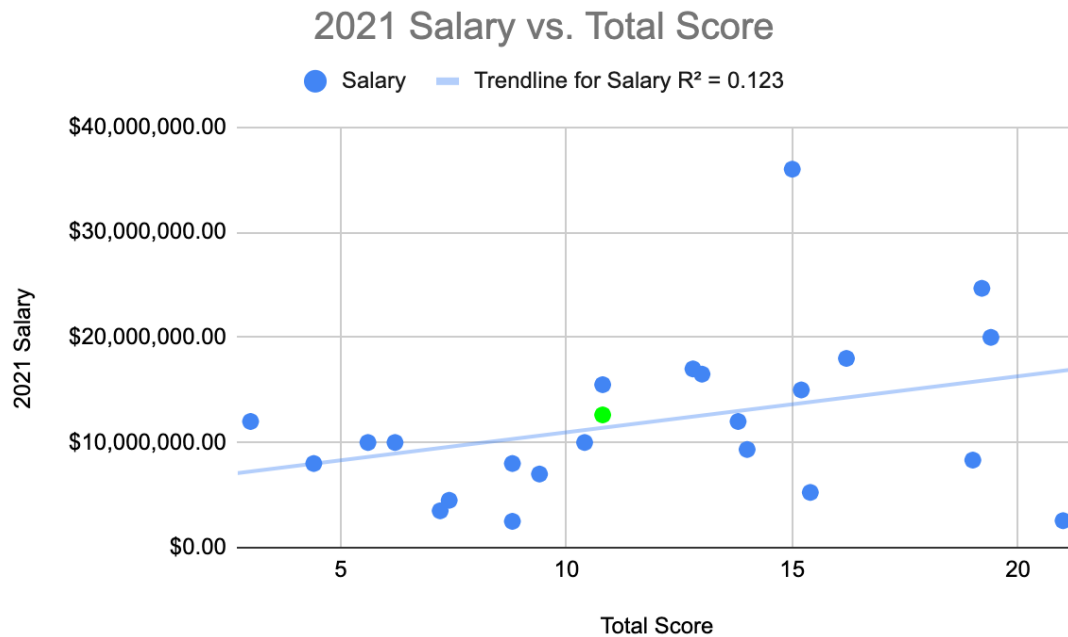
Martin Maldonado
(The green dot represents the optimal salary for Maldonado)



Eduardo Escobar
(The green dot represents the optimal salary for Escobar)



Danny Duffy
(The green dot represents the optimal salary for Duffy)



The average age of a free agent player within the MLB is 32.2, and the average age of MLB player peak performance is 29.4, which leads to players getting paid less once they pass the age of 30. The 3 players we selected (Maldonado, Escobar, and Duffy) are 35, 33, and 33 respectively. All 3 of them are above the age of 30.

Metric Selection

Position Players: Aside from using the instructed metrics of wRC+ and wOBA, we also wanted to incorporate two other metrics that would evaluate the overall performance and value of a position player. Therefore, to not account just for hitting value, we chose to evaluate players on their defensive and baserunning value through the statistics DEF and BsR. The DEF statistic measures a player's defensive value relative to league average, and adjusts for positional placement as well. An average DEF score standardized to 0 with anything above that as a positive to player value. BsR is "FanGraphs' all encompassing base running statistic that turns stolen bases, caught stealings, and other base running plays (taking extra bases, being thrown out on the bases, etc) into runs above and below average" ([Fangraphs](#)). As with DEF, the average BsR score is 0 and above that is extra runs above average to the player's respective team. We think the combination of wRC+, wOBA, DEF, and BsR should be a great way to account for all 5 tools of value a player provides (hit contact, hit power, run speed, arm strength, fielding ability).

Pitchers: After being given instruction to use FIP and strikeouts as two of our metrics, we needed two more to round out our analysis. FIP basically measures how well a pitcher is at preventing runs, independent of his team's fielding ability. Strikeouts are good at showing how well pitchers are at deceiving opposing batters. With those two categories covered, we wanted statistics that could measure a pitcher's control and how well their pitches are being hit. For this reason, we chose BB% and Hard% for our data. A low BB% signifies that a pitcher can locate his pitches well and a low Hard% means that pitchers are preventing hitters from getting good wood on the ball.

Weighting Logic

Hitters: For hitters, we put a clear emphasis on weighting hitting as more valuable than the fielding and base running metrics of DEF and BsR. With fielding, though DEF accounts for positional adjustment, MLB front offices still value positions differently based on different skill sets, e.g. the fielding of a SS or CF is more important than that of a 1B, so we decided to weigh this metric less (15%) and focus more on hitting since it is more valuable for position players. Therefore, hitting metrics wRC+ and wOBA were weighted more heavily, both with 35% of our scoring. Both are great metrics that value the success of a hitter very accurately, so we decided to weigh them evenly. Finally, BsR was also weighted on a relatively low scale of 15%. Since speed is not a required strength of many of the greatest players' games, and since most baserunners are relatively conservative on the basepaths unless speed is their top attribute, it is more of a luxury to have speed in today's game. This is why BsR was weighted less than hitting.

Pitchers: In regards to pitching, we thought FIP was the most important statistic since the name of the game for pitching is to prevent the other team from scoring the ball which FIP does. Because of this, we weighted FIP the most at 40%. The other three statistics, K's, BB%, and Hard%, are not extremely more valuable when compared to each other, so we rounded up our dataset ranking system by weighting them equally at 20% each.

Salary Logic

To determine the optimal salary for our selected free agent players, we took a moving average of the salaries from two players above and two players below the free agent from our ranking lists. For example, to find Martin Maldonado optimal salary we took the average salaries of Chance Sisco, Jason Castro, Mike Zunino, and Alex Avila. Their salaries respectively are \$580,500, \$3,500,000, \$2,000,000, and \$1,500,000. The average of those four salaries is \$1,895,125. From the graph above you can see Martin Maldonado is below the trend line. We feel like this is right because Maldonado will be 35 years old in the 2022 season. An increasing age of an athlete past age 30 correlates to salary reduction.

For Eduardo Escobar, he was near the bottom for our National League third basemen list. Because of this, we could not do a moving average of two players above and two players below. Instead we took the moving average of the player above him salary, his current salary, and the player below him salary. The average of those three salaries was \$3,472,222. From the graph above he is below the trend line as well. We think this is because he had a poor 2020 season and is at the bottom of our ranking list from our formula.

Lastly, to determine Danny Duffy's optimal salary, we did the same process as we did for Maldonado: taking the moving average of two players above and two players below salaries. The average for that was \$12,625,000. From the graph above you can see that Duffy is above the trend line. We think this is because Duffy will be 33 years of age in the 2022 season and is still probably young enough to get another free agent contract before he retires. He is still young enough to produce positive output for a team in the next few years, so a team could pay him slightly more than trend.