After downloading the batting, pitching career in 2021 from fangraphs.com, I realized that I had to download different Starting Pitching data and Relievers Pitching data from the website. If I downloaded players’ state with all selection, then it will only provide the players of 39 which is very limited, and based on my research, all 39 pitchers have held the position of Starting Pitching for the position. Therefore, I downloaded differently for the RP (relievers pitching) and SP (starting Pitching). After downloading, I combined all data into the given Excel file. I started to clean and fill some blank sections.

As given the word document, the data that I downloaded had some blank for the Team. Therefore, I filtered out to list only “---” in the team column and started to search individual players in google to see the previous team in 2021. In fact, it was a simple process since I only had to copy and paste the name on google. On top of that, I started to decide on each player who had two different positions. For example, Shohei Ohtani had two different positions which were pitcher and designated hitter (DH) and I decided him to be as DH rather than Pitcher because I couldn’t see his pitching from the downloaded pitching data.

After cleaning the dataset, I started to check the minimum Pas and Ips from the 2021 season data. From this process, I couldn’t find any value that is below or equal to 300 PAs for the offense and 90 IPs and 30 IPs for the Starting Pitchers and Relief Pitchers. Once we checked and removed unqualified data from the dataset, I started to Normalization for WAR. To speed up the calculation, I decided to use the proportional expression for individual cells.

Since finished the WAR normalization, I combined all three different sheets of different position’ players into one sheet. Once I decided to use an Optimization tool called Solver the maximize the average of WAR, I had to minimize the players to less than 200 people. So, I removed some of the relief pitchers who had negative WAR and very small WAR values because the purpose of the whole project is maximizing the WAR. Also removed some of Starting Pitchers and players in Outfield because there were a total of 43 players in Outfield players which is about double or triple of different players in the dataset. On top of that, I decided to make higher WAR, I merged three different Outfield players into one position as Outfield Player.

To select the 25 players for the roster, I decided to use the tool called Solver. From the constraints, I put a limit of 25 players, and the salary must be below 132 million dollars, and each position has to be 5 starters, 7 relievers, 2 catchers, 6 infielders, and 5 outfielders. On top of that, I set up the constraints for the salary range and a list of teams that has to be less than or equal to 2 members in each team. All the constraints were basically into binary numbers to calculate and make shorter calculation time for the solver.

From the result of calculation, and selecting 25 people with constraints, I used a total of $126,209,042 out of $132,000,000 which means that we have $5,790,958 left. I used 19 teams to make a roster out of 31 teams. I selected 2 players from LAD, CLE, STL, TBR, KCR, and MIL and selected one player from TOR, PHI, WSN, SKP, OAK, NYY, PIT, SFG, BOS, LAA, CHW, ATL, and CHC. On top of that, I selected 5 players who are receiving higher than 10 million dollars for the salary and selected only 3 players who are receiving between 5 million and 10 million, selected 4 people who are between 2 million to 5 million dollars for the salary lastly, I selected 13 players who are receiving lower than 2 million dollars for salary.

From the result, I received 4.4 for the average of normalization WAR which is also can be 4.5 for WAR without normalization.