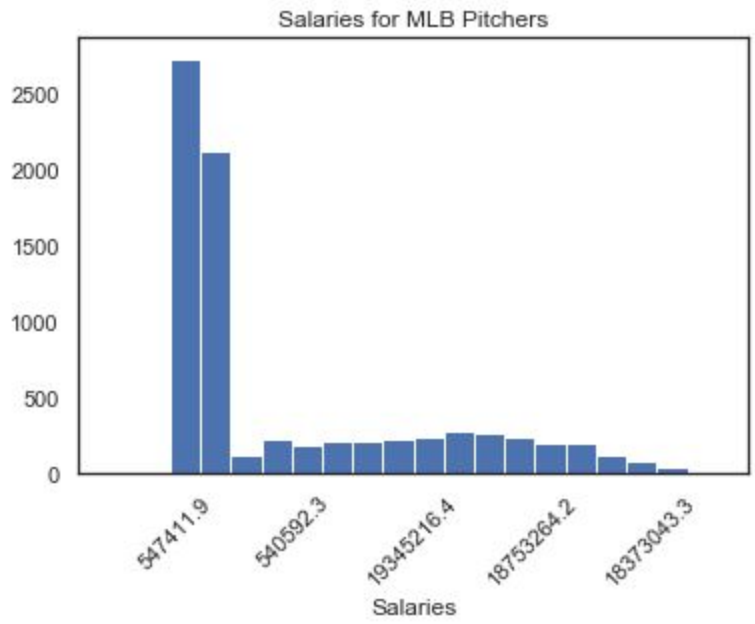
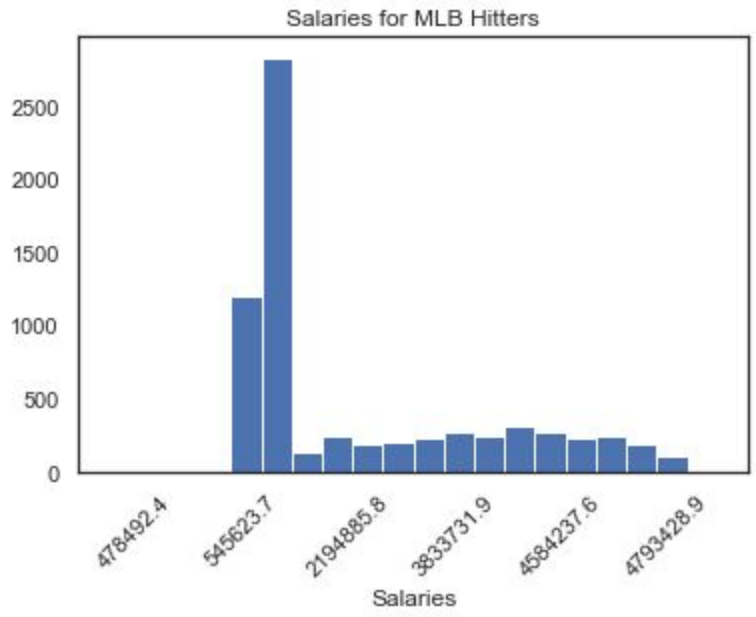
Exploratory Data Analysis

Distribution of Salaries(Target Variable) for Pitchers and Hitters



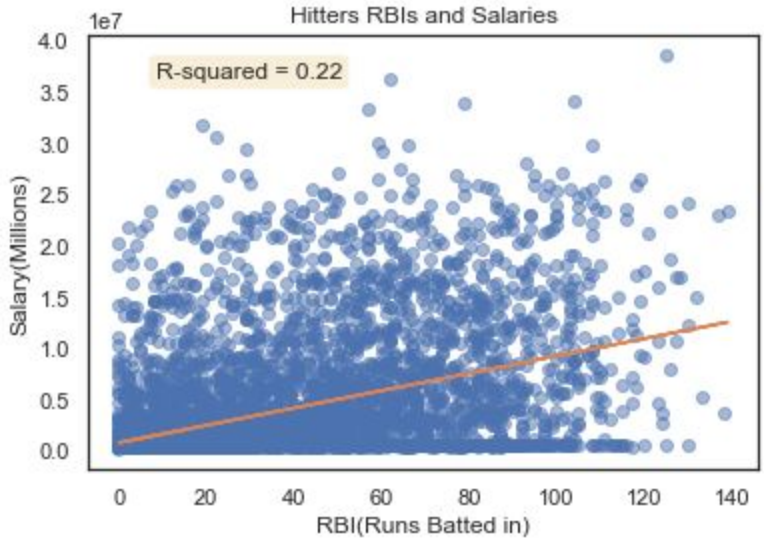
The chart above depicts the salary distribution for pitchers in the MLB. According to the chart most of the pitchers earn a salary around $550,000, which is close to the minimum salary amount. As we move to the right, we can see a drop off for players earning more than the league minimum. Towards the end of the chart there are less than 100 players earning a salary of $20 million or more.



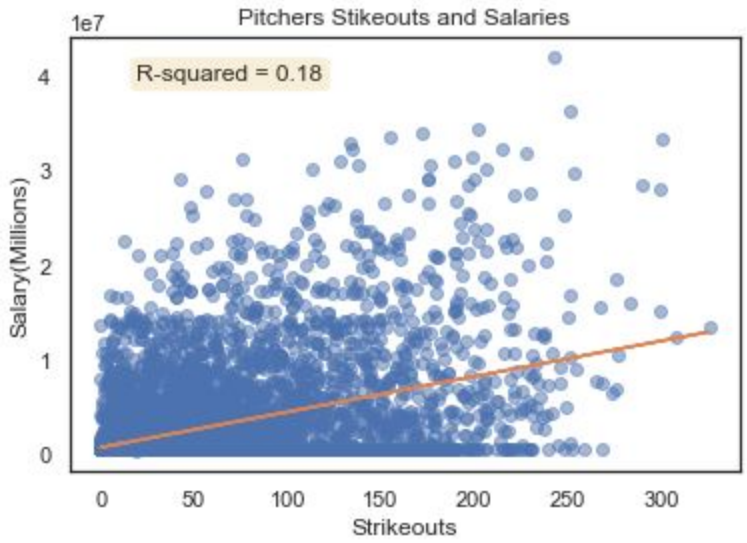
Similarly to the pitchers distribution of salaries, hitters too mostly earn the league minimum around $550,000. As we move to the right, we can see a drop off for players earning more than the league minimum. Towards the end of the chart we can see that the distribution of salaries for hitters goes all the way up $50 million.

Independent Variables Vs. Dependant Variable

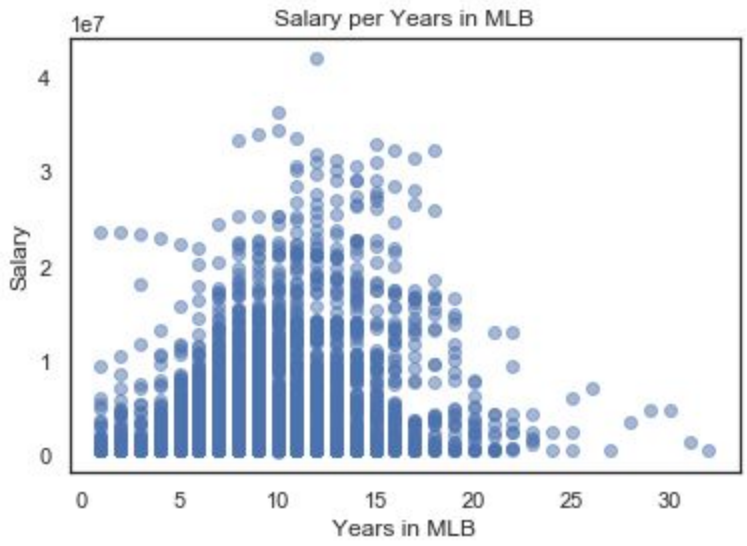
The following charts will be a comparison between multiple independent variables and our target variable ‘salary’. These relationships will be depicted by a scatter plot.



The chart above represents a positive correlation between a player's RBI(runs batted in) and their salary. This feature also has the strongest level of correlation with the target variable within the hitter dataset. As the number of RBI's increases so does a player's salary. We can also see a couple of data points with high RBI totals and a low salary. This could be due to exceptionally good rookies just joining the league and being paid the minimum.



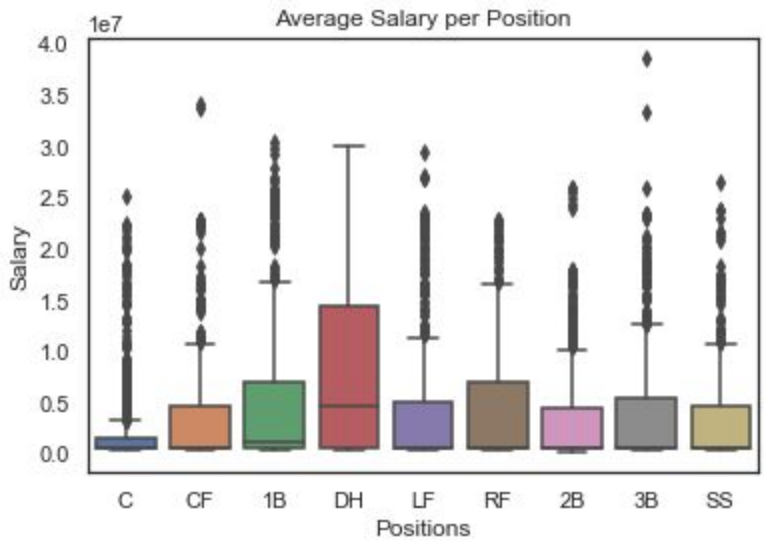
The chart above represents a positive correlation between a pitcher's strikeouts and their salary. This feature also has the strongest level of correlation with the target variable within the pitcher dataset. As the number of strikeouts increases so does a player's salary. The data includes different types of pitchers such as starting pitchers, relief pitchers, and closers. Number of strikeouts is one of the top measured pitching metrics per season.



The chart above depicts a weak or non-existent positive correlation between ‘Years in MLB’ and ‘Salary’. During the first couple of years in the MLB we can see a gradual increase in salary. We can also see that a player's maximum salary tends to peak after 7-10 years of MLB service. After 15 years of MLB service a player's salary begins to decrease rapidly. This could be due to aging and below average performance.

Categorical Data vs Target Variable

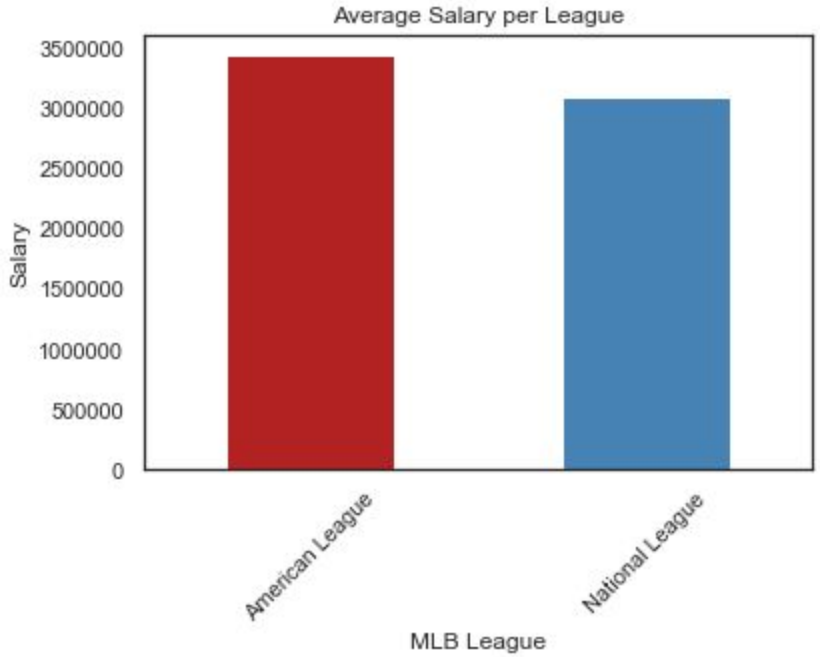
- ➔ In the following graph I will be creating a box and whisker plot, depicting the salaries paid to each position.



The chart above depicts the average salaries per player position. Beginning with catchers, it looks like they are the least paid on the diamond. We can also see that there are a couple of outlier points reaching up to salaries of $2.5million. This was rather interesting to see because catchers are in charge of controlling the game, and calling every pitch. On the other we can see that DH or Designated hitters are paid on average a higher salary. The salary paid to a DH seems to be more consistent than the salaries paid to other position players, due to the presence of less outliers.

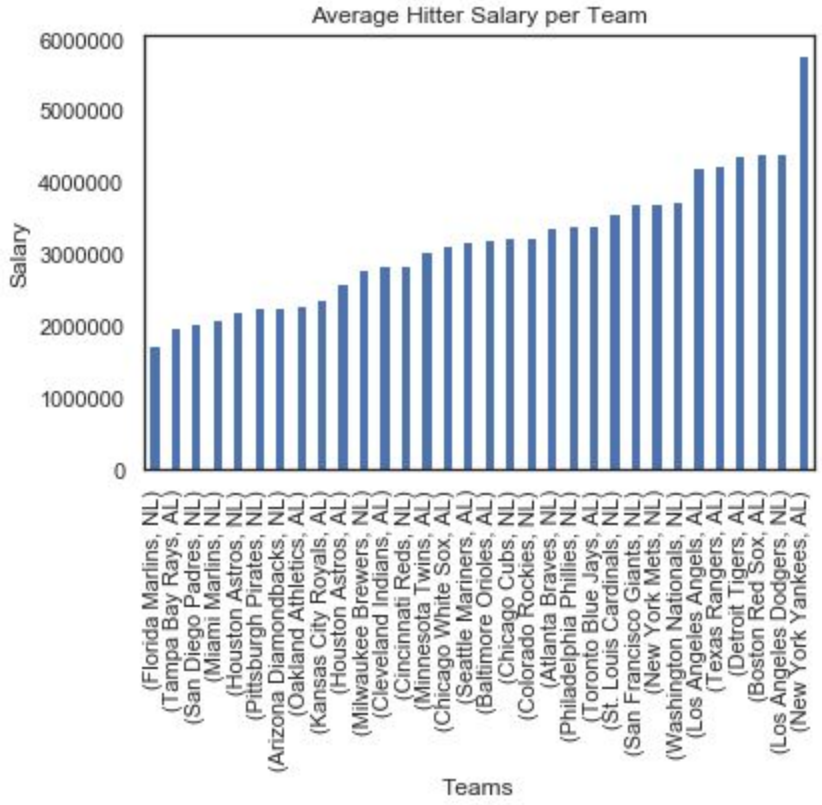
- ➔ In the following graphs I will be creating multiple box plots, depicting the distribution of salaries paid to each categorical data.

From a players perspective what league pays the most?



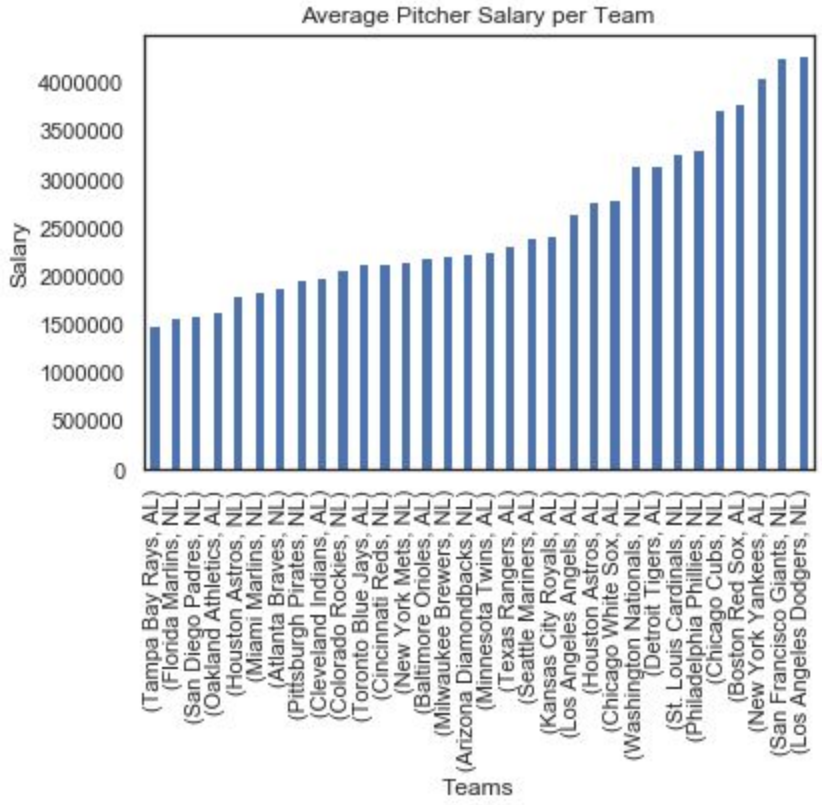
In the ‘Average Salary per League’ chart, we grouped the average player salaries for each league. Players in the American League appear to have a slightly higher salary than their National League counterparts. The slight increase in player salaries within the American League might be due to a couple of reasons. I believe one of the biggest factors is because 4 out of the 5 top paying teams are within the American League. This can be seen in the following chart.

From a hitters perspective what team pays the most?



In the above chart we can see the average salaries for hitters paid by each team. The New York Yankees tend to pay a higher salary, followed by the Los Angeles Dodgers. The teams with lower average salaries appear to be the Florida Marlins and San Diego Padres. It is important to note that the Miami Marlins and Florida Marlins each appear in the chart. They are the same team but were renamed over the last couple of years.

From a pitchers perspective what team pays the most?



In the above chart we can see the average salaries for pitchers paid by each team. The Los Angeles Dodgers tend to pay a higher salary to pitchers, followed by the San Francisco Giants. It seems like California is a good market for pitchers. The teams with lower average salaries appear to be the Florida Marlins and Tampa Bay Rays. As mentioned earlier, It is important to note that the Miami Marlins and Florida Marlins each appear in the chart. They are the same team but were renamed over the last couple of years.