ITCS 3162

Ryan Malmfelt

9/13/2024

Project 1

Exit Velocity in MLB

INTRODUCTION:

In the past few years, Major League Baseball has begun to look at more advanced statistics in the game. One of those statistics is exit velocity. Exit velocity measures the speed of the baseball as it comes off the bat, immediately after a batter makes contact. This stat is tracked for every batted ball. Achieving a high exit velocity is one of a hitter's primary goals. A hard hit ball may result in an out, but the defense will have less time to react, so the batter's chance of reaching base will always be higher. The overall problem that I will be looking at is does having a higher exit velocity make you a better player in Major League Baseball? Using this dataset I will try and solve problems related to exit velocity. Questions I plan to lean into and answer are something like this; Does higher exit velocity lead to a higher or lower launch angle off the bat? Piggybacking off of that question I will look if having a higher launch angle leads to farther home runs. Another question I look to answer is what players have a higher average exit velocity and how that correlates to launch angle and average home run distance. I plan to also focus primarily on 2021-2022 data and compare those years to see if exit velocity had increased between those years.

INTRODUCING THE DATA:

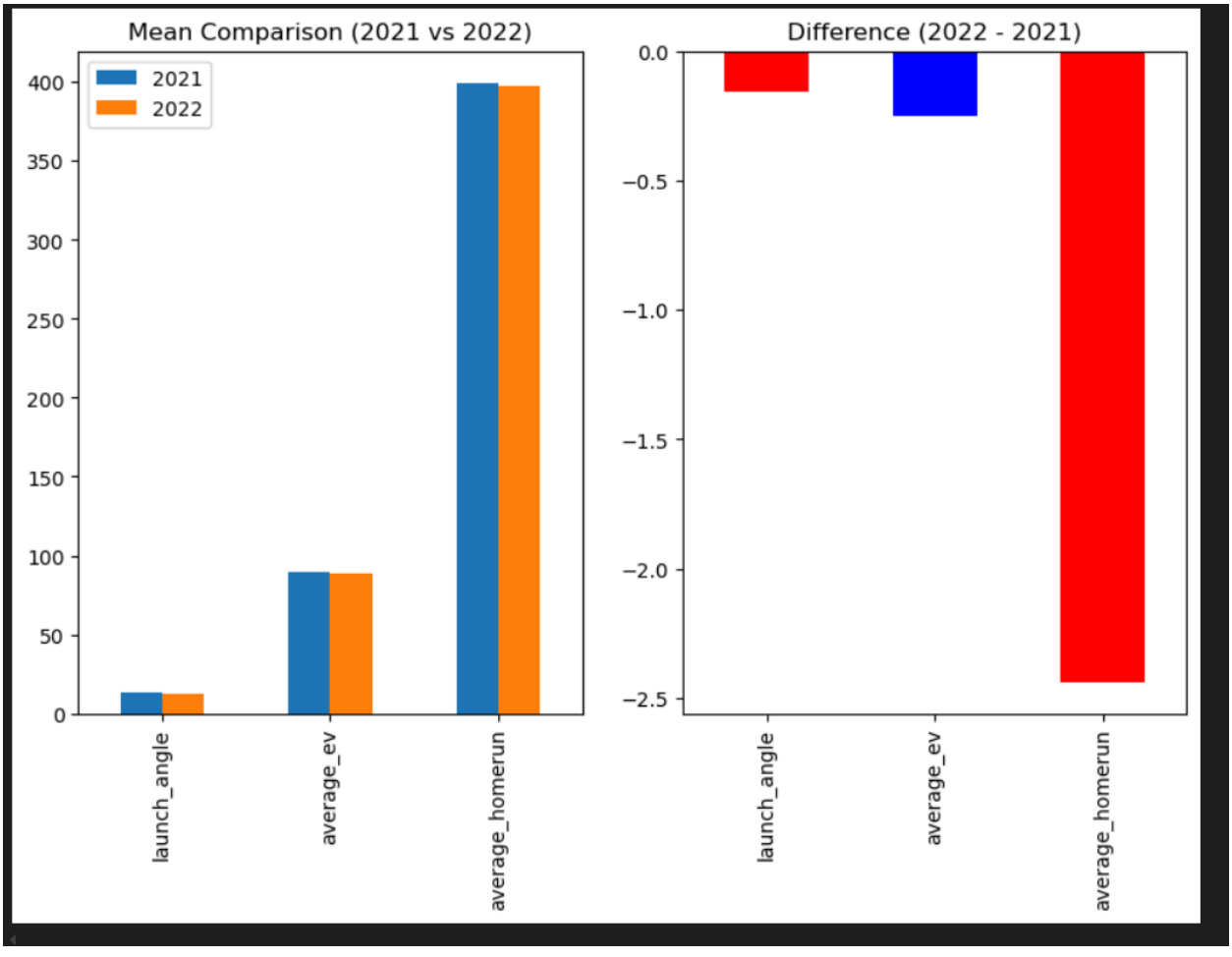
The dataset that I decided to use comes from the Kaggle website and it contains Major League Baseball batting exit velocity data from 2015–2022. It has around 2,000 rows of data and exactly 20 columns of different features. Some of the more important features that I will be focusing on are the players, average_ev, which is average exit velocity, launch_angle, which is the angle that the ball comes off the bat at, and average_homerun, which is how far the average home run ball is for the players of this dataset. Using this dataset I hope to answer the questions that I asked above.

DATA PREPROCESSING:

I began my data preprocessing by making sure the usability of the dataset I chose was high. I wanted to check the usability of the dataset because if the usability of the dataset is low, then how are you gonna know if you're going to get good and accurate results? The usability of the dataset I pulled from Kaggle was 9.71/10; so the dataset is very well functional and good to use for the problems that I intend to answer. Next I wanted to make sure that I was able to pull the CSV file and copy it into my notebook in order to have all the features that the file included. This step is needed to make sure you have the correct data in order to answer your questions. Another step that I chose to take was to look for null values in the dataset. It is important to check for null values because

having null values means that you have unknown values in your dataset and could mix up your potential results. After checking the dataset, I determined that it had zero null values in all the columns. Then in order to answer the specific questions that I wanted too, I separated the columns that I needed. I felt that this was paramount because having data that you do not need can mess up the results you want to look for. I selected the columns player, year, average_ev, launch_angle, and average_homerun because those were the main features that I plan to focus on. Then I selected the first 480 rows because those consisted of data from the years 2021-2022. I did this because I plan to center all of my visualizations on these years and compare them to see trends in exit velocity.

VISUALIZATIONS:



The first model that I created compares the mean values for launch angle, average exit velocity, and average home run distance along with the differences between them from the years 2021-2022. Starting with the graph on the left. The graph on the left shows the mean comparisons of the features I focused on. The average launch angle for the 2021 MLB season was 12.97 degrees and in 2022 it was 12.82 degrees. This shows that the average launch angle decreased by .15 from 2021-2022. Now to look at the average exit velocity. In 2021 the average exit velocity was 89.24 MPH and in 2022 the average was 89 MPH. This shows that there was another decrease in statistical analysis. The average exit velocity decreased by .24 MPH from 2021-2022, which is very little but can still be significant when looking at if a player hits the ball hard enough. The last thing these models show is the average home run distance and their difference. The model shows that in 2021 the average home run distance was around 399.62 feet and in 2022 it was around 397.18 feet. This is where I see the largest difference with around 2.44 feet between 2021-2022. These models primarily show that on average a MLB player hit the ball higher, harder, and farther in the 2021 season than the 2022 baseball season.