Quality of Contact Algorithm - Austin Lee

Data Exploration/Data Cleaning

Using the attached csv, I removed all the columns that would have no correlation to the contact of the ball and only kept columns of information that would be a "Result" of contact with the baseball. The columns I chose were: **Exit Speed, Launch Angle, Direction, Distance, and Bearing.** I also kept the column **Play Result** because I wanted to calculate if there was any correlation of the 5 categories that would contribute to any of the "play results".

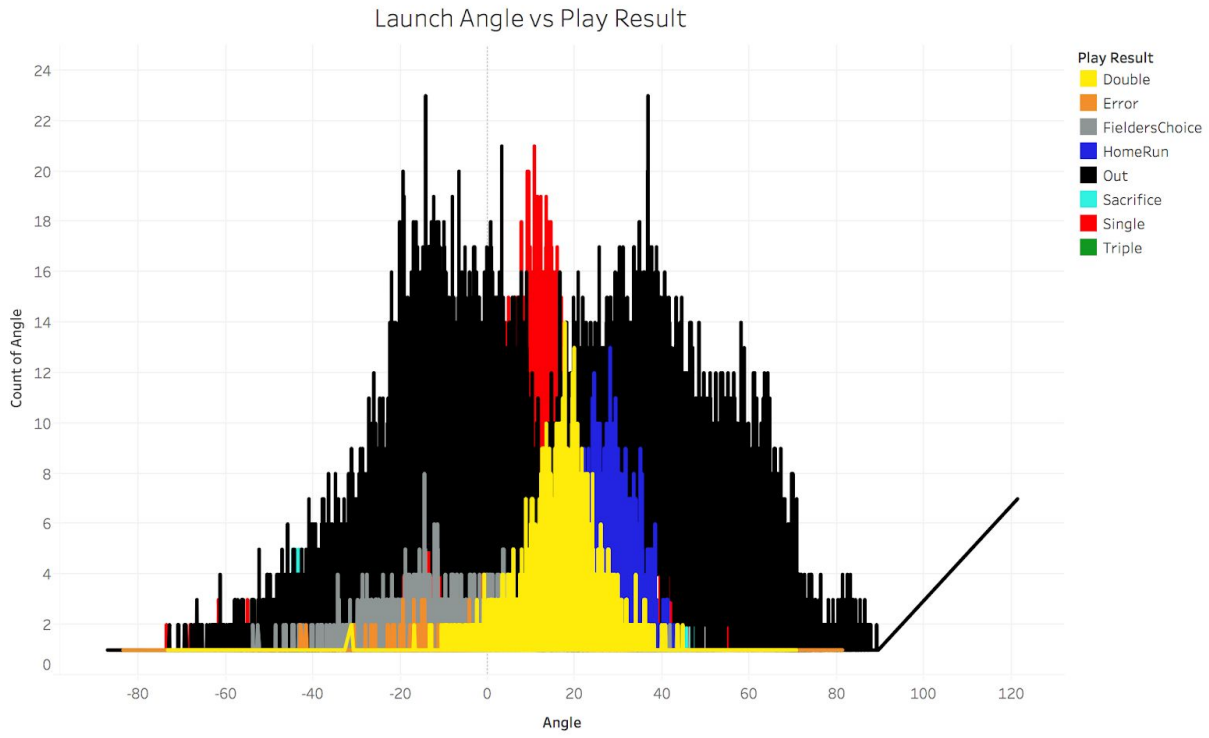
Trend #1 - A larger distance and exit speed are correlated to a better "play result".

This was calculated using a Jupyter Notebook, grouping the data by play result, and then taking the average of each 5 columns. There is a slight discrepancy with distance because an average out had a higher distance than a single, but is less valuable so this is not as strong of an indicator compared to exit speed.

play_result	exit_speed	angle	direction	distance	bearing
HomeRun	103.564584	28.213486	-2.028699	400.118244	-1.747968
Triple	97.586745	20.120352	9.951198	303.221003	13.664727
Double	97.326549	16.626925	-3.817449	264.312806	-3.909124
Single	89.751320	6.264910	-2.406049	143.365378	-1.765964
Out	85.252010	13.953233	0.007042	150.991733	0.458863
Error	84.726466	-6.287160	-6.955740	58.893099	-5.490504
Undefined	83.093043	16.716667	-5.996377	182.933333	-6.318551
FieldersChoice	82.568008	-12.783992	-7.218497	28.561605	-6.123943
Sacrifice	70.643226	3.295348	1.996409	195.952374	2.272519

Trend #2 - Range of Launch Angle of Positive Plays in Certain Plays

Though there is no linear correlation, between increasing launch angle and the "success" of a play result, there is an extreme subset of launch angles where the "positive plays" vs the outs, where it generally is bimodal around -14 and 36 degrees.

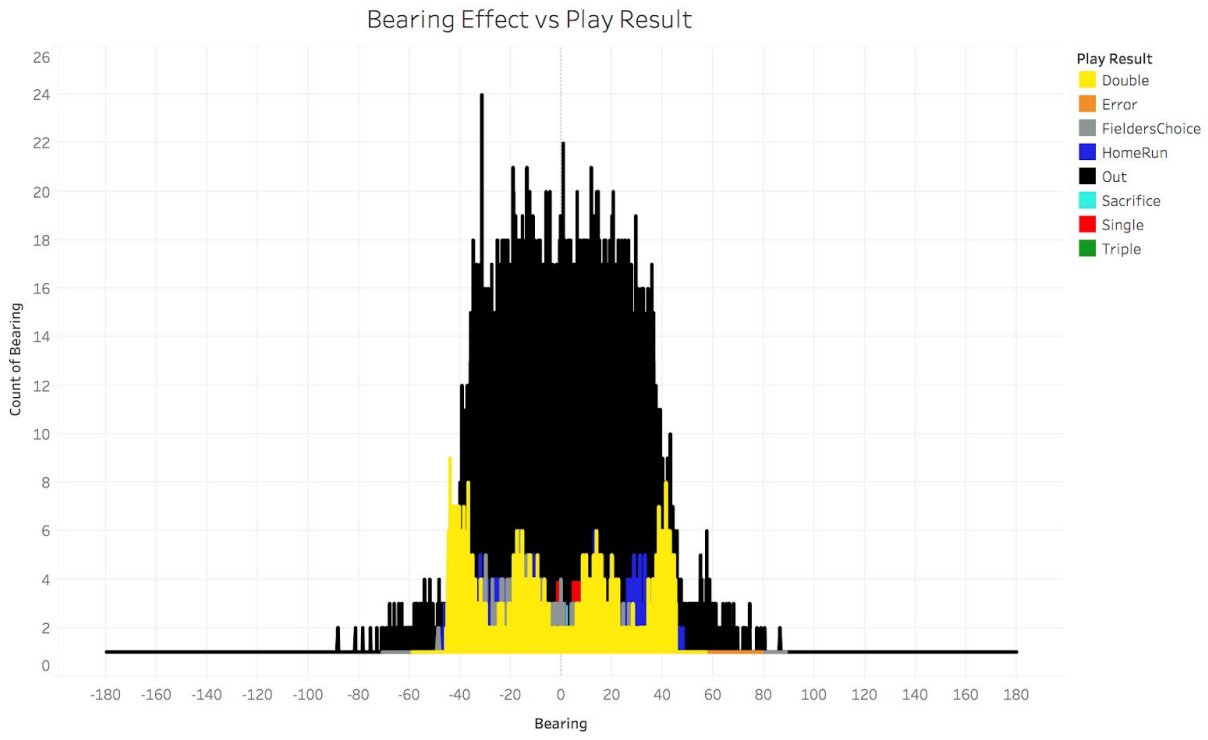


The trend of count of Angle for Angle. Color shows details about Play Result. The view is filtered on Play Result, which excludes Undefined.

Therefore, I wanted to take note of the launch angle, and store the range where 95% of the hits occurred. I calculated this by adding and subtracting 2 standard deviations of the average. Standard Deviation was retrieved through Tableau functions.

Play	Range of Launch Angles (95%)
Home Run	17.81 - 38.61
Triple	-0.54 - 40.76
Double	-5.02 - 38.27
Single	-26.42 - 38.94

Trend #3 - Certain Bearings can lend to certain types of plays versus completely random. Though the correlation isn't super strong compared to the other attributes, the graphs when plotted show that certain directions/bearings will more likely lead to a specific play result. The peaks are recorded below.

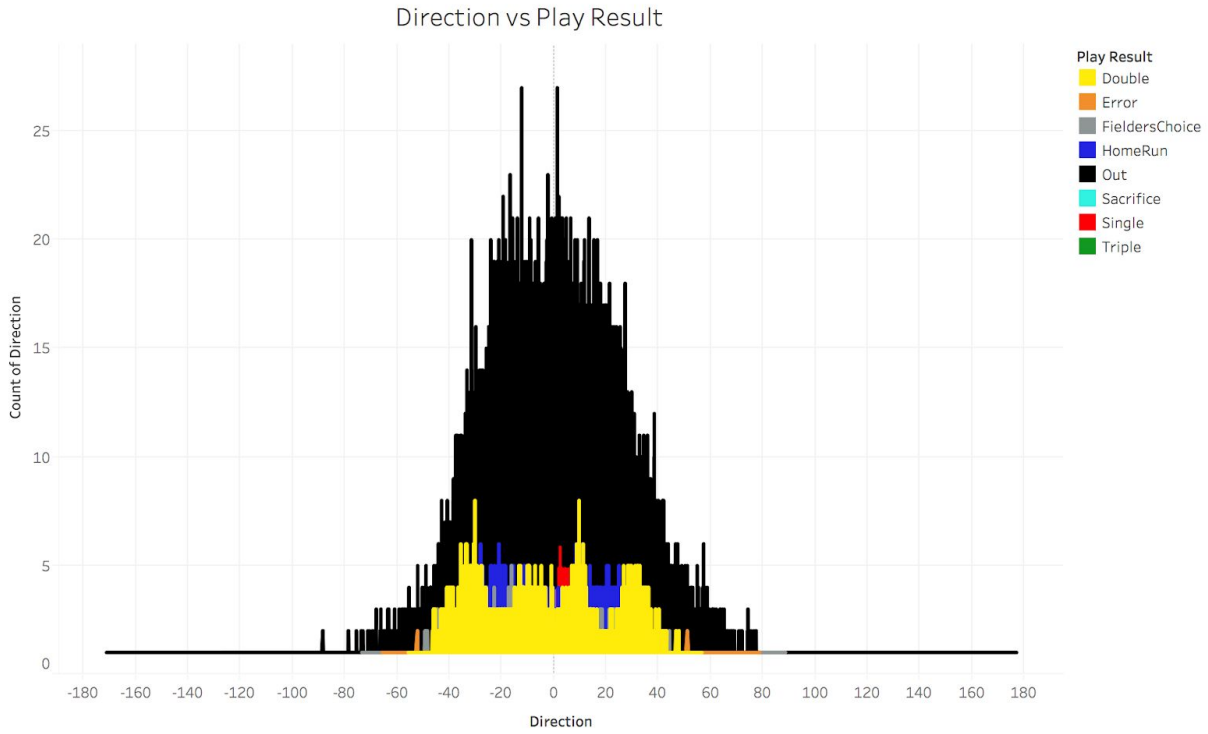


The trend of count of Bearing for Bearing. Color shows details about Play Result. The view is filtered on Play Result, which excludes Undefined.

Play	Peaks of Bearing
Home Run	-30, 30
Triple	43
Double	-44, -17.8, 13, 41
Single	0

Trend #4 - Most hits are in smaller range subset of directions. Though the correlation isn't super strong compared to the other attributes, the graphs when plotted show that almost all positive hits certain range of directions. So this range is important

to keep note of.



The trend of count of Direction for Direction. Color shows details about Play Result. The view is filtered on Play Result, which excludes Undefined.

Play	Range of Direction (95%)
Home Run	-41.5 -37.5
Triple	9 - 49
Double	-53 - 47
Single	-41.5 -37.5

Scoring

Using the 5 pieces of criteria, I created a scoring model for each set of criteria. Essentially I had a starting point of 3 for each attribute and based on how far it would be from the average of a 'HR' in this case an ideal hit, I would drive the section a grade. I added extra weight into exit speed (5) and distance (3) because of their stronger correlation to quality hits

Exit Speed Score = 5* (3 - (Avg Exit Speed of HR - Exit Speed)/ SD of HR Exit Speed)

Distance Score* = 3* (3 - (Avg Distance of HR - Distance)/100)
**Used 100 as the divider, because it seemed increments of 100 correlated to an extra base.*

Launch Angle Score** = 3 - |(Avg Launch Angle of HR - Angle)|/10

Bearing Angle Score** = 3 - |(Avg Bearing of HR - Bearing)|/ SD of Bearing of HR

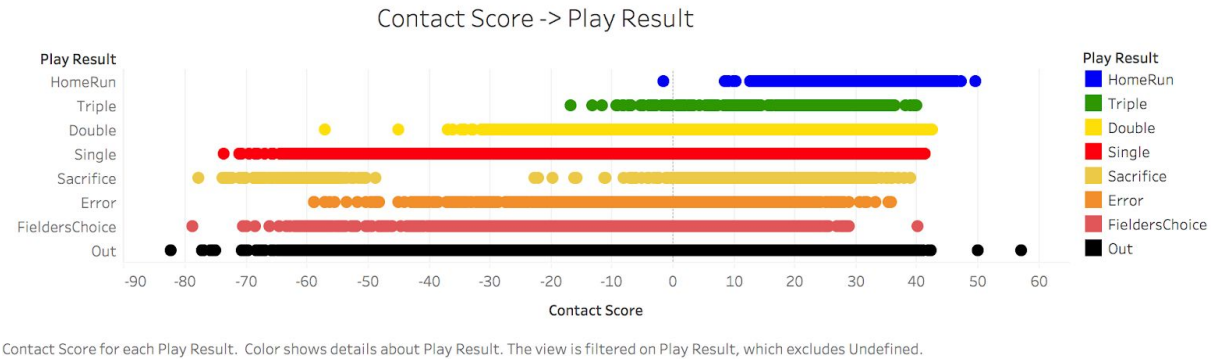
Direction Score** = 3 - |(Avg Direction of HR - Direction)|/ SD of HR Direction

*** Since these are radial, not linear, had to use absolute value.*

Final Score = Exit Speed + Distance + Launch Angle + Bearing + Direction

Results

I chose a gantt chart to show the range of contact scores for all possible play results.



Contact Score for each Play Result. Color shows details about Play Result. The view is filtered on Play Result, which excludes Undefined.

From the results, I think the contact score metric was a success because the higher contact score an at-bat has, the chances that this will result in a positive play increase by a lot, showing that it is measuring the quality of an at-bat better. Unfortunately, there are very few ranges that directly correlate to a specific play result, but I think that is realistic in baseball, as even a ball that is hit with the ideal launch angle and extremely long distance can still result in an out.