Name: Conor Desmond

Detroit Tigers Baseball Analytics Questionnaire

1. Do you have any experience working with a sports team before? If so, please elaborate on your experiences especially any interactions with coaches or players. [250 words]

   I made scouting reports on opponents for the Marist College baseball team with guidelines from the coaches. The players would look at these. I used Synergy to put stats and visuals on the scouting report. The visuals were of where the opposing pitchers would pitch the ball, where in the field the opposing hitters would hit the ball to, and a tree-like image of what pitches the pitcher would throw in each count and what the percentages were on how often he threw the pitches for each count.

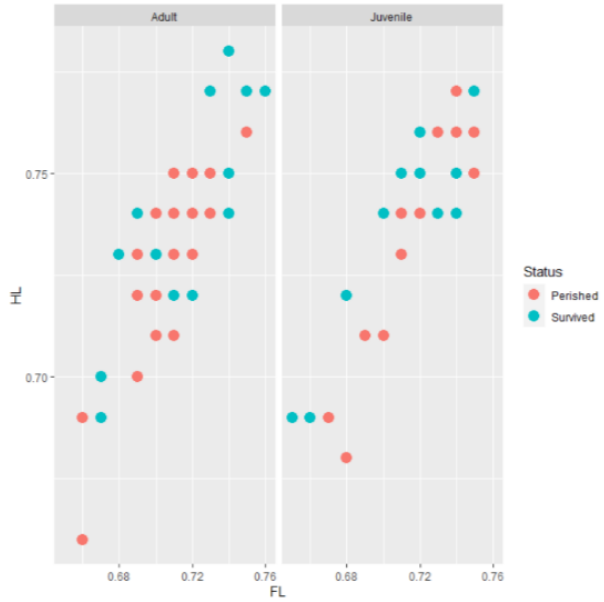
   I scored baseball games for Perfect Game USA using DiamondKast software. I recorded the substitution changes, took scouting notes on players, and recorded each pitch, marking the velocity and type of pitch. I also made sure everything ran smoothly on the field: confirmed umpires were on the field prior to games, confirmed both teams were ready to start play, and made sure we didn't fall far behind schedule.

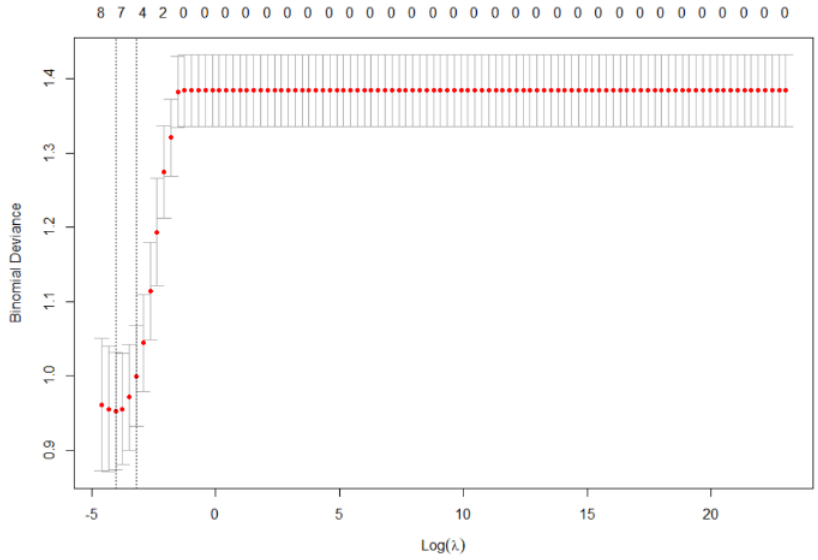
2. Please describe your experience with creating data visualizations and developing Shiny apps. Include any examples if you have them (they do not need to be baseball related). [250 words]

I made a Word Cloud from feedback on a data hierarchy at Regeneron using JMP, I made graphs in Excel at Regeneron, my RISP Research has graphs using Python and Excel, and in classes at Marist I learned how to make ROC curves in Python and other graphs using ggplot in R and seaborn and matplotlib in Python. I made scatter plots, bar charts, stacked bar charts, animated line graphs, map graphs, heatmaps, boxplots, balloon plots, and more.

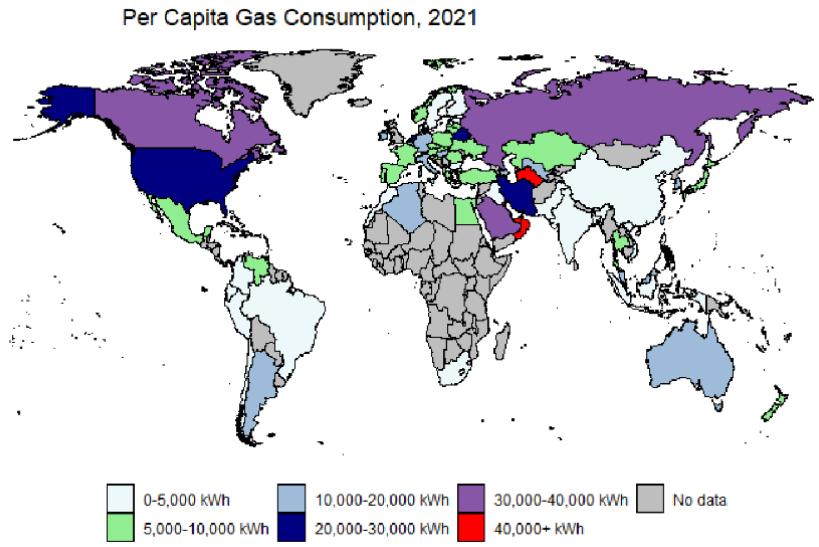
Applied Stats. Class Graphs:

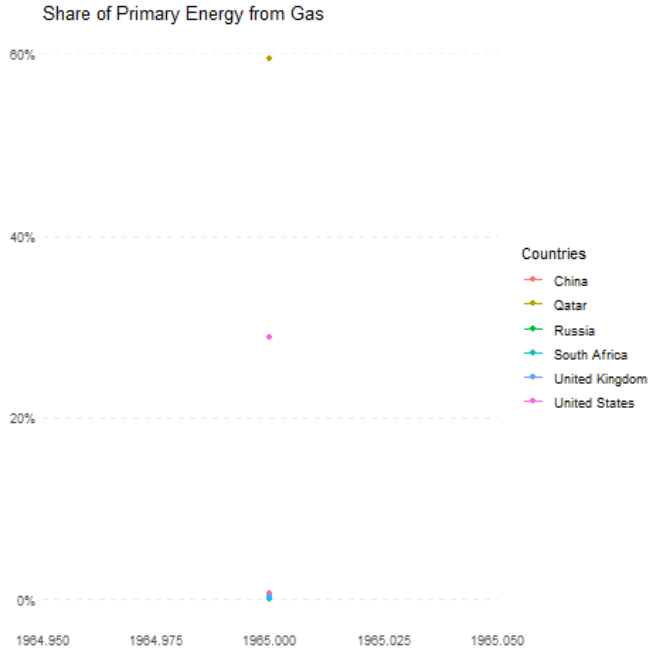
```
> ggplot(sparrows, aes(FL, HL, color = Status)) + geom_point(size = 4)+
facet_wrap(~AG)
```



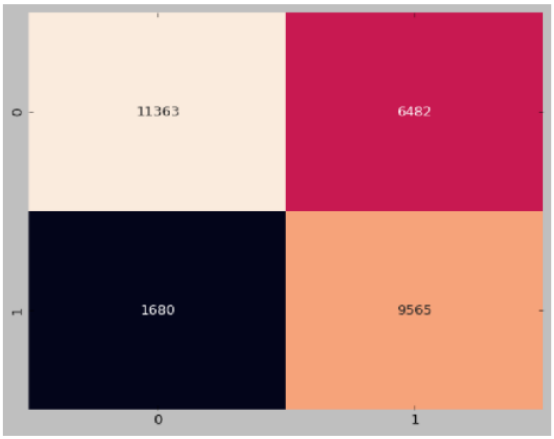


Data Visualization Class Graphs:

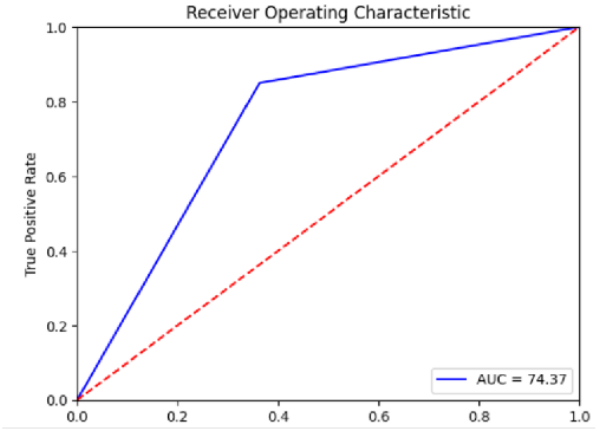




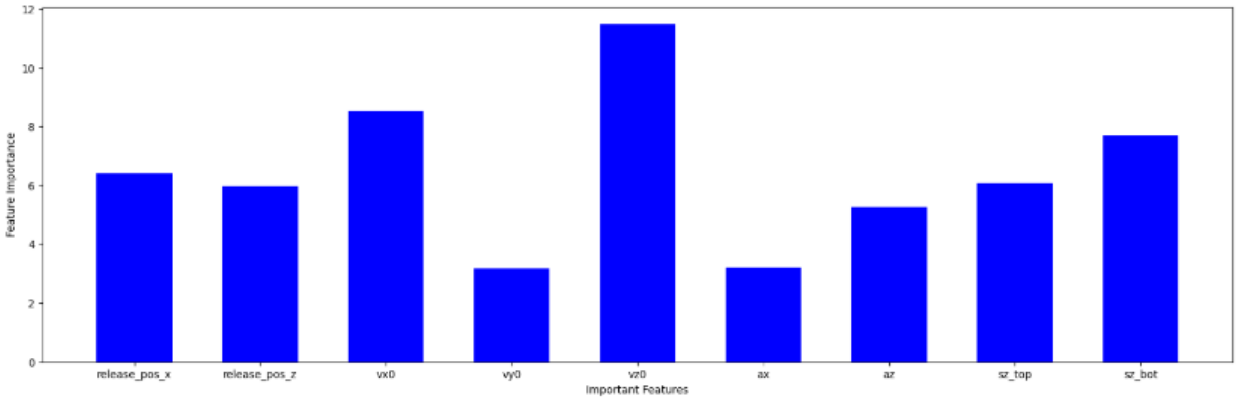
Data Science Capstone Class Visuals:



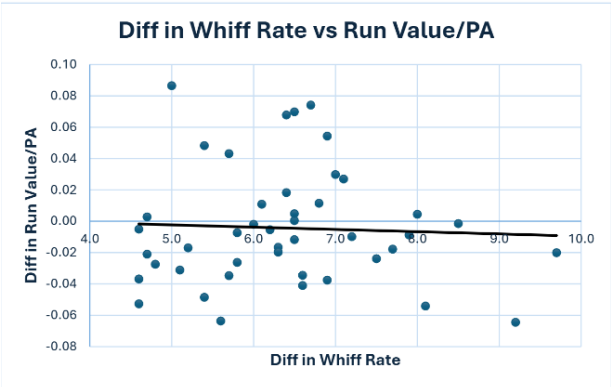
precision: 59.6%
recall: 85.1%
specificity: 63.7%
roc: 74.4%
f1: 70.1%

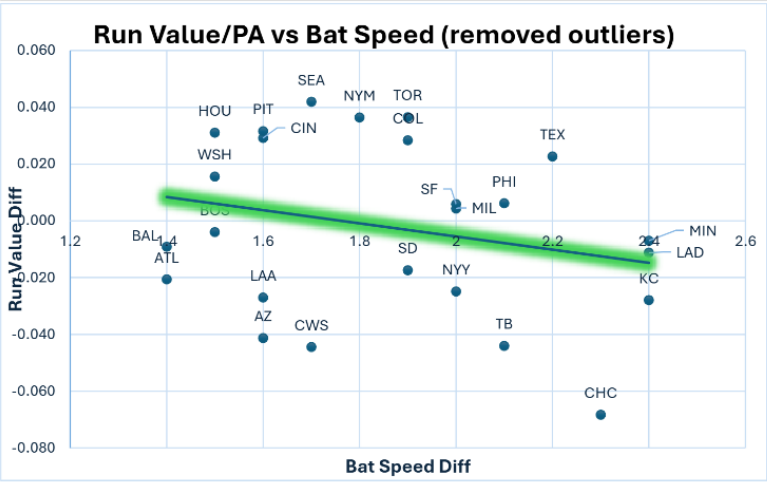


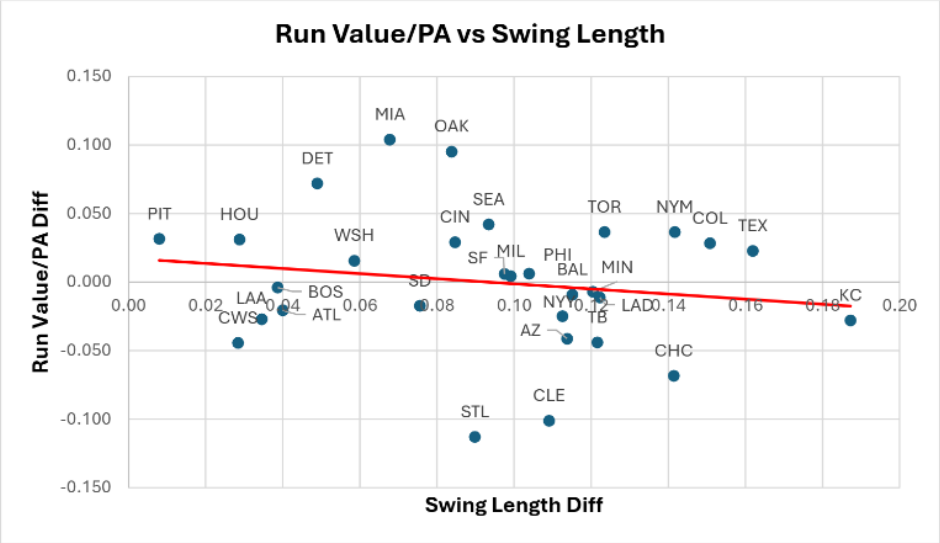
Reds Hackathon Graph:



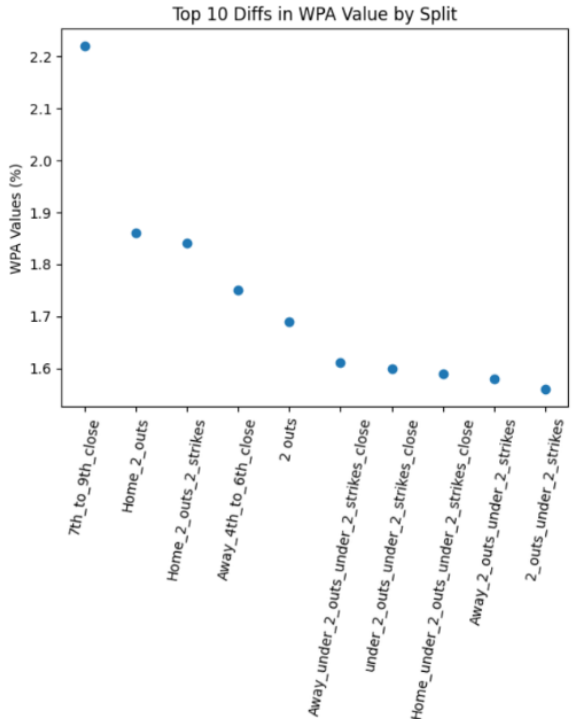
Excel Graphs from RISP Research project (these 3 graphs include with and without RISP):

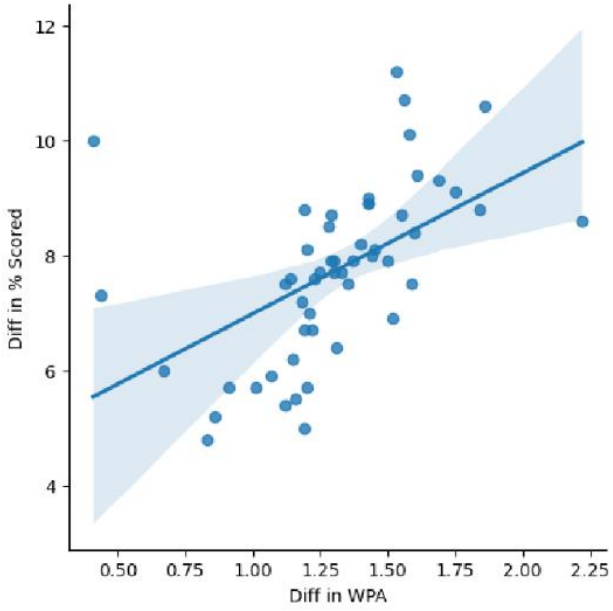






Graphs in Python from RISP Project (these two graphs are with RISP):





For the RISP project, I was comparing under 2 strikes results to 2 strike results to get the differences. I'll talk more about the project in question 4.

3. Please give a brief overview of your experience with predictive modeling as well as a description of an analytical project that you have completed. Include the purpose of the project, the methods or models chosen, any additional methods or models that were tested, and what tools or programming languages you used to complete the project. This project does not need to relate to baseball. [250 words]

   I gathered data on past teams that went to the World Series to predict who would win the World Series at the end of the season and see what patterns there were with the teams. I learned basic linear regression in R for predictive modeling. At first, I tried using that and DataRobot to predict the World Series winner. DataRobot is a site that uses auto-ML to make a lot of models for you once you upload a data set. I was able to predict the winner in 2022. I took machine learning in college where I learned about different ML algorithms in Python that use scikit-learn and TensorFlow for predictive modeling. I tried using some of those algorithms and auto-ML to predict the World Series winner and World Series teams in 2023. I did not predict the winner or the teams, but was close on predicting some playoff team stats. The ML algorithms I used were Elastic and Lasso regression and the ones that auto-ML came up with. Other ones I learned about were Ridge regression, Keras neural network, logistic regression, stochastic gradient descent, gradient boosting regressor, support vector machines, and tree based algorithms.

   Link to 2023 code: World-Series-Research/ws_research_github.ipynb at main · ConorD28/World-Series-Research

   Sports predictions on X: The Playoff Predictor (@conord282) / X

4. What is a project that you believe would add substantial value to a baseball team? Please describe the project and provide an overview of how you would complete it. [250 words]

My project on RISP, which looks into if hitters should change their approach with RISP. I looked at data from hitters from 2021-2023 using Savant's search and I used the bat speed and swing length metrics.

I identified hitters that change their approach with 2 strikes by looking at the difference in their whiff rate without 2 strikes and with 2 strikes and the league average whiff rate. I then looked at how their stats differed from with 2 strikes to without 2 strikes. For K% and walk%, I instead looked at similar players, similar based on their run value/PA without 2 strikes and whiff rate without 2 strikes and saw how their stats compared to these players. I then applied some of those differences to the rest of the league to see how the WPA, Run Value, and percent of times a run scores would change when using a different approach. For every split, WPA, Run Value and percent of times a run scores were better with a different approach. I looked at 3 teams to look at it from a team perspective instead of league perspective and the teams were better with a different approach with RISP too. Looking at the new bat speed and swing length metrics this year, run value/PA would get better the more teams shorten their swings and slow down their bat speed. A team having its hitters change their approach with RISP could help them to win more.

Project code/write up: [ConorD28/RISP-Research (github.com)]

5. For Question 5, please refer to the table below. [250 words]
    - The following pitches all come from the same pitcher. Rank them in regards to quality ("Stuff") from 1 to 3 (1 = Best, 3 = Worst) and explain your reasoning. Assume a consistent release point across all three.

1 for changeup, 2 for slider, and 3 for fastball. The changeup 1 since it moves the most, has a high spin efficiency, and a low spin rate. The slider 2 because has a high spin rate and it moves a lot horizontally. The fastball 3 because even though it moves the most, the spin efficiency should be higher, close to 100%, the spin rate is not high, and the velocity is not high either.

- What adjustments would you recommend, whether to individual pitches or the arsenal as a whole? What existing pitch would you adjust?

I would recommend adding a curveball with a lot of vertical movement, so then there would be a pitch in the arsenal that moves the other way vertically. The fastball I would adjust so the spin efficiency is higher. The pitcher needs to get on top of the ball more, then it would have a higher spin efficiency, spin rate, and velocity.

Pitch Type	Velocity	Spin Rate	Observed Spin Dir.	Spin Eff.	Horizontal Movement	Vertical Movement
FB	92.7	2145	1:25 (228°)	80%	8.5	13.4
SL	81.4	2675	8:47 (84°)	38%	-8.5	-1.6
CH	84.5	1760	1:43 (230°)	86%	13.3	9.1

6. For Question 6, please refer to the table below. [150 words]
    - Players A, B, and C are available to acquire (for this exercise assume positions are inconsequential, they are all the same handedness, that they are the same

age and of similar cost). Please rank them from the player you are most interested in, to least interested in. Explain your reasoning.

I rank them as Player C, Player B, and Player A. I am most interested in Player C because of his high exit velocity. His contact is low, but when he does make contact, he can have a lot of success. Player B because he swings a lot at pitches in the zone, chases less than Player A, and has a better EV and LA than Player A. Player B chasing less should allow him to put pitches in play that are better to hit. Player A has great contact, but he only swings at pitches in the zone 10% of the time and has worse EV and LA than player B, so his success will be determined more by chance compared to Player B. He also may get weak outs from chasing and making contact with low EV.

	EV	LA	Swing%	Z-Contact%	O-Swing%	O-Contact%
Player A	88.5	12.5	40%	95%	30%	90%
Player B	89.0	13.0	45%	85%	25%	60%
Player C	93.0	12.0	50%	75%	33%	55%