Sprint Challenge 1

By: Nathan Ensley

1. Today's sports businesses fight for an edge on their competitors as they always have; however, they have more tools at their disposal than in the past. These tools come in the form of analytics, countering some of the cognitive biases experienced in the past and informing industry leaders and medical professionals in making better decisions.

- In particular, coaches and managers may overlook certain strategic flaws if the adjustments conflict with their intuition – a cognitive bias known as confirmation bias (McNerney). Analytics adoption and acceptance within sports business could reduce the impact of confirmation bias by relying on statistical data to drive strategy.
- In addition to confirmation bias, motivated reasoning – a tendency to scrutinize ideas that counter one's intuition more intensively – may slow a coach's adoption of data-driven strategic changes (McNerney). By integrating analytics as a core part in a sport businesses' decisions, with data dashboards and weekly reports, the power of a coach's motivated reasoning is reduced.
- The Atlanta Hawks collected and analyzed data from ticket scanning, traffic patterns, and consumer sentiment to push back game start times by 30 minutes between 2014 and 2015 (Harrison 1-2). This data-driven decision resulted in more fans being in their arena seats at game start time. Similar business decisions that effectively improve fan experience can be made by incorporating analytics into business strategy.
- By having sustained business analytics within a sports organization, data is transformed into meaningful and actionable information that can enable professionals to make quicker and better business decisions that best utilize their human resources and reduce costs (Harrison 3).
- A 2016 empirical study reported that sport organizations saw a 7.2% increase in revenue growth year over year following adoption of business analytics, showing that a pivot to business analytics drives a direct return on investment (Harrison 3-4).
- Applying analytics into sports business can enhance the fan's experience. For the 2016 Australian Open, Tennis Australia leveraged analytics to send customized push notifications to fans through their event app, which allowed these fans to stay better informed and potentially become more passionate about the sport (Harrison 12).
- Market research provides insight into how a sports business functions, analyzing ticket, merch, and food sales to locate where revenue is generated. This market research can

be even more effective using analytics. A business analytics team in 2015 led a focus group that identified purchase convenience as a primary concern for consumers, resulting in a self-order kiosk being installed at the NHL All Star Game that allowed fans to purchase merchandise that could be delivered to their house (Harrison 17).

- Teams can use analytics to reduce the risk of injury by measuring certain movement statistics periodically. An athlete that shows a significant decline in their load ability when using their legs is at greater risk of injury (Fried 53). Training staff can then manage their work load appropriately to avoid serious leg injuries, like ACL injuries, that take months to recover from.
- New wearable technology has opened the door for analytics to inform substitutions in team sports. This wearable technology includes GPS data that can measure the distance and intensity a player has traveled (Fried 55). A coach or manager can then use this data to statistically identify what players are most likely to be past their ideal performance level and substitute them out with a fresher player.
- Baseball and Cricket teams can use sensors within balls to analyze the effectiveness of different pitching techniques (Fried 56). By analyzing delivery techniques and resulting spin rates, pitchers can make their pitches more difficult to hit, resulting in higher strikeout and win rates.
- Following an injury, players must rehabilitate and gradually increase their work load over time until they are healthy enough to play at their peak performance level. Using movement data analytics, trainers and coaches can design a rehabilitation program that allows a player to recover gradually without exceeding certain safety thresholds, reducing the risk of re-injury (Fried 61).

Works Cited (from class)

Fried, Gil, and Ceyda Mumcu. *Sport Analytics: A Data-Driven Approach to Sport Business and Management*. Routledge, an Imprint of the Taylor & Francis Group, 2017.

Harrison, C. Keith, and Scott Bukstein. *Sport Business Analytics: Using Data to Increase Revenue and Improve Operational Efficiency*. CRC Press, Taylor & Francis Group, 2017.

McNerney, Samuel. “Cognitive Biases in Sports: The Irrationality of Coaches, Commentators and Fans.” *Scientific American*, Scientific American, 20 Feb. 2024, www.scientificamerican.com/blog/guest-blog/cognitive-biases-in-sports-the-irrationality-of-coaches-commentators-and-fans/.

2. CRISP-DM and Analytics Frameworks

CRISP-DM is a framework for structuring a data mining solution. This process evolves from defining a problem, collecting relevant data, and finally developing a model that answers the proposed question. This process centers entirely on the project data and how it is utilized. When formulating the problem, the data scientist needs to understand the relevant industry and business context (business understanding phase) to consider what data needs to be collected and what that data means (data understanding phase). Once relevant data is collected, the data scientist must aggregate, transform, and clean the data (data preparation phase), so it is ready for use in a model. Once the data is prepared, the data scientist can proceed with using various analytical techniques and machine learning models in evaluating the established business question (model generation phase). After running these models, the data scientist must assess the accuracy and reliability of these models, establishing whether they have sufficiently answered the proposed business question (model evaluation phase). From this model, the data scientist applies the gained business insights into a business solution that effectively solves some business problem (solution deployment phase). The model insights can also develop one’s own business understanding, which may spur anotherl data-driven exploration that repeats this data mining cycle.

The descriptive, predictive, and prescriptive frameworks are ways of structuring analytical questions. Descriptive analytics measures the current state of a business, measuring what has happened in the past. Predictive analytics attempts to predict the future, using past data and tendencies to extrapolate future outcomes. Finally, prescriptive analytics evaluates the best future course of action to achieve a desired outcome. This stage of analytics enables businesses to make proactive business decisions using past data and future predictions to direct their actions.

3. **Transformer-Based Baseball Modeling for Pitch Outcome Prediction & Strategy Optimization**

Business Understanding

- Pitch calling is very complicated in baseball. Coaches or players have to make a pitch decisions within seconds (due to modern pitch clocks) while accounting for many factors: their available pitch arsenal, the batter's hitting tendencies, the batter's previous at-bats, and the current game situation.
- With limited time to make these pitching decisions, the project looked at whether an analytical system could be developed that accounts for game-time factors and makes a real-time pitch recommendation.

Data Understanding

- The project used statcast pitching data from 2015 to 2023. Each pitch had 87 continuous and categorical dimensions describing its attributes (pitch speed, pitch type, etc.) and outcome (strike, ball, etc.).

Data Preparation:

- The team then grouped these pitches into chronological sequences for each batter based on the last 400 pitches that batter has faced.

Model Generation:

- Developed a transformer-based prediction model that predicts the outcome of an upcoming pitch to rank the most optimal pitches.
- The pitch sequence is first input into the model, where it goes through different embedding, positional encoding, and transformers to produce a probability distribution for the pitch outcome and a separate distribution for hit location.
- Each simulated pitch receives a pitch score based on a weighted formula maximizing desired outcomes (strikes/groundouts) while minimizing undesired outcomes (hits/home runs).

Model Evaluation:

- The team compared the transformer model to a historical average baseline and XGBoost Model. The historical average made predictions based on historical distribution trends,

while the XGBoost Model used the same features as the transformer model without sequential context.

- The transformer model performed the best at predicting uncommon but important outcomes like singles, doubles, and home runs compared to the historical average and XGBoost baseline.
- The transformer model also showed better player adaptability than the XGBoost model at analyzing hitting styles. The transformer model better predicted the comparatively higher number of singles and lower number of home runs of a contact-based hitter (Luis Arraez).

Solution Deployment:

- The top 5% of pitch scores can be filtered on a scatterplot of pitch locations showing where the optimal pitch type and location based on pitcher, batter, and game situation.
- Following a thrown pitch, the model can use the context of that pitch to derive a graph of optimal pitches for the pitcher/coach to quickly analyze and make a pitch decision.

4. Analytics Questions for College Sports Teams

- How much monetary value do particular college players provide to the team? Thus, how much should college sports teams pay these players in a new competitive revenue-sharing platform, which has no past historical data to rely on?
- What factors lead to the greatest attendance for the college soccer team? (Are there certain factors like time and marketing that lead to greater attendance?)
- What type of players should a college sports team target to see greatest performance value returned to the team, when accounting for changes in conference difficulty and type of coaching system?
- What current donors are most undecided about donating for another season and what factors most influence their current reason to donate? (This can affect who the sales team targets in their donating pitches and how they interact with them.)
- What types of organizations are most likely to sponsor the team and how much would these organizations be willing to pay? (This looks at what organization types are currently sponsoring the team to influence what companies should be contacted for greatest sponsorship likelihood.)

5. Career Goals

My primary career goal is to secure a data analytics job in the Indianapolis metro area, preferably in the sports industry. I am willing to take jobs in business analytics, data science, or statistics in other industries as well, especially in the government or tech sectors. Specific sports I have a particular interest in include motorsports, soccer, and lacrosse, although I enjoy nearly every sport besides fighting. In particular, I want to trailblaze analytics for a specific sport and not pursue sports like baseball, where many of the analytical applications have been solved. Additionally, I have an interest in joining a smaller college sports program or the NCAA, where I may need to evaluate both business-related and sports performance data in the same job position. Overall, location is more important to me than industry type. Thus, I'm more interested in getting a job based near Indianapolis than getting my dream role elsewhere, although I'm willing to travel with the team for games if necessary.

In soccer and lacrosse, geospatial data plays a crucial role in analyzing how tactics influence the flow of the game and certain outcomes, such as goals and pass success. In motorsports, time data is key, with variations in lap times and gaps between cars being intensely studied. Over the summer, I specifically analyzed pit stop times, where time can be gained and lost off the track.